Bruce K. Driver

Analysis Tools with Examples

May 7, 2012 *File:newanal.tex*

Contents

Part I Prequel

1 Introduction / User Guide ... 3

2 Set Operations ... 5
2.1 Exercises ... 6
2.2 Appendix: Zorn's Lemma and the Hausdorff Maximal Principle ... 7

3 A Brief Review of Real and Complex Numbers ... 9
3.1 The Real Numbers ... 9
3.1.1 The Decimal Representation of a Real Number ... 11
3.2 The Complex Numbers ... 13
3.3 Exercises ... 14

4 Limits and Sums ... 15
4.1 Limsups, Liminfs and Extended Limits ... 15
4.2 Sums of positive functions ... 16
4.3 Sums of complex functions ... 19
4.4 Iterated sums and the Fubini and Tonelli Theorems ... 21
4.5 ℓ^p – spaces, Minkowski and Holder Inequalities ... 21
4.6 Exercises ... 24
4.6.1 Limit Problems ... 24
4.6.2 Monotone and Dominated Convergence Theorem Problems ... 25
4.6.3 ℓ^p Exercises ... 26

Part II Measure Theory I.

5 What are measures and why "measurable" sets ... 31

6 Set Operations . . . 35
6.1 Exercises . . . 36
6.2 Algebraic sub-structures of sets . . . 37

7 Finitely Additive Measures / Integration . . . 41
7.1 Examples of Measures . . . 42
7.2 Simple Random Variables . . . 44
7.2.1 The algebraic structure of simple functions* . . . 44
7.3 Simple Integration . . . 45
7.3.1 Appendix: Bonferroni Inequalities . . . 49
7.3.2 Appendix: Riemann Stieljtes integral . . . 50
7.4 Simple Independence and the Weak Law of Large Numbers . . . 51
7.4.1 Complex Weierstrass Approximation Theorem . . . 54
7.4.2 Product Measures and Fubini's Theorem . . . 56
7.5 Appendix: A Multi-dimensional Weirstrass Approximation Theorem . . . 57

8 Countably Additive Measures . . . 59
8.1 Overview . . . 59
8.2 $\pi - \lambda$ Theorem . . . 60
8.2.1 A Density Result* . . . 62
8.3 Construction of Measures . . . 62
8.4 Radon Measures on $\mathbb{R}$ . . . 66
8.4.1 Lebesgue Measure . . . 68
8.5 A Discrete Kolmogorov's Extension Theorem . . . 68
8.6 Appendix: Regularity and Uniqueness Results* . . . 71
8.7 Appendix: Completions of Measure Spaces* . . . 72
8.8 Appendix Monotone Class Theorems* . . . 72

9 Measurable Functions (Random Variables) . . . 75
9.1 Measurable Functions . . . 75
9.2 Factoring Random Variables . . . 81
9.3 Summary of Measurability Statements . . . 82
9.4 Distributions / Laws of Random Vectors . . . 83
9.5 Generating All Distributions from the Uniform Distribution . . . 84

10 Integration Theory . . . 87
10.1 Integrals of positive functions . . . 87
10.2 Integrals of Complex Valued Functions . . . 91
10.2.1 Square Integrable Random Variables and Correlations . . . 96
10.2.2 Some Discrete Distributions . . . 97
10.3 Integration on $\mathbb{R}$ . . . 98
10.4 Densities and Change of Variables Theorems . . . 102
10.5 Normal (Gaussian) Random Variables . . . 103

10.6 Stirling's Formula . . . 105
10.6.1 Two applications of Stirling's formula . . . 108
10.6.2 A primitive Stirling type approximation . . . 110
10.7 Comparison of the Lebesgue and the Riemann Integral* . . . 111
10.8 Measurability on Complete Measure Spaces* . . . 113
10.9 More Exercises . . . 114

11 Functional Forms of the $\pi - \lambda$ Theorem . . . 117
11.1 Multiplicative System Theorems . . . 117
11.2 Exercises . . . 121
11.3 A Strengthening of the Multiplicative System Theorem* . . . 122
11.4 The Bounded Approximation Theorem* . . . 123
11.5 Exercises . . . 124
11.6 σ – Function Algebras (Older Version of above notes from 280 Notes!!) . . . 125
11.6.1 Another (Better) Multiplicative System Theorem . . . 128
11.7 Exercises . . . 130

12 Multiple and Iterated Integrals . . . 131
12.1 Iterated Integrals . . . 131
12.2 Tonelli's Theorem and Product Measure . . . 131
12.3 Fubini's Theorem . . . 133
12.4 Fubini's Theorem and Completions* . . . 136
12.5 Exercises . . . 137

Part III Metric, Banach, and Hilbert Space Basics

13 Metric Spaces . . . 141
13.1 Metric spaces as topological spaces . . . 142
13.1.1 Continuity . . . 143
13.2 Completeness in Metric Spaces . . . 144
13.3 Supplementary Remarks . . . 145
13.3.1 Word of Caution . . . 145
13.3.2 Riemannian Metrics . . . 145
13.4 Exercises . . . 146

14 Banach Spaces . . . 149
14.1 Bounded Linear Operators Basics . . . 151
14.2 General Sums in Banach Spaces . . . 156
14.3 Inverting Elements in $L(X)$ . . . 157
14.4 Exercises . . . 158

15 Hölder Spaces as Banach Spaces ... 161
15.1 Exercises ... 163

16 L^p-Space Basics ... 165
16.1 $L^p(\mu)$ is a Banach Space ... 165
16.2 Density Results ... 169
16.3 Convergence in Measure ... 171

17 Bochner Integral ... 175
17.1 Banach Valued L^p – Spaces ... 175
17.2 Bochner Integral ... 177

18 Hilbert Space Basics ... 181
18.1 Hilbert Space Basis ... 187
18.2 L^2-Orthonormal Basis Examples ... 191
18.3 Fourier Series ... 192
18.3.1 Fourier Series Exercises ... 193
18.4 Exercises ... 195
18.5 Supplement 1: Converse of the Parallelogram Law ... 196
18.6 Supplement 2. Non-complete inner product spaces ... 197

19 Compactness in Metric Space ... 199
19.1 General Topological Notions ... 199
19.2 Compactness in Metric Spaces ... 200
19.3 Locally Compact Banach Spaces ... 202
19.4 Compact Operators ... 204
19.4.1 Compact Operators on a Hilbert Space ... 205
19.5 Hilbert Schmidt Operators ... 206
19.6 Weak Convergence and Compactness in Hilbert Spaces ... 208
19.7 Arzela-Ascoli Theorem (Compactness in $C(X)$) ... 210
19.8 Exercises ... 212
19.8.1 Metric Spaces as Topological Spaces ... 212
19.8.2 Arzela-Ascoli Theorem Problems ... 213

20 Some Spectral Theory ... 215
20.1 The Spectral Theorem for Self Adjoint Compact Operators ... 218

Part IV More on Banach Spaces

21 The Hahn-Banach Theorem 223
21.0.1 Hahn – Banach Theorem Problems 228
21.0.2 *Quotient spaces, adjoints, and more reflexivity 229
21.0.3 Hahn-Banach Theorem Problems 231

22 Baire Category Theorem 233
22.1 Metric Space Baire Category Theorem 233
22.2 Locally Compact Hausdorff Space Baire Category Theorem 234
22.3 Baire Category Theorem Exercises 237

23 The Baire Category Theorem applied to Banach Sapces 239
23.1 The Open Mapping Theorem 239
23.2 Closed Graph Theorem 239
23.3 Uniform Boundedness Principle 241
23.3.1 Applications to Fourier Series 243
23.4 Exercises 244
23.4.1 More Examples of Banach Spaces 244
23.4.2 Open Mapping and Closed Operator Problems 244
23.4.3 Weak Topology and Convergence Problems 245

Part V The Structure of Measures

24 Complex Measures, Radon-Nikodym Theorem 249
24.1 The Radon-Nikodym Theorem 249
24.2 The Structure of Signed Measures 252
24.2.1 Hahn Decomposition Theorem 252
24.2.2 Jordan Decomposition 253
24.3 Complex Measures 255
24.4 Absolute Continuity on an Algebra 257
24.5 Exercises 258

25 Lebesgue Differentiation and the Fundamental Theorem of Calculus 261
25.1 A Covering Lemma and Averaging Operators 261
25.2 Maximal Functions 262
25.3 Lebesque Set 263
25.4 The Fundamental Theorem of Calculus 265
25.4.1 Increasing Functions 265
25.4.2 Functions of Bounded Variation 267
25.4.3 Alternative method to Proving Theorem 25.29 273
25.5 The connection of Weak and pointwise derivatives 274
25.6 Exercises 277
25.7 Summary of B.V. and A.C. Functions 278

25.7.1 Absolute Continuity . . . 279

26 Geometric Integration . . . 283
26.1 Surface Integrals . . . 284
26.2 More spherical coordinates . . . 289
26.3 n – dimensional manifolds with boundaries . . . 291
26.4 Divergence Theorem . . . 293
26.5 The proof of the Divergence Theorem . . . 294
26.5.1 The Proof of the Divergence Theorem 26.23 . . . 296
26.5.2 Extensions of the Divergence Theorem to Lipschitz domains . . . 297
26.6 Application to Holomorphic functions . . . 297
26.7 Dirichlet Problems on D . . . 299
26.7.1 Appendix: More Proofs of Proposition 26.32 . . . 301
26.8 Exercises . . . 302

27 Sard's Theorem . . . 305

Part VI More on L^p – Spaces

28 More Inequalities . . . 311
28.1 Jensen's Inequality . . . 311
28.2 Interpolation of L^p – spaces . . . 313
28.2.1 Summary: . . . 315
28.3 Exercises . . . 315

29 The Dual of L^p . . . 317
29.1 Converse of Hölder's Inequality . . . 317

30 Uniform Integrability . . . 323
30.1 Exercises . . . 327

31 Approximation Theorems and Convolutions . . . 329
31.1 Density Theorems . . . 329
31.2 Convolution and Young's Inequalities . . . 332
31.2.1 Smooth Partitions of Unity . . . 340
31.3 Classical Weierstrass Approximation Theorem . . . 340
31.4 Exercises . . . 341
31.5 Consider putting Chapter 40 here. . . . 343

Part VII Calculus and Ordinary Differential Equations in Banach Spaces

32 The Riemann Integral . . . 347
32.1 The Fundamental Theorem of Calculus . . . 349
32.2 Integral Operators as Examples of Bounded Operators . . . 351
32.3 Linear Ordinary Differential Equations . . . 352
32.4 Operator Logarithms . . . 354
32.5 Classical Weierstrass Approximation Theorem . . . 356
32.6 Iterated Integrals . . . 359
32.7 Exercises . . . 360

33 Ordinary Differential Equations in a Banach Space . . . 363
33.1 Examples . . . 363
33.2 Uniqueness Theorem and Continuous Dependence on Initial Data . . . 364
33.3 Local Existence (Non-Linear ODE) . . . 365
33.4 Global Properties . . . 367
33.5 Semi-Group Properties of time independent flows . . . 370
33.6 Exercises . . . 371

34 Banach Space Calculus . . . 373
34.1 The Differential . . . 373
34.2 Product and Chain Rules . . . 374
34.3 Partial Derivatives . . . 376
34.4 Higher Order Derivatives . . . 377
34.5 Inverse and Implicit Function Theorems . . . 379
34.6 Smooth Dependence of ODE's on Initial Conditions* . . . 383
34.7 Existence of Periodic Solutions . . . 385
34.8 Contraction Mapping Principle . . . 386
34.9 Exercises . . . 387
34.9.1 Alternate construction of g. To be made into an exercise. . . . 388

Part VIII Topological Spaces

35 Topological Space Basics . . . 391
35.1 Constructing Topologies and Checking Continuity . . . 392
35.2 Product Spaces I . . . 394
35.3 Closure operations . . . 396
35.4 Countability Axioms . . . 397
35.5 Connectedness . . . 398
35.6 Compactness . . . 401
35.7 Exercises . . . 405
35.7.1 General Topological Space Problems . . . 405
35.7.2 Connectedness Problems . . . 406
35.7.3 Metric Spaces as Topological Spaces . . . 406

35.7.4 Compactness Problems ... 407

36 Compactness ... 409
36.1 Local and σ − Compactness ... 409
36.2 Function Space Compactness Criteria ... 410
36.3 Tychonoff's Theorem ... 412
36.4 Banach – Alaoglu's Theorem ... 415
36.4.1 Weak and Strong Topologies ... 415
36.5 Weak Convergence in Hilbert Spaces ... 417
36.6 Exercises ... 419
36.6.1 Weak Convergence Problems ... 419
36.6.2 Arzela-Ascoli Theorem Problems ... 419

37 Locally Compact Hausdorff Spaces ... 421
37.1 Locally compact form of Urysohn's Metrization Theorem ... 423
37.2 Partitions of Unity ... 425
37.3 $C_0(X)$ and the Alexanderov Compactification ... 427
37.4 Stone-Weierstrass Theorem ... 429
37.5 *More on Separation Axioms: Normal Spaces ... 431
37.6 Exercises ... 433

38 Examples of Measures ... 435
38.1 The Riesz-Markov Theorem ... 435
38.2 Proof of the Riesz-Markov Theorem 38.9 ... 437
38.2.1 Rudin's Proof of the Riesz-Markov Theorem 38.9 ... 439
38.2.2 Regularity Results For Radon Measures ... 441
38.2.3 The dual of $C_0(X)$ ... 443
38.3 Classifying Radon Measures on $\mathbb{R}$ ... 445
38.3.1 Classifying Radon Measures on $\mathbb{R}$ using Theorem **??** ... 445
38.3.2 Classifying Radon Measures on $\mathbb{R}$ using the Riesz-Markov Theorem 38.9 ... 445
38.3.3 The Lebesgue-Stieljtes Integral ... 445
38.4 Kolmogorov's Existence of Measure on Products Spaces ... 446

39 Probability Measures on Lusin Spaces ... 449
39.1 Weak Convergence Results ... 449
39.2 Haar Measures ... 451
39.3 Hausdorff Measure ... 451
39.4 Exercises ... 451
39.4.1 The Laws of Large Number Exercises ... 452

Part IX Further Hilbert and Banach Space Techniques

40 $\mathbf{L^2}$ - Hilbert Spaces Techniques and Fourier Series 455
40.1 Fourier Series Considerations 455
40.1.1 Dirichlet, Fejér and Kernels 455
40.1.2 The Dirichlet Problems on D and the Poisson Kernel 458
40.2 Weak L^2-Derivatives 459
40.3 *Conditional Expectation 460
40.4 Exercises 462
40.5 Conditional Expectation Exercises 462

41 Weak and Strong Derivatives 463
41.1 Basic Definitions and Properties 463
41.2 Exercises 471

42 Conditional Expectation 473
42.1 Basic Martingale Results 475

Part X The Fourier Transform and Distributions

43 Fourier Transform 481
43.1 Motivation 481
43.2 Fourier Transform formal development 482
43.3 Schwartz Test Functions 483
43.4 Fourier Inversion Formula 485
43.5 Summary of Basic Properties of $\mathcal{F}$ and $\mathcal{F}^{-1}$ 487
43.6 Fourier Transforms of Measures and Bochner's Theorem 487
43.7 Supplement: Heisenberg Uncertainty Principle 490
43.7.1 Exercises 491
43.7.2 More Proofs of the Fourier Inversion Theorem 491
43.7.3 Disregard this subsection on old variants of Corollary 43.17 493

44 Constant Coefficient partial differential equations 495
44.1 Elliptic examples 495
44.2 Poisson Semi-Group 496
44.3 Heat Equation on $\mathbb{R}^n$ 497
44.4 Wave Equation on $\mathbb{R}^n$ 499
44.5 Elliptic Regularity 503

45 Elementary Generalized Functions / Distribution Theory 507
45.1 Distributions on $U \subset_o \mathbb{R}^n$ 507
45.2 Examples of distributions and related computations 507
45.3 Other classes of test functions 511
45.4 Compactly supported distributions 514

45.5 Tempered Distributions and the Fourier Transform 515
45.6 Wave Equation 519
45.7 Appendix: Topology on $C_c^\infty(U)$ 522

46 Convolutions involving distributions 525
46.1 Tensor Product of Distributions 525
46.2 Elliptic Regularity 530
46.3 Appendix: Old Proof of Theorem 46.4 532

References 535

Index 537

Part I

Prequel

1

Introduction / User Guide

Not written as of yet. Topics to mention.

1. A better and more general integral.
   a) Convergence Theorems
   b) Integration over diverse collection of sets. (See probability theory.)
   c) Integration relative to different weights or densities including singular weights.
   d) Characterization of dual spaces.
   e) Completeness.
2. Infinite dimensional Linear algebra.
3. ODE and PDE.
4. Harmonic and Fourier Analysis.
5. Probability Theory

Set Operations

Let $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be the non-negative integers and $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$ – the positive and negative integers including 0, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers (see Chapter 3 below), and $\mathbb{C}$ the complex numbers. We will also use $\mathbb{F}$ to stand for either of the fields $\mathbb{R}$ or $\mathbb{C}$.

Notation 2.1 *Given two sets X and Y, let Y^X denote the collection of all functions $f : X \to Y$. If $X = \mathbb{N}$, we will say that $f \in Y^{\mathbb{N}}$ is a sequence with values in Y and often write f_n for $f(n)$ and express f as $\{f_n\}_{n=1}^{\infty}$. If $X = \{1, 2, \ldots, N\}$, we will write Y^N in place of $Y^{\{1,2,\ldots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \ldots, f_N)$ where $f_n = f(n)$.*

Notation 2.2 *More generally if $\{X_\alpha : \alpha \in A\}$ is a collection of non-empty sets, let $X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map defined by $\pi_\alpha(x) = x_\alpha$. If If $X_\alpha = X$ for some fixed space X, then we will write $\prod_{\alpha \in A} X_\alpha$ as X^A rather than X_A.*

Recall that an element $x \in X_A$ is a "**choice function**," i.e. an assignment $x_\alpha := x(\alpha) \in X_\alpha$ for each $\alpha \in A$. The **axiom of choice** (see Appendix **??**.) states that $X_A \neq \emptyset$ provided that $X_\alpha \neq \emptyset$ for each $\alpha \in A$.

Notation 2.3 *Given a set X, let 2^X denote the **power set** of X – the collection of all subsets of X including the empty set.*

The reason for writing the power set of X as 2^X is that if we think of 2 meaning $\{0, 1\}$, then an element of $a \in 2^X = \{0, 1\}^X$ is completely determined by the set

$$A := \{x \in X : a(x) = 1\} \subset X.$$

In this way elements in $\{0, 1\}^X$ are in one to one correspondence with subsets of X.

For $A \in 2^X$ let

$$A^c := X \setminus A = \{x \in X : x \notin A\}$$

and more generally if $A, B \subset X$ let

$$B \setminus A := \{x \in B : x \notin A\} = A \cap B^c.$$

We also define the symmetric difference of A and B by

$$A \triangle B := (B \setminus A) \cup (A \setminus B).$$

As usual if $\{A_\alpha\}_{\alpha \in I}$ is an indexed collection of subsets of X we define the union and the intersection of this collection by

$$\cup_{\alpha \in I} A_\alpha := \{x \in X : \exists\, \alpha \in I \ni x \in A_\alpha\} \text{ and}$$
$$\cap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \,\forall\, \alpha \in I\}.$$

Notation 2.4 *We will also write $\coprod_{\alpha \in I} A_\alpha$ for $\cup_{\alpha \in I} A_\alpha$ in the case that $\{A_\alpha\}_{\alpha \in I}$ are pairwise disjoint, i.e. $A_\alpha \cap A_\beta = \emptyset$ if $\alpha \neq \beta$.*

Notice that $\cup$ is closely related to $\exists$ and $\cap$ is closely related to $\forall$. For example let $\{A_n\}_{n=1}^{\infty}$ be a sequence of subsets from X and define

$$\{A_n \text{ i.o.}\} := \{x \in X : \#\{n : x \in A_n\} = \infty\} \text{ and}$$
$$\{A_n \text{ a.a.}\} := \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\}.$$

(One should read $\{A_n \text{ i.o.}\}$ as A_n infinitely often and $\{A_n \text{ a.a.}\}$ as A_n almost always.) Then $x \in \{A_n \text{ i.o.}\}$ iff

$$\forall N \in \mathbb{N}\, \exists\, n \geq N \ni x \in A_n$$

and this may be expressed as

$$\{A_n \text{ i.o.}\} = \cap_{N=1}^{\infty} \cup_{n \geq N} A_n.$$

Similarly, $x \in \{A_n \text{ a.a.}\}$ iff

$$\exists\, N \in \mathbb{N} \ni \forall\, n \geq N, \;\; x \in A_n$$

which may be written as

$$\{A_n \text{ a.a.}\} = \cup_{N=1}^{\infty} \cap_{n \geq N} A_n.$$

Definition 2.5. *A set X is said to be **countable** if is empty or there is an injective function $f : X \to \mathbb{N}$, otherwise X is said to be **uncountable.***

Lemma 2.6 (Basic Properties of Countable Sets).

1. *If $A \subset X$ is a subset of a countable set X then A is countable.*
2. *Any infinite subset $\Lambda \subset \mathbb{N}$ is in one to one correspondence with $\mathbb{N}$.*
3. *A non-empty set X is countable iff there exists a surjective map, $g : \mathbb{N} \to X$.*
4. *If X and Y are countable then $X \times Y$ is countable.*
5. *Suppose for each $m \in \mathbb{N}$ that A_m is a countable subset of a set X, then $A = \cup_{m=1}^{\infty} A_m$ is countable. In short, the countable union of countable sets is still countable.*
6. *If X is an infinite set and Y is a set with at least two elements, then Y^X is uncountable. In particular 2^X is uncountable for any infinite set X.*

Proof. 1. If $f : X \to N$ is an injective map then so is the restriction, $f|_A$, of f to the subset A. 2. Let $f(1) = \min \Lambda$ and define f inductively by

$$f(n+1) = \min \left(\Lambda \setminus \{f(1), \dots, f(n)\}\right).$$

Since Λ is infinite the process continues indefinitely. The function $f : \mathbb{N} \to \Lambda$ defined this way is a bijection.

3. If $g : \mathbb{N} \to X$ is a surjective map, let

$$f(x) = \min g^{-1}(\{x\}) = \min \{n \in \mathbb{N} : f(n) = x\}.$$

Then $f : X \to \mathbb{N}$ is injective which combined with item

2. (taking $\Lambda = f(X)$) shows X is countable. Conversely if $f : X \to \mathbb{N}$ is injective let $x_0 \in X$ be a fixed point and define $g : \mathbb{N} \to X$ by $g(n) = f^{-1}(n)$ for $n \in f(X)$ and $g(n) = x_0$ otherwise.

4. Let us first construct a bijection, h, from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$. To do this put the elements of $\mathbb{N} \times \mathbb{N}$ into an array of the form

$$\begin{pmatrix} (1,1) & (1,2) & (1,3) & \dots \\ (2,1) & (2,2) & (2,3) & \dots \\ (3,1) & (3,2) & (3,3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and then "count" these elements by counting the sets $\{(i,j) : i + j = k\}$ one at a time. For example let $h(1) = (1,1)$, $h(2) = (2,1)$, $h(3) = (1,2)$, $h(4) = (3,1)$, $h(5) = (2,2)$, $h(6) = (1,3)$ and so on. If $f : \mathbb{N} \to X$ and $g : \mathbb{N} \to Y$ are surjective functions, then the function $(f \times g) \circ h : \mathbb{N} \to X \times Y$ is surjective where $(f \times g)(m,n) := (f(m), g(n))$ for all $(m,n) \in \mathbb{N} \times \mathbb{N}$.

5. If $A = \emptyset$ then A is countable by definition so we may assume $A \neq \emptyset$. With out loss of generality we may assume $A_1 \neq \emptyset$ and by replacing A_m by A_1 if necessary we may also assume $A_m \neq \emptyset$ for all m. For each $m \in \mathbb{N}$ let $a_m : \mathbb{N} \to A_m$ be a surjective function and then define $f : \mathbb{N} \times \mathbb{N} \to \cup_{m=1}^{\infty} A_m$ by $f(m,n) := a_m(n)$. The function f is surjective and hence so is the composition, $f \circ h : \mathbb{N} \to \cup_{m=1}^{\infty} A_m$, where $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is the bijection defined above.

6. Let us begin by showing $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$ is uncountable. For sake of contradiction suppose $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is a surjection and write $f(n)$ as $(f_1(n), f_2(n), f_3(n), \dots)$. Now define $a \in \{0,1\}^{\mathbb{N}}$ by $a_n := 1 - f_n(n)$. By construction $f_n(n) \neq a_n$ for all n and so $a \notin f(\mathbb{N})$. This contradicts the assumption that f is surjective and shows $2^{\mathbb{N}}$ is uncountable. For the general case, since $Y_0^X \subset Y^X$ for any subset $Y_0 \subset Y$, if Y_0^X is uncountable then so is Y^X. In this way we may assume Y_0 is a two point set which may as well be $Y_0 = \{0,1\}$. Moreover, since X is an infinite set we may find an injective map $x : \mathbb{N} \to X$ and use this to set up an injection, $i : 2^{\mathbb{N}} \to 2^X$ by setting $i(A) := \{x_n : n \in \mathbb{N}\} \subset X$ for all $A \subset \mathbb{N}$. If 2^X were countable we could find a surjective map $f : 2^X \to \mathbb{N}$ in which case $f \circ i : 2^{\mathbb{N}} \to \mathbb{N}$ would be surjective as well. However this is impossible since we have already seed that $2^{\mathbb{N}}$ is uncountable. ∎

We end this section with some notation which will be used frequently in the sequel.

Notation 2.7 *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E) | E \in \mathcal{E}\}.$$

If $\mathcal{G} \subset 2^X$, let

$$f_*\mathcal{G} := \{A \in 2^Y | f^{-1}(A) \in \mathcal{G}\}.$$

Definition 2.8. *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the* ***inclusion map*** *($i_A(x) = x$ for all $x \in A$) and*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

2.1 Exercises

Let $f : X \to Y$ be a function and $\{A_i\}_{i \in I}$ be an indexed family of subsets of Y, verify the following assertions.

Exercise 2.1. $(\cap_{i \in I} A_i)^c = \cup_{i \in I} A_i^c$.

Exercise 2.2. Suppose that $B \subset Y$, show that $B \setminus (\cup_{i \in I} A_i) = \cap_{i \in I}(B \setminus A_i)$.

Exercise 2.3. $f^{-1}(\cup_{i \in I} A_i) = \cup_{i \in I} f^{-1}(A_i)$.

Exercise 2.4. $f^{-1}(\cap_{i \in I} A_i) = \cap_{i \in I} f^{-1}(A_i)$.

Exercise 2.5. Find a counterexample which shows that $f(C \cap D) = f(C) \cap f(D)$ need not hold.

2.2 Appendix: Zorn's Lemma and the Hausdorff Maximal Principle

Definition 2.9. *A partial order $\leq$ on X is a relation with following properties;*

1. *If $x \leq y$ and $y \leq z$ then $x \leq z$.*
2. *If $x \leq y$ and $y \leq x$ then $x = y$.*
3. *$x \leq x$ for all $x \in X$.*

Example 2.10. Let Y be a set and $X = 2^Y$. There are two natural partial orders on X.

1. Ordered by inclusion, $A \leq B$ is $A \subset B$ and
2. Ordered by reverse inclusion, $A \leq B$ if $B \subset A$.

Definition 2.11. *Let $(X, \leq)$ be a partially ordered set we say X is **linearly** or **totally** ordered if for all $x, y \in X$ either $x \leq y$ or $y \leq x$. The real numbers $\mathbb{R}$ with the usual order $\leq$ is a typical example.*

Definition 2.12. *Let $(X, \leq)$ be a partial ordered set. We say $x \in X$ is a **maximal** element if for all $y \in X$ such that $y \geq x$ implies $y = x$, i.e. there is no element larger than x. An **upper bound** for a subset E of X is an element $x \in X$ such that $x \geq y$ for all $y \in E$.*

Example 2.13. Let

$$X = \left\{ a = \{1\}\ b = \{1,2\}\ c = \{3\}\ d = \{2,4\}\ e = \{2\} \right\}$$

ordered by set inclusion. Then b and d are maximal elements despite that fact that $b \nleq d$ and $d \nleq b$. We also have;

1. If $E = \{a, c, e\}$, then E has **no** upper bound.
2. If $E = \{a, e\}$, then b is an upper bound.
3. If $E = \{e\}$, then b and d are upper bounds.

Theorem 2.14. *The following are equivalent.*

1. ***The axiom of choice:** to each collection, $\{X_\alpha\}_{\alpha \in A}$, of non-empty sets there exists a "choice function," $x : A \to \coprod_{\alpha \in A} X_\alpha$ such that $x(\alpha) \in X_\alpha$ for all $\alpha \in A$, i.e. $\prod_{\alpha \in A} X_\alpha \neq \emptyset$.*
2. ***The Hausdorff Maximal Principle**: Every partially ordered set has a **maximal** (relative to the inclusion order) linearly ordered subset.*

3. ***Zorn's Lemma**: If X is partially ordered set such that every linearly ordered subset of X has an upper bound, then X has a maximal element.*[1]

Proof. ($2 \Rightarrow 3$) Let X be a partially ordered subset as in 3 and let $\mathcal{F} = \{E \subset X : E \text{ is linearly ordered}\}$ which we equip with the inclusion partial ordering. By 2. there exist a maximal element $E \in \mathcal{F}$. By assumption, the linearly ordered set E has an upper bound $x \in X$. The element x is maximal, for if $y \in Y$ and $y \geq x$, then $E \cup \{y\}$ is still an linearly ordered set containing E. So by maximality of E, $E = E \cup \{y\}$, i.e. $y \in E$ and therefore $y \leq x$ showing which combined with $y \geq x$ implies that $y = x$.[2]

($3 \Rightarrow 1$) Let $\{X_\alpha\}_{\alpha \in A}$ be a collection of non-empty sets, we must show $\prod_{\alpha \in A} X_\alpha$ is not empty. Let $\mathcal{G}$ denote the collection of functions $g : D(g) \to \coprod_{\alpha \in A} X_\alpha$ such that $D(g)$ is a subset of A, and for all $\alpha \in D(g)$, $g(\alpha) \in X_\alpha$. Notice that $\mathcal{G}$ is not empty, for we may let $\alpha_0 \in A$ and $x_0 \in X_\alpha$ and then set $D(g) = \{\alpha_0\}$ and $g(\alpha_0) = x_0$ to construct an element of $\mathcal{G}$. We now put a partial order on $\mathcal{G}$ as follows. We say that $f \leq g$ for $f, g \in \mathcal{G}$ provided that $D(f) \subset D(g)$ and $f = g|_{D(f)}$. If $\Phi \subset \mathcal{G}$ is a linearly ordered set, let $D(h) = \cup_{g \in \Phi} D(g)$ and for $\alpha \in D(g)$ let $h(\alpha) = g(\alpha)$. Then $h \in \mathcal{G}$ is an upper bound for Φ. So by Zorn's Lemma there exists a maximal element $h \in \mathcal{G}$. To finish the proof we need only show that $D(h) = A$. If this were not the case, then let $\alpha_0 \in A \setminus D(h)$ and $x_0 \in X_{\alpha_0}$. We may now define $D(\tilde{h}) = D(h) \cup \{\alpha_0\}$ and

$$\tilde{h}(\alpha) = \begin{cases} h(\alpha) & \text{if } \alpha \in D(h) \\ x_0 & \text{if } \alpha = \alpha_0. \end{cases}$$

Then $h \leq \tilde{h}$ while $h \neq \tilde{h}$ violating the fact that h was a maximal element.

($1 \Rightarrow 2$) Let $(X, \leq)$ be a partially ordered set. Let $\mathcal{F}$ be the collection of linearly ordered subsets of X which we order by set inclusion. Given $x_0 \in X$,

[1] If X is a countable set we may prove Zorn's Lemma by induction. Let $\{x_n\}_{n=1}^\infty$ be an enumeration of X, and define $E_n \subset X$ inductively as follows. For $n = 1$ let $E_1 = \{x_1\}$, and if E_n have been chosen, let $E_{n+1} = E_n \cup \{x_{n+1}\}$ if x_{n+1} is an upper bound for E_n otherwise let $E_{n+1} = E_n$. The set $E = \cup_{n=1}^\infty E_n$ is a linearly ordered (you check) subset of X and hence by assumption E has an upper bound, $x \in X$. I claim that his element is maximal, for if there exists $y = x_m \in X$ such that $y \geq x$, then x_m would be an upper bound for E_{m-1} and therefore $y = x_m \in E_m \subset E$. That is to say if $y \geq x$, then $y \in E$ and hence $y \leq x$, so $y = x$. (Hence we may view Zorn's lemma as a " jazzed" up version of induction.)

[2] Similarly one may show that $3 \Rightarrow 2$. Let $\mathcal{F} = \{E \subset X : E \text{ is linearly ordered}\}$ and order $\mathcal{F}$ by inclusion. If $\mathcal{M} \subset \mathcal{F}$ is linearly ordered, let $E = \cup\mathcal{M} = \bigcup_{A \in \mathcal{M}} A$. If $x, y \in E$ then $x \in A$ and $y \in B$ for some $A, B \subset \mathcal{M}$. Now $\mathcal{M}$ is linearly ordered by set inclusion so $A \subset B$ or $B \subset A$ i.e. $x, y \in A$ or $x, y \in B$. Since A and B are linearly order we must have either $x \leq y$ or $y \leq x$, that is to say E is linearly ordered. Hence by 3. there exists a maximal element $E \in \mathcal{F}$ which is the assertion in 2.

$\{x_0\} \in \mathcal{F}$ is linearly ordered set so that $\mathcal{F} \neq \emptyset$. Fix an element $P_0 \in \mathcal{F}$. If P_0 is not maximal there exists $P_1 \in \mathcal{F}$ such that $P_0 \subsetneq P_1$. In particular we may choose $x \notin P_0$ such that $P_0 \cup \{x\} \in \mathcal{F}$. The idea now is to keep repeating this process of adding points $x \in X$ until we construct a maximal element P of $\mathcal{F}$. We now have to take care of some details. We may assume with out loss of generality that $\tilde{\mathcal{F}} = \{P \in \mathcal{F} : P \text{ is not maximal}\}$ is a non-empty set. For $P \in \tilde{\mathcal{F}}$, let $P^* = \{x \in X : P \cup \{x\} \in \mathcal{F}\}$. As the above argument shows, $P^* \neq \emptyset$ for all $P \in \tilde{\mathcal{F}}$. Using the axiom of choice, there exists $f \in \prod_{P \in \tilde{\mathcal{F}}} P^*$. We now define $g : \mathcal{F} \to \mathcal{F}$ by

$$g(P) = \begin{cases} P & \text{if } \quad P \text{ is maximal} \\ P \cup \{f(x)\} & \text{if } P \text{ is not maximal.} \end{cases} \tag{2.1}$$

The proof is completed by Lemma 2.15 below which shows that g must have a fixed point $P \in \mathcal{F}$. This fixed point is maximal by construction of g. ■

Lemma 2.15. *The function $g : \mathcal{F} \to \mathcal{F}$ defined in Eq. (2.1) has a fixed point.*[3]

Proof. The **idea of the proof** is as follows. Let $P_0 \in \mathcal{F}$ be chosen arbitrarily. Notice that $\Phi = \left\{g^{(n)}(P_0)\right\}_{n=0}^{\infty} \subset \mathcal{F}$ is a linearly ordered set and it is therefore easily verified that $P_1 = \bigcup_{n=0}^{\infty} g^{(n)}(P_0) \in \mathcal{F}$. Similarly we may repeat the process to construct $P_2 = \bigcup_{n=0}^{\infty} g^{(n)}(P_1) \in \mathcal{F}$ and $P_3 = \bigcup_{n=0}^{\infty} g^{(n)}(P_2) \in \mathcal{F}$, etc. etc. Then take $P_\infty = \cup_{n=0}^{\infty} P_n$ and start again with P_0 replaced by P_∞. Then keep going this way until eventually the sets stop increasing in size, in which case we have found our fixed point. The problem with this strategy is that we may never win. (This is very reminiscent of constructing measurable sets and the way out is to use measure theoretic like arguments.)

Let us now start the **formal proof.** Again let $P_0 \in \mathcal{F}$ and let $\mathcal{F}_1 = \{P \in \mathcal{F} : P_0 \subset P\}$. Notice that $\mathcal{F}_1$ has the following properties:

1. $P_0 \in \mathcal{F}_1$.
2. If $\Phi \subset \mathcal{F}_1$ is a totally ordered (by set inclusion) subset then $\cup \Phi \in \mathcal{F}_1$.
3. If $P \in \mathcal{F}_1$ then $g(P) \in \mathcal{F}_1$.

Let us call a general subset $\mathcal{F}' \subset \mathcal{F}$ satisfying these three conditions a tower and let

$$\mathcal{F}_0 = \cap \{\mathcal{F}' : \mathcal{F}' \text{ is a tower}\}.$$

[3] Here is an easy proof if the elements of $\mathcal{F}$ happened to all be finite sets and there existed a set $P \in \mathcal{F}$ with a maximal number of elements. In this case the condition that $P \subset g(P)$ would imply that $P = g(P)$, otherwise $g(P)$ would have more elements than P.

Standard arguments show that $\mathcal{F}_0$ is still a tower and clearly is the smallest tower containing P_0. (Morally speaking $\mathcal{F}_0$ consists of all of the sets we were trying to constructed in the "idea section" of the proof.) We now claim that $\mathcal{F}_0$ is a linearly ordered subset of $\mathcal{F}$. To prove this let $\Gamma \subset \mathcal{F}_0$ be the linearly ordered set

$$\Gamma = \{C \in \mathcal{F}_0 : \text{ for all } A \in \mathcal{F}_0 \text{ either } A \subset C \text{ or } C \subset A\}.$$

Shortly we will show that $\Gamma \subset \mathcal{F}_0$ is a tower and hence that $\mathcal{F}_0 = \Gamma$. That is to say $\mathcal{F}_0$ is linearly ordered. Assuming this for the moment let us finish the proof.

Let $P \equiv \cup \mathcal{F}_0$ which is in $\mathcal{F}_0$ by property 2 and is clearly the largest element in $\mathcal{F}_0$. By 3. it now follows that $P \subset g(P) \in \mathcal{F}_0$ and by maximality of P, we have $g(P) = P$, the desired fixed point. So to finish the proof, we must show that Γ is a tower. First off it is clear that $P_0 \in \Gamma$ so in particular Γ is not empty. For each $C \in \Gamma$ let

$$\Phi_C := \{A \in \mathcal{F}_0 : \text{either } A \subset C \text{ or } g(C) \subset A\}.$$

We will begin by showing that $\Phi_C \subset \mathcal{F}_0$ is a tower and therefore that $\Phi_C = \mathcal{F}_0$. 1. $P_0 \in \Phi_C$ since $P_0 \subset C$ for all $C \in \Gamma \subset \mathcal{F}_0$. 2. If $\Phi \subset \Phi_C \subset \mathcal{F}_0$ is totally ordered by set inclusion, then $A_\Phi := \cup \Phi \in \mathcal{F}_0$. We must show $A_\Phi \in \Phi_C$, that is that $A_\Phi \subset C$ or $C \subset A_\Phi$. Now if $A \subset C$ for all $A \in \Phi$, then $A_\Phi \subset C$ and hence $A_\Phi \in \Phi_C$. On the other hand if there is some $A \in \Phi$ such that $g(C) \subset A$ then clearly $g(C) \subset A_\Phi$ and again $A_\Phi \in \Phi_C$. 3. Given $A \in \Phi_C$ we must show $g(A) \in \Phi_C$, i.e. that

$$g(A) \subset C \text{ or } g(C) \subset g(A). \tag{2.2}$$

There are three cases to consider: either $A \subsetneq C$, $A = C$, or $g(C) \subset A$. In the case $A = C$, $g(C) = g(A) \subset g(A)$ and if $g(C) \subset A$ then $g(C) \subset A \subset g(A)$ and Eq. (2.2) holds in either of these cases. So assume that $A \subsetneq C$. Since $C \in \Gamma$, either $g(A) \subset C$ (in which case we are done) or $C \subset g(A)$. Hence we may assume that

$$A \subsetneq C \subset g(A).$$

Now if C were a proper subset of $g(A)$ it would then follow that $g(A) \setminus A$ would consist of at least two points which contradicts the definition of g. Hence we must have $g(A) = C \subset C$ and again Eq. (2.2) holds, so Φ_C is a tower. It is now easy to show Γ is a tower. It is again clear that $P_0 \in \Gamma$ and Property 2. may be checked for Γ in the same way as it was done for Φ_C above. For Property 3., if $C \in \Gamma$ we may use $\Phi_C = \mathcal{F}_0$ to conclude for all $A \in \mathcal{F}_0$, either $A \subset C \subset g(C)$ or $g(C) \subset A$, i.e. $g(C) \in \Gamma$. Thus Γ is a tower and we are done. ■

A Brief Review of Real and Complex Numbers

Although it is assumed that the reader of this book is familiar with the properties of the real numbers, $\mathbb{R}$, nevertheless I feel it is instructive to define them here and sketch the development of their basic properties. It will most certainly be assumed that the reader is familiar with basic algebraic properties of the natural numbers $\mathbb{N}$ and the ordered field of rational numbers,

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z} : n \neq 0 \right\}.$$

As usual, for $q \in \mathbb{Q}$, we define

$$|q| = \begin{cases} q & \text{if } q \geq 0 \\ -q & \text{if } q \leq 0. \end{cases}$$

Notice that if $q \in \mathbb{Q}$ and $|q| \leq n^{-1} := \frac{1}{n}$ for all n, then $q = 0$. Since if $q \neq 0$, then $|q| = \frac{m}{n}$ for some $m, n \in \mathbb{N}$ and hence $|q| \geq \frac{1}{n}$. A similar argument shows $q \geq 0$ iff $q \geq -\frac{1}{n}$ for all $n \in \mathbb{N}$. These trivial remarks will be used in the future without further reference.

Definition 3.1. *A sequence $\{q_n\}_{n=1}^\infty \subset \mathbb{Q}$ **converges** to $q \in \mathbb{Q}$ if $|q - q_n| \to 0$ as $n \to \infty$, i.e. if for all $N \in \mathbb{N}$, $|q - q_n| \leq \frac{1}{N}$ for a.a. n. As usual if $\{q_n\}_{n=1}^\infty$ converges to q we will write $q_n \to q$ as $n \to \infty$ or $q = \lim_{n\to\infty} q_n$.*

Definition 3.2. *A sequence $\{q_n\}_{n=1}^\infty \subset \mathbb{Q}$ is **Cauchy** if $|q_n - q_m| \to 0$ as $m, n \to \infty$. More precisely we require for each $N \in \mathbb{N}$ that $|q_m - q_n| \leq \frac{1}{N}$ for a.a. pairs (m, n).*

Exercise 3.1. Show that all convergent sequences $\{q_n\}_{n=1}^\infty \subset \mathbb{Q}$ are Cauchy and that all Cauchy sequences $\{q_n\}_{n=1}^\infty$ are bounded – i.e. there exists $M \in \mathbb{N}$ such that

$$|q_n| \leq M \text{ for all } n \in \mathbb{N}.$$

Exercise 3.2. Suppose $\{q_n\}_{n=1}^\infty$ and $\{r_n\}_{n=1}^\infty$ are Cauchy sequences in $\mathbb{Q}$.

1. Show $\{q_n + r_n\}_{n=1}^\infty$ and $\{q_n \cdot r_n\}_{n=1}^\infty$ are Cauchy.
   Now assume that $\{q_n\}_{n=1}^\infty$ and $\{r_n\}_{n=1}^\infty$ are convergent sequences in $\mathbb{Q}$.
2. Show $\{q_n + r_n\}_{n=1}^\infty$ $\{q_n \cdot r_n\}_{n=1}^\infty$ are convergent in $\mathbb{Q}$ and
$$\lim_{n\to\infty} (q_n + r_n) = \lim_{n\to\infty} q_n + \lim_{n\to\infty} r_n \text{ and}$$
$$\lim_{n\to\infty} (q_n r_n) = \lim_{n\to\infty} q_n \cdot \lim_{n\to\infty} r_n.$$
3. If we further assume $q_n \leq r_n$ for all n, show $\lim_{n\to\infty} q_n \leq \lim_{n\to\infty} r_n$. (It suffices to consider the case where $q_n = 0$ for all n.)

The rational numbers $\mathbb{Q}$ suffer from the defect that they are not complete, i.e. not all Cauchy sequences are convergent. In fact, according to Corollary 3.14 below, "most" Cauchy sequences of rational numbers do not converge to a rational number.

Exercise 3.3. Use the following outline to construct a Cauchy sequence $\{q_n\}_{n=1}^\infty \subset \mathbb{Q}$ which is **not** convergent in $\mathbb{Q}$.

1. Recall that there is no element $q \in \mathbb{Q}$ such that $q^2 = 2$.[1] To each $n \in \mathbb{N}$ let $m_n \in \mathbb{N}$ be chosen so that
$$\frac{m_n^2}{n^2} < 2 < \frac{(m_n + 1)^2}{n^2} \tag{3.1}$$
   and let $q_n := \frac{m_n}{n}$.
2. Verify that $q_n^2 \to 2$ as $n \to \infty$ and that $\{q_n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{Q}$.
3. Show $\{q_n\}_{n=1}^\infty$ does not have a limit in $\mathbb{Q}$.

3.1 The Real Numbers

Let $\mathcal{C}$ denote the collection of Cauchy sequences $a = \{a_n\}_{n=1}^\infty \subset \mathbb{Q}$ and say $a, b \in \mathcal{C}$ are equivalent (write $a \sim b$) iff $\lim_{n\to\infty} |a_n - b_n| = 0$. (The reader should check that " $\sim$ " is an equivalence relation.)

Definition 3.3. *A **real number** is an equivalence class, $\bar{a} := \{b \in \mathcal{C} : b \sim a\}$ associated to some element $a \in \mathcal{C}$. The collection of real numbers will be denoted by $\mathbb{R}$. For $q \in \mathbb{Q}$, let $i(q) = \bar{a}$ where a is the constant sequence $a_n = q$ for all $n \in \mathbb{N}$. We will simply write 0 for $i(0)$ and 1 for $i(1)$.*

[1] This fact also shows that the intermediate value theorem, (see Theorem 35.50 below.) fails when working with continuous functions defined over $\mathbb{Q}$.

Exercise 3.4. Given $\bar{a}, \bar{b} \in \mathbb{R}$ show that the definitions

$$-\bar{a} = \overline{(-a)},\ \bar{a} + \bar{b} := \overline{(a+b)} \text{ and } \bar{a}\cdot\bar{b} := \overline{a\cdot b}$$

are well defined. Here $-a$, $a+b$ and $a \cdot b$ denote the sequences $\{-a_n\}_{n=1}^\infty$, $\{a_n + b_n\}_{n=1}^\infty$ and $\{a_n \cdot b_n\}_{n=1}^\infty$ respectively. Further verify that with these operations, $\mathbb{R}$ becomes a field and the map $i : \mathbb{Q} \to \mathbb{R}$ is injective homomorphism of fields. **Hint:** if $\bar{a} \neq 0$ show that $\bar{a}$ may be represented by a sequence $a \in \mathcal{C}$ with $|a_n| \geq \frac{1}{N}$ for all n and some $N \in \mathbb{N}$. For this representative show the sequence $a^{-1} := \left\{a_n^{-1}\right\}_{n=1}^\infty \in \mathcal{C}$. The multiplicative inverse to $\bar{a}$ may now be constructed as: $\frac{1}{\bar{a}} = \bar{a}^{-1} := \overline{\left\{a_n^{-1}\right\}_{n=1}^\infty}$.

Definition 3.4. *Let $\bar{a}, \bar{b} \in \mathbb{R}$. Then*

1. *$\bar{a} > 0$ if there exists an $N \in \mathbb{N}$ such that $a_n > \frac{1}{N}$ for a.a. n.*
2. *$\bar{a} \geq 0$ iff either $\bar{a} > 0$ or $\bar{a} = 0$. Equivalently (as the reader should verify), $\bar{a} \geq 0$ iff for all $N \in \mathbb{N}$, $a_n \geq -\frac{1}{N}$ for a.a. n.*
3. *Write $\bar{a} > \bar{b}$ or $\bar{b} < \bar{a}$ if $\bar{a} - \bar{b} > 0$*
4. *Write $\bar{a} \geq \bar{b}$ or $\bar{b} \leq \bar{a}$ if $\bar{a} - \bar{b} \geq 0$.*

Exercise 3.5. Show " $\geq$ " make $\mathbb{R}$ into a linearly ordered field and the map $i : \mathbb{Q} \to \mathbb{R}$ preserves order. Namely if $\bar{a}, \bar{b} \in \mathbb{R}$ then

1. exactly one of the following relations hold: $\bar{a} < \bar{b}$ or $\bar{a} > \bar{b}$ or $\bar{a} = \bar{b}$.
2. If $\bar{a} \geq 0$ and $\bar{b} \geq 0$ then $\bar{a} + \bar{b} \geq 0$ and $\bar{a} \cdot \bar{b} \geq 0$.
3. If $q, r \in Q$ then $q \leq r$ iff $i(q) \leq i(r)$.

The **absolute value** of a real number $\bar{a}$ is defined analogously to that of a rational number by

$$|\bar{a}| = \begin{cases} \bar{a} & \text{if } \bar{a} \geq 0 \\ -\bar{a} & \text{if } \bar{a} < 0 \end{cases}.$$

Observe this definition is consistent with our previous definition of the absolute value on $\mathbb{Q}$, namely $i(|q|) = |i(q)|$. Also notice that $\bar{a} = 0$ (i.e. $a \sim 0$ where 0 denotes the constant sequence of all zeros) iff for all $N \in \mathbb{N}$, $|a_n| \leq \frac{1}{N}$ for a.a. n. This is equivalent to saying $|\bar{a}| \leq i\left(\frac{1}{N}\right)$ for all $N \in \mathbb{N}$ iff $\bar{a} = 0$.

Definition 3.5. *A sequence $\{\bar{a}_n\}_{n=1}^\infty \subset \mathbb{R}$ **converges** to $\bar{a} \in \mathbb{R}$ if $|\bar{a} - \bar{a}_n| \to 0$ as $n \to \infty$, i.e. if for all $N \in \mathbb{N}$, $|\bar{a} - \bar{a}_n| \leq i\left(\frac{1}{N}\right)$ for a.a. n. As before (for rational numbers) if $\{\bar{a}_n\}_{n=1}^\infty$ converges to $\bar{a}$ we will write $\bar{a}_n \to \bar{a}$ as $n \to \infty$ or $\bar{a} = \lim_{n\to\infty} \bar{a}_n$.*

Exercise 3.6. Given $\bar{a}, \bar{b} \in \mathbb{R}$ show

$$\left|\bar{a}\bar{b}\right| = |\bar{a}|\left|\bar{b}\right| \text{ and } \left|\bar{a} + \bar{b}\right| \leq |\bar{a}| + \left|\bar{b}\right|.$$

The latter inequality being referred to as the **triangle inequality**.

By exercise 3.6,

$$|\bar{a}| = \left|\bar{a} - \bar{b} + \bar{b}\right| \leq \left|\bar{a} - \bar{b}\right| + \left|\bar{b}\right|$$

and hence

$$|\bar{a}| - \left|\bar{b}\right| \leq \left|\bar{a} - \bar{b}\right|$$

and by reversing the roles of $\bar{a}$ and $\bar{b}$ we also have

$$-\left(|\bar{a}| - \left|\bar{b}\right|\right) = \left|\bar{b}\right| - |\bar{a}| \leq \left|\bar{b} - \bar{a}\right| = \left|\bar{a} - \bar{b}\right|.$$

Therefore,

$$\left||\bar{a}| - \left|\bar{b}\right|\right| \leq \left|\bar{a} - \bar{b}\right|$$

and consequently if $\{\bar{a}_n\}_{n=1}^\infty \subset \mathbb{R}$ converges to $\bar{a} \in \mathbb{R}$ then

$$||\bar{a}_n| - |\bar{a}|| \leq |\bar{a}_n - \bar{a}| \to 0 \text{ as } n \to \infty.$$

Remark 3.6. The field $i(\mathbb{Q})$ is **dense** in $\mathbb{R}$ in the sense that if $\bar{a} \in \mathbb{R}$ there exists $\{q_n\}_{n=1}^\infty \subset \mathbb{Q}$ such that $i(q_n) \to \bar{a}$ as $n \to \infty$. Indeed, simply let $q_n = a_n$ where a represents $\bar{a}$. Since a is a Cauchy sequence, to any $N \in \mathbb{N}$ there exists $M \in \mathbb{N}$ such that

$$-\frac{1}{N} \leq a_m - a_n \leq \frac{1}{N} \text{ for all } m, n \geq M$$

and therefore

$$-i\left(\frac{1}{N}\right) \leq i(a_m) - \bar{a} \leq i\left(\frac{1}{N}\right) \text{ for all } m \geq M.$$

This shows

$$|i(q_m) - \bar{a}| = |i(a_m) - \bar{a}| \leq i\left(\frac{1}{N}\right) \text{ for all } m \geq M$$

and since N is arbitrary it follows that $i(q_m) \to \bar{a}$ as $m \to \infty$.

Definition 3.7. *A sequence $\{\bar{a}_n\}_{n=1}^\infty \subset \mathbb{R}$ is **Cauchy** if $|\bar{a}_n - \bar{a}_m| \to 0$ as $m, n \to \infty$. More precisely we require for each $N \in \mathbb{N}$ that $|\bar{a}_m - \bar{a}_n| \leq i\left(\frac{1}{N}\right)$ for a.a. pairs (m, n).*

Exercise 3.7. The analogues of the results in Exercises 3.1 and 3.2 hold with $\mathbb{Q}$ replaced by $\mathbb{R}$. (We now say a subset $\Lambda \subset \mathbb{R}$ is bounded if there exists $M \in \mathbb{N}$ such that $|\lambda| \leq i(M)$ for all $\lambda \in \Lambda$.)

For the purposes of real analysis the most important property of $\mathbb{R}$ is that it is "complete."

Theorem 3.8. *The ordered field $\mathbb{R}$ is **complete**, i.e. all Cauchy sequences in $\mathbb{R}$ are convergent.*

Proof. Suppose that $\{\bar{a}(m)\}_{m=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$. By Remark 3.6, we may choose $q_m \in \mathbb{Q}$ such that

$$|\bar{a}(m) - i(q_m)| \le i\left(m^{-1}\right) \text{ for all } m \in \mathbb{N}.$$

Given $N \in \mathbb{N}$, choose $M \in \mathbb{N}$ such that $|\bar{a}(m) - \bar{a}(n)| \le i\left(N^{-1}\right)$ for all $m, n \ge M$. Then

$$\begin{aligned}|i(q_m) - i(q_n)| &\le |i(q_m) - \bar{a}(m)| + |\bar{a}(m) - \bar{a}(n)| + |\bar{a}(n) - i(q_n)| \\ &\le i\left(m^{-1}\right) + i\left(n^{-1}\right) + i\left(N^{-1}\right)\end{aligned}$$

and therefore

$$|q_m - q_n| \le m^{-1} + n^{-1} + N^{-1} \text{ for all } m, n \ge M.$$

It now follows that $q = \{q_m\}_{m=1}^{\infty} \in \mathcal{C}$ and therefore q represents a point $\bar{q} \in \mathbb{R}$. Using Remark 3.6 and the triangle inequality,

$$\begin{aligned}|\bar{a}(m) - \bar{q}| &\le |\bar{a}(m) - i(q_m)| + |i(q_m) - \bar{q}| \\ &\le i\left(m^{-1}\right) + |i(q_m) - \bar{q}| \to 0 \text{ as } m \to \infty\end{aligned}$$

and therefore $\lim_{m\to\infty} \bar{a}(m) = \bar{q}$. ■

Definition 3.9. *A number $M \in \mathbb{R}$ is an* ***upper bound*** *for a set $\Lambda \subset \mathbb{R}$ if $\lambda \le M$ for all $\lambda \in \Lambda$ and a number $m \in \mathbb{R}$ is an* ***lower bound*** *for a set $\Lambda \subset \mathbb{R}$ if $\lambda \ge m$ for all $\lambda \in \Lambda$. Upper and lower bounds need not exist. If Λ has an upper (lower) bound, Λ is said to be* ***bounded from above (below)***.

Theorem 3.10. *To each non-empty set $\Lambda \subset \mathbb{R}$ which is bounded from above (below) there is a unique* ***least upper bound*** *denoted by* $\sup \Lambda \in \mathbb{R}$ *(respectively* ***greatest lower bound*** *denoted by* $\inf \Lambda \in \mathbb{R}$).

Proof. Suppose Λ is bounded from above and for each $n \in \mathbb{N}$, let $m_n \in \mathbb{Z}$ be the smallest integer such that $i\left(\frac{m_n}{2^n}\right)$ is an upper bound for Λ. The sequence $q_n := \frac{m_n}{2^n}$ is Cauchy because $q_m \in [q_n - 2^{-n}, q_n] \cap \mathbb{Q}$ for all $m \ge n$, i.e.

$$|q_m - q_n| \le 2^{-\min(m,n)} \to 0 \text{ as } m, n \to \infty.$$

Passing to the limit, $n \to \infty$, in the inequality $i(q_n) \ge \lambda$, which is valid for all $\lambda \in \Lambda$ implies

$$\bar{q} = \lim_{n\to\infty} i(q_n) \ge \lambda \text{ for all } \lambda \in \Lambda.$$

Thus $\bar{q}$ is an upper bound for Λ. If there were another upper bound $M \in \mathbb{R}$ for Λ such that $M < \bar{q}$, it would follow that $M \le i(q_n) < \bar{q}$ for some n. But this is a contradiction because $\{q_n\}_{n=1}^{\infty}$ is a decreasing sequence, $i(q_n) \ge i(q_m)$ for all $m \ge n$ and therefore $i(q_n) \ge \bar{q}$ for all n. Therefore $\bar{q}$ is the unique least upper bound for Λ. The existence of lower bounds is proved analogously. ■

Proposition 3.11. *If $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ is an increasing (decreasing) sequence which is bounded from above (below), then $\{a_n\}_{n=1}^{\infty}$ is convergent and*

$$\lim_{n\to\infty} a_n = \sup\{a_n : n \in \mathbb{N}\} \quad \left(\lim_{n\to\infty} a_n = \inf\{a_n : n \in \mathbb{N}\}\right).$$

If $\Lambda \subset \mathbb{R}$ is a set bounded from above then there exists $\{\lambda_n\} \subset \Lambda$ such that $\lambda_n \uparrow M := \sup \Lambda$, as $n \to \infty$, i.e. $\{\lambda_n\}$ is increasing and $\lim_{n\to\infty} \lambda_n = M$.

Proof. Let $M := \sup\{a_n : n \in \mathbb{N}\}$, then for each $N \in \mathbb{N}$ there must exist $m \in \mathbb{N}$ such that $M - i\left(N^{-1}\right) < a_m \le M$. Since a_n is increasing, it follows that

$$M - i\left(N^{-1}\right) < a_n \le M \text{ for all } n \ge m.$$

From this we conclude that $\lim a_n$ exists and $\lim a_n = M$. If $M = \sup \Lambda$, for each $n \in \mathbb{N}$ we may choose $\lambda_n \in \Lambda$ such that

$$M - i\left(n^{-1}\right) < \lambda_n \le M. \tag{3.2}$$

By replacing λ_n by $\max\{\lambda_1, \dots, \lambda_n\}$[2] if necessary we may assume that λ_n is increasing in n. It now follows easily from Eq. (3.2) that $\lim_{n\to\infty} \lambda_n = M$. ■

3.1.1 The Decimal Representation of a Real Number

Let $\alpha \in \mathbb{R}$ or $\alpha \in \mathbb{Q}$, $m, n \in \mathbb{Z}$ and $S := \sum_{k=n}^{m} \alpha^k$. If $\alpha = 1$ then $\sum_{k=n}^{m} \alpha^k = m - n + 1$ while for $\alpha \ne 1$,

$$\alpha S - S = \alpha^{m+1} - \alpha^n$$

and solving for S gives the important geometric summation formula,

$$\sum_{k=n}^{m} \alpha^k = \frac{\alpha^{m+1} - \alpha^n}{\alpha - 1} \text{ if } \alpha \ne 1. \tag{3.3}$$

Taking $\alpha = 10^{-1}$ in Eq. (3.3) implies

$$\sum_{k=n}^{m} 10^{-k} = \frac{10^{-(m+1)} - 10^{-n}}{10^{-1} - 1} = \frac{1}{10^{n-1}} \frac{1 - 10^{-(m-n+1)}}{9}$$

and in particular, for all $M \ge n$,

$$\lim_{m\to\infty} \sum_{k=n}^{m} 10^{-k} = \frac{1}{9 \cdot 10^{n-1}} \ge \sum_{k=n}^{M} 10^{-k}.$$

Let $\mathbb{D}$ denote those sequences $\alpha \in \{0, 1, 2, \dots, 9\}^{\mathbb{Z}}$ with the following properties:

[2] The notation, $\max \Lambda$, denotes $\sup \Lambda$ along with the assertion that $\sup \Lambda \in \Lambda$. Similarly, $\min \Lambda = \inf \Lambda$ along with the assertion that $\inf \Lambda \in \Lambda$.

1. there exists $N \in \mathbb{N}$ such that $\alpha_{-n} = 0$ for all $n \geq N$ and
2. $\alpha_n \neq 0$ for some $n \in \mathbb{Z}$.

Associated to each $\alpha \in \mathbb{D}$ is the sequence $a = a(\alpha)$ defined by

$$a_n := \sum_{k=-\infty}^{n} \alpha_k 10^{-k}.$$

Since for $m > n$,

$$|a_m - a_n| = \left| \sum_{k=n+1}^{m} \alpha_k 10^{-k} \right| \leq 9 \sum_{k=n+1}^{m} 10^{-k} \leq 9 \frac{1}{9 \cdot 10^n} = \frac{1}{10^n},$$

it follows that

$$|a_m - a_n| \leq \frac{1}{10^{\min(m,n)}} \to 0 \text{ as } m, n \to \infty.$$

Therefore $a = \underline{a(\alpha)} \in \mathcal{C}$ and we may define a map $D : \{\pm 1\} \times \mathbb{D} \to \mathbb{R}$ defined by $D(\varepsilon, \alpha) = \varepsilon \overline{a(\alpha)}$. As is customary we will denote $D(\varepsilon, \alpha) = \varepsilon \overline{a(\alpha)}$ as

$$\varepsilon \cdot \alpha_m \dots \alpha_0 . \alpha_1 \alpha_2 \dots \alpha_n \dots \tag{3.4}$$

where m is the largest integer in $\mathbb{Z}$ such that $\alpha_k = 0$ for all $k < m$. If $m > 0$ the expression in Eq. (3.4) should be interpreted as

$$\varepsilon \cdot 0.0 \dots 0 \alpha_m \alpha_{m+1} \dots .$$

An element $\alpha \in \mathbb{D}$ has a tail of all 9's starting at $N \in \mathbb{N}$ if $\alpha_n = 9$ and for all $n \geq N$ and $\alpha_{N-1} \neq 9$. If α has a tail of 9's starting at $N \in \mathbb{N}$, then for $n > N$,

$$\begin{aligned} a_n(\alpha) &= \sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + 9 \sum_{k=N}^{n} 10^{-k} \\ &= \sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + \frac{9}{10^{N-1}} \cdot \frac{1 - 10^{-(n-N)}}{9} \\ &\to \sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + 10^{-(N-1)} \text{ as } n \to \infty. \end{aligned}$$

If α' is the digits in the decimal expansion of $\sum_{k=-\infty}^{N-1} \alpha_k 10^{-k} + 10^{-(N-1)}$, then

$$\alpha' \in \mathbb{D}' := \{\alpha \in \mathbb{D} : \alpha \text{ does not have a tail of all 9's}\}.$$

and we have just shown that $D(\varepsilon, \alpha) = D(\varepsilon, \alpha')$. In particular this implies

$$D(\{\pm 1\} \times \mathbb{D}') = D(\{\pm 1\} \times \mathbb{D}). \tag{3.5}$$

Theorem 3.12 (Decimal Representation). *The map*

$$D : \{\pm 1\} \times \mathbb{D}' \to \mathbb{R} \setminus \{0\}$$

is a bijection.

Proof. Suppose $D(\varepsilon, \alpha) = D(\delta, \beta)$ for some (ε, α) and (δ, β) in $\{\pm 1\} \times \mathbb{D}$. Since $D(\varepsilon, \alpha) > 0$ if $\varepsilon = 1$ and $D(\varepsilon, \alpha) < 0$ if $\varepsilon = -1$ it follows that $\varepsilon = \delta$. Let $a = a(\alpha)$ and $b = a(\beta)$ be the sequences associated to α and β respectively. Suppose that $\alpha \neq \beta$ and let $j \in \mathbb{Z}$ be the position where α and β first disagree, i.e. $\alpha_n = \beta_n$ for all $n < j$ while $\alpha_j \neq \beta_j$. For sake of definiteness suppose $\beta_j > \alpha_j$. Then for $n > j$ we have

$$\begin{aligned} b_n - a_n &= (\beta_j - \alpha_j) 10^{-j} + \sum_{k=j+1}^{n} (\beta_k - \alpha_k) 10^{-k} \\ &\geq 10^{-j} - 9 \sum_{k=j+1}^{n} 10^{-k} \geq 10^{-j} - 9 \frac{1}{9 \cdot 10^j} = 0. \end{aligned}$$

Therefore $b_n - a_n \geq 0$ for all n and $\lim (b_n - a_n) = 0$ iff $\beta_j = \alpha_j + 1$ and $\beta_k = 9$ and $\alpha_k = 0$ for all $k > j$. In summary, $D(\varepsilon, \alpha) = D(\delta, \beta)$ with $\alpha \neq \beta$ implies either α or β has an infinite tail of nines which shows that D is injective when restricted to $\{\pm 1\} \times \mathbb{D}'$. To see that D is surjective it suffices to show any $\bar{b} \in \mathbb{R}$ with $0 < \bar{b} < 1$ is in the range of D. For each $n \in \mathbb{N}$, let $a_n = .\alpha_1 \dots \alpha_n$ with $\alpha_i \in \{0, 1, 2, \dots, 9\}$ such that

$$i(a_n) < \bar{b} \leq i(a_n) + i\left(10^{-n}\right). \tag{3.6}$$

Since $a_{n+1} = a_n + \alpha_{n+1} 10^{-(n+1)}$ for some $\alpha_{n+1} \in \{0, 1, 2, \dots, 9\}$, we see that $a_{n+1} = .\alpha_1 \dots \alpha_n \alpha_{n+1}$, i.e. the first n digits in the decimal expansion of a_{n+1} are the same as in the decimal expansion of a_n. Hence this defines α_n uniquely for all $n \geq 1$. By setting $\alpha_n = 0$ when $n \leq 0$, we have constructed from $\bar{b}$ an element $\alpha \in \mathbb{D}$. Because of Eq. (3.6), $D(1, \alpha) = \bar{b}$. ■

Notation 3.13 *From now on we will identify $\mathbb{Q}$ with $i(\mathbb{Q}) \subset \mathbb{R}$ and elements in $\mathbb{R}$ with their decimal expansions.*

To summarize, we have constructed a complete ordered field $\mathbb{R}$ "containing" $\mathbb{Q}$ as a dense subset. Moreover every element in $\mathbb{R}$ (modulo those of the form $m10^{-n}$ for some $m \in \mathbb{Z}$ and $n \in \mathbb{N}$) has a unique decimal expansion.

Corollary 3.14. *The set $(0,1) := \{a \in \mathbb{R} : 0 < a < 1\}$ is uncountable while $\mathbb{Q} \cap (0,1)$ is countable.*

Proof. By Theorem 3.12, the set $\{0,1,2\ldots,8\}^{\mathbb{N}}$ can be mapped injectively into $(0,1)$ and therefore it follows from Lemma 2.6 that $(0,1)$ is uncountable. For each $m\in\mathbb{N}$, let $A_m := \left\{\frac{n}{m} : n\in\mathbb{N} \text{ with } n<m\right\}$. Since $\mathbb{Q}\cap(0,1) = \cup_{m=1}^\infty A_m$ and $\#(A_m)<\infty$ for all m, another application of Lemma 2.6 shows $\mathbb{Q}\cap(0,1)$ is countable. ■

3.2 The Complex Numbers

Definition 3.15 (Complex Numbers). *Let $\mathbb{C}=\mathbb{R}^2$ equipped with multiplication rule*

$$(a,b)(c,d) := (ac-bd, bc+ad) \tag{3.7}$$

and the usual rule for vector addition. As is standard we will write $0=(0,0)$, $1=(1,0)$ and $i=(0,1)$ so that every element z of $\mathbb{C}$ may be written as $z=(x,y)=x1+yi$ which in the future will be written simply as $z=x+iy$. If $z=x+iy$, let $\operatorname{Re} z = x$ and $\operatorname{Im} z = y$.

Writing $z=a+ib$ and $w=c+id$, the multiplication rule in Eq. (3.7) becomes

$$(a+ib)(c+id) := (ac-bd)+i(bc+ad) \tag{3.8}$$

and in particular $1^2=1$ and $i^2=-1$.

Proposition 3.16. *The complex numbers $\mathbb{C}$ with the above multiplication rule satisfies the usual definitions of a field. For example $wz=zw$ and $z(w_1+w_2) = zw_1+zw_2$, etc. Moreover if $z\neq 0$, z has a multiplicative inverse given by*

$$z^{-1} = \frac{a}{a^2+b^2} - i\frac{b}{a^2+b^2}. \tag{3.9}$$

Proof. The proof is a straightforward verification. Only the last assertion will be verified here. Suppose $z=a+ib\neq 0$, we wish to find $w=c+id$ such that $zw=1$ and this happens by Eq. (3.8) iff

$$ac-bd=1 \text{ and} \tag{3.10}$$
$$bc+ad=0. \tag{3.11}$$

Solving these equations for c and d gives $c=\frac{a}{a^2+b^2}$ and $d=-\frac{b}{a^2+b^2}$ as claimed. ■

Notation 3.17 (Conjugation and Modulus) *If $z=a+ib$ with $a,b\in\mathbb{R}$ let $\bar z = a-ib$ and*

$$|z| := \sqrt{z\bar z} = \sqrt{a^2+b^2} = \sqrt{|\operatorname{Re} z|^2 + |\operatorname{Im} z|^2}.$$

See Exercise 3.8 for the existence of the square root as a positive real number.

Notice that

$$\operatorname{Re} z = \frac{1}{2}(z+\bar z) \text{ and } \operatorname{Im} z = \frac{1}{2i}(z-\bar z). \tag{3.12}$$

Proposition 3.18. *Complex conjugation and the modulus operators satisfy the following properties.*

1. $\bar{\bar z} = z$,
2. $\overline{zw} = \bar z\bar w$ *and* $\bar z + \bar w = \overline{z+w}$.
3. $|\bar z| = |z|$
4. $|zw| = |z|\,|w|$ *and in particular* $|z^n| = |z|^n$ *for all* $n\in\mathbb{N}$.
5. $|\operatorname{Re} z|\le|z|$ *and* $|\operatorname{Im} z|\le|z|$
6. $|z+w|\le|z|+|w|$.
7. $z=0$ *iff* $|z|=0$.
8. *If* $z\neq 0$ *then* $z^{-1} := \frac{\bar z}{|z|^2}$ *(also written as* $\frac{1}{z}$*) is the inverse of* z.
9. $\left|z^{-1}\right| = |z|^{-1}$ *and more generally* $|z^n| = |z|^n$ *for all* $n\in\mathbb{Z}$.

Proof. All of these properties are direct computations except for possibly the triangle inequality in item 6 which is verified by the following computation;

$$\begin{aligned}|z+w|^2 &= (z+w)\left(\overline{z+w}\right) = |z|^2+|w|^2+w\bar z+\bar w z\\ &= |z|^2+|w|^2+w\bar z+\overline{w\bar z}\\ &= |z|^2+|w|^2+2\operatorname{Re}(w\bar z) \le |z|^2+|w|^2+2|z|\,|w|\\ &= (|z|+|w|)^2.\end{aligned}$$

■

Definition 3.19. *A sequence $\{z_n\}_{n=1}^\infty\subset\mathbb{C}$ is **Cauchy** if $|z_n-z_m|\to 0$ as $m,n\to\infty$ and is **convergent** to $z\in\mathbb{C}$ if $|z-z_n|\to 0$ as $n\to\infty$. As usual if $\{z_n\}_{n=1}^\infty$ converges to z we will write $z_n\to z$ as $n\to\infty$ or $z=\lim_{n\to\infty} z_n$.*

Theorem 3.20. *The complex numbers are complete,i.e. all Cauchy sequences are convergent.*

Proof. This follows from the completeness of real numbers and the easily proved observations that if $z_n = a_n+ib_n\in\mathbb{C}$, then

1. $\{z_n\}_{n=1}^\infty\subset\mathbb{C}$ is Cauchy iff $\{a_n\}_{n=1}^\infty\subset\mathbb{R}$ and $\{b_n\}_{n=1}^\infty\subset\mathbb{R}$ are Cauchy and
2. $z_n\to z=a+ib$ as $n\to\infty$ iff $a_n\to a$ and $b_n\to b$ as $n\to\infty$.

■

3.3 Exercises

Exercise 3.8. Show to every $a \in \mathbb{R}$ with $a \geq 0$ there exists a unique number $b \in \mathbb{R}$ such that $b \geq 0$ and $b^2 = a$. Of course we will call $b = \sqrt{a}$. Also show that $a \to \sqrt{a}$ is an increasing function on $[0, \infty)$. **Hint:** To construct $b = \sqrt{a}$ for $a > 0$, to each $n \in \mathbb{N}$ let $m_n \in \mathbb{N}_0$ be chosen so that

$$\frac{m_n^2}{n^2} < a \leq \frac{(m_n + 1)^2}{n^2} \text{ i.e. } i\left(\frac{m_n^2}{n^2}\right) < a \leq i\left(\frac{(m_n + 1)^2}{n^2}\right)$$

and let $q_n := \frac{m_n}{n}$. Then show $b = \overline{\{q_n\}_{n=1}^{\infty}} \in \mathbb{R}$ satisfies $b > 0$ and $b^2 = a$.

Limits and Sums

4.1 Limsups, Liminfs and Extended Limits

Notation 4.1 *The* ***extended real numbers*** *is the set* $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, *i.e. it is* $\mathbb{R}$ *with two new points called* ∞ *and* $-\infty$. *We use the following conventions,* $\pm\infty \cdot 0 = 0$, $\pm\infty \cdot a = \pm\infty$ *if* $a \in \mathbb{R}$ *with* $a > 0$, $\pm\infty \cdot a = \mp\infty$ *if* $a \in \mathbb{R}$ *with* $a < 0$, $\pm\infty + a = \pm\infty$ *for any* $a \in \mathbb{R}$, $\infty + \infty = \infty$ *and* $-\infty - \infty = -\infty$ *while* $\infty - \infty$ *is not defined. A sequence* $a_n \in \bar{\mathbb{R}}$ *is said to converge to* ∞ $(-\infty)$ *if for all* $M \in \mathbb{R}$ *there exists* $m \in \mathbb{N}$ *such that* $a_n \geq M$ $(a_n \leq M)$ *for all* $n \geq m$.

Lemma 4.2. *Suppose* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *are convergent sequences in* $\bar{\mathbb{R}}$, *then:*

1. *If* $a_n \leq b_n$ *for a.a.* n *then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.
2. *If* $c \in \mathbb{R}$, $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$.
3. *If* $\{a_n + b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \tag{4.1}$$
provided the right side is not of the form $\infty - \infty$.
4. $\{a_n b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{4.2}$$
provided the right hand side is not of the for $\pm\infty \cdot 0$ *of* $0 \cdot (\pm\infty)$.

Before going to the proof consider the simple example where $a_n = n$ and $b_n = -\alpha n$ with $\alpha > 0$. Then

$$\lim (a_n + b_n) = \begin{cases} \infty & \text{if } \alpha < 1 \\ 0 & \text{if } \alpha = 1 \\ -\infty & \text{if } \alpha > 1 \end{cases}$$

while

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{``} = \text{''} \infty - \infty.$$

This shows that the requirement that the right side of Eq. (4.1) is not of form $\infty - \infty$ is necessary in Lemma 4.2. Similarly by considering the examples $a_n = n$ and $b_n = n^{-\alpha}$ with $\alpha > 0$ shows the necessity for assuming right hand side of Eq. (4.2) is not of the form $\infty \cdot 0$.

Proof. The proofs of items 1. and 2. are left to the reader.

Proof of Eq. (4.1). Let $a := \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Case 1., suppose $b = \infty$ in which case we must assume $a > -\infty$. In this case, for every $M > 0$, there exists N such that $b_n \geq M$ and $a_n \geq a - 1$ for all $n \geq N$ and this implies

$$a_n + b_n \geq M + a - 1 \text{ for all } n \geq N.$$

Since M is arbitrary it follows that $a_n + b_n \to \infty$ as $n \to \infty$. The cases where $b = -\infty$ or $a = \pm\infty$ are handled similarly. Case 2. If $a, b \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|a - a_n| \leq \varepsilon \text{ and } |b - b_n| \leq \varepsilon \text{ for all } n \geq N.$$

Therefore,

$$|a + b - (a_n + b_n)| = |a - a_n + b - b_n| \leq |a - a_n| + |b - b_n| \leq 2\varepsilon$$

for all $n \geq N$. Since n is arbitrary, it follows that $\lim_{n\to\infty} (a_n + b_n) = a + b$.

Proof of Eq. (4.2). It will be left to the reader to prove the case where $\lim a_n$ and $\lim b_n$ exist in $\mathbb{R}$. I will only consider the case where $a = \lim_{n\to\infty} a_n \neq 0$ and $\lim_{n\to\infty} b_n = \infty$ here. Let us also suppose that $a > 0$ (the case $a < 0$ is handled similarly) and let $\alpha := \min\left(\frac{a}{2}, 1\right)$. Given any $M < \infty$, there exists $N \in \mathbb{N}$ such that $a_n \geq \alpha$ and $b_n \geq M$ for all $n \geq N$ and for this choice of N, $a_n b_n \geq M\alpha$ for all $n \geq N$. Since $\alpha > 0$ is fixed and M is arbitrary it follows that $\lim_{n\to\infty} (a_n b_n) = \infty$ as desired. ■

For any subset $\Lambda \subset \bar{\mathbb{R}}$, let $\sup \Lambda$ and $\inf \Lambda$ denote the least upper bound and greatest lower bound of Λ respectively. The convention being that $\sup \Lambda = \infty$ if $\infty \in \Lambda$ or Λ is not bounded from above and $\inf \Lambda = -\infty$ if $-\infty \in \Lambda$ or Λ is not bounded from below. We will also use the **conventions** that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

Notation 4.3 *Suppose that* $\{x_n\}_{n=1}^\infty \subset \bar{\mathbb{R}}$ *is a sequence of numbers. Then*

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k : k \geq n\} \text{ and} \tag{4.3}$$

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k \geq n\}. \tag{4.4}$$

We will also write $\underline{\lim}$ *for* $\liminf$ *and* $\overline{\lim}$ *for* $\limsup$.

Remark 4.4. Notice that if $a_k := \inf\{x_k : k \geq n\}$ and $b_k := \sup\{x_k : k \geq n\}$, then $\{a_k\}$ is an increasing sequence while $\{b_k\}$ is a decreasing sequence. Therefore the limits in Eq. (4.3) and Eq. (4.4) always exist in $\bar{\mathbb{R}}$ and

$$\liminf_{n\to\infty} x_n = \sup_n \inf\{x_k : k \geq n\} \text{ and}$$
$$\limsup_{n\to\infty} x_n = \inf_n \sup\{x_k : k \geq n\}.$$

The following proposition contains some basic properties of liminfs and limsups.

Proposition 4.5. *Let $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ be two sequences of real numbers. Then*

1. $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} a_n$ *exists in* $\bar{\mathbb{R}}$ *iff*
$$\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n \in \bar{\mathbb{R}}.$$
2. *There is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \limsup_{n\to\infty} a_n$. *Similarly, there is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \liminf_{n\to\infty} a_n$.
3. $$\limsup_{n\to\infty} (a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{4.5}$$
whenever the right side of this equation is not of the form $\infty - \infty$.
4. *If* $a_n \geq 0$ *and* $b_n \geq 0$ *for all* $n \in \mathbb{N}$, *then*
$$\limsup_{n\to\infty} (a_n b_n) \leq \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{4.6}$$
provided the right hand side of (4.6) is not of the form $0 \cdot \infty$ *or* $\infty \cdot 0$.

Proof. Item 1. will be proved here leaving the remaining items as an exercise to the reader. Since

$$\inf\{a_k : k \geq n\} \leq \sup\{a_k : k \geq n\} \ \forall\, n,$$

$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

Now suppose that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = a \in \mathbb{R}$. Then for all $\varepsilon > 0$, there is an integer N such that

$$a - \varepsilon \leq \inf\{a_k : k \geq N\} \leq \sup\{a_k : k \geq N\} \leq a + \varepsilon,$$

i.e.

$$a - \varepsilon \leq a_k \leq a + \varepsilon \text{ for all } k \geq N.$$

Hence by the definition of the limit, $\lim_{k\to\infty} a_k = a$. If $\liminf_{n\to\infty} a_n = \infty$, then we know for all $M \in (0, \infty)$ there is an integer N such that

$$M \leq \inf\{a_k : k \geq N\}$$

and hence $\lim_{n\to\infty} a_n = \infty$. The case where $\limsup_{n\to\infty} a_n = -\infty$ is handled similarly.

Conversely, suppose that $\lim_{n\to\infty} a_n = A \in \bar{\mathbb{R}}$ exists. If $A \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that $|A - a_n| \leq \varepsilon$ for all $n \geq N(\varepsilon)$, i.e.

$$A - \varepsilon \leq a_n \leq A + \varepsilon \text{ for all } n \geq N(\varepsilon).$$

From this we learn that

$$A - \varepsilon \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq A + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$A \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq A,$$

i.e. that $A = \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$. If $A = \infty$, then for all $M > 0$ there exists $N = N(M)$ such that $a_n \geq M$ for all $n \geq N$. This show that $\liminf_{n\to\infty} a_n \geq M$ and since M is arbitrary it follows that

$$\infty \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

The proof for the case $A = -\infty$ is analogous to the $A = \infty$ case. ■

4.2 Sums of positive functions

In this and the next few sections, let X and Y be two sets. We will write $\alpha \subset\subset X$ to denote that α is a **finite** subset of X and write 2_f^X for those $\alpha \subset\subset X$.

Definition 4.6. *Suppose that $a : X \to [0, \infty]$ is a function and $F \subset X$ is a subset, then*

$$\sum_F a = \sum_{x\in F} a(x) := \sup\left\{\sum_{x\in\alpha} a(x) : \alpha \subset\subset F\right\}.$$

Remark 4.7. Suppose that $X = \mathbb{N} = \{1, 2, 3, \dots\}$ and $a : X \to [0, \infty]$, then

$$\sum_{\mathbb{N}} a = \sum_{n=1}^\infty a(n) := \lim_{N\to\infty} \sum_{n=1}^N a(n).$$

Indeed for all N, $\sum_{n=1}^N a(n) \le \sum_{\mathbb{N}} a$, and thus passing to the limit we learn that

$$\sum_{n=1}^{\infty} a(n) \le \sum_{\mathbb{N}} a.$$

Conversely, if $\alpha \subset\subset \mathbb{N}$, then for all N large enough so that $\alpha \subset \{1, 2, \dots, N\}$, we have $\sum_\alpha a \le \sum_{n=1}^N a(n)$ which upon passing to the limit implies that

$$\sum_\alpha a \le \sum_{n=1}^{\infty} a(n).$$

Taking the supremum over α in the previous equation shows

$$\sum_{\mathbb{N}} a \le \sum_{n=1}^{\infty} a(n).$$

Remark 4.8. Suppose $a : X \to [0, \infty]$ and $\sum_X a < \infty$, then $\{x \in X : a(x) > 0\}$ is at most countable. To see this first notice that for any $\varepsilon > 0$, the set $\{x : a(x) \ge \varepsilon\}$ must be finite for otherwise $\sum_X a = \infty$. Thus

$$\{x \in X : a(x) > 0\} = \bigcup\nolimits_{k=1}^{\infty} \{x : a(x) \ge 1/k\}$$

which shows that $\{x \in X : a(x) > 0\}$ is a countable union of finite sets and thus countable by Lemma 2.6.

Lemma 4.9. *Suppose that $a, b : X \to [0, \infty]$ are two functions, then*

$$\sum_X (a + b) = \sum_X a + \sum_X b \text{ and}$$
$$\sum_X \lambda a = \lambda \sum_X a$$

for all $\lambda \ge 0$.

I will only prove the first assertion, the second being easy. Let $\alpha \subset\subset X$ be a finite set, then

$$\sum_\alpha (a + b) = \sum_\alpha a + \sum_\alpha b \le \sum_X a + \sum_X b$$

which after taking sups over α shows that

$$\sum_X (a + b) \le \sum_X a + \sum_X b.$$

Similarly, if $\alpha, \beta \subset\subset X$, then

$$\sum_\alpha a + \sum_\beta b \le \sum_{\alpha\cup\beta} a + \sum_{\alpha\cup\beta} b = \sum_{\alpha\cup\beta} (a + b) \le \sum_X (a + b).$$

Taking sups over α and β then shows that

$$\sum_X a + \sum_X b \le \sum_X (a + b).$$

Lemma 4.10. *Let X and Y be sets, $R \subset X \times Y$ and suppose that $a : R \to \bar{\mathbb{R}}$ is a function. Let ${}_xR := \{y \in Y : (x, y) \in R\}$ and $R_y := \{x \in X : (x, y) \in R\}$. Then*

$$\sup_{(x,y)\in R} a(x, y) = \sup_{x\in X} \sup_{y\in_x R} a(x, y) = \sup_{y\in Y} \sup_{x\in R_y} a(x, y) \text{ and}$$
$$\inf_{(x,y)\in R} a(x, y) = \inf_{x\in X} \inf_{y\in_x R} a(x, y) = \inf_{y\in Y} \inf_{x\in R_y} a(x, y).$$

(Recall the conventions: $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.)

Proof. Let $M = \sup_{(x,y)\in R} a(x, y)$, $N_x := \sup_{y\in_x R} a(x, y)$. Then $a(x, y) \le M$ for all $(x, y) \in R$ implies $N_x = \sup_{y\in_x R} a(x, y) \le M$ and therefore that

$$\sup_{x\in X} \sup_{y\in_x R} a(x, y) = \sup_{x\in X} N_x \le M. \tag{4.7}$$

Similarly for any $(x, y) \in R$,

$$a(x, y) \le N_x \le \sup_{x\in X} N_x = \sup_{x\in X} \sup_{y\in_x R} a(x, y)$$

and therefore

$$M = \sup_{(x,y)\in R} a(x, y) \le \sup_{x\in X} \sup_{y\in_x R} a(x, y) \tag{4.8}$$

Equations (4.7) and (4.8) show that

$$\sup_{(x,y)\in R} a(x, y) = \sup_{x\in X} \sup_{y\in_x R} a(x, y).$$

The assertions involving infimums are proved analogously or follow from what we have just proved applied to the function $-a$. ■

Theorem 4.11 (Monotone Convergence Theorem for Sums). *Suppose that $f_n : X \to [0, \infty]$ is an increasing sequence of functions and*

$$f(x) := \lim_{n\to\infty} f_n(x) = \sup_n f_n(x).$$

Then

$$\lim_{n\to\infty} \sum_X f_n = \sum_X f$$

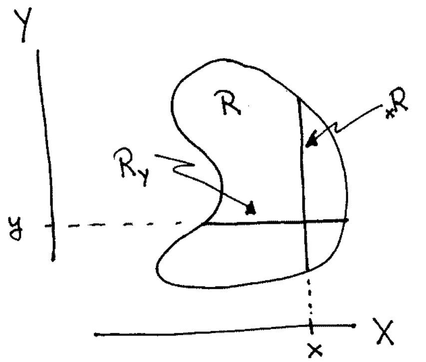


Fig. 4.1. The x and y – slices of a set $R\subset X\times Y$.

Proof. We will give two proofs.
First proof. Let

$$2_f^X := \{A \subset X : A \subset\subset X\}.$$

Then

$$\begin{aligned}\lim_{n\to\infty}\sum_X f_n &= \sup_n \sum_X f_n = \sup_n \sup_{\alpha\in 2_f^X}\sum_\alpha f_n = \sup_{\alpha\in 2_f^X}\sup_n \sum_\alpha f_n \\ &= \sup_{\alpha\in 2_f^X}\lim_{n\to\infty}\sum_\alpha f_n = \sup_{\alpha\in 2_f^X}\sum_\alpha \lim_{n\to\infty} f_n \\ &= \sup_{\alpha\in 2_f^X}\sum_\alpha f = \sum_X f.\end{aligned}$$

Second Proof. Let $S_n = \sum_X f_n$ and $S = \sum_X f$. Since $f_n \le f_m \le f$ for all $n \le m$, it follows that

$$S_n \le S_m \le S$$

which shows that $\lim_{n\to\infty} S_n$ exists and is less that S, i.e.

$$A := \lim_{n\to\infty}\sum_X f_n \le \sum_X f. \tag{4.9}$$

Noting that $\sum_\alpha f_n \le \sum_X f_n = S_n \le A$ for all $\alpha \subset\subset X$ and in particular,

$$\sum_\alpha f_n \le A \text{ for all } n \text{ and } \alpha \subset\subset X.$$

Letting n tend to infinity in this equation shows that

$$\sum_\alpha f \le A \text{ for all } \alpha \subset\subset X$$

and then taking the sup over all $\alpha \subset\subset X$ gives

$$\sum_X f \le A = \lim_{n\to\infty}\sum_X f_n \tag{4.10}$$

which combined with Eq. (4.9) proves the theorem. ∎

Lemma 4.12 (Fatou's Lemma for Sums). *Suppose that $f_n : X \to [0,\infty]$ is a sequence of functions, then*

$$\sum_X \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty}\sum_X f_n.$$

Proof. Define $g_k := \inf_{n\ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k\to\infty$. Since $g_k \le f_n$ for all $n \ge k$,

$$\sum_X g_k \le \sum_X f_n \text{ for all } n \ge k$$

and therefore

$$\sum_X g_k \le \liminf_{n\to\infty}\sum_X f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k\to\infty$ to find

$$\sum_X \liminf_{n\to\infty} f_n = \sum_X \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty}\sum_X g_k \le \liminf_{n\to\infty}\sum_X f_n.$$

∎

Remark 4.13. If $A = \sum_X a < \infty$, then for all $\varepsilon > 0$ there exists $\alpha_\varepsilon \subset\subset X$ such that

$$A \ge \sum_\alpha a \ge A - \varepsilon$$

for all $\alpha \subset\subset X$ containing α_ε or equivalently,

$$\left|A - \sum_\alpha a\right| \le \varepsilon \tag{4.11}$$

for all $\alpha \subset\subset X$ containing α_ε. Indeed, choose α_ε so that $\sum_{\alpha_\varepsilon} a \ge A - \varepsilon$.

4.3 Sums of complex functions

Definition 4.14. *Suppose that $a : X \to \mathbb{C}$ is a function, we say that*

$$\sum_X a = \sum_{x\in X} a(x)$$

exists and is equal to $A \in \mathbb{C}$, if for all $\varepsilon > 0$ there is a finite subset $\alpha_\varepsilon \subset X$ such that for all $\alpha \subset\subset X$ containing α_ε we have

$$\left| A - \sum_\alpha a \right| \le \varepsilon.$$

The following lemma is left as an exercise to the reader.

Lemma 4.15. *Suppose that $a, b : X \to \mathbb{C}$ are two functions such that $\sum_X a$ and $\sum_X b$ exist, then $\sum_X (a + \lambda b)$ exists for all $\lambda \in \mathbb{C}$ and*

$$\sum_X (a + \lambda b) = \sum_X a + \lambda \sum_X b.$$

Definition 4.16 (Summable). *We call a function $a : X \to \mathbb{C}$ **summable** if*

$$\sum_X |a| < \infty.$$

Proposition 4.17. *Let $a : X \to \mathbb{C}$ be a function, then $\sum_X a$ exists iff $\sum_X |a| < \infty$, i.e. iff a is summable. Moreover if a is summable, then*

$$\left| \sum_X a \right| \le \sum_X |a|.$$

Proof. If $\sum_X |a| < \infty$, then $\sum_X (\operatorname{Re} a)^\pm < \infty$ and $\sum_X (\operatorname{Im} a)^\pm < \infty$ and hence by Remark 4.13 these sums exists in the sense of Definition 4.14. Therefore by Lemma 4.15, $\sum_X a$ exists and

$$\sum_X a = \sum_X (\operatorname{Re} a)^+ - \sum_X (\operatorname{Re} a)^- + i \left(\sum_X (\operatorname{Im} a)^+ - \sum_X (\operatorname{Im} a)^- \right).$$

Conversely, if $\sum_X |a| = \infty$ then, because $|a| \le |\operatorname{Re} a| + |\operatorname{Im} a|$, we must have

$$\sum_X |\operatorname{Re} a| = \infty \text{ or } \sum_X |\operatorname{Im} a| = \infty.$$

Thus it suffices to consider the case where $a : X \to \mathbb{R}$ is a real function. Write $a = a^+ - a^-$ where

$$a^+(x) = \max(a(x), 0) \text{ and } a^-(x) = \max(-a(x), 0). \tag{4.12}$$

Then $|a| = a^+ + a^-$ and

$$\infty = \sum_X |a| = \sum_X a^+ + \sum_X a^-$$

which shows that either $\sum_X a^+ = \infty$ or $\sum_X a^- = \infty$. Suppose, with out loss of generality, that $\sum_X a^+ = \infty$. Let $X' := \{x \in X : a(x) \ge 0\}$, then we know that $\sum_{X'} a = \infty$ which means there are finite subsets $\alpha_n \subset X' \subset X$ such that $\sum_{\alpha_n} a \ge n$ for all n. Thus if $\alpha \subset\subset X$ is any finite set, it follows that $\lim_{n\to\infty} \sum_{\alpha_n \cup \alpha} a = \infty$, and therefore $\sum_X a$ can not exist as a number in $\mathbb{R}$.
Finally if a is summable, write $\sum_X a = \rho e^{i\theta}$ with $\rho \ge 0$ and $\theta \in \mathbb{R}$, then

$$\begin{aligned}
\left| \sum_X a \right| &= \rho = e^{-i\theta} \sum_X a = \sum_X e^{-i\theta} a \\
&= \sum_X \operatorname{Re}\left[e^{-i\theta} a\right] \le \sum_X \left(\operatorname{Re}\left[e^{-i\theta} a\right]\right)^+ \\
&\le \sum_X \left|\operatorname{Re}\left[e^{-i\theta} a\right]\right| \le \sum_X \left|e^{-i\theta} a\right| \le \sum_X |a|.
\end{aligned}$$

Alternatively, this may be proved by approximating $\sum_X a$ by a finite sum and then using the triangle inequality of $|\cdot|$. ■

Remark 4.18. Suppose that $X = \mathbb{N}$ and $a : \mathbb{N} \to \mathbb{C}$ is a sequence, then it is not necessarily true that

$$\sum_{n=1}^\infty a(n) = \sum_{n\in\mathbb{N}} a(n). \tag{4.13}$$

This is because

$$\sum_{n=1}^\infty a(n) = \lim_{N\to\infty} \sum_{n=1}^N a(n)$$

depends on the ordering of the sequence a where as $\sum_{n\in\mathbb{N}} a(n)$ does not. For example, take $a(n) = (-1)^n/n$ then $\sum_{n\in\mathbb{N}} |a(n)| = \infty$ i.e. $\sum_{n\in\mathbb{N}} a(n)$ does **not** exist while $\sum_{n=1}^\infty a(n)$ does exist. On the other hand, if

$$\sum_{n\in\mathbb{N}} |a(n)| = \sum_{n=1}^\infty |a(n)| < \infty$$

then Eq. (4.13) is valid.

Theorem 4.19 (Dominated Convergence Theorem for Sums). *Suppose that $f_n : X \to \mathbb{C}$ is a sequence of functions on X such that $f(x) = \lim_{n\to\infty} f_n(x) \in \mathbb{C}$ exists for all $x \in X$. Further assume there is a **dominating function** $g : X \to [0,\infty)$ such that*

$$|f_n(x)| \le g(x) \text{ for all } x \in X \text{ and } n \in \mathbb{N} \tag{4.14}$$

and that g is summable. Then

$$\lim_{n\to\infty} \sum_{x\in X} f_n(x) = \sum_{x\in X} f(x). \tag{4.15}$$

Proof. Notice that $|f| = \lim |f_n| \le g$ so that f is summable. By considering the real and imaginary parts of f separately, it suffices to prove the theorem in the case where f is real. By Fatou's Lemma,

$$\sum_X (g \pm f) = \sum_X \liminf_{n\to\infty} (g \pm f_n) \le \liminf_{n\to\infty} \sum_X (g \pm f_n)$$
$$= \sum_X g + \liminf_{n\to\infty} \left(\pm \sum_X f_n \right).$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\sum_X g \pm \sum_X f \le \sum_X g + \begin{cases} \liminf_{n\to\infty} \sum_X f_n \\ -\limsup_{n\to\infty} \sum_X f_n \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \sum_X f_n \le \sum_X f \le \liminf_{n\to\infty} \sum_X f_n.$$

This shows that $\lim\limits_{n\to\infty} \sum_X f_n$ exists and is equal to $\sum_X f$. ∎

Proof. (Second Proof.) Passing to the limit in Eq. (4.14) shows that $|f| \le g$ and in particular that f is summable. Given $\varepsilon > 0$, let $\alpha \subset\subset X$ such that

$$\sum_{X\setminus\alpha} g \le \varepsilon.$$

Then for $\beta \subset\subset X$ such that $\alpha \subset \beta$,

$$\left| \sum_\beta f - \sum_\beta f_n \right| = \left| \sum_\beta (f - f_n) \right|$$
$$\le \sum_\beta |f - f_n| = \sum_\alpha |f - f_n| + \sum_{\beta\setminus\alpha} |f - f_n|$$
$$\le \sum_\alpha |f - f_n| + 2 \sum_{\beta\setminus\alpha} g$$
$$\le \sum_\alpha |f - f_n| + 2\varepsilon.$$

and hence that

$$\left| \sum_\beta f - \sum_\beta f_n \right| \le \sum_\alpha |f - f_n| + 2\varepsilon.$$

Since this last equation is true for all such $\beta \subset\subset X$, we learn that

$$\left| \sum_X f - \sum_X f_n \right| \le \sum_\alpha |f - f_n| + 2\varepsilon$$

which then implies that

$$\limsup_{n\to\infty} \left| \sum_X f - \sum_X f_n \right| \le \limsup_{n\to\infty} \sum_\alpha |f - f_n| + 2\varepsilon$$
$$= 2\varepsilon.$$

Because $\varepsilon > 0$ is arbitrary we conclude that

$$\limsup_{n\to\infty} \left| \sum_X f - \sum_X f_n \right| = 0.$$

which is the same as Eq. (4.15). ∎

Remark 4.20. Theorem 4.19 may easily be generalized as follows. Suppose f_n, g_n, g are summable functions on X such that $f_n \to f$ and $g_n \to g$ pointwise, $|f_n| \le g_n$ and $\sum_X g_n \to \sum_X g$ as $n \to \infty$. Then f is summable and Eq. (4.15) still holds. For the proof we use Fatou's Lemma to again conclude

$$\sum_X (g \pm f) = \sum_X \liminf_{n\to\infty} (g_n \pm f_n) \le \liminf_{n\to\infty} \sum_X (g_n \pm f_n)$$
$$= \sum_X g + \liminf_{n\to\infty} \left(\pm \sum_X f_n \right)$$

and then proceed exactly as in the first proof of Theorem 4.19.

4.4 Iterated sums and the Fubini and Tonelli Theorems

Let X and Y be two sets. The proof of the following lemma is left to the reader.

Lemma 4.21. *Suppose that $a : X \to \mathbb{C}$ is function and $F \subset X$ is a subset such that $a(x) = 0$ for all $x \notin F$. Then $\sum_F a$ exists iff $\sum_X a$ exists and when the sums exists,*
$$\sum_X a = \sum_F a.$$

Theorem 4.22 (Tonelli's Theorem for Sums). *Suppose that $a : X \times Y \to [0, \infty]$, then*
$$\sum_{X\times Y} a = \sum_X \sum_Y a = \sum_Y \sum_X a.$$

Proof. It suffices to show, by symmetry, that
$$\sum_{X\times Y} a = \sum_X \sum_Y a$$
Let $\Lambda \subset\subset X \times Y$. Then for any $\alpha \subset\subset X$ and $\beta \subset\subset Y$ such that $\Lambda \subset \alpha \times \beta$, we have
$$\sum_\Lambda a \le \sum_{\alpha\times\beta} a = \sum_\alpha \sum_\beta a \le \sum_\alpha \sum_Y a \le \sum_X \sum_Y a,$$
i.e. $\sum_\Lambda a \le \sum_X \sum_Y a$. Taking the sup over Λ in this last equation shows
$$\sum_{X\times Y} a \le \sum_X \sum_Y a.$$
For the reverse inequality, for each $x \in X$ choose $\beta_n^x \subset\subset Y$ such that $\beta_n^x \uparrow Y$ as $n \uparrow \infty$ and
$$\sum_{y\in Y} a(x, y) = \lim_{n\to\infty} \sum_{y\in\beta_n^x} a(x, y).$$
If $\alpha \subset\subset X$ is a given finite subset of X, then
$$\sum_{y\in Y} a(x, y) = \lim_{n\to\infty} \sum_{y\in\beta_n} a(x, y) \text{ for all } x \in \alpha$$
where $\beta_n := \cup_{x\in\alpha}\beta_n^x \subset\subset Y$. Hence
$$\sum_{x\in\alpha} \sum_{y\in Y} a(x, y) = \sum_{x\in\alpha} \lim_{n\to\infty} \sum_{y\in\beta_n} a(x, y) = \lim_{n\to\infty} \sum_{x\in\alpha} \sum_{y\in\beta_n} a(x, y)$$
$$= \lim_{n\to\infty} \sum_{(x,y)\in\alpha\times\beta_n} a(x, y) \le \sum_{X\times Y} a.$$

Since α is arbitrary, it follows that
$$\sum_{x\in X} \sum_{y\in Y} a(x, y) = \sup_{\alpha\subset\subset X} \sum_{x\in\alpha} \sum_{y\in Y} a(x, y) \le \sum_{X\times Y} a$$
which completes the proof. ■

Theorem 4.23 (Fubini's Theorem for Sums). *Now suppose that $a : X \times Y \to \mathbb{C}$ is a summable function, i.e. by Theorem 4.22 any one of the following equivalent conditions hold:*

1. *$\sum_{X\times Y} |a| < \infty$,*
2. *$\sum_X \sum_Y |a| < \infty$ or*
3. *$\sum_Y \sum_X |a| < \infty$.*
   Then
$$\sum_{X\times Y} a = \sum_X \sum_Y a = \sum_Y \sum_X a.$$

Proof. If $a : X \to \mathbb{R}$ is real valued the theorem follows by applying Theorem 4.22 to $a^\pm$ – the positive and negative parts of a. The general result holds for complex valued functions a by applying the real version just proved to the real and imaginary parts of a. ■

4.5 ℓ^p – spaces, Minkowski and Holder Inequalities

In this chapter, let $\mu : X \to (0, \infty)$ be a given function. Let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. For $p \in (0, \infty)$ and $f : X \to \mathbb{F}$, let
$$\|f\|_p := \left(\sum_{x\in X} |f(x)|^p \mu(x)\right)^{1/p}$$
and for $p = \infty$ let
$$\|f\|_\infty = \sup\{|f(x)| : x \in X\}.$$
Also, for $p > 0$, let
$$\ell^p(\mu) = \{f : X \to \mathbb{F} : \|f\|_p < \infty\}.$$
In the case where $\mu(x) = 1$ for all $x \in X$ we will simply write $\ell^p(X)$ for $\ell^p(\mu)$.

Definition 4.24. *A **norm** on a vector space Z is a function $\|\cdot\| : Z \to [0, \infty)$ such that*

1. *(Homogeneity) $\|\lambda f\| = |\lambda| \|f\|$ for all $\lambda \in \mathbb{F}$ and $f \in Z$.*
2. *(Triangle inequality) $\|f + g\| \le \|f\| + \|g\|$ for all $f, g \in Z$.*

3. (Positive definite) $\|f\| = 0$ *implies* $f = 0$.

A function $p : Z \to [0, \infty)$ *satisfying properties 1. and 2. but not necessarily 3. above will be called a* ***semi-norm*** *on* Z.

A pair $(Z, \|\cdot\|)$ *where* Z *is a vector space and* $\|\cdot\|$ *is a norm on* Z *is called a* ***normed vector space.***

The rest of this section is devoted to the proof of the following theorem.

Theorem 4.25. *For* $p \in [1, \infty]$, $(\ell^p(\mu), \|\cdot\|_p)$ *is a normed vector space.*

Proof. The only difficulty is the proof of the triangle inequality which is the content of Minkowski's Inequality proved in Theorem 4.31 below. ■

Proposition 4.26. *Let* $f : [0, \infty) \to [0, \infty)$ *be a continuous strictly increasing function such that* $f(0) = 0$ *(for simplicity) and* $\lim_{s \to \infty} f(s) = \infty$. *Let* $g = f^{-1}$ *and for* $s, t \geq 0$ *let*

$$F(s) = \int_0^s f(s')ds' \text{ and } G(t) = \int_0^t g(t')dt'.$$

Then for all $s, t \geq 0$,

$$st \leq F(s) + G(t)$$

and equality holds iff $t = f(s)$.

Proof. Let

$$A_s := \{(\sigma, \tau) : 0 \leq \tau \leq f(\sigma) \text{ for } 0 \leq \sigma \leq s\} \text{ and}$$
$$B_t := \{(\sigma, \tau) : 0 \leq \sigma \leq g(\tau) \text{ for } 0 \leq \tau \leq t\}$$

then as one sees from Figure 4.2, $[0, s] \times [0, t] \subset A_s \cup B_t$. (In the figure: $s = 3$, $t = 1$, A_3 is the region under $t = f(s)$ for $0 \leq s \leq 3$ and B_1 is the region to the left of the curve $s = g(t)$ for $0 \leq t \leq 1$.) Hence if m denotes the area of a region in the plane, then

$$st = m([0, s] \times [0, t]) \leq m(A_s) + m(B_t) = F(s) + G(t).$$

As it stands, this proof is a bit on the intuitive side. However, it will become rigorous if one takes m to be "Lebesgue measure" on the plane which will be introduced later.

We can also give a calculus proof of this theorem under the additional assumption that f is C^1. (This restricted version of the theorem is all we need in this section.) To do this fix $t \geq 0$ and let

$$h(s) = st - F(s) = \int_0^s (t - f(\sigma))d\sigma.$$

If $\sigma > g(t) = f^{-1}(t)$, then $t - f(\sigma) < 0$ and hence if $s > g(t)$, we have

$$\begin{aligned} h(s) = \int_0^s (t - f(\sigma))d\sigma &= \int_0^{g(t)} (t - f(\sigma))d\sigma + \int_{g(t)}^s (t - f(\sigma))d\sigma \\ &\leq \int_0^{g(t)} (t - f(\sigma))d\sigma = h(g(t)). \end{aligned}$$

Combining this with $h(0) = 0$ we see that $h(s)$ takes its maximum at some point $s \in (0, g(t)]$ and hence at a point where $0 = h'(s) = t - f(s)$. The only solution to this equation is $s = g(t)$ and we have thus shown

$$st - F(s) = h(s) \leq \int_0^{g(t)} (t - f(\sigma))d\sigma = h(g(t))$$

with equality when $s = g(t)$. To finish the proof we must show $\int_0^{g(t)} (t - f(\sigma))d\sigma = G(t)$. This is verified by making the change of variables $\sigma = g(\tau)$ and then integrating by parts as follows:

$$\begin{aligned} \int_0^{g(t)} (t - f(\sigma))d\sigma &= \int_0^t (t - f(g(\tau)))g'(\tau)d\tau = \int_0^t (t - \tau)g'(\tau)d\tau \\ &= \int_0^t g(\tau)d\tau = G(t). \end{aligned}$$

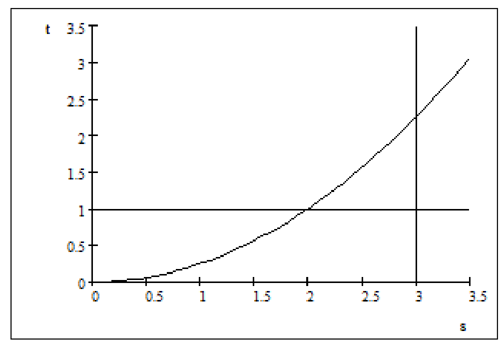


Fig. 4.2. A picture proof of Proposition 4.26.

■

Definition 4.27. *The conjugate exponent* $q \in [1, \infty]$ *to* $p \in [1, \infty]$ *is* $q := \frac{p}{p-1}$ *with the conventions that* $q = \infty$ *if* $p = 1$ *and* $q = 1$ *if* $p = \infty$. *Notice that* q *is characterized by any of the following identities:*

$$\frac{1}{p}+\frac{1}{q}=1,\ 1+\frac{q}{p}=q,\ p-\frac{p}{q}=1 \text{ and } q(p-1)=p. \tag{4.16}$$

Lemma 4.28. *Let $p\in(1,\infty)$ and $q:=\frac{p}{p-1}\in(1,\infty)$ be the conjugate exponent. Then*

$$st\le\frac{s^p}{p}+\frac{t^q}{q} \text{ for all } s,t\ge 0 \tag{4.17}$$

with equality if and only if $t^q=s^p$. (See Example 28.9 below for a generalization of the inequality in Eq. (4.17).)

Proof. Let $F(s)=\frac{s^p}{p}$ for $p>1$. Then $f(s)=s^{p-1}=t$ and $g(t)=t^{\frac{1}{p-1}}=t^{q-1}$, wherein we have used $q-1=p/(p-1)-1=1/(p-1)$. Therefore $G(t)=t^q/q$ and hence by Proposition 4.26,

$$st\le\frac{s^p}{p}+\frac{t^q}{q}$$

with equality iff $t=s^{p-1}$, i.e. $t^q=s^{q(p-1)}=s^p$.

** For those who do not want to use Proposition 4.26, here is a direct calculus proof. Fix $t>0$ and let

$$h(s):=st-\frac{s^p}{p}.$$

Then $h(0)=0$, $\lim_{s\to\infty}h(s)=-\infty$ and $h'(s)=t-s^{p-1}$ which equals zero iff $s=t^{\frac{1}{p-1}}$. Since

$$h\left(t^{\frac{1}{p-1}}\right)=t^{\frac{1}{p-1}}t-\frac{t^{\frac{p}{p-1}}}{p}=t^{\frac{p}{p-1}}-\frac{t^{\frac{p}{p-1}}}{p}=t^q\left(1-\frac{1}{p}\right)=\frac{t^q}{q},$$

it follows from the first derivative test that

$$\max h=\max\left\{h(0),h\left(t^{\frac{1}{p-1}}\right)\right\}=\max\left\{0,\frac{t^q}{q}\right\}=\frac{t^q}{q}.$$

So we have shown

$$st-\frac{s^p}{p}\le\frac{t^q}{q} \text{ with equality iff } t=s^{p-1}.$$

■

Theorem 4.29 (Hölder's inequality). *Let $p,q\in[1,\infty]$ be conjugate exponents. For all $f,g:X\to\mathbb{F}$,*

$$\|fg\|_1\le\|f\|_p\cdot\|g\|_q. \tag{4.18}$$

If $p\in(1,\infty)$ and f and g are not identically zero, then equality holds in Eq. (4.18) iff

$$\left(\frac{|f|}{\|f\|_p}\right)^p=\left(\frac{|g|}{\|g\|_q}\right)^q. \tag{4.19}$$

Proof. The proof of Eq. (4.18) for $p\in\{1,\infty\}$ is easy and will be left to the reader. The cases where $\|f\|_p=0$ or ∞ or $\|g\|_q=0$ or ∞ are easily dealt with and are also left to the reader. So we will assume that $p\in(1,\infty)$ and $0<\|f\|_p,\|g\|_q<\infty$. Letting $s=|f(x)|/\|f\|_p$ and $t=|g|/\|g\|_q$ in Lemma 4.28 implies

$$\frac{|f(x)g(x)|}{\|f\|_p\|g\|_q}\le\frac{1}{p}\ \frac{|f(x)|^p}{\|f\|_p^p}+\frac{1}{q}\ \frac{|g(x)|^q}{\|g\|_q^q}$$

with equality iff

$$\frac{|f(x)|^p}{\|f\|_p^p}=s^p=t^q=\frac{|g(x)|^q}{\|g\|_q^q}. \tag{4.20}$$

Multiplying this equation by $\mu(x)$ and then summing on x gives

$$\frac{\|fg\|_1}{\|f\|_p\|g\|_q}\le\frac{1}{p}+\frac{1}{q}=1$$

with equality iff Eq. (4.20) holds for all $x\in X$, i.e. iff Eq. (4.19) holds. ■

Definition 4.30. *For a complex number $\lambda\in\mathbb{C}$, let*

$$\operatorname{sgn}(\lambda)=\begin{cases}\frac{\lambda}{|\lambda|} & \text{if } \lambda\ne 0\\ 0 & \text{if } \lambda=0.\end{cases}$$

For $\lambda,\mu\in\mathbb{C}$ we will write $\operatorname{sgn}(\lambda)\mathrel{\hat{=}}\operatorname{sgn}(\mu)$ if $\operatorname{sgn}(\lambda)=\operatorname{sgn}(\mu)$ or $\lambda\mu=0$.

Theorem 4.31 (Minkowski's Inequality). *If $1\le p\le\infty$ and $f,g\in\ell^p(\mu)$ then*

$$\|f+g\|_p\le\|f\|_p+\|g\|_p. \tag{4.21}$$

Moreover, assuming f and g are not identically zero, equality holds in Eq. (4.21) iff

$$\operatorname{sgn}(f)\mathrel{\hat{=}}\operatorname{sgn}(g) \text{ when } p=1 \text{ and}$$
$$f=cg \text{ for some } c>0 \text{ when } p\in(1,\infty).$$

Proof. For $p=1$,

$$\|f+g\|_1=\sum_X|f+g|\mu\le\sum_X(|f|\mu+|g|\mu)=\sum_X|f|\mu+\sum_X|g|\mu$$

with equality iff

$$|f|+|g|=|f+g| \iff \operatorname{sgn}(f)\mathrel{\hat{=}}\operatorname{sgn}(g).$$

For $p=\infty$,

$$\|f+g\|_\infty = \sup_X |f+g| \le \sup_X (|f|+|g|) \\ \le \sup_X |f| + \sup_X |g| = \|f\|_\infty + \|g\|_\infty.$$

Now assume that $p \in (1,\infty)$. Since

$$|f+g|^p \le (2\max(|f|,|g|))^p = 2^p \max(|f|^p,|g|^p) \le 2^p(|f|^p+|g|^p)$$

it follows that

$$\|f+g\|_p^p \le 2^p\left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

Eq. (4.21) is easily verified if $\|f+g\|_p = 0$, so we may assume $\|f+g\|_p > 0$. Multiplying the inequality,

$$|f+g|^p = |f+g||f+g|^{p-1} \le |f|\,|f+g|^{p-1} + |g||f+g|^{p-1} \tag{4.22}$$

by μ, then summing on x and applying Holder's inequality on each term gives

$$\sum_X |f+g|^p\mu \le \sum_X |f|\;\; |f+g|^{p-1}\mu + \sum_X |g|\;\; |f+g|^{p-1}\mu \\ \le (\|f\|_p + \|g\|_p)\left\| |f+g|^{p-1}\right\|_q. \tag{4.23}$$

Since $q(p-1) = p$, as in Eq. (4.16),

$$\||f+g|^{p-1}\|_q^q = \sum_X (|f+g|^{p-1})^q\mu = \sum_X |f+g|^p\mu = \|f+g\|_p^p. \tag{4.24}$$

Combining Eqs. (4.23) and (4.24) shows

$$\|f+g\|_p^p \le (\|f\|_p + \|g\|_p)\,\|f+g\|_p^{p/q} \tag{4.25}$$

and solving this equation for $\|f+g\|_p$ (making use of Eq. (4.16)) implies Eq. (4.21). Now suppose that f and g are not identically zero and $p \in (1,\infty)$. Equality holds in Eq. (4.21) iff equality holds in Eq. (4.25) iff equality holds in Eq. (4.23) and Eq. (4.22). The latter happens iff

$$\operatorname{sgn}(f) \doteq \operatorname{sgn}(g) \text{ and} \\ \left(\frac{|f|}{\|f\|_p}\right)^p = \frac{|f+g|^p}{\|f+g\|_p^p} = \left(\frac{|g|}{\|g\|_p}\right)^p. \tag{4.26}$$

wherein we have used

$$\left(\frac{|f+g|^{p-1}}{\||f+g|^{p-1}\|_q}\right)^q = \frac{|f+g|^p}{\|f+g\|_p^p}.$$

Finally Eq. (4.26) is equivalent to $|f| = c\,|g|$ with $c = (\|f\|_p/\|g\|_p) > 0$ and this equality along with $\operatorname{sgn}(f) \doteq \operatorname{sgn}(g)$ implies $f = cg$. ■

4.6 Exercises

Exercise 4.1. Now suppose for each $n \in \mathbb{N} := \{1,2,\dots\}$ that $f_n : X \to \mathbb{R}$ is a function. Let

$$D := \{x \in X : \lim_{n\to\infty} f_n(x) = +\infty\}$$

show that

$$D = \cap_{M=1}^\infty \cup_{N=1}^\infty \cap_{n\ge N}\{x \in X : f_n(x) \ge M\}. \tag{4.27}$$

Exercise 4.2. Let $f_n : X \to \mathbb{R}$ be as in the last problem. Let

$$C := \{x \in X : \lim_{n\to\infty} f_n(x) \text{ exists in } \mathbb{R}\}.$$

Find an expression for C similar to the expression for D in (4.27). (Hint: use the Cauchy criteria for convergence.)

4.6.1 Limit Problems

Exercise 4.3. Show $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$.

Exercise 4.4. Suppose that $\limsup_{n\to\infty} a_n = M \in \bar{\mathbb{R}}$, show that there is a subsequence $\{a_{n_k}\}_{k=1}^\infty$ of $\{a_n\}_{n=1}^\infty$ such that $\lim_{k\to\infty} a_{n_k} = M$.

Exercise 4.5. Show that

$$\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{4.28}$$

provided that the right side of Eq. (4.28) is well defined, i.e. no $\infty - \infty$ or $-\infty + \infty$ type expressions. (It is OK to have $\infty + \infty = \infty$ or $-\infty - \infty = -\infty$, etc.)

Exercise 4.6. Suppose that $a_n \ge 0$ and $b_n \ge 0$ for all $n \in \mathbb{N}$. Show

$$\limsup_{n\to\infty}(a_n b_n) \le \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{4.29}$$

provided the right hand side of (4.29) is not of the form $0\cdot\infty$ or $\infty\cdot 0$.

Exercise 4.7. Prove Lemma 4.15.

Exercise 4.8. Prove Lemma 4.21.

4.6.2 Monotone and Dominated Convergence Theorem Problems

Exercise 4.9. Let $M < \infty$, show there are polynomials $p_n(t)$ and $q_n(t)$ for $n \in \mathbb{N}$ such that

$$\lim_{n\to\infty} \sup_{0\le t\le M} \left|\sqrt{t} - q_n(t)\right| = 0 \tag{4.30}$$

and

$$\lim_{n\to\infty} \sup_{|t|\le M} \left||t| - p_n(t)\right| = 0 \tag{4.31}$$

using the following outline.

1. Let $f(x) = \sqrt{1-x}$ for $|x| \le 1$ and use Taylor's theorem with integral remainder (see Eq. **??** of Appendix **??**), or analytic function theory if you know it, to show there are constants[1] $c_n > 0$ for $n \in \mathbb{N}$ such that
$$\sqrt{1-x} = 1 - \sum_{n=1}^{\infty} c_n x^n \text{ for all } |x| < 1. \tag{4.32}$$
2. Let $\tilde{q}_m(x) := 1 - \sum_{n=1}^{m} c_n x^n$. Use (4.32) to show $\sum_{n=1}^{\infty} c_n = 1$ and conclude from this that
$$\lim_{m\to\infty} \sup_{|x|\le 1} \left|\sqrt{1-x} - \tilde{q}_m(x)\right| = 0. \tag{4.33}$$
3. Conclude that $q_n(t) := \sqrt{M}\tilde{q}_n(1 - t/M)$ and $p_n(t) := q_n(t^2)$ for $n \in \mathbb{N}$ are polynomials verifying Eqs. (4.30) and (4.31) respectively.

Notation 4.32 *For $u_0 \in \mathbb{R}^n$ and $\delta > 0$, let $B_{u_0}(\delta) := \{x \in \mathbb{R}^n : |x - u_0| < \delta\}$ be the ball in $\mathbb{R}^n$ centered at u_0 with radius δ.*

Exercise 4.10. Suppose $U \subset \mathbb{R}^n$ is a set and $u_0 \in U$ is a point such that $U \cap (B_{u_0}(\delta) \setminus \{u_0\}) \neq \emptyset$ for all $\delta > 0$. Let $G : U \setminus \{u_0\} \to \mathbb{C}$ be a function on $U \setminus \{u_0\}$. Show that $\lim_{u\to u_0} G(u)$ exists and is equal to $\lambda \in \mathbb{C}$,[2] iff for all sequences $\{u_n\}_{n=1}^{\infty} \subset U \setminus \{u_0\}$ which converge to u_0 (i.e. $\lim_{n\to\infty} u_n = u_0$) we have $\lim_{n\to\infty} G(u_n) = \lambda$.

Exercise 4.11. Suppose that Y is a set, $U \subset \mathbb{R}^n$ is a set, and $f : U \times Y \to \mathbb{C}$ is a function satisfying:

[1] In fact $c_n := \frac{(2n-3)!!}{2^n n!}$, but this is not needed.

[2] More explicitly, $\lim_{u\to u_0} G(u) = \lambda$ means for every every $\epsilon > 0$ there exists a $\delta > 0$ such that
$$|G(u) - \lambda| < \epsilon \text{ whenever } u \in U \cap (B_{u_0}(\delta) \setminus \{u_0\}).$$

1. For each $y \in Y$, the function $u \in U \to f(u,y)$ is continuous on U.[3]
2. There is a summable function $g : Y \to [0,\infty)$ such that
$$|f(u,y)| \le g(y) \text{ for all } y \in Y \text{ and } u \in U.$$

Show that
$$F(u) := \sum_{y\in Y} f(u,y) \tag{4.34}$$
is a continuous function for $u \in U$.

Exercise 4.12. Suppose that Y is a set, $J = (a,b) \subset \mathbb{R}$ is an interval, and $f : J \times Y \to \mathbb{C}$ is a function satisfying:

1. For each $y \in Y$, the function $u \to f(u,y)$ is differentiable on J,
2. There is a summable function $g : Y \to [0,\infty)$ such that
$$\left|\frac{\partial}{\partial u} f(u,y)\right| \le g(y) \text{ for all } y \in Y \text{ and } u \in J.$$
3. There is a $u_0 \in J$ such that $\sum_{y\in Y} |f(u_0,y)| < \infty$.

Show:

a) for all $u \in J$ that $\sum_{y\in Y} |f(u,y)| < \infty$.
b) Let $F(u) := \sum_{y\in Y} f(u,y)$, show F is differentiable on J and that
$$\dot{F}(u) = \sum_{y\in Y} \frac{\partial}{\partial u} f(u,y).$$

(Hint: Use the mean value theorem.)

Exercise 4.13 (Differentiation of Power Series). Suppose $R > 0$ and $\{a_n\}_{n=0}^{\infty}$ is a sequence of complex numbers such that $\sum_{n=0}^{\infty} |a_n| r^n < \infty$ for all $r \in (0,R)$. Show, using Exercise 4.12, $f(x) := \sum_{n=0}^{\infty} a_n x^n$ is continuously differentiable for $x \in (-R,R)$ and

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1} = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

[3] To say $g := f(\cdot, y)$ is continuous on U means that $g : U \to \mathbb{C}$ is continuous relative to the metric on $\mathbb{R}^n$ restricted to U.

Exercise 4.14. Show the functions

$$e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!}, \tag{4.35}$$

$$\sin x := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \text{ and} \tag{4.36}$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \tag{4.37}$$

are infinitely differentiable and they satisfy

$$\frac{d}{dx} e^x = e^x \text{ with } e^0 = 1$$
$$\frac{d}{dx} \sin x = \cos x \text{ with } \sin(0) = 0$$
$$\frac{d}{dx} \cos x = -\sin x \text{ with } \cos(0) = 1.$$

Exercise 4.15. Continue the notation of Exercise 4.14.

1. Use the product and the chain rule to show,
$$\frac{d}{dx}\left[e^{-x} e^{(x+y)}\right] = 0$$
and conclude from this, that $e^{-x}e^{(x+y)} = e^y$ for all $x, y \in \mathbb{R}$. In particular taking $y = 0$ this implies that $e^{-x} = 1/e^x$ and hence that $e^{(x+y)} = e^x e^y$. Use this result to show $e^x \uparrow \infty$ as $x \uparrow \infty$ and $e^x \downarrow 0$ as $x \downarrow -\infty$.
Remark: since $e^x \geq \sum_{n=0}^{N} \frac{x^n}{n!}$ when $x \geq 0$, it follows that $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$ for any $n \in \mathbb{N}$, i.e. e^x grows at a rate faster than any polynomial in x as $x \to \infty$.
2. Use the product rule to show
$$\frac{d}{dx}\left(\cos^2 x + \sin^2 x\right) = 0$$
and use this to conclude that $\cos^2 x + \sin^2 x = 1$ for all $x \in \mathbb{R}$.

Exercise 4.16. Let $\{a_n\}_{n=-\infty}^{\infty}$ be a summable sequence of complex numbers, i.e. $\sum_{n=-\infty}^{\infty} |a_n| < \infty$. For $t \geq 0$ and $x \in \mathbb{R}$, define

$$F(t,x) = \sum_{n=-\infty}^{\infty} a_n e^{-tn^2} e^{inx},$$

where as usual $e^{ix} = \cos(x) + i\sin(x)$, this is motivated by replacing x in Eq. (4.35) by ix and comparing the result to Eqs. (4.36) and (4.37).

1. $F(t,x)$ is continuous for $(t,x) \in [0,\infty)\times\mathbb{R}$. **Hint:** Let $Y = \mathbb{Z}$ and $u = (t,x)$ and use Exercise 4.11.
2. $\partial F(t,x)/\partial t$, $\partial F(t,x)/\partial x$ and $\partial^2 F(t,x)/\partial x^2$ exist for $t > 0$ and $x \in \mathbb{R}$. **Hint:** Let $Y = \mathbb{Z}$ and $u = t$ for computing $\partial F(t,x)/\partial t$ and $u = x$ for computing $\partial F(t,x)/\partial x$ and $\partial^2 F(t,x)/\partial x^2$ via Exercise 4.12. In computing the t derivative, you should let $\varepsilon > 0$ and apply Exercise 4.12 with $t = u > \varepsilon$ and then afterwards let $\varepsilon \downarrow 0$.
3. F satisfies the heat equation, namely
$$\partial F(t,x)/\partial t = \partial^2 F(t,x)/\partial x^2 \text{ for } t > 0 \text{ and } x \in \mathbb{R}.$$

4.6.3 ℓ^p Exercises

Exercise 4.17. Generalize Proposition 4.26 as follows. Let $a \in [-\infty, 0]$ and $f : \mathbb{R} \cap [a,\infty) \to [0,\infty)$ be a continuous strictly increasing function such that $\lim_{s\to\infty} f(s) = \infty$, $f(a) = 0$ if $a > -\infty$ or $\lim_{s\to-\infty} f(s) = 0$ if $a = -\infty$. Also let $g = f^{-1}$, $b = f(0) \geq 0$,

$$F(s) = \int_0^s f(s')ds' \text{ and } G(t) = \int_0^t g(t')dt'.$$

Then for all $s, t \geq 0$,

$$st \leq F(s) + G(t \vee b) \leq F(s) + G(t)$$

and equality holds iff $t = f(s)$. In particular, taking $f(s) = e^s$, prove Young's inequality stating

$$st \leq e^s + (t \vee 1)\ln(t \vee 1) - (t \vee 1) \leq e^s + t\ln t - t,$$

where $s \vee t := \min(s,t)$. **Hint:** Refer to Figures 4.3 and 4.4..

Exercise 4.18. Using differential calculus, prove the following inequalities

1. For $y > 0$, let $g(x) := xy - e^x$ for $x \in \mathbb{R}$. Use calculus to compute the maximum of $g(x)$ and use this prove Young's inequality;
$$xy \leq e^x + y\ln y - y \text{ for } x \in \mathbb{R} \text{ and } y > 0.$$
2. For $p > 1$ and $y \geq 0$, let $g(x) := xy - x^p/p$ for $x \geq 0$. Again use calculus to compute the maximum of $g(x)$ and show that your result gives the following inequality;
$$xy \leq \frac{x^p}{p} + \frac{y^q}{q} \text{ for all } x, y \geq 0.$$
where $q = \frac{p}{p-1}$, i.e. $\frac{1}{q} = 1 - \frac{1}{p}$.

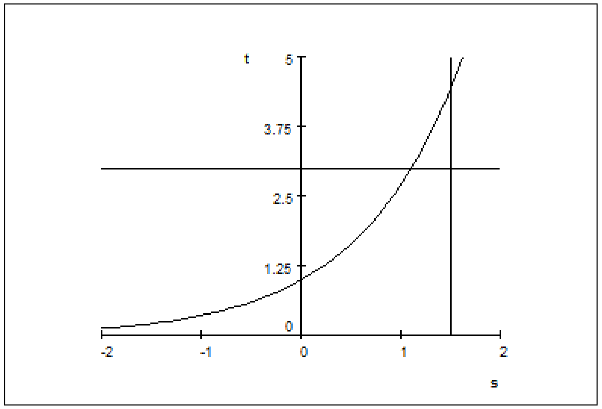


Fig. 4.3. Comparing areas when $t \geq b$ goes the same way as in the text.

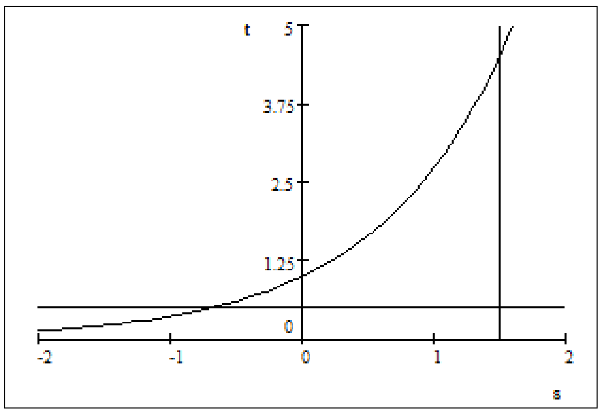


Fig. 4.4. When $t \leq b$, notice that $g(t) \leq 0$ but $G(t) \geq 0$. Also notice that $G(t)$ is no longer needed to estimate st.

3. Suppose now that $u : [0, \infty) \to [0, \infty)$ is a C^1 - function such that: $u(0) = 0$, $\lim_{x\to\infty} \frac{u(x)}{x} = \infty$, and $u'(x) > 0$ for all $x > 0$. Show

$$xy \leq u(x) + v(y) \text{ for all } x, y \geq 0,$$

where $v(y) = y(u')^{-1}(y) - u\left((u')^{-1}(y)\right)$. **Hint:** consider the function, $g(x) := xy - u(x)$.

Part II

Measure Theory I.

5

What are measures and why “measurable” sets

Throughout this chapter, we will let X and Ω be sets. Our goal is to study “measures” and their related integrals on these sets. Before giving a (preliminary) definition of a measure let me give some “physical” examples;

1. Suppose that Ω is a region in space filled with some material. To each subset $A \subset \Omega$ we might let $\mu(A)$ denote the weight (or volume, or monetary value, heat energy contained in A) of the material in Ω.
2. Suppose that Ω is a region in space filled with charged particles, to each subset $A \subset \Omega$ we might let $\mu(A)$ denote the total charge of the particles contained in A. (This is an example of a signed measure, i.e. it might take both positive and negative values.)
3. Perhaps Ω represents the face of a dart board at which drunk patrons are attempting to hit in the center. For $A \subset \Omega$, we might let $P(A)$ denote the total number of darts which landed in A.

With these examples in mind let us formalize (preliminarliy) the notion of a measure on a set X. (Given the physical examples just mentioned, I hope the axioms in the next Definition 5.1 look reasonable to the reader.)

Definition 5.1 (Preliminary). *A **positive**[1] **measure** μ “on” a set X is a function $\mu : 2^X \to [0, \infty]$ such that*

1. $\mu(\emptyset) = 0$
2. ***Additivity.*** *If A and B are disjoint subsets of X, i.e. $A \cap B = AB = \emptyset$, then $\mu(A \cup B) = \mu(A) + \mu(B)$.*
3. ***Continuity.*** *Suppose that $\{A_n\}_{n=1}^\infty \subset 2^X$ with $A_n \uparrow$ (i.e. $A_n \subset A_{n+1}$ for all n), then*
$$\mu(\cup_{n=1}^\infty A_n) = \lim_{n\to\infty} \mu(A_n).$$

Notation 5.2 *Given $\{A_n\}_{n=1}^\infty \subset 2^X$, we write $\sum_{n=1}^\infty A_n$ to denote $\cup_{n=1}^\infty A_n$ under the additional assumption that $A_n \cap A_m = \emptyset$ for all $m \neq n$.*

Lemma 5.3 (Reformulations of Continuity). *If $\mu : 2^X \to [0, \infty]$ satisfies items 1. and 2. of Definition 5.1 then μ satisfies item 3 Definition 5.1 iff*

$$\mu(A) = \sum_{n=1}^\infty \mu(A_n) \text{ whenever } A = \sum_{n=1}^\infty A_n. \tag{5.1}$$

Moroever if $\mu(X) < \infty$ then μ satisfies item 3. iff $\lim_{n\to\infty} \mu(A_n) = \mu(\cap_{n=1}^\infty A_n)$ whenever $\{A_n\}_{n=1}^\infty \subset 2^X$ with $A_n \downarrow$, i.e. $A_n \supset A_{n+1}$ for all n.

Proof. First observe that if $A \subset B$ then $B = (B \setminus A) \cup A$ with $(B \setminus A) \cap A = \emptyset$ and hence $\mu(B) = \mu(B \setminus A) + \mu(A)$, i.e.

$$\mu(B \setminus A) = \mu(B) - \mu(A) \text{ for all } A \subset B \subset X.$$

Now suppose that μ satisfies item 3. Definition 5.1 (continuity). If $A := \sum_{n=1}^\infty A_n$ and $B_k := \sum_{n=1}^k A_n$, then $B_k \uparrow A$ as $k \uparrow \infty$ and therefore

$$\begin{aligned}\mu(A) &= \lim_{k\to\infty} \mu(B_k) = \lim_{k\to\infty} \mu\left(\sum_{n=1}^k A_n\right) \\ &= \lim_{k\to\infty} \sum_{n=1}^k \mu(A_n) \text{ (by finite additivity and induction)} \\ &= \sum_{n=1}^\infty \mu(A_n).\end{aligned}$$

Conversiely, suppose that Eq. (5.1) holds whenever $A_n \cap A_m = \emptyset$ for all $m \neq n$. Given $B_n \uparrow B$, let $A_1 = B_1$ and define A_n inductively by $A_n = B_n \setminus A_{n-1}$. Then $B = \sum_{n=1}^\infty A_n$ and therfore

$$\mu(B) = \sum_{n=1}^\infty \mu(A_n) = \lim_{k\to\infty} \sum_{n=1}^k \mu(A_n) = \lim_{k\to\infty} \mu\left(\sum_{n=1}^k A_n\right) = \lim_{k\to\infty} \mu(B_k),$$

i.e. μ satisfies item 3. Definition 5.1. The second assertion is left as an exercise to the reader. ■

Example 5.4 (Counting Type Measures). Let $\mu(A) = \#(A)$ – the number of points in A. Then μ is a measure on X. More generally, if $\Lambda \subset X$ is any fixed subset of X, then $\mu_\Lambda(A) := \#(\Lambda \cap A)$ defines a measure on X. Even more generally, if $\lambda : X \to [0, \infty]$ is any function, then

[1] We will deal with signed measures later.

$$\mu_\lambda(A) := \sum_{x\in A} \lambda(x) := \sup_{\Lambda \subset_f A} \sum_{x\in\Lambda} \lambda(x)$$

defined a measure on X.

The measures we often most want to understand are those measure lengths, areas, or more generally n – dimensional volumes. For example, suppose we take $X = \mathbb{R}^2$ and let $\mu(A)$ denote the “area” of a subset $A \subset X$. I think we would all agreee that

$$\mu((a,b] \times (c,d]) = (b-a)(d-c) \text{ for } -\infty < a < b < \infty \text{ and } -\infty < c < d < \infty.$$

With this basic building block we might want to compute the area the unit disk. One way to try to do this is (see Figure 5.1) to approximate the disk by finite unions of disjoint rectangles. By the additivity axiom of μ we can compute the area of the approximations and then by using the continuity axiom we could take a limit of these approximations to find the area of the unit disk.

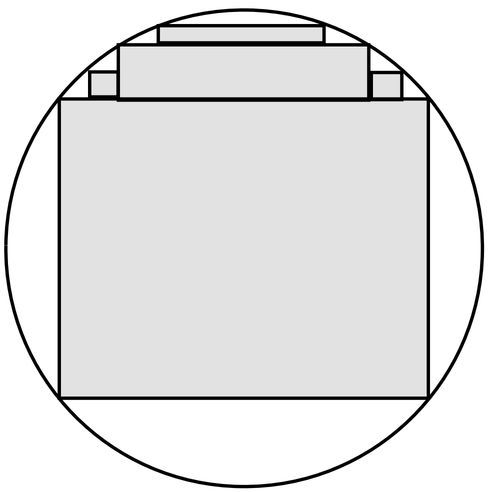


Fig. 5.1. Here is an indication of how one might approximate a disk by finite disjoint union of rectangles.

Definition 5.1 is all well fine except for the unfortunate fact that measures (like areas and volumes) with vary natural and desirable properties often do not exist. We give a couple of example illustrating this point now.

Theorem 5.5 (No-Go Theorem 1). *Let $S = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle. Then there is no measure $\mu : 2^S \to [0,1]$ such that $0 < \mu(S) < \infty$ that is invariant under rotations.*

Proof. We are going to use the fact proved below in Proposition 8.3 (of Lemma 5.3 above), that the continuity condition on μ is equivalent to the σ – additivity of μ. For $z \in S$ and $N \subset S$ let

$$zN := \{zn \in S : n \in N\}, \tag{5.2}$$

that is to say $e^{i\theta}N$ is the set N rotated counter clockwise by angle θ. By assumption, we are supposing that

$$\mu(zN) = \mu(N) \tag{5.3}$$

for all $z \in S$ and $N \subset S$.

Let

$$R := \{z = e^{i2\pi t} : t \in \mathbb{Q}\} = \{z = e^{i2\pi t} : t \in [0,1) \cap \mathbb{Q}\}$$

– a countable subgroup of S. As above R acts on S by rotations and divides S up into equivalence classes, where $z, w \in S$ are equivalent if $z = rw$ for some $r \in R$. Choose (using the axiom of choice) one representative point n from each of these equivalence classes and let $N \subset S$ be the set of these representative points. Then every point $z \in S$ may be uniquely written as $z = nr$ with $n \in N$ and $r \in R$. That is to say

$$S = \sum_{r\in R} (rN) \tag{5.4}$$

where $\sum_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (5.3) and (5.4),

$$1 = \mu(S) = \sum_{r\in R} \mu(rN) = \sum_{r\in R} \mu(N). \tag{5.5}$$

We have thus arrived at a contradiction, since the right side of Eq. (5.5) is either equal to 0 or to ∞ depending on whether $\mu(N) = 0$ or $\mu(N) > 0$. ∎

Theorem 5.6. *There is no measure $\mu : 2^{\mathbb{R}} \to [0,\infty]$ such that*

1. *$\mu([a,b)) = (b-a)$ for all $a < b$ and*
2. *is translation invariant, i.e. $\mu(A+x) = \mu(A)$ for all $x \in \mathbb{R}$ and $A \in 2^{\mathbb{R}}$, where*

$$A + x := \{y + x : y \in A\} \subset \mathbb{R}.$$

In fact the theorem is still true even if (1) is replaced by the weaker condition that $0 < \mu((0,1]) < \infty$.

The counting measure $\mu(A) = \#(A)$ is translation invariant. However $\mu((0,1]) = \infty$ in this case and so μ does not satisfy condition 1.

Proof. First proof. Let us identify $[0,1)$ with the unit circle $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ by the map

$$\phi(t) = e^{i2\pi t} = (\cos 2\pi t + i \sin 2\pi t) \in S^1$$

for $t \in [0,1)$. Using this identification we may use μ to define a function ν on 2^{S^1} by $\nu(\phi(A)) = \mu(A)$ for all $A \subset [0,1)$. This new function is a measure on S^1 with the property that $0 < \nu((0,1]) < \infty$. For $z \in S^1$ and $N \subset S^1$ let

$$zN := \{zn \in S^1 : n \in N\}, \tag{5.6}$$

that is to say $e^{i\theta}N$ is N rotated counter clockwise by angle θ. We now claim that ν is invariant under these rotations, i.e.

$$\nu(zN) = \nu(N) \tag{5.7}$$

for all $z \in S^1$ and $N \subset S^1$. To verify this, write $N = \phi(A)$ and $z = \phi(t)$ for some $t \in [0,1)$ and $A \subset [0,1)$. Then

$$\phi(t)\phi(A) = \phi(t + A \bmod 1)$$

where for $A \subset [0,1)$ and $t \in [0,1)$,

$$\begin{aligned} t + A \bmod 1 &:= \{a + t \bmod 1 \in [0,1) : a \in A\} \\ &= ((t + A) \cap \{a < 1 - t\}) \cup ([(t-1) + A] \cap \{a \geq 1 - t\}). \end{aligned}$$

Thus

$$\begin{aligned} \nu(\phi(t)\phi(A)) &= \mu(t + A \bmod 1) \\ &= \mu\left((a + A \cap \{a < 1 - t\}) \cup ((t-1) + A \cap \{a \geq 1 - t\})\right) \\ &= \mu\left((a + A \cap \{a < 1 - t\})\right) + \mu\left(((t-1) + A \cap \{a \geq 1 - t\})\right) \\ &= \mu\left(A \cap \{a < 1 - t\}\right) + \mu\left(A \cap \{a \geq 1 - t\}\right) \\ &= \mu\left((A \cap \{a < 1 - t\}) \cup (A \cap \{a \geq 1 - t\})\right) \\ &= \mu(A) = \nu(\phi(A)). \end{aligned}$$

Therefore it suffices to prove that no finite non-trivial measure ν on S^1 such that Eq. (5.7) holds. To do this we will "construct" a non-measurable set $N = \phi(A)$ for some $A \subset [0,1)$. Let

$$R := \{z = e^{i2\pi t} : t \in \mathbb{Q}\} = \{z = e^{i2\pi t} : t \in [0,1) \cap \mathbb{Q}\}$$

– a countable subgroup of S^1. As above R acts on S^1 by rotations and divides S^1 up into equivalence classes, where $z, w \in S^1$ are equivalent if $z = rw$ for some $r \in R$. Choose (using the axiom of choice) one representative point n from each of these equivalence classes and let $N \subset S^1$ be the set of these representative points. Then every point $z \in S^1$ may be uniquely written as $z = nr$ with $n \in N$ and $r \in R$. That is to say

$$S^1 = \coprod_{r \in R} (rN) \tag{5.8}$$

where $\coprod_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (5.7) and (5.8),

$$\nu(S^1) = \sum_{r \in R} \nu(rN) = \sum_{r \in R} \nu(N).$$

The right member from this equation is either 0 or ∞, 0 if $\nu(N) = 0$ and ∞ if $\nu(N) > 0$. In either case it is not equal $\nu(S^1) \in (0,1)$. Thus we have reached the desired contradiction. ■

Proof. Second proof of Theorem 5.6. For $N \subset [0,1)$ and $\alpha \in [0,1)$, let

$$\begin{aligned} N^\alpha &= N + \alpha \bmod 1 \\ &= \{a + \alpha \bmod 1 \in [0,1) : a \in N\} \\ &= (\alpha + N \cap \{a < 1 - \alpha\}) \cup ((\alpha - 1) + N \cap \{a \geq 1 - \alpha\}). \end{aligned}$$

Then

$$\begin{aligned} \mu(N^\alpha) &= \mu(\alpha + N \cap \{a < 1 - \alpha\}) + \mu((\alpha - 1) + N \cap \{a \geq 1 - \alpha\}) \\ &= \mu(N \cap \{a < 1 - \alpha\}) + \mu(N \cap \{a \geq 1 - \alpha\}) \\ &= \mu(N \cap \{a < 1 - \alpha\} \cup (N \cap \{a \geq 1 - \alpha\})) \\ &= \mu(N). \end{aligned} \tag{5.9}$$

We will now construct a bad set N which coupled with Eq. (5.9) will lead to a contradiction. Set

$$Q_x := \{x + r \in \mathbb{R} : r \in \mathbb{Q}\} = x + \mathbb{Q}.$$

Notice that $Q_x \cap Q_y \neq \emptyset$ implies that $Q_x = Q_y$. Let $\mathcal{O} = \{Q_x : x \in \mathbb{R}\}$ – the orbit space of the $\mathbb{Q}$ action. For all $A \in \mathcal{O}$ choose $f(A) \in [0,1/3) \cap A$[2] and define $N = f(\mathcal{O})$. Then observe:

1. $f(A) = f(B)$ implies that $A \cap B \neq \emptyset$ which implies that $A = B$ so that f is injective.
2. $\mathcal{O} = \{Q_n : n \in N\}$.

Let R be the countable set,

$$R := \mathbb{Q} \cap [0,1).$$

We now claim that

[2] We have used the Axiom of choice here, i.e. $\prod_{A \in \mathcal{F}} (A \cap [0,1/3]) \neq \emptyset$

$$N^r \cap N^s = \emptyset \text{ if } r \neq s \text{ and} \tag{5.10}$$

$$[0,1) = \cup_{r\in R} N^r. \tag{5.11}$$

Indeed, if $x \in N^r \cap N^s \neq \emptyset$ then $x = r + n \bmod 1$ and $x = s + n' \bmod 1$, then $n - n' \in \mathbb{Q}$, i.e. $Q_n = Q_{n'}$. That is to say, $n = f(Q_n) = f(Q_{n'}) = n'$ and hence that $s = r \bmod 1$, but $s, r \in [0,1)$ implies that $s = r$. Furthermore, if $x \in [0,1)$ and $n := f(Q_x)$, then $x - n = r \in \mathbb{Q}$ and $x \in N^{r \bmod 1}$. Now that we have constructed N, we are ready for the contradiction. By Equations (5.9–5.11) we find

$$\begin{aligned} 1 = \mu([0,1)) &= \sum_{r\in R} \mu(N^r) = \sum_{r\in R} \mu(N) \\ &= \begin{cases} \infty & \text{if } \mu(N) > 0 \\ 0 & \text{if } \mu(N) = 0 \end{cases}. \end{aligned}$$

which is certainly inconsistent. Incidentally we have just produced an example of so called "non – measurable" set. ■

Because of Theorems 5.5 and 5.6, we have to in general relinquish the idea that measure μ can be defined on all of 2^X. In other words we are going to have to restrict our attention to only measuring some sub-collection, $\mathcal{B} \subset 2^X$, of all subsets of X. We will refer to $\mathcal{B}$ as the collection of **measurable** sets. We will developed this below., it is necessary to modify Definition 5.1. Our revised notion of a measure will appear in Definition **??** of Chapter **??** below.

Set Operations

Let $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be the non-negative integers and $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$ – the positive and negative integers including 0, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers, and $\mathbb{C}$ the complex numbers. We will also use $\mathbb{F}$ to stand for either of the fields $\mathbb{R}$ or $\mathbb{C}$.

Notation 6.1 *Given two sets X and Y, let Y^X denote the collection of all functions $f : X \to Y$. If $X = \mathbb{N}$, we will say that $f \in Y^{\mathbb{N}}$ is a sequence with values in Y and often write f_n for $f(n)$ and express f as $\{f_n\}_{n=1}^{\infty}$. If $X = \{1, 2, \ldots, N\}$, we will write Y^N in place of $Y^{\{1,2,\ldots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \ldots, f_N)$ where $f_n = f(n)$.*

Notation 6.2 *More generally if $\{X_\alpha : \alpha \in A\}$ is a collection of non-empty sets, let $X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map defined by $\pi_\alpha(x) = x_\alpha$. If If $X_\alpha = X$ for some fixed space X, then we will write $\prod_{\alpha \in A} X_\alpha$ as X^A rather than X_A.*

Recall that an element $x \in X_A$ is a "**choice function**," i.e. an assignment $x_\alpha := x(\alpha) \in X_\alpha$ for each $\alpha \in A$. The **axiom of choice** states that $X_A \neq \emptyset$ provided that $X_\alpha \neq \emptyset$ for each $\alpha \in A$.

Notation 6.3 *Given a set X, let 2^X denote the* ***power set*** *of X – the collection of all subsets of X including the empty set.*

The reason for writing the power set of X as 2^X is that if we think of 2 meaning $\{0, 1\}$, then an element of $a \in 2^X = \{0, 1\}^X$ is completely determined by the set

$$A := \{x \in X : a(x) = 1\} \subset X.$$

In this way elements in $\{0, 1\}^X$ are in one to one correspondence with subsets of X.

For $A \in 2^X$ let

$$A^c := X \setminus A = \{x \in X : x \notin A\}$$

and more generally if $A, B \subset X$ let

$$B \setminus A := \{x \in B : x \notin A\} = B \cap A^c.$$

We also define the symmetric difference of A and B by

$$A \triangle B := (B \setminus A) \cup (A \setminus B).$$

As usual if $\{A_\alpha\}_{\alpha \in I}$ is an indexed collection of subsets of X we define the union and the intersection of this collection by

$$\cup_{\alpha \in I} A_\alpha := \{x \in X : \exists\, \alpha \in I \ni x \in A_\alpha\} \text{ and}$$
$$\cap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \,\forall\, \alpha \in I\}.$$

Notation 6.4 *We will also write $\sum_{\alpha \in I} A_\alpha$ for $\cup_{\alpha \in I} A_\alpha$ in the case that $\{A_\alpha\}_{\alpha \in I}$ are pairwise disjoint, i.e. $A_\alpha \cap A_\beta = \emptyset$ if $\alpha \neq \beta$.*

Notice that $\cup$ is closely related to $\exists$ and $\cap$ is closely related to $\forall$. For example let $\{A_n\}_{n=1}^{\infty}$ be a sequence of subsets from X and define

$$\inf_{k \geq n} A_n := \cap_{k \geq n} A_k, \quad \sup_{k \geq n} A_n := \cup_{k \geq n} A_k,$$

$$\limsup_{n \to \infty} A_n := \inf_n \sup_{k \geq n} A_k = \{x \in X : \#\{n : x \in A_n\} = \infty\} =: \{A_n \text{ i.o.}\}$$

and

$$\liminf_{n \to \infty} A_n := \sup_n \inf_{k \geq n} A_k = \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\} =: \{A_n \text{ a.a.}\}.$$

(One should read $\{A_n \text{ i.o.}\}$ as A_n infinitely often and $\{A_n \text{ a.a.}\}$ as A_n almost always.) Then $x \in \{A_n \text{ i.o.}\}$ iff

$$\forall N \in \mathbb{N}\ \exists\ n \geq N \ni x \in A_n$$

and this may be expressed as

$$\{A_n \text{ i.o.}\} = \cap_{N=1}^{\infty} \cup_{n \geq N} A_n.$$

Similarly, $x \in \{A_n \text{ a.a.}\}$ iff

$$\exists\ N \in \mathbb{N} \ni \forall\ n \geq N, \ \ x \in A_n$$

which may be written as

$$\{A_n \text{ a.a.}\} = \cup_{N=1}^{\infty} \cap_{n \geq N} A_n.$$

Definition 6.5. *Given a set $A \subset X$, let*

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

*be the **indicator function** of A.*

Lemma 6.6 (Properties of inf and sup). *We have:*

1. $\left(\cup_n A_n\right)^c = \cap_n A_n^c$,
2. $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ a.a.}\}$,
3. $\limsup_{n\to\infty} A_n = \{x \in X : \sum_{n=1}^\infty 1_{A_n}(x) = \infty\}$,
4. $\liminf_{n\to\infty} A_n = \left\{x \in X : \sum_{n=1}^\infty 1_{A_n^c}(x) < \infty\right\}$,
5. $\sup_{k\geq n} 1_{A_k}(x) = 1_{\cup_{k\geq n} A_k} = 1_{\sup_{k\geq n} A_k}$,
6. $\inf_{k\geq n} 1_{A_k}(x) = 1_{\cap_{k\geq n} A_k} = 1_{\inf_{k\geq n} A_k}$,
7. $1_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} 1_{A_n}$, and
8. $1_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} 1_{A_n}$.

Proof. These results follow fairly directly from the definitions and so the proof is left to the reader. (The reader should definitely provide a proof for herself.) ■

Definition 6.7. *A set X is said to be **countable** if is empty or there is an injective function $f : X \to \mathbb{N}$, otherwise X is said to be **uncountable.***

Lemma 6.8 (Basic Properties of Countable Sets).

1. *If $A \subset X$ is a subset of a countable set X then A is countable.*
2. *Any infinite subset $\Lambda \subset \mathbb{N}$ is in one to one correspondence with $\mathbb{N}$.*
3. *A non-empty set X is countable iff there exists a surjective map, $g : \mathbb{N} \to X$.*
4. *If X and Y are countable then $X \times Y$ is countable.*
5. *Suppose for each $m \in \mathbb{N}$ that A_m is a countable subset of a set X, then $A = \cup_{m=1}^\infty A_m$ is countable. In short, the countable union of countable sets is still countable.*
6. *If X is an infinite set and Y is a set with at least two elements, then Y^X is uncountable. In particular 2^X is uncountable for any infinite set X.*

Proof. 1. If $f : X \to \mathbb{N}$ is an injective map then so is the restriction, $f|_A$, of f to the subset A. 2. Let $f(1) = \min \Lambda$ and define f inductively by

$$f(n+1) = \min\left(\Lambda \setminus \{f(1), \dots, f(n)\}\right).$$

Since Λ is infinite the process continues indefinitely. The function $f : \mathbb{N} \to \Lambda$ defined this way is a bijection.

3. If $g : \mathbb{N} \to X$ is a surjective map, let

$$f(x) = \min g^{-1}(\{x\}) = \min\{n \in \mathbb{N} : f(n) = x\}.$$

Then $f : X \to \mathbb{N}$ is injective which combined with item

2. (taking $\Lambda = f(X)$) shows X is countable. Conversely if $f : X \to \mathbb{N}$ is injective let $x_0 \in X$ be a fixed point and define $g : \mathbb{N} \to X$ by $g(n) = f^{-1}(n)$ for $n \in f(X)$ and $g(n) = x_0$ otherwise.

4. Let us first construct a bijection, h, from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$. To do this put the elements of $\mathbb{N} \times \mathbb{N}$ into an array of the form

$$\begin{pmatrix} (1,1) & (1,2) & (1,3) & \dots \\ (2,1) & (2,2) & (2,3) & \dots \\ (3,1) & (3,2) & (3,3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and then "count" these elements by counting the sets $\{(i,j) : i + j = k\}$ one at a time. For example let $h(1) = (1,1)$, $h(2) = (2,1)$, $h(3) = (1,2)$, $h(4) = (3,1)$, $h(5) = (2,2)$, $h(6) = (1,3)$ and so on. If $f : \mathbb{N} \to X$ and $g : \mathbb{N} \to Y$ are surjective functions, then the function $(f \times g) \circ h : \mathbb{N} \to X \times Y$ is surjective where $(f \times g)(m,n) := (f(m), g(n))$ for all $(m,n) \in \mathbb{N} \times \mathbb{N}$.

5. If $A = \emptyset$ then A is countable by definition so we may assume $A \neq \emptyset$. With out loss of generality we may assume $A_1 \neq \emptyset$ and by replacing A_m by A_1 if necessary we may also assume $A_m \neq \emptyset$ for all m. For each $m \in \mathbb{N}$ let $a_m : \mathbb{N} \to A_m$ be a surjective function and then define $f : \mathbb{N} \times \mathbb{N} \to \cup_{m=1}^\infty A_m$ by $f(m,n) := a_m(n)$. The function f is surjective and hence so is the composition, $f \circ h : \mathbb{N} \to \cup_{m=1}^\infty A_m$, where $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is the bijection defined above.

6. Let us begin by showing $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$ is uncountable. For sake of contradiction suppose $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is a surjection and write $f(n)$ as $(f_1(n), f_2(n), f_3(n), \dots)$. Now define $a \in \{0,1\}^{\mathbb{N}}$ by $a_n := 1 - f_n(n)$. By construction $f_n(n) \neq a_n$ for all n and so $a \notin f(\mathbb{N})$. This contradicts the assumption that f is surjective and shows $2^{\mathbb{N}}$ is uncountable. For the general case, since $Y_0^X \subset Y^X$ for any subset $Y_0 \subset Y$, if Y_0^X is uncountable then so is Y^X. In this way we may assume Y_0 is a two point set which may as well be $Y_0 = \{0,1\}$. Moreover, since X is an infinite set we may find an injective map $x : \mathbb{N} \to X$ and use this to set up an injection, $i : 2^{\mathbb{N}} \to 2^X$ by setting $i(A) := \{x_n : n \in \mathbb{N}\} \subset X$ for all $A \subset \mathbb{N}$. If 2^X were countable we could find a surjective map $f : 2^X \to \mathbb{N}$ in which case $f \circ i : 2^{\mathbb{N}} \to \mathbb{N}$ would be surjective as well. However this is impossible since we have already seed that $2^{\mathbb{N}}$ is uncountable. ■

6.1 Exercises

Let $f : X \to Y$ be a function and $\{A_i\}_{i\in I}$ be an indexed family of subsets of Y, verify the following assertions.

Exercise 6.1. $(\cap_{i\in I}A_i)^c = \cup_{i\in I}A_i^c$.

Exercise 6.2. Suppose that $B \subset Y$, show that $B \setminus (\cup_{i\in I}A_i) = \cap_{i\in I}(B \setminus A_i)$.

Exercise 6.3. $f^{-1}(\cup_{i\in I}A_i) = \cup_{i\in I}f^{-1}(A_i)$.

Exercise 6.4. $f^{-1}(\cap_{i\in I}A_i) = \cap_{i\in I}f^{-1}(A_i)$.

Exercise 6.5. Find a counterexample which shows that $f(C \cap D) = f(C) \cap f(D)$ need not hold.

Example 6.9. Let $X = \{a, b, c\}$ and $Y = \{1, 2\}$ and define $f(a) = f(b) = 1$ and $f(c) = 2$. Then $\emptyset = f(\{a\} \cap \{b\}) \neq f(\{a\}) \cap f(\{b\}) = \{1\}$ and $\{1, 2\} = f(\{a\}^c) \neq f(\{a\})^c = \{2\}$.

6.2 Algebraic sub-structures of sets

Definition 6.10. *A collection of subsets $\mathcal{A}$ of a set X is a π – **system** or **multiplicative system** if $\mathcal{A}$ is closed under taking finite intersections.*

Definition 6.11. *A collection of subsets $\mathcal{A}$ of a set X is an **algebra (Field)** if*

1. *$\emptyset, X \in \mathcal{A}$*
2. *$A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$*
3. *$\mathcal{A}$ is closed under finite unions, i.e. if $A_1, \ldots, A_n \in \mathcal{A}$ then $A_1 \cup \cdots \cup A_n \in \mathcal{A}$. In view of conditions 1. and 2., 3. is equivalent to*

3′. *$\mathcal{A}$ is closed under finite intersections.*

Definition 6.12. *A collection of subsets $\mathcal{B}$ of X is a σ – **algebra** (or sometimes called a σ – **field**) if $\mathcal{B}$ is an algebra which also closed under countable unions, i.e. if $\{A_i\}_{i=1}^{\infty} \subset \mathcal{B}$, then $\cup_{i=1}^{\infty}A_i \in \mathcal{B}$. (Notice that since $\mathcal{B}$ is also closed under taking complements, $\mathcal{B}$ is also closed under taking countable intersections.)*

Example 6.13. Here are some examples of algebras.

1. $\mathcal{B} = 2^X$, then $\mathcal{B}$ is a σ – algebra.
2. $\mathcal{B} = \{\emptyset, X\}$ is a σ – algebra called the trivial σ – field.
3. Let $X = \{1, 2, 3\}$, then $\mathcal{A} = \{\emptyset, X, \{1\}, \{2, 3\}\}$ is an algebra while, $\mathcal{S} := \{\emptyset, X, \{2, 3\}\}$ is a not an algebra but is a π – system.

Proposition 6.14. *Let $\mathcal{E}$ be any collection of subsets of X. Then there exists a unique smallest algebra $\mathcal{A}(\mathcal{E})$ and σ – algebra $\sigma(\mathcal{E})$ which contains $\mathcal{E}$.*

Proof. Simply take

$$\mathcal{A}(\mathcal{E}) := \bigcap\{\mathcal{A} : \mathcal{A} \text{ is an algebra such that } \mathcal{E} \subset \mathcal{A}\}$$

and

$$\sigma(\mathcal{E}) := \bigcap\{\mathcal{M} : \mathcal{M} \text{ is a } \sigma - \text{algebra such that } \mathcal{E} \subset \mathcal{M}\}.$$

■

Example 6.15. Suppose $X = \{1, 2, 3\}$ and $\mathcal{E} = \{\emptyset, X, \{1, 2\}, \{1, 3\}\}$, see Figure 6.1. Then

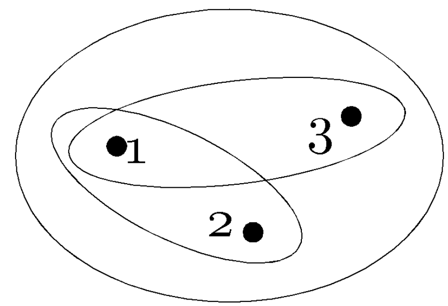


Fig. 6.1. A collection of subsets.

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = 2^X.$$

On the other hand if $\mathcal{E} = \{\{1, 2\}\}$, then $\mathcal{A}(\mathcal{E}) = \{\emptyset, X, \{1, 2\}, \{3\}\}$.

Exercise 6.6. Suppose that $\mathcal{E}_i \subset 2^X$ for $i = 1, 2$. Show that $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \mathcal{A}(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \mathcal{A}(\mathcal{E}_1)$. Similarly show, $\sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \sigma(\mathcal{E}_1)$. Give a simple example where $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ while $\mathcal{E}_1 \neq \mathcal{E}_2$.

In this course we will often be interested in the Borel σ – algebra on a topological space.

Definition 6.16 (Borel σ – field). *The **Borel σ – algebra**, $\mathcal{B} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$, on $\mathbb{R}$ is the smallest σ -field containing all of the open subsets of $\mathbb{R}$. More generally if (X, τ) is a topological space, the Borel σ – algebra on X is $\mathcal{B}_X := \sigma(\tau)$ – i.e. the smallest σ – algebra containing all open (closed) subsets of X.*

Exercise 6.7. Verify the Borel σ – algebra, $\mathcal{B}_{\mathbb{R}}$, is generated by any of the following collection of sets:

$$1.\ \{(a,\infty) : a \in \mathbb{R}\},\ 2.\ \{(a,\infty) : a \in \mathbb{Q}\} \text{ or } 3.\ \{[a,\infty) : a \in \mathbb{Q}\}.$$

Hint: make use of Exercise 6.6.

We will postpone a more in depth study of σ – algebras until later. For now, let us concentrate on understanding the the simpler notion of an algebra.

Definition 6.17. *Let X be a set. We say that a family of sets $\mathcal{F} \subset 2^X$ is a* ***partition*** *of X if distinct members of $\mathcal{F}$ are disjoint and if X is the union of the sets in $\mathcal{F}$.*

Example 6.18. Let X be a set and $\mathcal{E} = \{A_1, \dots, A_n\}$ where $A_1, \dots, A_n$ is a partition of X. In this case

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset \{1,2,\dots,n\}\}$$

where $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. Notice that

$$\#\left(\mathcal{A}(\mathcal{E})\right) = \#(2^{\{1,2,\dots,n\}}) = 2^n.$$

Example 6.19. Suppose that X is a set and that $\mathcal{A} \subset 2^X$ is a finite algebra, i.e. $\#\left(\mathcal{A}\right) < \infty$. For each $x \in X$ let

$$A_x = \cap \{A \in \mathcal{A} : x \in A\} \in \mathcal{A},$$

wherein we have used $\mathcal{A}$ is finite to insure $A_x \in \mathcal{A}$. Hence A_x is the smallest set in $\mathcal{A}$ which contains x.

Now suppose that $y \in X$. If $x \in A_y$ then $A_x \subset A_y$ so that $A_x \cap A_y = A_x$. On the other hand, if $x \notin A_y$ then $x \in A_x \setminus A_y$ and therefore $A_x \subset A_x \setminus A_y$, i.e. $A_x \cap A_y = \emptyset$. Therefore we have shown, either $A_x \cap A_y = \emptyset$ or $A_x \cap A_y = A_x$. By reversing the roles of x and y it also follows that either $A_y \cap A_x = \emptyset$ or $A_y \cap A_x = A_y$. Therefore we may conclude, either $A_x = A_y$ or $A_x \cap A_y = \emptyset$ for all $x, y \in X$.

Let us now define $\{B_i\}_{i=1}^k$ to be an enumeration of $\{A_x\}_{x\in X}$. It is a straightforward to conclude that

$$\mathcal{A} = \{\cup_{i\in\Lambda} B_i : \Lambda \subset \{1,2,\dots,k\}\}.$$

For example observe that for any $A \in \mathcal{A}$, we have $A = \cup_{x\in A} A_x = \cup_{i\in\Lambda} B_i$ where $\Lambda := \{i : B_i \subset A\}$.

Proposition 6.20. *Suppose that $\mathcal{B} \subset 2^X$ is a σ – algebra and $\mathcal{B}$ is at most a countable set. Then there exists a unique* ***finite*** *partition $\mathcal{F}$ of X such that $\mathcal{F} \subset \mathcal{B}$ and every element $B \in \mathcal{B}$ is of the form*

$$B = \cup \{A \in \mathcal{F} : A \subset B\}. \tag{6.1}$$

In particular $\mathcal{B}$ is actually a finite set and $\#\left(\mathcal{B}\right) = 2^n$ for some $n \in \mathbb{N}$.

Proof. We proceed as in Example 6.19. For each $x \in X$ let

$$A_x = \cap \{A \in \mathcal{B} : x \in A\} \in \mathcal{B},$$

wherein we have used $\mathcal{B}$ is a countable σ – algebra to insure $A_x \in \mathcal{B}$. Just as above either $A_x \cap A_y = \emptyset$ or $A_x = A_y$ and therefore $\mathcal{F} = \{A_x : x \in X\} \subset \mathcal{B}$ is a (necessarily countable) partition of X for which Eq. (6.1) holds for all $B \in \mathcal{B}$.

Enumerate the elements of $\mathcal{F}$ as $\mathcal{F} = \{P_n\}_{n=1}^N$ where $N \in \mathbb{N}$ or $N = \infty$. If $N = \infty$, then the correspondence

$$a \in \{0,1\}^{\mathbb{N}} \to A_a = \cup\{P_n : a_n = 1\} \in \mathcal{B}$$

is bijective and therefore, by Lemma 6.8, $\mathcal{B}$ is uncountable. Thus any countable σ – algebra is necessarily finite. This finishes the proof modulo the uniqueness assertion which is left as an exercise to the reader. ∎

Example 6.21 (Countable/Co-countable σ – Field). Let $X = \mathbb{R}$ and $\mathcal{E} := \{\{x\} : x \in \mathbb{R}\}$. Then $\sigma\left(\mathcal{E}\right)$ consists of those subsets, $A \subset \mathbb{R}$, such that A is countable or A^c is countable. Similarly, $\mathcal{A}\left(\mathcal{E}\right)$ consists of those subsets, $A \subset \mathbb{R}$, such that A is finite or A^c is finite. More generally we have the following exercise.

Exercise 6.8. Let X be a set, I be an **infinite** index set, and $\mathcal{E} = \{A_i\}_{i\in I}$ be a partition of X. Prove the algebra, $\mathcal{A}\left(\mathcal{E}\right)$, and that σ – algebra, $\sigma\left(\mathcal{E}\right)$, generated by $\mathcal{E}$ are given by

$$\mathcal{A}(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \#\left(\Lambda\right) < \infty \text{ or } \#\left(\Lambda^c\right) < \infty\}$$

and

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \Lambda \text{ countable or } \Lambda^c \text{ countable}\}$$

respectively. Here we are using the convention that $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. In particular if I is countable, then

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I\}.$$

Proposition 6.22. *Let X be a set and $\mathcal{E} \subset 2^X$. Let $\mathcal{E}^c := \{A^c : A \in \mathcal{E}\}$ and $\mathcal{E}_c := \mathcal{E} \cup \{X, \emptyset\} \cup \mathcal{E}^c$ Then*

$$\mathcal{A}(\mathcal{E}) := \{\textit{finite unions of finite intersections of elements from } \mathcal{E}_c\}. \tag{6.2}$$

Proof. Let $\mathcal{A}$ denote the right member of Eq. (6.2). From the definition of an algebra, it is clear that $\mathcal{E} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{E})$. Hence to finish that proof it suffices to show $\mathcal{A}$ is an algebra. The proof of these assertions are routine except for possibly showing that $\mathcal{A}$ is closed under complementation. To check $\mathcal{A}$ is closed under complementation, let $Z \in \mathcal{A}$ be expressed as

$$Z = \bigcup_{i=1}^{N} \bigcap_{j=1}^{K} A_{ij}$$

where $A_{ij} \in \mathcal{E}_c$. Therefore, writing $B_{ij} = A_{ij}^c \in \mathcal{E}_c$, we find that

$$Z^c = \bigcap_{i=1}^{N} \bigcup_{j=1}^{K} B_{ij} = \bigcup_{j_1,\dots,j_N=1}^{K} \left(B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}\right) \in \mathcal{A}$$

wherein we have used the fact that $B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}$ is a finite intersection of sets from $\mathcal{E}_c$. ∎

Remark 6.23. One might think that in general $\sigma(\mathcal{E})$ may be described as the countable unions of countable intersections of sets in $\mathcal{E}^c$. However this is in general **false,** since if

$$Z = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} A_{ij}$$

with $A_{ij} \in \mathcal{E}_c$, then

$$Z^c = \bigcup_{j_1=1, j_2=1, \dots j_N=1, \dots}^{\infty} \left(\bigcap_{\ell=1}^{\infty} A_{\ell, j_\ell}^c \right)$$

which is now an **uncountable** union. Thus the above description is not correct. In general it is complicated to explicitly describe $\sigma(\mathcal{E})$, see Proposition 1.23 on page 39 of Folland for details. Also see Proposition 6.20.

Exercise 6.9. Let τ be a topology on a set X and $\mathcal{A} = \mathcal{A}(\tau)$ be the algebra generated by τ. Show $\mathcal{A}$ is the collection of subsets of X which may be written as finite union of sets of the form $F \cap V$ where F is closed and V is open.

Definition 6.24. *A set $\mathcal{S} \subset 2^X$ is said to be an **semialgebra or elementary class** provided that*

- $\emptyset \in \mathcal{S}$
- *$\mathcal{S}$ is closed under finite intersections*
- *if $E \in \mathcal{S}$, then E^c is a finite disjoint union of sets from $\mathcal{S}$. (In particular $X = \emptyset^c$ is a finite disjoint union of elements from $\mathcal{S}$.)*

Proposition 6.25. *Suppose $\mathcal{S} \subset 2^X$ is a elementary class, then $\mathcal{A} = \mathcal{A}(\mathcal{S})$ consists of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$.*

Proof. (Although it is possible to give a proof using Proposition 6.22, it is just as simple to give a direct proof.) Let $\mathcal{A}$ denote the collection of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$. Clearly $\mathcal{S} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{S})$ so it suffices to show $\mathcal{A}$ is an algebra since $\mathcal{A}(\mathcal{S})$ is the smallest algebra containing $\mathcal{S}$. By the properties of $\mathcal{S}$, we know that $\emptyset, X \in \mathcal{A}$. The following two steps now finish the proof.

1. ($\mathcal{A}$ is closed under finite intersections.) Suppose that $A_i = \sum_{F \in \Lambda_i} F \in \mathcal{A}$ where, for $i = 1, 2, \dots, n$, Λ_i is a finite collection of disjoint sets from $\mathcal{S}$. Then

$$\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \left(\sum_{F \in \Lambda_i} F \right) = \bigcup_{(F_1,\dots,F_n) \in \Lambda_1 \times \dots \times \Lambda_n} (F_1 \cap F_2 \cap \dots \cap F_n)$$

and this is a disjoint (you check) union of elements from $\mathcal{S}$. Therefore $\mathcal{A}$ is closed under finite intersections.

2. ($\mathcal{A}$ is closed under complementation.) If $A = \sum_{F \in \Lambda} F$ with Λ being a finite collection of disjoint sets from $\mathcal{S}$, then $A^c = \bigcap_{F \in \Lambda} F^c$. Since, by assumption, $F^c \in \mathcal{A}$ for all $F \in \Lambda \subset \mathcal{S}$ and $\mathcal{A}$ is closed under finite intersections by step 1., it follows that $A^c \in \mathcal{A}$. ∎

Example 6.26. Let $X = \mathbb{R}$, then

$$\begin{aligned} \mathcal{S} &:= \left\{ (a,b] \cap \mathbb{R} : a, b \in \bar{\mathbb{R}} \right\} \\ &= \{(a,b] : a \in [-\infty, \infty) \text{ and } a < b < \infty\} \cup \{\emptyset, \mathbb{R}\} \end{aligned}$$

is a elementary class. The algebra, $\mathcal{A}(\mathcal{S})$, generated by $\mathcal{S}$ consists of finite disjoint unions of sets from $\mathcal{S}$. For example,

$$A = (0, \pi] \cup (2\pi, 7] \cup (11, \infty) \in \mathcal{A}(\mathcal{S}).$$

Exercise 6.10. Let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be elementary classs. Show the collection

$$\mathcal{S} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$$

is also a elementary class.

Finitely Additive Measures / Integration

Definition 7.1. *Suppose that $\mathcal{E} \subset 2^X$ is a collection of subsets of X and $\mu : \mathcal{E} \to [0, \infty]$ is a function. Then*

1. *μ is **additive or finitely additive on** $\mathcal{E}$ if*
$$\mu(E) = \sum_{i=1}^{n} \mu(E_i) \tag{7.1}$$
whenever $E = \sum_{i=1}^{n} E_i \in \mathcal{E}$ with $E_i \in \mathcal{E}$ for $i = 1, 2, \ldots, n < \infty$.
2. *μ is σ – **additive (or countable additive) on** $\mathcal{E}$ if Eq. (7.1) holds even when $n = \infty$.*
3. *μ is **sub-additive (finitely sub-additive) on** $\mathcal{E}$ if*
$$\mu(E) \le \sum_{i=1}^{n} \mu(E_i)$$
whenever $E = \bigcup_{i=1}^{n} E_i \in \mathcal{E}$ with $n \in \mathbb{N} \cup \{\infty\}$ $(n \in \mathbb{N})$.
4. *μ is a **finitely additive measure** if $\mathcal{E} = \mathcal{A}$ is an algebra, $\mu(\emptyset) = 0$, and μ is finitely additive on $\mathcal{A}$.*
5. *μ is a **premeasure** if μ is a finitely additive measure which is σ – additive on $\mathcal{A}$.*
6. *μ is a **measure** if μ is a premeasure on a σ – algebra. Furthermore if $\mu(X) = 1$, we say μ is a **probability measure** on X.*

Proposition 7.2 (Basic properties of finitely additive measures). *Suppose μ is a finitely additive measure on an algebra, $\mathcal{A} \subset 2^X$, $A, B \in \mathcal{A}$ with $A \subset B$and $\{A_j\}_{j=1}^n \subset \mathcal{A}$, then :*

1. *(μ is **monotone**) $\mu(A) \le \mu(B)$ if $A \subset B$.*
2. *For $A, B \in \mathcal{A}$, the following **strong additivity formula** holds;*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{7.2}$$
3. *(μ is **finitely subbadditive**) $\mu(\cup_{j=1}^n A_j) \le \sum_{j=1}^n \mu(A_j)$.*
4. *μ is sub-additive on $\mathcal{A}$ iff*
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A = \sum_{i=1}^{\infty} A_i \tag{7.3}$$
where $A \in \mathcal{A}$ and $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ are pairwise disjoint sets.
5. *(μ is **countably superadditive**) If $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then*
$$\mu\left(\sum_{i=1}^{\infty} A_i\right) \ge \sum_{i=1}^{\infty} \mu(A_i). \tag{7.4}$$
(See Remark 7.9 for example where this inequality is strict.)
6. *A finitely additive measure, μ, is a premeasure iff μ is subadditive.*

Proof.

1. Since B is the disjoint union of A and $(B \setminus A)$ and $B \setminus A = B \cap A^c \in \mathcal{A}$ it follows that
$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A).$$
2. Since
$$A \cup B = [A \setminus (A \cap B)] \coprod [B \setminus (A \cap B)] \coprod A \cap B,$$
$$\begin{aligned}\mu(A \cup B) &= \mu(A \cup B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A \setminus (A \cap B)) + \mu(B \setminus (A \cap B)) + \mu(A \cap B).\end{aligned}$$
Adding $\mu(A \cap B)$ to both sides of this equation proves Eq. (7.2).
3. Let $\widetilde{E}_j = E_j \setminus (E_1 \cup \cdots \cup E_{j-1})$ so that the $\tilde{E}_j$'s are pair-wise disjoint and $E = \cup_{j=1}^n \widetilde{E}_j$. Since $\tilde{E}_j \subset E_j$ it follows from the monotonicity of μ that
$$\mu(E) = \sum_{j=1}^{n} \mu(\widetilde{E}_j) \le \sum_{j=1}^{n} \mu(E_j).$$
4. If $A = \bigcup_{i=1}^{\infty} B_i$ with $A \in \mathcal{A}$ and $B_i \in \mathcal{A}$, then $A = \sum_{i=1}^{\infty} A_i$ where $A_i := B_i \setminus (B_1 \cup \ldots B_{i-1}) \in \mathcal{A}$ and $B_0 = \emptyset$. Therefore using the monotonicity of μ and Eq. (7.3)
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \le \sum_{i=1}^{\infty} \mu(B_i).$$
5. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then $\sum_{i=1}^{n} A_i \subset A$ for all n and so by the monotonicity and finite additivity of μ, $\sum_{i=1}^{n} \mu(A_i) \le \mu(A)$. Letting $n \to \infty$ in this equation shows μ is superadditive.
6. This is a combination of items 5. and 6.

■

7.1 Examples of Measures

Most $\sigma-$algebras and σ-additive measures are somewhat difficult to describe and define. However, there are a few special cases where we can describe explicitly what is going on.

Example 7.3. Suppose that Ω is a finite set, $\mathcal{B} := 2^\Omega$, and $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

Then

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \subset \Omega$$

defines a measure on 2^Ω.

Example 7.4. Suppose that X is any set and $x \in X$ is a point. For $A \subset X$, let

$$\delta_x(A) = \begin{cases} 1 \text{ if } & x \in A \\ 0 \text{ if } & x \notin A. \end{cases}$$

Then $\mu = \delta_x$ is a measure on X called the Dirac delta measure at x.

Example 7.5. Suppose $\mathcal{B} \subset 2^X$ is a σ algebra, μ is a measure on $\mathcal{B}$, and $\lambda > 0$, then $\lambda \cdot \mu$ is also a measure on $\mathcal{B}$. Moreover, if J is an index set and $\{\mu_j\}_{j\in J}$ are all measures on $\mathcal{B}$, then $\mu = \sum_{j=1}^\infty \mu_j$, i.e.

$$\mu(A) := \sum_{j=1}^\infty \mu_j(A) \text{ for all } A \in \mathcal{B},$$

defines another measure on $\mathcal{B}$. To prove this we must show that μ is countably additive. Suppose that $A = \sum_{i=1}^\infty A_i$ with $A_i \in \mathcal{B}$, then (using Tonelli for sums, Theorem 4.22),

$$\begin{aligned}\mu(A) &= \sum_{j=1}^\infty \mu_j(A) = \sum_{j=1}^\infty \sum_{i=1}^\infty \mu_j(A_i) \\ &= \sum_{i=1}^\infty \sum_{j=1}^\infty \mu_j(A_i) = \sum_{i=1}^\infty \mu(A_i).\end{aligned}$$

Example 7.6. Suppose that X is a countable set and $\lambda : X \to [0,\infty]$ is a function. Let $X = \{x_n\}_{n=1}^\infty$ be an enumeration of X and then we may define a measure μ on 2^X by,

$$\mu = \mu_\lambda := \sum_{n=1}^\infty \lambda(x_n)\delta_{x_n}.$$

We will now show this measure is independent of our choice of enumeration of X by showing,

$$\mu(A) = \sum_{x\in A} \lambda(x) := \sup_{\Lambda\subset\subset A} \sum_{x\in\Lambda} \lambda(x) \ \forall\ A \subset X. \tag{7.5}$$

Here we are using the notation, $\Lambda \subset\subset A$ to indicate that Λ is a finite subset of A.

To verify Eq. (7.5), let $M := \sup_{\Lambda\subset\subset A} \sum_{x\in\Lambda} \lambda(x)$ and for each $N \in \mathbb{N}$ let

$$\Lambda_N := \{x_n : x_n \in A \text{ and } 1 \le n \le N\}.$$

Then by definition of μ,

$$\begin{aligned}\mu(A) &= \sum_{n=1}^\infty \lambda(x_n)\delta_{x_n}(A) = \lim_{N\to\infty} \sum_{n=1}^N \lambda(x_n) 1_{x_n\in A} \\ &= \lim_{N\to\infty} \sum_{x\in\Lambda_N} \lambda(x) \le M.\end{aligned}$$

On the other hand if $\Lambda \subset\subset A$, then

$$\sum_{x\in\Lambda} \lambda(x) = \sum_{n:\ x_n\in\Lambda} \lambda(x_n) = \mu(\Lambda) \le \mu(A)$$

from which it follows that $M \le \mu(A)$. This shows that μ is independent of how we enumerate X.

The above example has a natural extension to the case where X is uncountable and $\lambda : X \to [0,\infty]$ is any function. In this setting we simply may define $\mu : 2^X \to [0,\infty]$ using Eq. (7.5). We leave it to the reader to verify that this is indeed a measure on 2^X.

We will construct many more measure in Chapter 8 below. The starting point of these constructions will be the construction of finitely additive measures using the next proposition.

Proposition 7.7 (Construction of Finitely Additive Measures). *Suppose $\mathcal{S} \subset 2^X$ is a semi-algebra (see Definition 6.24) and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ is the algebra generated by $\mathcal{S}$. Then every additive function $\mu : \mathcal{S} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ extends uniquely to an additive measure (which we still denote by μ) on $\mathcal{A}$.*

Proof. Since (by Proposition 6.25) every element $A \in \mathcal{A}$ is of the form $A = \sum_i E_i$ for a finite collection of $E_i \in \mathcal{S}$, it is clear that if μ extends to a measure then the extension is unique and must be given by

$$\mu(A) = \sum_i \mu(E_i). \tag{7.6}$$

To prove existence, the main point is to show that $\mu(A)$ in Eq. (7.6) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in \mathcal{S}$, then we must show

$$\sum_i \mu(E_i) = \sum_j \mu(F_j). \tag{7.7}$$

But $E_i = \sum_j (E_i \cap F_j)$ and the additivity of μ on $\mathcal{S}$ implies $\mu(E_i) = \sum_j \mu(E_i \cap F_j)$ and hence

$$\sum_i \mu(E_i) = \sum_i \sum_j \mu(E_i \cap F_j) = \sum_{i,j} \mu(E_i \cap F_j).$$

Similarly,

$$\sum_j \mu(F_j) = \sum_{i,j} \mu(E_i \cap F_j)$$

which combined with the previous equation shows that Eq. (7.7) holds. It is now easy to verify that μ extended to $\mathcal{A}$ as in Eq. (7.6) is an additive measure on $\mathcal{A}$. ■

Proposition 7.8. *Let $X = \mathbb{R}$, $\mathcal{S}$ be the semi-algebra,*

$$\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}, \tag{7.8}$$

and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ be the algebra formed by taking finite disjoint unions of elements from $\mathcal{S}$, see Proposition 6.25. To each finitely additive probability measures $\mu : \mathcal{A} \to [0, \infty]$, there is a unique increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ and

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) \ \forall \ a \le b \text{ in } \bar{\mathbb{R}}. \tag{7.9}$$

Conversely, given an increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ there is a unique finitely additive measure $\mu = \mu_F$ on $\mathcal{A}$ such that the relation in Eq. (7.9) holds. (Eventually we will only be interested in the case where $F(-\infty) = \lim_{a \downarrow -\infty} F(a)$ and $F(\infty) = \lim_{b \uparrow \infty} F(b)$.)

Proof. Given a finitely additive probability measure μ, let

$$F(x) := \mu((-\infty, x] \cap \mathbb{R}) \text{ for all } x \in \bar{\mathbb{R}}.$$

Then $F(\infty) = 1$, $F(-\infty) = 0$ and for $b > a$,

$$F(b) - F(a) = \mu((-\infty, b] \cap \mathbb{R}) - \mu((-\infty, a]) = \mu((a, b] \cap \mathbb{R}).$$

Conversely, suppose $F : \bar{\mathbb{R}} \to [0, 1]$ as in the statement of the theorem is given. Define μ on $\mathcal{S}$ using the formula in Eq. (7.9). The argument will be completed by showing μ is additive on $\mathcal{S}$ and hence, by Proposition 7.7, has a unique extension to a finitely additive measure on $\mathcal{A}$. Suppose that

$$(a, b] = \sum_{i=1}^n (a_i, b_i].$$

By reordering $(a_i, b_i]$ if necessary, we may assume that

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < b_{n-1} = a_n < b_n = b.$$

Therefore, by the telescoping series argument,

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) = \sum_{i=1}^n [F(b_i) - F(a_i)] = \sum_{i=1}^n \mu((a_i, b_i] \cap \mathbb{R}).$$

■

Remark 7.9. Suppose that $F : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ is any non-decreasing function such that $F(\mathbb{R}) \subset \mathbb{R}$. Then the same methods used in the proof of Proposition 7.8 shows that there exists a unique finitely additive measure, $\mu = \mu_F$, on $\mathcal{A} = \mathcal{A}(\mathcal{S})$ such that Eq. (7.9) holds. If $F(\infty) > \lim_{b \uparrow \infty} F(b)$ and $A_i = (i, i+1]$ for $i \in \mathbb{N}$, then

$$\begin{aligned}\sum_{i=1}^\infty \mu_F(A_i) &= \sum_{i=1}^\infty (F(i+1) - F(i)) = \lim_{N \to \infty} \sum_{i=1}^N (F(i+1) - F(i)) \\ &= \lim_{N \to \infty} (F(N+1) - F(1)) < F(\infty) - F(1) = \mu_F(\cup_{i=1}^\infty A_i).\end{aligned}$$

This shows that strict inequality can hold in Eq. (7.4) and that μ_F is **not** a premeasure. Similarly one shows μ_F is **not** a premeasure if $F(-\infty) < \lim_{a \downarrow -\infty} F(a)$ or if F is **not** right continuous at some point $a \in \mathbb{R}$. Indeed, in the latter case consider

$$(a, a+1] = \sum_{n=1}^\infty (a + \frac{1}{n+1}, a + \frac{1}{n}].$$

Working as above we find,

$$\sum_{n=1}^\infty \mu_F\left((a + \frac{1}{n+1}, a + \frac{1}{n}]\right) = F(a+1) - F(a+)$$

while $\mu_F((a, a+1]) = F(a+1) - F(a)$. We will eventually show in Chapter 8 below that μ_F extends uniquely to a σ − additive measure on $\mathcal{B}_\mathbb{R}$ whenever F is increasing, right continuous, and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x)$.

Before constructing σ – additive measures (see Chapter 8 below), we are going to pause to discuss a preliminary notion of integration and develop some of its properties. Hopefully this will help the reader to develop the necessary intuition before heading to the general theory. First we need to describe the functions we are (currently) able to integrate.

7.2 Simple Random Variables

Definition 7.10 (Simple random variables). *A function, $f : \Omega \to Y$ is said to be* ***simple*** *if $f(\Omega) \subset Y$ is a finite set. If $\mathcal{A} \subset 2^{\Omega}$ is an algebra, we say that a simple function $f : \Omega \to Y$ is* ***measurable*** *if $\{f = y\} := f^{-1}(\{y\}) \in \mathcal{A}$ for all $y \in Y$. A measurable simple function, $f : \Omega \to \mathbb{C}$, is called a* ***simple random variable*** *relative to $\mathcal{A}$.*

Notation 7.11 *Given an algebra, $\mathcal{A} \subset 2^{\Omega}$, let $\mathbb{S}(\mathcal{A})$ denote the collection of simple random variables from Ω to $\mathbb{C}$. For example if $A \in \mathcal{A}$, then $1_A \in \mathbb{S}(\mathcal{A})$ is a measurable simple function.*

Lemma 7.12. *Let $\mathcal{A} \subset 2^{\Omega}$ be an algebra, then;*

1. *$\mathbb{S}(\mathcal{A})$ is a sub-algebra of all functions from Ω to $\mathbb{C}$.*
2. *$f : \Omega \to \mathbb{C}$, is a $\mathcal{A}$ – simple random variable iff there exists $\alpha_i \in \mathbb{C}$ and $A_i \in \mathcal{A}$ for $1 \le i \le n$ for some $n \in \mathbb{N}$ such that*
$$f = \sum_{i=1}^{n} \alpha_i 1_{A_i}. \tag{7.10}$$
3. *For any function, $F : \mathbb{C} \to \mathbb{C}$, $F \circ f \in \mathbb{S}(\mathcal{A})$ for all $f \in \mathbb{S}(\mathcal{A})$. In particular, $|f| \in \mathbb{S}(\mathcal{A})$ if $f \in \mathbb{S}(\mathcal{A})$.*

Proof. 1. Let us observe that $1_{\Omega} = 1$ and $1_{\emptyset} = 0$ are in $\mathbb{S}(\mathcal{A})$. If $f, g \in \mathbb{S}(\mathcal{A})$ and $c \in \mathbb{C} \setminus \{0\}$, then
$$\{f + cg = \lambda\} = \bigcup_{a,b \in \mathbb{C}: a + cb = \lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{7.11}$$
and
$$\{f \cdot g = \lambda\} = \bigcup_{a,b \in \mathbb{C}: a \cdot b = \lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{7.12}$$
from which it follows that $f + cg$ and $f \cdot g$ are back in $\mathbb{S}(\mathcal{A})$.

2. Since $\mathbb{S}(\mathcal{A})$ is an algebra, every f of the form in Eq. (7.10) is in $\mathbb{S}(\mathcal{A})$. Conversely if $f \in \mathbb{S}(\mathcal{A})$ it follows by definition that $f = \sum_{\alpha \in f(\Omega)} \alpha 1_{\{f = \alpha\}}$ which is of the form in Eq. (7.10).

3. If $F : \mathbb{C} \to \mathbb{C}$, then
$$F \circ f = \sum_{\alpha \in f(\Omega)} F(\alpha) \cdot 1_{\{f = \alpha\}} \in \mathbb{S}(\mathcal{A}).$$
■

Exercise 7.1 ($\mathcal{A}$ – measurable simple functions). As in Example 6.19, let $\mathcal{A} \subset 2^X$ be a finite algebra and $\{B_1, \dots, B_k\}$ be the partition of X associated to $\mathcal{A}$. Show that a function, $f : X \to \mathbb{C}$, is an $\mathcal{A}$ – simple function iff f is constant on B_i for each i. Thus any $\mathcal{A}$ – simple function is of the form,
$$f = \sum_{i=1}^{k} \alpha_i 1_{B_i} \tag{7.13}$$
for some $\alpha_i \in \mathbb{C}$.

Corollary 7.13. *Suppose that Λ is a finite set and $Z : X \to \Lambda$ is a function. Let*
$$\mathcal{A} := \mathcal{A}(Z) := Z^{-1}\left(2^{\Lambda}\right) := \left\{Z^{-1}(E) : E \subset \Lambda\right\}.$$
Then $\mathcal{A}$ is an algebra and $f : X \to \mathbb{C}$ is an $\mathcal{A}$ – simple function iff $f = F \circ Z$ for some function $F : \Lambda \to \mathbb{C}$.

Proof. For $\lambda \in \Lambda$, let
$$A_{\lambda} := \{Z = \lambda\} = \{x \in X : Z(x) = \lambda\}.$$
The $\{A_{\lambda}\}_{\lambda \in \Lambda}$ is the partition of X determined by $\mathcal{A}$. Therefore f is an $\mathcal{A}$ – simple function iff $f|_{A_{\lambda}}$ is constant for each $\lambda \in \Lambda$. Let us denote this constant value by $F(\lambda)$. As $Z = \lambda$ on A_{λ}, $F : \Lambda \to \mathbb{C}$ is a function such that $f = F \circ Z$.

Conversely if $F : \Lambda \to \mathbb{C}$ is a function and $f = F \circ Z$, then $f = F(\lambda)$ on A_{λ}, i.e. f is an $\mathcal{A}$ – simple function. ■

7.2.1 The algebraic structure of simple functions*

Definition 7.14. *A* ***simple function algebra***, *$\mathbb{S}$, is a subalgebra[1] of the bounded complex functions on X such that $1 \in \mathbb{S}$ and each function in $\mathbb{S}$ is a simple function. If $\mathbb{S}$ is a simple function algebra, let*
$$\mathcal{A}(\mathbb{S}) := \{A \subset X : 1_A \in \mathbb{S}\}.$$
(It is easily checked that $\mathcal{A}(\mathbb{S})$ is a sub-algebra of 2^X.)

[1] To be more explicit we are assuming that $\mathbb{S}$ is a linear subspace of bounded functions which is closed under pointwise multiplication.

Lemma 7.15. *Suppose that* $\mathbb{S}$ *is a simple function algebra,* $f \in \mathbb{S}$ *and* $\alpha \in f(X)$ *– the range of* f. *Then* $\{f = \alpha\} \in \mathcal{A}(\mathbb{S})$.

Proof. Let $\{\lambda_i\}_{i=0}^n$ be an enumeration of $f(X)$ with $\lambda_0 = \alpha$. Then

$$g := \left[\prod_{i=1}^n (\alpha - \lambda_i)\right]^{-1} \prod_{i=1}^n (f - \lambda_i 1) \in \mathbb{S}.$$

Moreover, we see that $g = 0$ on $\cup_{i=1}^n \{f = \lambda_i\}$ while $g = 1$ on $\{f = \alpha\}$. So we have shown $g = 1_{\{f=\alpha\}} \in \mathbb{S}$ and therefore that $\{f = \alpha\} \in \mathcal{A}(\mathbb{S})$. ∎

Exercise 7.2. Continuing the notation introduced above:

1. Show $\mathcal{A}(\mathbb{S})$ is an algebra of sets.
2. Show $\mathbb{S}(\mathcal{A})$ is a simple function algebra.
3. Show that the map

$$\mathcal{A} \in \left\{\text{Algebras} \subset 2^X\right\} \to \mathbb{S}(\mathcal{A}) \in \{\textbf{simple function algebras on } X\}$$

is bijective and the map, $\mathbb{S} \to \mathcal{A}(\mathbb{S})$, is the inverse map.

7.3 Simple Integration

Definition 7.16 (Simple Integral). *Suppose now that* P *is a finitely additive probability measure on an algebra* $\mathcal{A} \subset 2^X$. *For* $f \in \mathbb{S}(\mathcal{A})$ *the* ***integral or expectation,*** $\mathbb{E}(f) = \mathbb{E}_P(f)$, *is defined by*

$$\mathbb{E}_P(f) = \int_X f dP = \sum_{y\in\mathbb{C}} y P(f = y). \tag{7.14}$$

Example 7.17. Suppose that $A \in \mathcal{A}$, then

$$\mathbb{E}1_A = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A). \tag{7.15}$$

Remark 7.18. Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (**??**) by

$$P(A) = \lim_{N\to\infty} \frac{1}{N} \sum 1_A(\omega(k))$$

where $\omega(k) \in \Omega$ was the result of the k^{th} "independent" experiment. If we use this interpretation back in Eq. (7.14) we arrive at,

$$\begin{aligned}
\mathbb{E}(f) &= \sum_{y\in\mathbb{C}} y P(f = y) = \sum_{y\in\mathbb{C}} y \cdot \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{f(\omega(k))=y} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{y\in\mathbb{C}} y \sum_{k=1}^N 1_{f(\omega(k))=y} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \sum_{y\in\mathbb{C}} f(\omega(k)) \cdot 1_{f(\omega(k))=y} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N f(\omega(k)).
\end{aligned}$$

Thus informally, $\mathbb{E}f$ should represent the limiting average of the values of f over many "independent" experiments. We will come back to this later when we study the strong law of large numbers.

Proposition 7.19. *The expectation operator,* $\mathbb{E} = \mathbb{E}_P : \mathbb{S}(\mathcal{A}) \to \mathbb{C}$, *satisfies:*

1. *If* $f \in \mathbb{S}(\mathcal{A})$ *and* $\lambda \in \mathbb{C}$, *then*
$$\mathbb{E}(\lambda f) = \lambda \mathbb{E}(f). \tag{7.16}$$
2. *If* $f, g \in \mathbb{S}(\mathcal{A})$, *then*
$$\mathbb{E}(f + g) = \mathbb{E}(g) + \mathbb{E}(f). \tag{7.17}$$
Items 1. and 2. say that $\mathbb{E}(\cdot)$ *is a linear functional on* $\mathbb{S}(\mathcal{A})$.
3. *If* $f = \sum_{j=1}^N \lambda_j 1_{A_j}$ *for some* $\lambda_j \in \mathbb{C}$ *and some* $A_j \in \mathbb{C}$, *then*
$$\mathbb{E}(f) = \sum_{j=1}^N \lambda_j P(A_j). \tag{7.18}$$
4. $\mathbb{E}$ *is* ***positive****, i.e.* $\mathbb{E}(f) \geq 0$ *for all* $0 \leq f \in \mathbb{S}(\mathcal{A})$. *More generally, if* $f, g \in \mathbb{S}(\mathcal{A})$ *and* $f \leq g$, *then* $\mathbb{E}(f) \leq \mathbb{E}(g)$.
5. *For all* $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}f| \leq \mathbb{E}|f|. \tag{7.19}$$

Proof.

1. If $\lambda \neq 0$, then

$$\begin{aligned}
\mathbb{E}(\lambda f) &= \sum_{y\in\mathbb{C}} y\ P(\lambda f = y) = \sum_{y\in\mathbb{C}} y\ P(f = y/\lambda) \\
&= \sum_{z\in\mathbb{C}} \lambda z\ P(f = z) = \lambda \mathbb{E}(f).
\end{aligned}$$

The case $\lambda = 0$ is trivial.

2. Writing $\{f=a,g=b\}$ for $f^{-1}(\{a\})\cap g^{-1}(\{b\})$, then

$$\begin{aligned}\mathbb{E}(f+g) &= \sum_{z\in\mathbb{C}} z\ P(f+g=z) \\ &= \sum_{z\in\mathbb{C}} z\ P\left(\sum_{a+b=z}\{f=a,\ g=b\}\right) \\ &= \sum_{z\in\mathbb{C}} z \sum_{a+b=z} P\left(\{f=a,\ g=b\}\right) \\ &= \sum_{z\in\mathbb{C}}\sum_{a+b=z}(a+b)\,P\left(\{f=a,\ g=b\}\right) \\ &= \sum_{a,b}(a+b)\,P\left(\{f=a,\ g=b\}\right).\end{aligned}$$

But

$$\begin{aligned}\sum_{a,b} aP\left(\{f=a,\ g=b\}\right) &= \sum_a a\sum_b P\left(\{f=a,\ g=b\}\right) \\ &= \sum_a aP\left(\cup_b\{f=a,\ g=b\}\right) \\ &= \sum_a aP\left(\{f=a\}\right) = \mathbb{E}f\end{aligned}$$

and similarly,

$$\sum_{a,b} bP\left(\{f=a,\ g=b\}\right) = \mathbb{E}g.$$

Equation (7.17) is now a consequence of the last three displayed equations.

3. If $f=\sum_{j=1}^{N}\lambda_j 1_{A_j}$, then

$$\mathbb{E}f = \mathbb{E}\left[\sum_{j=1}^{N}\lambda_j 1_{A_j}\right] = \sum_{j=1}^{N}\lambda_j\mathbb{E}1_{A_j} = \sum_{j=1}^{N}\lambda_j P\left(A_j\right).$$

4. If $f\geq 0$ then

$$\mathbb{E}(f) = \sum_{a\geq 0} aP(f=a)\geq 0$$

and if $f\leq g$, then $g-f\geq 0$ so that

$$\mathbb{E}\left(g\right)-\mathbb{E}\left(f\right)=\mathbb{E}\left(g-f\right)\geq 0.$$

5. By the triangle inequality,

$$\left|\mathbb{E}f\right| = \left|\sum_{\lambda\in\mathbb{C}}\lambda P\left(f=\lambda\right)\right| \leq \sum_{\lambda\in\mathbb{C}}\left|\lambda\right| P\left(f=\lambda\right) = \mathbb{E}\left|f\right|,$$

wherein the last equality we have used Eq. (7.18) and the fact that $|f| = \sum_{\lambda\in\mathbb{C}}|\lambda|\,1_{f=\lambda}$.

■

Remark 7.20. If Ω is a finite set and $\mathcal{A}=2^{\Omega}$, then

$$f\left(\cdot\right)=\sum_{\omega\in\Omega} f\left(\omega\right)1_{\{\omega\}}$$

and hence

$$\mathbb{E}_P f = \sum_{\omega\in\Omega} f\left(\omega\right)P\left(\{\omega\}\right).$$

Remark 7.21. All of the results in Proposition 7.19 and Remark 7.20 remain valid when P is replaced by a finite measure, $\mu:\mathcal{A}\to[0,\infty)$, i.e. it is enough to assume $\mu\left(X\right)<\infty$.

Exercise 7.3. Let P is a finitely additive probability measure on an algebra $\mathcal{A}\subset 2^X$ and for $A,B\in\mathcal{A}$ let $\rho\left(A,B\right):=P\left(A\Delta B\right)$ where $A\Delta B=\left(A\setminus B\right)\cup\left(B\setminus A\right)$. Show;

1. $\rho\left(A,B\right)=\mathbb{E}\left|1_A-1_B\right|$ and then use this (or not) to show
2. $\rho\left(A,C\right)\leq\rho\left(A,B\right)+\rho\left(B,C\right)$ for all $A,B,C\in\mathcal{A}$.

Remark: it is now easy to see that $\rho:\mathcal{A}\times\mathcal{A}\to[0,1]$ satisfies the axioms of a metric except for the condition that $\rho\left(A,B\right)=0$ does not imply that $A=B$ but only that $A=B$ modulo a set of probability zero.

Remark 7.22 (Chebyshev's Inequality). Suppose that $f\in\mathbb{S}(\mathcal{A})$, $\varepsilon>0$, and $p>0$, then

$$1_{|f|\geq\varepsilon}\leq\frac{|f|^p}{\varepsilon^p}1_{|f|\geq\varepsilon}\leq\varepsilon^{-p}\left|f\right|^p$$

and therefore, see item 4. of Proposition 7.19,

$$P\left(\{|f|\geq\varepsilon\}\right)=\mathbb{E}\left[1_{|f|\geq\varepsilon}\right]\leq\mathbb{E}\left[\frac{|f|^p}{\varepsilon^p}1_{|f|\geq\varepsilon}\right]\leq\varepsilon^{-p}\mathbb{E}\left|f\right|^p. \tag{7.20}$$

Observe that

$$\left|f\right|^p=\sum_{\lambda\in\mathbb{C}}\left|\lambda\right|^p 1_{\{f=\lambda\}}$$

is a simple random variable and $\{|f|\geq\varepsilon\}=\sum_{|\lambda|\geq\varepsilon}\{f=\lambda\}\in\mathcal{A}$ as well. Therefore, $\frac{|f|^p}{\varepsilon^p}1_{|f|\geq\varepsilon}$ is still a simple random variable.

Lemma 7.23 (Inclusion Exclusion Formula). *If* $A_n \in \mathcal{A}$ *for* $n = 1, 2, \ldots, M$ *such that* $\mu\left(\cup_{n=1}^M A_n\right) < \infty$, *then*

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right). \tag{7.21}$$

Proof. This may be proved inductively from Eq. (7.2). We will give a different and perhaps more illuminating proof here. Let $A := \cup_{n=1}^M A_n$.

Since $A^c = \left(\cup_{n=1}^M A_n\right)^c = \cap_{n=1}^M A_n^c$, we have

$$\begin{aligned}1 - 1_A = 1_{A^c} &= \prod_{n=1}^M 1_{A_n^c} = \prod_{n=1}^M (1 - 1_{A_n}) \\ &= 1 + \sum_{k=1}^M (-1)^k \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1}} \cdots 1_{A_{n_k}} \\ &= 1 + \sum_{k=1}^M (-1)^k \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}\end{aligned}$$

from which it follows that

$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}. \tag{7.22}$$

Integrating this identity with respect to μ gives Eq. (7.21). ■

Remark 7.24. The following identity holds even when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$,

$$\begin{aligned}\mu\left(\cup_{n=1}^M A_n\right) + &\sum_{k=2 \text{ \& } k \text{ even}}^M \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right) \\ &= \sum_{k=1 \text{ \& } k \text{ odd}}^M \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right).\end{aligned} \tag{7.23}$$

This can be proved by moving every term with a negative sign on the right side of Eq. (7.22) to the left side and then integrate the resulting identity. Alternatively, Eq. (7.23) follows directly from Eq. (7.21) if $\mu\left(\cup_{n=1}^M A_n\right) < \infty$ and when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$ one easily verifies that both sides of Eq. (7.23) are infinite.

To better understand Eq. (7.22), consider the case $M = 3$ where,

$$\begin{aligned}1 - 1_A &= (1 - 1_{A_1})(1 - 1_{A_2})(1 - 1_{A_3}) \\ &= 1 - (1_{A_1} + 1_{A_2} + 1_{A_3}) \\ &\quad + 1_{A_1}1_{A_2} + 1_{A_1}1_{A_3} + 1_{A_2}1_{A_3} - 1_{A_1}1_{A_2}1_{A_3}\end{aligned}$$

so that

$$1_{A_1 \cup A_2 \cup A_3} = 1_{A_1} + 1_{A_2} + 1_{A_3} - (1_{A_1 \cap A_2} + 1_{A_1 \cap A_3} + 1_{A_2 \cap A_3}) + 1_{A_1 \cap A_2 \cap A_3}$$

Here is an alternate proof of Eq. (7.22). Let $\omega \in \Omega$ and by relabeling the sets $\{A_n\}$ if necessary, we may assume that $\omega \in A_1 \cap \cdots \cap A_m$ and $\omega \notin A_{m+1} \cup \cdots \cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (7.22) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have

$$\begin{aligned}&\sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}(\omega) \\ &= \sum_{k=1}^m (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le m} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}(\omega) \\ &= \sum_{k=1}^m (-1)^{k+1} \binom{m}{k} \\ &= 1 - \sum_{k=0}^m (-1)^k (1)^{n-k} \binom{m}{k} \\ &= 1 - (1-1)^m = 1.\end{aligned}$$

This verifies Eq. (7.22) since $1_{\cup_{n=1}^M A_n}(\omega) = 1$.

Example 7.25 (Coincidences). Let Ω be the set of permutations (think of card shuffling), $\omega : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$, and define $P(A) := \frac{\#(A)}{n!}$ to be the uniform distribution (Haar measure) on Ω. We wish to compute the probability of the event, B, that a random permutation fixes some index i. To do this, let $A_i := \{\omega \in \Omega : \omega(i) = i\}$ and observe that $B = \cup_{i=1}^n A_i$. So by the Inclusion Exclusion Formula, we have

$$P(B) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < i_2 < i_3 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}).$$

Since

$$\begin{aligned}P(A_{i_1} \cap \cdots \cap A_{i_k}) &= P(\{\omega \in \Omega : \omega(i_1) = i_1, \ldots, \omega(i_k) = i_k\}) \\ &= \frac{(n-k)!}{n!}\end{aligned}$$

and

$$\#\{1 \le i_1 < i_2 < i_3 < \cdots < i_k \le n\} = \binom{n}{k},$$

we find

$$P(B) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}. \tag{7.24}$$

For large n this gives,

$$P(B) = -\sum_{k=1}^{n} \frac{1}{k!} (-1)^k \cong 1 - \sum_{k=0}^{\infty} \frac{1}{k!} (-1)^k = 1 - e^{-1} \cong 0.632.$$

Example 7.26 (Expected number of coincidences). Continue the notation in Example 7.25. We now wish to compute the expected number of fixed points of a random permutation, ω, i.e. how many cards in the shuffled stack have not moved on average. To this end, let

$$X_i = 1_{A_i}$$

and observe that

$$N(\omega) = \sum_{i=1}^{n} X_i(\omega) = \sum_{i=1}^{n} 1_{\omega(i)=i} = \#\{i : \omega(i) = i\}.$$

denote the number of fixed points of ω. Hence we have

$$\mathbb{E}N = \sum_{i=1}^{n} \mathbb{E}X_i = \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{(n-1)!}{n!} = 1.$$

Let us check the above formulas when $n = 3$. In this case we have

ω	$N(\omega)$
1 2 3	3
1 3 2	1
2 1 3	1
2 3 1	0
3 1 2	0
3 2 1	1

and so

$$P(\exists \text{ a fixed point}) = \frac{4}{6} = \frac{2}{3} \cong 0.67 \cong 0.632$$

while

$$\sum_{k=1}^{3} (-1)^{k+1} \frac{1}{k!} = 1 - \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

and

$$\mathbb{E}N = \frac{1}{6}(3 + 1 + 1 + 0 + 0 + 1) = 1.$$

The next three problems generalize the results above. The following notation will be used throughout these exercises.

1. $(\Omega, \mathcal{A}, P)$ is a finitely additive probability space, so $P(\Omega) = 1$,
2. $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots, n$,
3. $N(\omega) := \sum_{i=1}^{n} 1_{A_i}(\omega) = \#\{i : \omega \in A_i\}$, and
4. $\{S_k\}_{k=1}^{n}$ are given by

$$S_k := \sum_{1 \le i_1 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) = \sum_{\Lambda \subset \{1,2,\ldots,n\} \ni |\Lambda| = k} P(\cap_{i \in \Lambda} A_i).$$

Exercise 7.4. For $1 \le k \le n$, show;

1. (as functions on Ω) that

$$\binom{N}{k} = \sum_{\Lambda \subset \{1,2,\ldots,n\} \ni |\Lambda| = k} 1_{\cap_{i \in \Lambda} A_i}, \tag{7.25}$$

where by definition

$$\binom{m}{k} = \begin{cases} 0 & \text{if } \quad k > m \\ \frac{m!}{k! \cdot (m-k)!} & \text{if } 1 \le k \le m \\ 1 & \text{if } \quad k = 0 \end{cases}. \tag{7.26}$$

2. Conclude from Eq. (7.25) that for all $z \in \mathbb{C}$,

$$(1+z)^N = 1 + \sum_{k=1}^{n} z^k \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} 1_{A_{i_1} \cap \cdots \cap A_{i_k}} \tag{7.27}$$

provided $(1+z)^0 = 1$ even when $z = -1$.
3. Conclude from Eq. (7.25) that $S_k = \mathbb{E}_P \binom{N}{k}$.

Exercise 7.5. Taking expectations of Eq. (7.27) implies,

$$\mathbb{E}\left[(1+z)^N\right] = 1 + \sum_{k=1}^{n} S_k z^k. \tag{7.28}$$

Show that setting $z = -1$ in Eq. (7.28) gives another proof of the inclusion exclusion formula. **Hint:** use the definition of the expectation to write out $\mathbb{E}\left[(1+z)^N\right]$ explicitly.

Exercise 7.6. Let $1 \le m \le n$. In this problem you are asked to compute the probability that there are exactly m – coincidences. Namely you should show,

$$\begin{aligned} P(N = m) &= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} S_k \\ &= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \sum_{1 \le i_1 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned}$$

Hint: differentiate Eq. (7.28) m times with respect to z and then evaluate the result at $z = -1$. In order to do this you will find it useful to derive formulas for;

$$\frac{d^m}{dz^m}|_{z=-1} (1+z)^n \text{ and } \frac{d^m}{dz^m}|_{z=-1} z^k.$$

Example 7.27. Let us again go back to Example 7.26 where we computed,

$$S_k = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Therefore it follows from Exercise 7.6 that

$$\begin{aligned} P(\exists \text{ exactly m fixed points}) &= P(N = m) \\ &= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \frac{1}{m!} \sum_{k=m}^{n} (-1)^{k-m} \frac{1}{(k-m)!}. \end{aligned}$$

So if n is much bigger than m we may conclude that

$$P(\exists \text{ exactly m fixed points}) \cong \frac{1}{m!} e^{-1}.$$

Let us check our results are consistent with Eq. (7.24);

$$\begin{aligned} P(\exists \text{ a fixed point}) &= \sum_{m=1}^{n} P(N = m) \\ &= \sum_{m=1}^{n} \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{1 \le m \le k \le n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{k=1}^{n} \sum_{m=1}^{k} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{k=1}^{n} \left[\sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} - (-1)^k \right] \frac{1}{k!} \\ &= -\sum_{k=1}^{n} (-1)^k \frac{1}{k!} \end{aligned}$$

wherein we have used,

$$\sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} = (1-1)^k = 0.$$

7.3.1 Appendix: Bonferroni Inequalities

In this appendix (see Feller Volume 1., p. 106-111 for more) we want to discuss what happens if we truncate the sums in the inclusion exclusion formula of Lemma 7.23. In order to do this we will need the following lemma whose combinatorial meaning was explained to me by Jeff Remmel.

Lemma 7.28. *Let $n \in \mathbb{N}_0$ and $0 \le k \le n$, then*

$$\sum_{l=0}^{k} (-1)^l \binom{n}{l} = (-1)^k \binom{n-1}{k} 1_{n>0} + 1_{n=0}. \tag{7.29}$$

Proof. The case $n = 0$ is trivial. We give two proofs for when $n \in \mathbb{N}$.

First proof. Just use induction on k. When $k = 0$, Eq. (7.29) holds since $1 = 1$. The induction step is as follows,

$$\sum_{l=0}^{k+1}(-1)^l\binom{n}{l} = (-1)^k\binom{n-1}{k}+\binom{n}{k+1}$$
$$= \frac{(-1)^{k+1}}{(k+1)!}\left[n(n-1)\dots(n-k)-(k+1)(n-1)\dots(n-k)\right]$$
$$= \frac{(-1)^{k+1}}{(k+1)!}\left[(n-1)\dots(n-k)(n-(k+1))\right] = (-1)^{k+1}\binom{n-1}{k+1}.$$

Second proof. Let $X=\{1,2,\dots,n\}$ and observe that

$$m_k := \sum_{l=0}^{k}(-1)^l\binom{n}{l} = \sum_{l=0}^{k}(-1)^l\cdot\#\left(\Lambda\in 2^X : \#(\Lambda)=l\right)$$
$$= \sum_{\Lambda\in 2^X:\ \#(\Lambda)\le k}(-1)^{\#(\Lambda)} \tag{7.30}$$

Define $T:2^X\to 2^X$ by

$$T(S) = \begin{cases} S\cup\{1\} \text{ if } 1\notin S \\ S\setminus\{1\} \text{ if } 1\in S\end{cases}.$$

Observe that T is a bijection of 2^X such that T takes even cardinality sets to odd cardinality sets and visa versa. Moreover, if we let

$$\Gamma_k := \left\{\Lambda\in 2^X : \#(\Lambda)\le k \text{ and } 1\in\Lambda \text{ if } \#(\Lambda)=k\right\},$$

then $T(\Gamma_k)=\Gamma_k$ for all $1\le k\le n$. Since

$$\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)} = \sum_{\Lambda\in\Gamma_k}(-1)^{\#(T(\Lambda))} = \sum_{\Lambda\in\Gamma_k}-(-1)^{\#(\Lambda)}$$

we see that $\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=0$. Using this observation with Eq. (7.30) implies

$$m_k = \sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)} + \sum_{\#(\Lambda)=k\ \&\ 1\notin\Lambda}(-1)^{\#(\Lambda)} = 0+(-1)^k\binom{n-1}{k}.$$

■

Corollary 7.29 (Bonferroni Inequalitites). *Let $\mu:\mathcal{A}\to[0,\mu(X)]$ be a finitely additive finite measure on $\mathcal{A}\subset 2^X$, $A_n\in\mathcal{A}$ for $n=1,2,\dots,M$, $N:=\sum_{n=1}^M 1_{A_n}$, and*

$$S_k := \sum_{1\le i_1<\dots<i_k\le M}\mu\left(A_{i_1}\cap\dots\cap A_{i_k}\right) = \mathbb{E}_\mu\left[\binom{N}{k}\right].$$

Then for $1\le k\le M$,

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{l=1}^{k}(-1)^{l+1}S_l + (-1)^k\,\mathbb{E}_\mu\left[\binom{N-1}{k}\right]. \tag{7.31}$$

This leads to the Bonferroni inequalities;

$$\mu\left(\cup_{n=1}^M A_n\right) \le \sum_{l=1}^{k}(-1)^{l+1}S_l \textit{ if } k \textit{ is odd}$$

and

$$\mu\left(\cup_{n=1}^M A_n\right) \ge \sum_{l=1}^{k}(-1)^{l+1}S_l \textit{ if } k \textit{ is even.}$$

Proof. By Lemma 7.28,

$$\sum_{l=0}^{k}(-1)^l\binom{N}{l} = (-1)^k\binom{N-1}{k}1_{N>0}+1_{N=0}.$$

Therefore integrating this equation with respect to μ gives,

$$\mu(X)+\sum_{l=1}^{k}(-1)^l S_l = \mu(N=0)+(-1)^k\,\mathbb{E}_\mu\binom{N-1}{k}$$

and therefore,

$$\mu\left(\cup_{n=1}^M A_n\right) = \mu(N>0) = \mu(X)-\mu(N=0)$$
$$= -\sum_{l=1}^{k}(-1)^l S_l + (-1)^k\,\mathbb{E}_\mu\binom{N-1}{k}.$$

The Bonferroni inequalities are a simple consequence of Eq. (7.31) and the fact that

$$\binom{N-1}{k}\ge 0 \implies \mathbb{E}_\mu\binom{N-1}{k}\ge 0.$$

■

7.3.2 Appendix: Riemann Stieljtes integral

In this subsection, let X be a set, $\mathcal{A}\subset 2^X$ be an algebra of sets, and $P:=\mu:\mathcal{A}\to[0,\infty)$ be a finitely additive measure with $\mu(X)<\infty$. As above let

$$\mathbb{E}_\mu f := \int_X f\,d\mu := \sum_{\lambda\in\mathbb{C}}\lambda\mu(f=\lambda)\ \forall\ f\in\mathbb{S}(\mathcal{A}). \tag{7.32}$$

Notation 7.30 *For any function,* $f : X \to \mathbb{C}$ *let* $\|f\|_u := \sup_{x\in X} |f(x)|$. *Further, let* $\bar{\mathbb{S}} := \overline{\mathbb{S}(\mathcal{A})}$ *denote those functions,* $f : X \to \mathbb{C}$ *such that there exists* $f_n \in \mathbb{S}(\mathcal{A})$ *such that* $\lim_{n\to\infty} \|f - f_n\|_u = 0$.

Exercise 7.7. Prove the following statements.

1. For all $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}_\mu f| \le \mu(X) \|f\|_u . \tag{7.33}$$
2. If $f \in \bar{\mathbb{S}}$ and $f_n \in \mathbb{S} := \mathbb{S}(\mathcal{A})$ such that $\lim_{n\to\infty} \|f - f_n\|_u = 0$, show $\lim_{n\to\infty} \mathbb{E}_\mu f_n$ exists. Also show that defining $\mathbb{E}_\mu f := \lim_{n\to\infty} \mathbb{E}_\mu f_n$ is well defined, i.e. you must show that $\lim_{n\to\infty} \mathbb{E}_\mu f_n = \lim_{n\to\infty} \mathbb{E}_\mu g_n$ if $g_n \in \mathbb{S}$ such that $\lim_{n\to\infty} \|f - g_n\|_u = 0$.
3. Show $\mathbb{E}_\mu : \bar{\mathbb{S}} \to \mathbb{C}$ is still linear and still satisfies Eq. (7.33).
4. Show $|f| \in \bar{\mathbb{S}}$ if $f \in \bar{\mathbb{S}}$ and that Eq. (7.19) is still valid, i.e. $|\mathbb{E}_\mu f| \le \mathbb{E}_\mu |f|$ for all $f \in \bar{\mathbb{S}}$.

Let us now specialize the above results to the case where $X = [0, T]$ for some $T < \infty$. Let $\mathcal{S} := \{(a, b] : 0 \le a \le b \le T\} \cup \{0\}$ which is easily seen to be a semi-algebra. The following proposition is fairly straightforward and will be left to the reader.

Proposition 7.31 (Riemann Stieljtes integral). *Let* $F : [0, T] \to \mathbb{R}$ *be an increasing function, then;*

1. *there exists a unique finitely additive measure,* μ_F, *on* $\mathcal{A} := \mathcal{A}(\mathcal{S})$ *such that* $\mu_F((a, b]) = F(b) - F(a)$ *for all* $0 \le a \le b \le T$ *and* $\mu_F(\{0\}) = 0$. *(In fact one could allow for* $\mu_F(\{0\}) = \lambda$ *for any* $\lambda \ge 0$, *but we would then have to write* $\mu_{F,\lambda}$ *rather than* μ_F.*)*
2. *Show* $C([0, 1], \mathbb{C}) \subset \overline{\mathbb{S}(\mathcal{A})}$. *More precisely, suppose* $\pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$ *is a partition of* $[0, T]$ *and* $c = (c_1, \ldots, c_n) \in [0, T]^n$ *with* $t_{i-1} \le c_i \le t_i$ *for each* i. *Then for* $f \in C([0, 1], \mathbb{C})$, *let*
$$f_{\pi,c} := f(0) 1_{\{0\}} + \sum_{i=1}^n f(c_i) 1_{(t_{i-1}, t_i]}. \tag{7.34}$$
Show that $\|f - f_{\pi,c}\|_u$ *is small provided,* $|\pi| := \max\{|t_i - t_{i-1}| : i = 1, 2, \ldots, n\}$ *is small.*
3. *Using the above results, show*
$$\int_{[0,T]} f d\mu_F = \lim_{|\pi|\to 0} \sum_{i=1}^n f(c_i)(F(t_i) - F(t_{i-1}))$$
where the c_i *may be chosen arbitrarily subject to the constraint that* $t_{i-1} \le c_i \le t_i$.

It is customary to write $\int_0^T f dF$ for $\int_{[0,T]} f d\mu_F$. This integral satisfies the estimates,
$$\left| \int_{[0,T]} f d\mu_F \right| \le \int_{[0,T]} |f| \, d\mu_F \le \|f\|_u (F(T) - F(0)) \ \forall \ f \in \overline{\mathbb{S}(\mathcal{A})}.$$
When $F(t) = t$,
$$\int_0^T f dF = \int_0^T f(t) \, dt,$$
is the usual Riemann integral.

Exercise 7.8. Let $a \in (0, T)$, $\lambda > 0$, and
$$G(x) = \lambda \cdot 1_{x\ge a} = \begin{cases} \lambda \text{ if } x \ge a \\ 0 \text{ if } x < a \end{cases}.$$

1. Explicitly compute $\int_{[0,T]} f d\mu_G$ for all $f \in C([0, 1], \mathbb{C})$.
2. If $F(x) = x + \lambda \cdot 1_{x\ge a}$ describe $\int_{[0,T]} f d\mu_F$ for all $f \in C([0, 1], \mathbb{C})$. **Hint:** if $F(x) = G(x) + H(x)$ where G and H are two increasing functions on $[0, T]$, show
$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G + \int_{[0,T]} f d\mu_H.$$

Exercise 7.9. Suppose that $F, G : [0, T] \to \mathbb{R}$ are two increasing functions such that $F(0) = G(0)$, $F(T) = G(T)$, and $F(x) \ne G(x)$ for at most countably many points, $x \in (0, T)$. Show
$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G \text{ for all } f \in C([0, 1], \mathbb{C}). \tag{7.35}$$
Note: given $F(0) = G(0)$, $\mu_F = \mu_G$ on $\mathcal{A}$ iff $F = G$.

One of the points of the previous exercise is to show that Eq. (7.35) holds when $G(x) := F(x+)$ – the right continuous version of F. The exercise applies since an increasing function can have at most countably many jumps, see Remark **??**. So if we only want to integrate continuous functions, we may always assume that $F : [0, T] \to \mathbb{R}$ is right continuous.

7.4 Simple Independence and the Weak Law of Large Numbers

To motivate the exercises in this section, let us imagine that we are following the outcomes of two "independent" experiments with values $\{\alpha_k\}_{k=1}^\infty \subset \Lambda_1$ and

$\{\beta_k\}_{k=1}^\infty \subset \Lambda_2$ where Λ_1 and Λ_2 are two finite set of outcomes. Here we are using term independent in an intuitive form to mean that knowing the outcome of one of the experiments gives us no information about outcome of the other.

As an example of independent experiments, suppose that one experiment is the outcome of spinning a roulette wheel and the second is the outcome of rolling a dice. We expect these two experiments will be independent.

As an example of dependent experiments, suppose that dice roller now has two dice – one red and one black. The person rolling dice throws his black or red dice after the roulette ball has stopped and landed on either black or red respectively. If the black and the red dice are weighted differently, we expect that these two experiments are no longer independent.

Lemma 7.32 (Heuristic). *Suppose that $\{\alpha_k\}_{k=1}^\infty \subset \Lambda_1$ and $\{\beta_k\}_{k=1}^\infty \subset \Lambda_2$ are the outcomes of repeatedly running two experiments independent of each other and for $x \in \Lambda_1$ and $y \in \Lambda_2$,*

$$\begin{aligned} p(x,y) &:= \lim_{N\to\infty} \frac{1}{N} \#\{1 \le k \le N : \alpha_k = x \text{ and } \beta_k = y\}, \\ p_1(x) &:= \lim_{N\to\infty} \frac{1}{N} \#\{1 \le k \le N : \alpha_k = x\}, \text{ and} \\ p_2(y) &:= \lim_{N\to\infty} \frac{1}{N} \#\{1 \le k \le N : \beta_k = y\}. \end{aligned} \tag{7.36}$$

Then $p(x,y) = p_1(x)\, p_2(y)$. In particular this then implies for any $h : \Lambda_1 \times \Lambda_2 \to \mathbb{R}$ we have,

$$\mathbb{E}h = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N h(\alpha_k, \beta_k) = \sum_{(x,y)\in\Lambda_1\times\Lambda_2} h(x,y)\, p_1(x)\, p_2(y).$$

Proof. (Heuristic.) Let us imagine running the first experiment repeatedly with the results being recorded as, $\left\{\alpha_k^\ell\right\}_{k=1}^\infty$, where $\ell \in \mathbb{N}$ indicates the ℓ^{th} – run of the experiment. Then we have postulated that, independent of ℓ,

$$p(x,y) := \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\left\{\alpha_k^\ell = x \text{ and } \beta_k = y\right\}} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\left\{\alpha_k^\ell = x\right\}} \cdot 1_{\{\beta_k = y\}}$$

So for any $L \in \mathbb{N}$ we must also have,

$$\begin{aligned} p(x,y) &= \frac{1}{L} \sum_{\ell=1}^L p(x,y) = \frac{1}{L} \sum_{\ell=1}^L \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\left\{\alpha_k^\ell = x\right\}} \cdot 1_{\{\beta_k = y\}} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \frac{1}{L} \sum_{\ell=1}^L 1_{\left\{\alpha_k^\ell = x\right\}} \cdot 1_{\{\beta_k = y\}}. \end{aligned}$$

Taking the limit of this equation as $L \to \infty$ and interchanging the order of the limits (this is faith based) implies,

$$p(x,y) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\beta_k = y\}} \cdot \lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^L 1_{\left\{\alpha_k^\ell = x\right\}}. \tag{7.37}$$

Since for fixed k, $\left\{\alpha_k^\ell\right\}_{\ell=1}^\infty$ is just another run of the first experiment, by our postulate, we conclude that

$$\lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^L 1_{\left\{\alpha_k^\ell = x\right\}} = p_1(x) \tag{7.38}$$

independent of the choice of k. Therefore combining Eqs. (7.36), (7.37), and (7.38) implies,

$$p(x,y) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\beta_k = y\}} \cdot p_1(x) = p_2(y)\, p_1(x).$$

■

To understand this "Lemma" in another but equivalent way, let $X_1 : \Lambda_1 \times \Lambda_2 \to \Lambda_1$ and $X_2 : \Lambda_1 \times \Lambda_2 \to \Lambda_2$ be the projection maps, $X_1(x,y) = x$ and $X_2(x,y) = y$ respectively. Further suppose that $f : \Lambda_1 \to \mathbb{R}$ and $g : \Lambda_2 \to \mathbb{R}$ are functions, then using the heuristics Lemma 7.32 implies,

$$\begin{aligned} \mathbb{E}\left[f(X_1)\, g(X_2)\right] &= \sum_{(x,y)\in\Lambda_1\times\Lambda_2} f(x)\, g(y)\, p_1(x)\, p_2(y) \\ &= \sum_{x\in\Lambda_1} f(x)\, p_1(x) \cdot \sum_{y\in\Lambda_2} g(y)\, p_2(y) = \mathbb{E}f(X_1) \cdot \mathbb{E}g(X_2). \end{aligned}$$

Hopefully these heuristic computations will convince you that the mathematical notion of independence developed below is relevant. In what follows, we will use the obvious generalization of our "results" above to the setting of n – independent experiments. For notational simplicity we will now assume that $\Lambda_1 = \Lambda_2 = \cdots = \Lambda_n = \Lambda$.

Let Λ be a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, and $X_i : \Omega \to \Lambda$ be defined by $X_i(\omega) = \omega_i$ for $\omega \in \Omega$ and $i = 1, 2, \dots, n$. We further suppose $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1$$

and $P : 2^\Omega \to [0,1]$ is the probability measure defined by

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \in 2^\Omega. \tag{7.39}$$

Exercise 7.10 (Simple Independence 1.). Suppose $q_i : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$ for $i = 1,2,\dots,n$ and now define $p(\omega) = \prod_{i=1}^n q_i(\omega_i)$. Show for any functions, $f_i : \Lambda \to \mathbb{R}$ that

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] = \prod_{i=1}^n \mathbb{E}_{Q_i} f_i$$

where Q_i is the measure on Λ defined by, $Q_i(\gamma) = \sum_{\lambda\in\gamma} q_i(\lambda)$ for all $\gamma \subset \Lambda$.

Exercise 7.11 (Simple Independence 2.). Prove the converse of the previous exercise. Namely, if

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] \tag{7.40}$$

for any functions, $f_i : \Lambda \to \mathbb{R}$, then there exists functions $q_i : \Lambda \to [0,1]$ with $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$, such that $p(\omega) = \prod_{i=1}^n q_i(\omega_i)$.

Definition 7.33 (Independence). *We say simple random variables, $X_1, \dots, X_n$ with values in Λ on some probability space, $(\Omega, \mathcal{A}, P)$ are independent (more precisely P – independent) if Eq. (7.40) holds for all functions, $f_i : \Lambda \to \mathbb{R}$.*

Exercise 7.12 (Simple Independence 3.). Let $X_1, \dots, X_n : \Omega \to \Lambda$ and $P : 2^\Omega \to [0,1]$ be as described before Exercise 7.10. Show $X_1, \dots, X_n$ are independent iff

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \dots P(X_n \in A_n) \tag{7.41}$$

for all choices of $A_i \subset \Lambda$. Also explain why it is enough to restrict the A_i to single point subsets of Λ.

Exercise 7.13 (A Weak Law of Large Numbers). Suppose that $\Lambda \subset \mathbb{R}$ is a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, $p(\omega) = \prod_{i=1}^n q(\omega_i)$ where $q : \Lambda \to [0,1]$ such that $\sum_{\lambda\in\Lambda} q(\lambda) = 1$, and let $P : 2^\Omega \to [0,1]$ be the probability measure defined as in Eq. (7.39). Further let $X_i(\omega) = \omega_i$ for $i = 1,2,\dots,n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, and

$$S_n = \frac{1}{n}(X_1 + \dots + X_n).$$

1. Show, $\xi = \sum_{\lambda\in\Lambda} \lambda\, q(\lambda)$ and
$$\sigma^2 = \sum_{\lambda\in\Lambda} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda\in\Lambda} \lambda^2 q(\lambda) - \xi^2. \tag{7.42}$$

2. Show, $\mathbb{E}S_n = \xi$.
3. Let $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Show
$$\mathbb{E}[(X_i - \xi)(X_j - \xi)] = \delta_{ij}\sigma^2.$$
4. Using $S_n - \xi$ may be expressed as, $\frac{1}{n}\sum_{i=1}^n (X_i - \xi)$, show
$$\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{7.43}$$
5. Conclude using Eq. (7.43) and Remark 7.22 that
$$P(|S_n - \xi| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2. \tag{7.44}$$

So for large n, S_n is concentrated near $\xi = \mathbb{E}X_i$ with probability approaching 1 for n large. This is a version of the weak law of large numbers.

Definition 7.34 (Covariance). *Let $(\Omega, \mathcal{B}, P)$ is a finitely additive probability. The* ***covariance,*** $\operatorname{Cov}(X,Y)$, *of $X, Y \in \mathbb{S}(\mathcal{B})$ is defined by*

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X - \xi_X)(Y - \xi_Y)] = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

where $\xi_X := \mathbb{E}X$ and $\xi_Y := \mathbb{E}Y$. The variance of X,

$$\operatorname{Var}(X) := \operatorname{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$$

We say that X and Y are ***uncorrelated*** *if $\operatorname{Cov}(X,Y) = 0$, i.e. $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$. More generally we say $\{X_k\}_{k=1}^n \subset \mathbb{S}(\mathcal{B})$ are uncorrelated iff $\operatorname{Cov}(X_i, X_j) = 0$ for all $i \neq j$.*

Remark 7.35. 1. Observe that X and Y are independent iff $f(X)$ and $g(Y)$ are uncorrelated for all functions, f and g on the range of X and Y respectively. In particular if X and Y are independent then $\operatorname{Cov}(X,Y) = 0$.

2. If you look at your proof of the weak law of large numbers in Exercise 7.13 you will see that it suffices to assume that $\{X_i\}_{i=1}^n$ are uncorrelated rather than the stronger condition of being independent.

Exercise 7.14 (Bernoulli Random Variables). Let $\Lambda = \{0,1\}$, $X : \Lambda \to \mathbb{R}$ be defined by $X(0) = 0$ and $X(1) = 1$, $x \in [0,1]$, and define $Q = x\delta_1 + (1-x)\delta_0$, i.e. $Q(\{0\}) = 1 - x$ and $Q(\{1\}) = x$. Verify,

$$\xi(x) := \mathbb{E}_Q X = x \text{ and}$$
$$\sigma^2(x) := \mathbb{E}_Q(X - x)^2 = (1-x)x \leq 1/4.$$

Theorem 7.36 (Weierstrass Approximation Theorem via Bernstein's Polynomials.). *Suppose that* $f \in C([0,1], \mathbb{C})$ *and*

$$p_n(x) := \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Then

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f(x) - p_n(x)| = 0.$$

Proof. Let $x \in [0,1]$, $\Lambda = \{0,1\}$, $q(0) = 1-x$, $q(1) = x$, $\Omega = \Lambda^n$, and

$$P_x(\{\omega\}) = q(\omega_1)\dots q(\omega_n) = x^{\sum_{i=1}^n \omega_i} \cdot (1-x)^{1-\sum_{i=1}^n \omega_i}.$$

As above, let $S_n = \frac{1}{n}(X_1 + \dots + X_n)$, where $X_i(\omega) = \omega_i$ and observe that

$$P_x\left(S_n = \frac{k}{n}\right) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Therefore, writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$, we have

$$\mathbb{E}_x[f(S_n)] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} = p_n(x).$$

Hence we find

$$\begin{aligned} |p_n(x) - f(x)| &= |\mathbb{E}_x f(S_n) - f(x)| = |\mathbb{E}_x[f(S_n) - f(x)]| \\ &\le \mathbb{E}_x|f(S_n) - f(x)| \\ &= \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| \ge \varepsilon] \\ &\quad + \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| < \varepsilon] \\ &\le 2M \cdot P_x(|S_n - x| \ge \varepsilon) + \delta(\varepsilon) \end{aligned}$$

where

$$\begin{aligned} M &:= \max_{y\in[0,1]} |f(y)| \text{ and} \\ \delta(\varepsilon) &:= \sup\{|f(y) - f(x)| : x, y \in [0,1] \text{ and } |y-x| \le \varepsilon\} \end{aligned}$$

is the modulus of continuity of f. Now by the above exercises,

$$P_x(|S_n - x| \ge \varepsilon) \le \frac{1}{4n\varepsilon^2} \quad \text{(see Figure 7.1)} \tag{7.45}$$

and hence we may conclude that

$$\max_{x\in[0,1]} |p_n(x) - f(x)| \le \frac{M}{2n\varepsilon^2} + \delta(\varepsilon)$$

and therefore, that

$$\limsup_{n\to\infty} \max_{x\in[0,1]} |p_n(x) - f(x)| \le \delta(\varepsilon).$$

This completes the proof, since by uniform continuity of f, $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$. ■

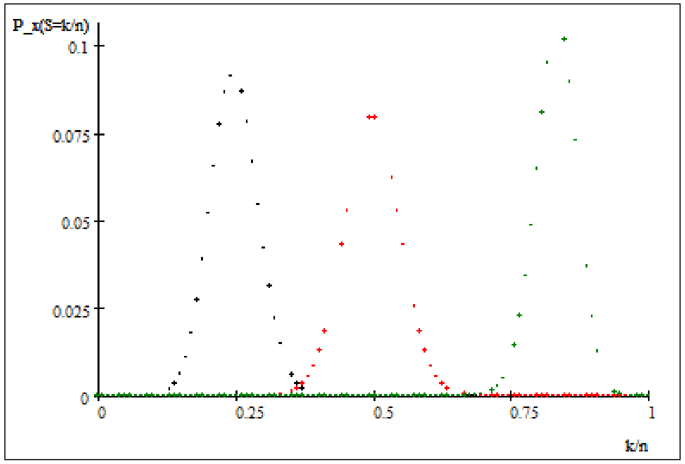


Fig. 7.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

7.4.1 Complex Weierstrass Approximation Theorem

The main goal of this subsection is to prove Theorem 7.42 which states that any continuous 2π – periodic function on $\mathbb{R}$ may be well approximated by trigonometric polynomials. The main ingredient is the following two dimensional generalization of Theorem 7.36. All of the results in this section have natural generalization to higher dimensions as well , see Theorem 7.46.

Theorem 7.37 (Weierstrass Approximation Theorem). *Suppose that* $K = [0,1]^2$, $f \in C(K, \mathbb{C})$, *and*

$$p_n(x,y) := \sum_{k,l=0}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right)\binom{n}{k}\binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}. \tag{7.46}$$

Then $p_n \to f$ *uniformly on* K.

Proof. We are going to follow the argument given in the proof of Theorem 7.36. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([0,1]^2, \mathbb{R})$. For $(x,y) \in K$ and $n \in \mathbb{N}$ we may choose a collection of independent Bernoulli simple random variables $\{X_i, Y_i\}_{i=1}^n$ such that $P(X_i = 1) = x$ and $P(Y_i = 1) = y$ for all $1 \le i \le n$. Then letting $S_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$, we have

$$\mathbb{E}[f(S_n, T_n)] = \sum_{k,l=0}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right) P(n \cdot S_n = k, n \cdot T_n = l) = p_n(x,y)$$

where $p_n(x,y)$ is the polynomial given in Eq. (7.46) wherein the assumed independence is needed to show,

$$P(n \cdot S_n = k, n \cdot T_n = l) = \binom{n}{k}\binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}.$$

Thus if $M = \sup\{|f(x,y)| : (x,y) \in K\}$, $\varepsilon > 0$,

$$\delta_\varepsilon = \sup\{|f(x', y') - f(x,y)| : (x,y), (x',y') \in K \text{ and } \|(x',y') - (x,y)\| \le \varepsilon\},$$

and

$$A := \{\|(S_n, T_n) - (x,y)\| > \varepsilon\},$$

we have,

$$\begin{aligned} |f(x,y) - p_n(x,y)| &= |\mathbb{E}(f(x,y) - f((S_n,T_n)))| \\ &\le \mathbb{E}|f(x,y) - f((S_n,T_n))| \\ &= \mathbb{E}[|f(x,y) - f(S_n,T_n)| : A] \\ &\quad + \mathbb{E}[|f(x,y) - f(S_n,T_n)| : A^c] \\ &\le 2M \cdot P(A) + \delta_\varepsilon \cdot P(A^c) \\ &\le 2M \cdot P(A) + \delta_\varepsilon. \end{aligned} \tag{7.47}$$

To estimate $P(A)$, observe that if

$$\|(S_n, T_n) - (x,y)\|^2 = (S_n - x)^2 + (T_n - y)^2 > \varepsilon^2,$$

then either,

$$(S_n - x)^2 > \varepsilon^2/2 \text{ or } (T_n - y)^2 > \varepsilon^2/2$$

and therefore by sub-additivity and Eq. (7.45) we know

$$\begin{aligned} P(A) &\le P\left(|S_n - x| > \varepsilon/\sqrt{2}\right) + P\left(|T_n - y| > \varepsilon/\sqrt{2}\right) \\ &\le \frac{1}{2n\varepsilon^2} + \frac{1}{2n\varepsilon^2} = \frac{1}{n\varepsilon^2}. \end{aligned} \tag{7.48}$$

Using this estimate in Eq. (7.47) gives,

$$|f(x,y) - p_n(x,y)| \le 2M \cdot \frac{1}{n\varepsilon^2} + \delta_\varepsilon$$

and as right is independent of $(x,y) \in K$ we may conclude,

$$\limsup_{n\to\infty} \sup_{(x,y)\in K} |f(x,y) - p_n(x,y)| \le \delta_\varepsilon$$

which completes the proof since $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$ because f is uniformly continuous on K. ∎

Remark 7.38. We can easily improve our estimate on $P(A)$ in Eq. (7.48) by a factor of two as follows. As in the proof of Theorem 7.36,

$$\begin{aligned} \mathbb{E}\left[\|(S_n,T_n) - (x,y)\|^2\right] &= \mathbb{E}\left[(S_n - x)^2 + (T_n - y)^2\right] \\ &= \mathrm{Var}(S_n) + \mathrm{Var}(T_n) \\ &= \frac{1}{n} x(1-x) + y(1-y) \le \frac{1}{2n}. \end{aligned}$$

Therefore by Chebyshev's inequality,

$$P(A) = P(\|(S_n,T_n) - (x,y)\| > \varepsilon) \le \frac{1}{\varepsilon^2}\mathbb{E}\|(S_n,T_n) - (x,y)\|^2 \le \frac{1}{2n\varepsilon^2}.$$

Corollary 7.39. *Suppose that $K = [a,b] \times [c,d]$ is any compact rectangle in $\mathbb{R}^2$. Then every function, $f \in C(K, \mathbb{C})$, may be uniformly approximated by polynomial functions in $(x,y) \in \mathbb{R}^2$.*

Proof. Let $F(x,y) := f(a + x(b-a), c + y(d-c))$ – a continuous function of $(x,y) \in [0,1]^2$. Given $\varepsilon > 0$, we may use Theorem Theorem 7.37 to find a polynomial, $p(x,y)$, such that $\sup_{(x,y)\in[0,1]^2} |F(x,y) - p(x,y)| \le \varepsilon$. Letting $\xi = a + x(b-a)$ and $\eta := c + y(d-c)$, it now follows that

$$\sup_{(\xi.\eta)\in K} \left| f(\xi,\eta) - p\left(\frac{\xi - a}{b-a}, \frac{\eta - c}{d-c}\right)\right| \le \varepsilon$$

which completes the proof since $p\left(\frac{\xi-a}{b-a}, \frac{\eta-c}{d-c}\right)$ is a polynomial in (ξ,η). ∎

Here is a version of the complex Weierstrass approximation theorem.

Theorem 7.40 (Complex Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{C}$ is a compact rectangle. Then there exists polynomials in $(z = x + iy, \bar{z} = x - iy)$, $p_n(z,\bar{z})$ for $z \in \mathbb{C}$, such that $\sup_{z\in K} |q_n(z,\bar{z}) - f(z)| \to 0$ as $n \to \infty$ for every $f \in C(K,\mathbb{C})$.*

Proof. The mapping $(x,y) \in \mathbb{R} \times \mathbb{R} \to z = x + iy \in \mathbb{C}$ is an isomorphism of vector spaces. Letting $\bar{z} = x - iy$ as usual, we have $x = \frac{z+\bar{z}}{2}$ and $y = \frac{z-\bar{z}}{2i}$. Therefore under this identification any polynomial $p(x,y)$ on $\mathbb{R} \times \mathbb{R}$ may be written as a polynomial q in $(z,\bar{z})$, namely

$$q(z,\bar{z}) = p\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i}\right).$$

Conversely a polynomial q in $(z,\bar{z})$ may be thought of as a polynomial p in (x,y), namely $p(x,y) = q(x+iy, x-iy)$. Hence the result now follows from Theorem 7.37. ■

Example 7.41. Let $K = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}$ be the set of polynomials in $(z,\bar{z})$ restricted to S^1. Then $\mathcal{A}$ is dense in $C(S^1)$. To prove this first observe if $f \in C\left(S^1\right)$ then $F(z) = |z| f\left(\frac{z}{|z|}\right)$ for $z \neq 0$ and $F(0) = 0$ defines $F \in C(\mathbb{C})$ such that $F|_{S^1} = f$. By applying Theorem 7.40 to F restricted to a compact rectangle containing S^1 we may find $q_n(z,\bar{z})$ converging uniformly to F on K and hence on S^1. Since $\bar{z} = z^{-1}$ on S^1, we have shown polynomials in z and z^{-1} are dense in $C(S^1)$.

Theorem 7.42 (Density of Trigonometric Polynomials). *Any 2π – periodic continuous function, $f : \mathbb{R} \to \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form*

$$p(x) = \sum_{\lambda \in \Lambda} a_\lambda e^{i\lambda \cdot x}$$

where Λ is a finite subset of $\mathbb{Z}$ and $a_\lambda \in \mathbb{C}$ for all $\lambda \in \Lambda$.

Proof. For $z \in S^1$, define $F(z) := f(\theta)$ where $\theta \in \mathbb{R}$ is chosen so that $z = e^{i\theta}$. Since f is 2π – periodic, F is well defined since if θ solves $e^{i\theta} = z$ then all other solutions are of the form $\{\theta + 2\pi n : n \in \mathbb{Z}\}$. Since the map $\theta \to e^{i\theta}$ is a local homeomorphism, i.e. for any $J = (a,b)$ with $b - a < 2\pi$, the map $\theta \in J \xrightarrow{\phi} \tilde{J} := \left\{e^{i\theta} : \theta \in J\right\} \subset S^1$ is a homeomorphism, it follows that $F(z) = f \circ \phi^{-1}(z)$ for $z \in \tilde{J}$. This shows F is continuous when restricted to $\tilde{J}$. Since such sets cover S^1, it follows that F is continuous.

By Example 7.41, the polynomials in z and $\bar{z} = z^{-1}$ are dense in $C(S^1)$. Hence for any $\varepsilon > 0$ there exists

$$p(z,\bar{z}) = \sum_{0 \le m,n \le N} a_{m,n} z^m \bar{z}^n$$

such that $|F(z) - p(z,\bar{z})| \le \varepsilon$ for all $z \in S^1$. Taking $z = e^{i\theta}$ then implies

$$\sup_\theta \left|f(\theta) - p\left(e^{i\theta}, e^{-i\theta}\right)\right| \le \varepsilon$$

where

$$p\left(e^{i\theta}, e^{-i\theta}\right) = \sum_{0 \le m,n \le N} a_{m,n} e^{i(m-n)\theta}$$

is the desired trigonometry polynomial. ■

7.4.2 Product Measures and Fubini's Theorem

In the last part of this section we will extend some of the above ideas to more general "finitely additive measure spaces." A **finitely additive measure space** is a triple, $(X, \mathcal{A}, \mu)$, where X is a set, $\mathcal{A} \subset 2^X$ is an algebra, and $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure. Let $(Y, \mathcal{B}, \nu)$ be another finitely additive measure space.

Definition 7.43. *Let $\mathcal{A} \odot \mathcal{B}$ be the smallest sub-algebra of $2^{X \times Y}$ containing all sets of the form $\mathcal{S} := \{A \times B : A \in \mathcal{A}$ and $B \in \mathcal{B}\}$. As we have seen in Exercise 6.10, $\mathcal{S}$ is a semi-algebra and therefore $\mathcal{A} \odot \mathcal{B}$ consists of subsets, $C \subset X \times Y$, which may be written as;*

$$C = \sum_{i=1}^n A_i \times B_i \text{ with } A_i \times B_i \in \mathcal{S}. \tag{7.49}$$

Theorem 7.44 (Product Measure and Fubini's Theorem). *Assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ for simplicity. Then there is a unique finitely additive measure, $\mu \odot \nu$, on $\mathcal{A} \odot \mathcal{B}$ such that $\mu \odot \nu (A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Moreover if $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$ then;*

1. *$y \to f(x,y)$ is in $\mathbb{S}(\mathcal{B})$ for all $x \in X$ and $x \to f(x,y)$ is in $\mathbb{S}(\mathcal{A})$ for all $y \in Y$.*
2. *$x \to \int_Y f(x,y)\,d\nu(y)$ is in $\mathbb{S}(\mathcal{A})$ and $y \to \int_X f(x,y)\,d\mu(x)$ is in $\mathbb{S}(\mathcal{B})$.*
3. *we have,*

$$\begin{aligned}\int_X \left[\int_Y f(x,y)\,d\nu(y)\right] d\mu(x) \\ &= \int_{X\times Y} f(x,y)\,d(\mu \odot \nu)(x,y) \\ &= \int_Y \left[\int_X f(x,y)\,d\mu(x)\right] d\nu(y).\end{aligned}$$

*We will refer to $\mu \odot \nu$ as the **product measure** of μ and ν.*

Proof. According to Eq. (7.49),

$$1_C(x,y) = \sum_{i=1}^n 1_{A_i \times B_i}(x,y) = \sum_{i=1}^n 1_{A_i}(x)\, 1_{B_i}(y)$$

from which it follows that $1_C(x,\cdot) \in \mathbb{S}(\mathcal{B})$ for each $x \in X$ and

$$\int_Y 1_C(x,y)\,d\nu(y) = \sum_{i=1}^n 1_{A_i}(x)\,\nu(B_i).$$

It now follows from this equation that $x \to \int_Y 1_C(x,y)\,d\nu(y) \in \mathbb{S}(\mathcal{A})$ and that

$$\int_X \left[\int_Y 1_C(x,y)\,d\nu(y)\right] d\mu(x) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i).$$

Similarly one shows that

$$\int_Y \left[\int_X 1_C(x,y)\,d\mu(x)\right] d\nu(y) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i).$$

In particular this shows that we may define

$$(\mu \odot \nu)(C) = \sum_{i=1}^n \mu(A_i)\,\nu(B_i)$$

and with this definition we have,

$$\int_X \left[\int_Y 1_C(x,y)\,d\nu(y)\right] d\mu(x) = (\mu \odot \nu)(C) = \int_Y \left[\int_X 1_C(x,y)\,d\mu(x)\right] d\nu(y).$$

From either of these representations it is easily seen that $\mu \odot \nu$ is a finitely additive measure on $\mathcal{A} \odot \mathcal{B}$ with the desired properties. Moreover, we have already verified the Theorem in the special case where $f = 1_C$ with $C \in \mathcal{A} \odot \mathcal{B}$. Since the general element, $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$, is a linear combination of such functions, it is easy to verify using the linearity of the integral and the fact that $\mathbb{S}(\mathcal{A})$ and $\mathbb{S}(\mathcal{B})$ are vector spaces that the theorem is true in general. ■

Example 7.45. Suppose that $f \in \mathbb{S}(\mathcal{A})$ and $g \in \mathbb{S}(\mathcal{B})$. Let $f \otimes g(x,y) := f(x)\,g(y)$. Since we have,

$$f \otimes g(x,y) = \left(\sum_a a 1_{f=a}(x)\right)\left(\sum_b b 1_{g=b}(y)\right) = \sum_{a,b} ab 1_{\{f=a\}\times\{g=b\}}(x,y)$$

it follows that $f \otimes g \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$. Moreover, using Fubini's Theorem 7.44 it follows that

$$\int_{X\times Y} f \otimes g \; d(\mu \odot \nu) = \left[\int_X f\; d\mu\right]\left[\int_Y g\; d\nu\right].$$

7.5 Appendix: A Multi-dimensional Weirstrass Approximation Theorem

The following theorem is the multi-dimensional generalization of Theorem 7.36.

Theorem 7.46 (Weierstrass Approximation Theorem). *Suppose that $K = [a_1,b_1] \times \dots [a_d,b_d]$ with $-\infty < a_i < b_i < \infty$ is a compact rectangle in $\mathbb{R}^d$. Then for every $f \in C(K,\mathbb{C})$, there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K.*

Proof. By a simple scaling and translation of the arguments of f we may assume without loss of generality that $K = [0,1]^d$. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([\mathbf{0},\mathbf{1}],\mathbb{R})$.

Given $x \in K$, let $\left\{X_n = \left(X_n^1,\dots,X_n^d\right)\right\}_{n=1}^\infty$ be i.i.d. random vectors with values in $\mathbb{R}^d$ such that

$$P(X_n = \eta) = \prod_{i=1}^d (1-x_i)^{1-\eta_i} x_i^{\eta_i}$$

for all $\eta = (\eta_1,\dots,\eta_d) \in \{0,1\}^d$. Since each X_n^j is a Bernoulli random variable with $P\left(X_n^j = 1\right) = x_j$, we know that

$$\mathbb{E}X_n = x \text{ and } \operatorname{Var}\left(X_n^j\right) = x_j - x_j^2 = x_j(1-x_j).$$

As usual let $S_n = S_n := X_1 + \dots + X_n \in \mathbb{R}^d$, then

$$\mathbb{E}\left[\frac{S_n}{n}\right] = x \text{ and}$$

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{S_n}{n} - x\right\|^2\right] &= \sum_{j=1}^d \mathbb{E}\left(\frac{S_n^j}{n} - x_j\right)^2 = \sum_{j=1}^d \operatorname{Var}\left(\frac{S_n^j}{n} - x_j\right) \\ &= \sum_{j=1}^d \operatorname{Var}\left(\frac{S_n^j}{n}\right) = \frac{1}{n^2} \cdot \sum_{j=1}^d \sum_{k=1}^n \operatorname{Var}\left(X_k^j\right) \\ &= \frac{1}{n}\sum_{j=1}^d x_j(1-x_j) \le \frac{d}{4n}.\end{aligned}$$

This shows $S_n/n \to x$ in $L^2(P)$ and hence by Chebyshev's inequality, $S_n/n \xrightarrow{P} x$ in and by a continuity theorem, $f\left(\frac{S_n}{n}\right) \xrightarrow{P} f(x)$ as $n\to\infty$. This along with the dominated convergence theorem shows

$$p_n(x) := \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \to f(x) \text{ as } n\to\infty, \tag{7.50}$$

where

$$\begin{aligned} p_n(x) &= \sum_{\eta:\{1,2,\dots,n\}\to\{0,1\}^d} f\left(\frac{\eta(1)+\cdots+\eta(n)}{n}\right) P(X_1=\eta(1),\dots,X_n=\eta(n)) \\ &= \sum_{\eta:\{1,2,\dots,n\}\to\{0,1\}^d} f\left(\frac{\eta(1)+\cdots+\eta(n)}{n}\right) \prod_{k=1}^n \prod_{i=1}^d (1-x_i)^{1-\eta_i(k)} x_i^{\eta_i(k)} \end{aligned}$$

is a polynomial of degree nd. In fact more is true.

Suppose $\varepsilon>0$ is given, $M=\sup\{|f(x)| : x\in K\}$, and

$$\delta_\varepsilon = \sup\{|f(y)-f(x)| : x,y\in K \text{ and } \|y-x\|\le\varepsilon\}.$$

By uniform continuity of f on K, $\lim_{\varepsilon\downarrow 0}\delta_\varepsilon = 0$. Therefore,

$$\begin{aligned} |f(x)-p_n(x)| &= \left|\mathbb{E}\left(f(x)-f\left(\frac{S_n}{n}\right)\right)\right| \le \mathbb{E}\left|f(x)-f\left(\frac{S_n}{n}\right)\right| \\ &\le \mathbb{E}\left[\left|f(x)-f\left(\frac{S_n}{n}\right)\right| : \|S_n-x\|>\varepsilon\right] \\ &\quad + \mathbb{E}\left[\left|f(x)-f\left(\frac{S_n}{n}\right)\right| : \|S_n-x\|\le\varepsilon\right] \\ &\le 2MP(\|S_n-x\|>\varepsilon)+\delta_\varepsilon. \end{aligned} \tag{7.51}$$

By Chebyshev's inequality,

$$P(\|S_n-x\|>\varepsilon) \le \frac{1}{\varepsilon^2}\mathbb{E}\|S_n-x\|^2 = \frac{d}{4n\varepsilon^2},$$

and therefore, Eq. (7.51) yields the estimate

$$\sup_{x\in K}|f(x)-p_n(x)| \le \frac{2dM}{n\varepsilon^2}+\delta_\varepsilon$$

and hence

$$\limsup_{n\to\infty}\sup_{x\in K}|f(x)-p_n(x)| \le \delta_\varepsilon \to 0 \text{ as } \varepsilon\downarrow 0.$$

■

Here is a version of the complex Weirstrass approximation theorem.

Theorem 7.47 (Complex Weierstrass Approximation Theorem). *Suppose that $K\subset\mathbb{C}^d\cong\mathbb{R}^d\times\mathbb{R}^d$ is a compact rectangle. Then there exists polynomials in $(z=x+iy,\bar z=x-iy)$, $p_n(z,\bar z)$ for $z\in\mathbb{C}^d$, such that $\sup_{z\in K}|q_n(z,\bar z)-f(z)|\to 0$ as $n\to\infty$ for every $f\in C(K,\mathbb{C})$.*

Proof. The mapping $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d\to z=x+iy\in\mathbb{C}^d$ is an isomorphism of vector spaces. Letting $\bar z=x-iy$ as usual, we have $x=\frac{z+\bar z}{2}$ and $y=\frac{z-\bar z}{2i}$. Therefore under this identification any polynomial $p(x,y)$ on $\mathbb{R}^d\times\mathbb{R}^d$ may be written as a polynomial q in $(z,\bar z)$, namely

$$q(z,\bar z) = p(\frac{z+\bar z}{2},\frac{z-\bar z}{2i}).$$

Conversely a polynomial q in $(z,\bar z)$ may be thought of as a polynomial p in (x,y), namely $p(x,y)=q(x+iy,x-iy)$. Hence the result now follows from Theorem 7.46. ■

Example 7.48. Let

$$K=\mathbb{T}^d=\left(S^1\right)^d=\{z\in\mathbb{C}^d : |z_i|=1 for 1\le i\le d\}$$

and $\mathcal{A}$ be the set of polynomials in $(z,\bar z)$ restricted to $\mathbb{T}^d$. Then $\mathcal{A}$ is dense in $C(\mathbb{T}^d)$. To prove this first observe if $f\in C(\mathbb{T}^d)$ then

$$F(z)=|z_1|\dots|z_d|\,f\left(\frac{z_1}{|z_1|},\dots,\frac{z_d}{|z_d|}\right)$$

for $z\ne 0$ and $F(0)=0$ defines $F\in C(\mathbb{C}^d)$ such that $F|_{\mathbb{T}^d}=f$. By applying Theorem 7.47 to F restricted to a compact rectangle containing $\mathbb{T}^d$ we may find $q_n(z,\bar z)$ converging uniformly to F on K and hence on $\mathbb{T}^d$. Since $\bar z=z^{-1}:=\left(z_1^{-1},\dots,z_d^{-1}\right)$ on $\mathbb{T}^d$, we have shown polynomials in z and z^{-1} are dense in $C(\mathbb{T}^d)$.

Exercise 7.15. Use Example 7.48 to show that any 2π – periodic continuous function, $g:\mathbb{R}^d\to\mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form

$$p(\theta)=\sum_{\lambda\in\Lambda}b_\lambda e^{i\lambda\cdot\theta}$$

where Λ is a finite subset of $\mathbb{Z}^d$ and $b_\lambda\in\mathbb{C}$ for all $\lambda\in\Lambda$. **Hint:** start by showing there exists a unique continuous function, $f:\left(S^1\right)^d\to\mathbb{C}$ such that $f\left(e^{i\theta_1},\dots,e^{i\theta_d}\right)=F(\theta)$ for all $\theta=(\theta_1,\dots,\theta_d)\in\mathbb{R}^d$.

Exercise 7.16. Suppose $f\in C(\mathbb{R},\mathbb{C})$ is a 2π – periodic function (i.e. $f(x+2\pi)=f(x)$ for all $x\in\mathbb{R}$) and

$$\int_0^{2\pi} f(x)\,e^{inx}dx=0 \text{ for all } n\in\mathbb{Z},$$

show again that $f\equiv 0$. **Hint**: Use Exercise 7.15.

Countably Additive Measures

Let $\mathcal{A} \subset 2^{\Omega}$ be an algebra and $\mu : \mathcal{A} \to [0, \infty]$ be a finitely additive measure. Recall that μ is a **premeasure** on $\mathcal{A}$ if μ is σ – additive on $\mathcal{A}$. If μ is a premeasure on $\mathcal{A}$ and $\mathcal{A}$ is a σ – algebra (Definition 6.12), we say that μ is a **measure** on $(\Omega, \mathcal{A})$ and that $(\Omega, \mathcal{A})$ is a **measurable space.**

Definition 8.1. *Let $(\Omega, \mathcal{B})$ be a measurable space. We say that $P : \mathcal{B} \to [0,1]$ is a **probability measure on** $(\Omega, \mathcal{B})$ if P is a measure on $\mathcal{B}$ such that $P(\Omega) = 1$. In this case we say that $(\Omega, \mathcal{B}, P)$ a probability space.*

8.1 Overview

The goal of this chapter is develop methods for proving the existence of probability measures with desirable properties. The main results of this chapter may are summarized in the following theorem.

Theorem 8.2. *A finitely additive probability measure P on an algebra, $\mathcal{A} \subset 2^{\Omega}$, extends to σ – additive measure on $\sigma(\mathcal{A})$ iff P is a premeasure on $\mathcal{A}$. If the extension exists it is unique.*

Proof. The uniqueness assertion is proved Proposition 8.15 below. The existence assertion of the theorem in the content of Theorem 8.27. ■

In order to use this theorem it is necessary to determine when a finitely additive probability measure in is in fact a premeasure. The following Proposition is sometimes useful in this regard.

Proposition 8.3 (Equivalent premeasure conditions). *Suppose that P is a finitely additive probability measure on an algebra, $\mathcal{A} \subset 2^{\Omega}$. Then the following are equivalent:*

1. *P is a premeasure on $\mathcal{A}$, i.e. P is σ – additive on $\mathcal{A}$.*
2. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, $P(A_n) \uparrow P(A)$.*
3. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, $P(A_n) \downarrow P(A)$.*
4. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow \Omega$, $P(A_n) \uparrow 1$.*
5. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow \emptyset$, $P(A_n) \downarrow 0$.*

Proof. We will start by showing $1 \iff 2 \iff 3$.

$1. \implies 2.$ Suppose $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Let $A_n' := A_n \setminus A_{n-1}$ with $A_0 := \emptyset$. Then $\{A_n'\}_{n=1}^{\infty}$ are disjoint, $A_n = \cup_{k=1}^{n} A_k'$ and $A = \cup_{k=1}^{\infty} A_k'$. Therefore,

$$P(A) = \sum_{k=1}^{\infty} P(A_k') = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k') = \lim_{n\to\infty} P(\cup_{k=1}^{n} A_k') = \lim_{n\to\infty} P(A_n).$$

$2. \implies 1.$ If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ are disjoint and $A := \cup_{n=1}^{\infty} A_n \in \mathcal{A}$, then $\cup_{n=1}^{N} A_n \uparrow A$. Therefore,

$$P(A) = \lim_{N\to\infty} P\left(\cup_{n=1}^{N} A_n\right) = \lim_{N\to\infty} \sum_{n=1}^{N} P(A_n) = \sum_{n=1}^{\infty} P(A_n).$$

$2. \implies 3.$ If $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, then $A_n^c \uparrow A^c$ and therefore,

$$\lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n^c) = P(A^c) = 1 - P(A).$$

$3. \implies 2.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A_n^c \downarrow A^c$ and therefore we again have,

$$\lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n^c) = P(A^c) = 1 - P(A).$$

The same proof used for $2. \iff 3.$ shows $4. \iff 5$ and it is clear that $3. \implies 5.$ To finish the proof we will show $5. \implies 2.$

$5. \implies 2.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A \setminus A_n \downarrow \emptyset$ and therefore

$$\lim_{n\to\infty} [P(A) - P(A_n)] = \lim_{n\to\infty} P(A \setminus A_n) = 0.$$

■

Remark 8.4. Observe that the equivalence of items 1. and 2. in the above proposition hold without the restriction that $P(\Omega) = 1$ and in fact $P(\Omega) = \infty$ may be allowed for this equivalence.

Lemma 8.5. *If $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure, then μ is countably sub-additive on $\mathcal{A}$.*

Proof. Suppose that $A_n \in \mathcal{A}$ with $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$. Let $A_1' := A_1$ and for $n \geq 2$, let $A_n' := A_n \setminus (A_1 \cup \dots A_{n-1}) \in \mathcal{A}$. Then $\cup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A_n'$ and therefore by the countable additivity and monotonicity of μ we have,

$$\mu\left(\cup_{n=1}^{\infty} A_n\right) = \mu\left(\sum_{n=1}^{\infty} A_n'\right) = \sum_{n=1}^{\infty} \mu\left(A_n'\right) \leq \sum_{n=1}^{\infty} \mu\left(A_n\right).$$

■

Let us now specialize to the case where $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{A}\left(\{(a,b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\}\right)$. In this case we will describe probability measures, P, on $\mathcal{B}_{\mathbb{R}}$ by their "cumulative distribution functions."

Definition 8.6. *Given a probability measure, P on $\mathcal{B}_{\mathbb{R}}$, the* ***cumulative distribution function*** *(CDF) of P is defined as the function, $F = F_P : \mathbb{R} \to [0,1]$ given as*

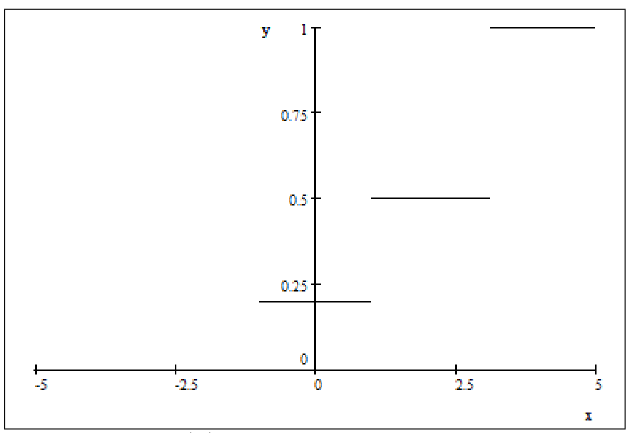
$$F(x) := P((-\infty, x]). \tag{8.1}$$

Example 8.7. Suppose that

$$P = p\delta_{-1} + q\delta_1 + r\delta_\pi$$

with $p, q, r > 0$ and $p + q + r = 1$. In this case,

$$F(x) = \begin{cases} 0 & \text{for } \quad x < -1 \\ p & \text{for } -1 \leq x < 1 \\ p+q & \text{for } \ 1 \leq x < \pi \\ 1 & \text{for } \pi \leq x < \infty \end{cases}.$$



A plot of $F(x)$ with $p = .2$, $q = .3$, and $r = .5$.

Lemma 8.8. *If $F = F_P : \mathbb{R} \to [0,1]$ is a distribution function for a probability measure, P, on $\mathcal{B}_{\mathbb{R}}$, then:*

1. *F is non-decreasing,*
2. *F is right continuous,*
3. *$F(-\infty) := \lim_{x \to -\infty} F(x) = 0$, and $F(\infty) := \lim_{x \to \infty} F(x) = 1$.*

Proof. The monotonicity of P shows that $F(x)$ in Eq. (8.1) is non-decreasing. For $b \in \mathbb{R}$ let $A_n = (-\infty, b_n]$ with $b_n \downarrow b$ as $n \to \infty$. The continuity of P implies

$$F(b_n) = P((-\infty, b_n]) \downarrow \mu((-\infty, b]) = F(b).$$

Since $\{b_n\}_{n=1}^{\infty}$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $F(b+) := \lim_{y \downarrow b} F(y) = F(b)$. This show that F is right continuous. Similar arguments show that $F(\infty) = 1$ and $F(-\infty) = 0$. ■

It turns out that Lemma 8.8 has the following important converse.

Theorem 8.9. *To each function $F : \mathbb{R} \to [0,1]$ satisfying properties 1. – 3.. in Lemma 8.8, there exists a unique probability measure, P_F, on $\mathcal{B}_{\mathbb{R}}$ such that*

$$P_F((a,b]) = F(b) - F(a) \ \text{ for all } -\infty < a \leq b < \infty.$$

Proof. The uniqueness assertion is proved in Corollary 8.17 below or see Exercises 8.2 and 8.11 below. The existence portion of the theorem is a special case of Theorem 8.33 below. ■

Example 8.10 (Uniform Distribution). The function,

$$F(x) := \begin{cases} 0 \text{ for } & x \leq 0 \\ x \text{ for } & 0 \leq x < 1 \\ 1 \text{ for } & 1 \leq x < \infty \end{cases},$$

is the distribution function for a measure, m on $\mathcal{B}_{\mathbb{R}}$ which is concentrated on $(0,1]$. The measure, m is called the **uniform distribution** or **Lebesgue measure** on $(0,1]$.

With this summary in hand, let us now start the formal development. We begin with uniqueness statement in Theorem 8.2.

8.2 $\pi - \lambda$ Theorem

Recall that a collection, $\mathcal{P} \subset 2^\Omega$, is a π – **class** or π – system if it is closed under finite intersections. We also need the notion of a λ –system.

Definition 8.11 (λ – system). *A collection of sets, $\mathcal{L} \subset 2^\Omega$, is λ –* ***class*** *or λ –* ***system*** *if*

a. $\Omega \in \mathcal{L}$

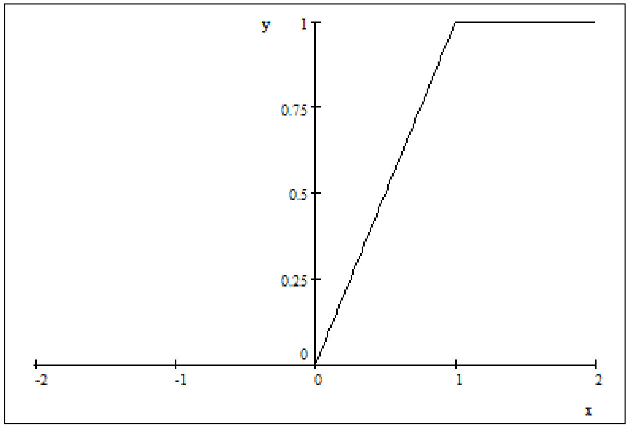


Fig. 8.1. The cumulative distribution function for the uniform distribution.

b. *If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A \in \mathcal{L}$. (Closed under proper differences.)*
c. *If $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$. (Closed under countable increasing unions.)*

Remark 8.12. If $\mathcal{L}$ is a collection of subsets of Ω which is both a λ – class and a π – system then $\mathcal{L}$ is a σ – algebra. Indeed, since $A^c = \Omega \setminus A$, we see that any λ - system is closed under complementation. If $\mathcal{L}$ is also a π – system, it is closed under intersections and therefore $\mathcal{L}$ is an algebra. Since $\mathcal{L}$ is also closed under increasing unions, $\mathcal{L}$ is a σ – algebra.

Lemma 8.13 (Alternate Axioms for a λ – System*). *Suppose that $\mathcal{L} \subset 2^\Omega$ is a collection of subsets Ω. Then $\mathcal{L}$ is a λ – class iff λ satisfies the following postulates:*

1. *$\Omega \in \mathcal{L}$*
2. *$A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$. (Closed under complementation.)*
3. *If $\{A_n\}_{n=1}^\infty \subset \mathcal{L}$ are disjoint, then $\sum_{n=1}^\infty A_n \in \mathcal{L}$. (Closed under disjoint unions.)*

Proof. Suppose that $\mathcal{L}$ satisfies a. – c. above. Clearly then postulates 1. and 2. hold. Suppose that $A, B \in \mathcal{L}$ such that $A \cap B = \emptyset$, then $A \subset B^c$ and

$$A^c \cap B^c = B^c \setminus A \in \mathcal{L}.$$

Taking complements of this result shows $A \cup B \in \mathcal{L}$ as well. So by induction, $B_m := \sum_{n=1}^m A_n \in \mathcal{L}$. Since $B_m \uparrow \sum_{n=1}^\infty A_n$ it follows from postulate c. that $\sum_{n=1}^\infty A_n \in \mathcal{L}$.

Now suppose that $\mathcal{L}$ satisfies postulates 1. – 3. above. Notice that $\emptyset \in \mathcal{L}$ and by postulate 3., $\mathcal{L}$ is closed under finite disjoint unions. Therefore if $A, B \in \mathcal{L}$ with $A \subset B$, then $B^c \in \mathcal{L}$ and $A \cap B^c = \emptyset$ allows us to conclude that $A \cup B^c \in \mathcal{L}$. Taking complements of this result shows $B \setminus A = A^c \cap B \in \mathcal{L}$ as well, i.e. postulate b. holds. If $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $B_n := A_n \setminus A_{n-1} \in \mathcal{L}$ for all n, where by convention $A_0 = \emptyset$. Hence it follows by postulate 3 that $\cup_{n=1}^\infty A_n = \sum_{n=1}^\infty B_n \in \mathcal{L}$. ∎

Theorem 8.14 (Dynkin's $\pi - \lambda$ Theorem). *If $\mathcal{L}$ is a λ class which contains a π – class, $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. We start by proving the following assertion; for any element $C \in \mathcal{L}$, the collection of sets,

$$\mathcal{L}^C := \{D \in \mathcal{L} : C \cap D \in \mathcal{L}\},$$

is a λ – system. To prove this claim, observe that: a. $\Omega \in \mathcal{L}^C$, b. if $A \subset B$ with $A, B \in \mathcal{L}^C$, then $A \cap C$, $B \cap C \in \mathcal{L}$ with $A \cap C \subset B \cap C$ and therefore,

$$(B \setminus A) \cap C = [B \cap C] \setminus A = [B \cap C] \setminus [A \cap C] \in \mathcal{L}.$$

This shows that $\mathcal{L}^C$ is closed under proper differences. c. If $A_n \in \mathcal{L}^C$ with $A_n \uparrow A$, then $A_n \cap C \in \mathcal{L}$ and $A_n \cap C \uparrow A \cap C \in \mathcal{L}$, i.e. $A \in \mathcal{L}^C$. Hence we have verified $\mathcal{L}^C$ is still a λ – system.

For the rest of the proof, we may assume without loss of generality that $\mathcal{L}$ is the smallest λ – class containing $\mathcal{P}$ – if not just replace $\mathcal{L}$ by the intersection of all λ – classes containing $\mathcal{P}$. Then for $C \in \mathcal{P}$ we know that $\mathcal{L}^C \subset \mathcal{L}$ is a λ - class containing $\mathcal{P}$ and hence $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{P}$ was arbitrary, we have shown, $C \cap D \in \mathcal{L}$ for all $C \in \mathcal{P}$ and $D \in \mathcal{L}$. We may now conclude that if $C \in \mathcal{L}$, then $\mathcal{P} \subset \mathcal{L}^C \subset \mathcal{L}$ and hence again $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{L}$ is arbitrary, we have shown $C \cap D \in \mathcal{L}$ for all $C, D \in \mathcal{L}$, i.e. $\mathcal{L}$ is a π – system. So by Remark 8.12, $\mathcal{L}$ is a σ algebra. Since $\sigma(\mathcal{P})$ is the smallest σ – algebra containing $\mathcal{P}$ it follows that $\sigma(\mathcal{P}) \subset \mathcal{L}$. ∎

As an immediate corollary, we have the following uniqueness result.

Proposition 8.15. *Suppose that $\mathcal{P} \subset 2^\Omega$ is a π – system. If P and Q are two probability*[1] *measures on $\sigma(\mathcal{P})$ such that $P = Q$ on $\mathcal{P}$, then $P = Q$ on $\sigma(\mathcal{P})$.*

Proof. Let $\mathcal{L} := \{A \in \sigma(\mathcal{P}) : P(A) = Q(A)\}$. One easily shows $\mathcal{L}$ is a λ – class which contains $\mathcal{P}$ by assumption. Indeed, $\Omega \in \mathcal{P} \subset \mathcal{L}$, if $A, B \in \mathcal{L}$ with $A \subset B$, then

$$P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$$

[1] More generally, P and Q could be two measures such that $P(\Omega) = Q(\Omega) < \infty$.

so that $B \setminus A \in \mathcal{L}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} Q(A_n) = Q(A)$ which shows $A \in \mathcal{L}$. Therefore $\sigma(\mathcal{P}) \subset \mathcal{L} = \sigma(\mathcal{P})$ and the proof is complete. ■

Example 8.16. Let $\Omega := \{a, b, c, d\}$ and let μ and ν be the probability measure on 2^Ω determined by, $\mu(\{x\}) = \frac{1}{4}$ for all $x \in \Omega$ and $\nu(\{a\}) = \nu(\{d\}) = \frac{1}{8}$ and $\nu(\{b\}) = \nu(\{c\}) = 3/8$. In this example,

$$\mathcal{L} := \left\{A \in 2^\Omega : P(A) = Q(A)\right\}$$

is λ – system which is not an algebra. Indeed, $A = \{a, b\}$ and $B = \{a, c\}$ are in $\mathcal{L}$ but $A \cap B \notin \mathcal{L}$.

Exercise 8.1. Suppose that μ and ν are two measures (not assumed to be finite) on a measure space, $(\Omega, \mathcal{B})$ such that $\mu = \nu$ on a π – system, $\mathcal{P}$. Further assume $\mathcal{B} = \sigma(\mathcal{P})$ and there exists $\Omega_n \in \mathcal{P}$ such that; i) $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ for all n and ii) $\Omega_n \uparrow \Omega$ as $n \uparrow \infty$. Show $\mu = \nu$ on $\mathcal{B}$.

Hint: Consider the measures, $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$.

Corollary 8.17. *A probability measure, P, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is uniquely determined by its cumulative distribution function,*

$$F(x) := P((-\infty, x]).$$

Proof. This follows from Proposition 8.15 wherein we use the fact that $\mathcal{P} := \{(-\infty, x] : x \in \mathbb{R}\}$ is a π – system such that $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{P})$. ■

Remark 8.18. Corollary 8.17 generalizes to $\mathbb{R}^n$. Namely a probability measure, P, on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is uniquely determined by its CDF,

$$F(x) := P((-\infty, x]) \text{ for all } x \in \mathbb{R}^n$$

where now

$$(-\infty, x] := (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n].$$

8.2.1 A Density Result*

Exercise 8.2 (Density of $\mathcal{A}$ in $\sigma(\mathcal{A})$). Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} := \sigma(\mathcal{A})$, and P is a probability measure on $\mathcal{B}$. Let $\rho(A, B) := P(A \Delta B)$. The goal of this exercise is to use the $\pi - \lambda$ theorem to show that $\mathcal{A}$ is dense in $\mathcal{B}$ relative to the "metric," ρ. More precisely you are to show using the following outline that for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that that $P(A \triangle B) < \varepsilon$.

1. Recall from Exercise 7.3 that $\rho(a, B) = P(A \triangle B) = \mathbb{E}|1_A - 1_B|$.
2. Observe; if $B = \cup B_i$ and $A = \cup_i A_i$, then
$$B \setminus A = \cup_i [B_i \setminus A] \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \triangle B_i \text{ and}$$
$$A \setminus B = \cup_i [A_i \setminus B] \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \triangle B_i$$
so that
$$A \triangle B \subset \cup_i (A_i \triangle B_i).$$
3. We also have
$$\begin{aligned}(B_2 \setminus B_1) \setminus (A_2 \setminus A_1) &= B_2 \cap B_1^c \cap (A_2 \setminus A_1)^c \\ &= B_2 \cap B_1^c \cap (A_2 \cap A_1^c)^c \\ &= B_2 \cap B_1^c \cap (A_2^c \cup A_1) \\ &= [B_2 \cap B_1^c \cap A_2^c] \cup [B_2 \cap B_1^c \cap A_1] \\ &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1)\end{aligned}$$
and similarly,
$$(A_2 \setminus A_1) \setminus (B_2 \setminus B_1) \subset (A_2 \setminus B_2) \cup (B_1 \setminus A_1)$$
so that
$$\begin{aligned}(A_2 \setminus A_1) \triangle (B_2 \setminus B_1) &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1) \cup (A_2 \setminus B_2) \cup (B_1 \setminus A_1) \\ &= (A_1 \triangle B_1) \cup (A_2 \triangle B_2).\end{aligned}$$
4. Observe that $A_n \in \mathcal{B}$ and $A_n \uparrow A$, then
$$\begin{aligned}P(B \triangle A_n) &= P(B \setminus A_n) + P(A_n \setminus B) \\ &\to P(B \setminus A) + P(A \setminus B) = P(A \triangle B).\end{aligned}$$
5. Let $\mathcal{L}$ be the collection of sets $B \in \mathcal{B}$ for which the assertion of the theorem holds. Show $\mathcal{L}$ is a λ – system which contains $\mathcal{A}$.

8.3 Construction of Measures

Definition 8.19. *Given a collection of subsets, $\mathcal{E}$, of Ω, let $\mathcal{E}_\sigma$ denote the collection of subsets of Ω which are finite or countable unions of sets from $\mathcal{E}$. Similarly let $\mathcal{E}_\delta$ denote the collection of subsets of Ω which are finite or countable intersections of sets from $\mathcal{E}$. We also write $\mathcal{E}_{\sigma\delta} = (\mathcal{E}_\sigma)_\delta$ and $\mathcal{E}_{\delta\sigma} = (\mathcal{E}_\delta)_\sigma$, etc.*

Lemma 8.20. *Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra. Then:*

1. $\mathcal{A}_\sigma$ *is closed under taking countable unions and finite intersections.*
2. $\mathcal{A}_\delta$ *is closed under taking countable intersections and finite unions.*
3. $\{A^c : A \in \mathcal{A}_\sigma\} = \mathcal{A}_\delta$ *and* $\{A^c : A \in \mathcal{A}_\delta\} = \mathcal{A}_\sigma$.

Proof. By construction $\mathcal{A}_\sigma$ is closed under countable unions. Moreover if $A = \cup_{i=1}^\infty A_i$ and $B = \cup_{j=1}^\infty B_j$ with $A_i, B_j \in \mathcal{A}$, then

$$A \cap B = \cup_{i,j=1}^\infty A_i \cap B_j \in \mathcal{A}_\sigma,$$

which shows that $\mathcal{A}_\sigma$ is also closed under finite intersections. Item 3. is straight forward and item 2. follows from items 1. and 3. ■

Remark 8.21. Let us recall from Proposition 8.3 and Remark 8.4 that a finitely additive measure $\mu : \mathcal{A} \to [0,\infty]$ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \uparrow \mu(A)$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Furthermore if $\mu(\Omega) < \infty$, then μ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \downarrow 0$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \downarrow \emptyset$.

Proposition 8.22. *Given a premeasure,* $\mu : \mathcal{A} \to [0,\infty]$, *we extend* μ *to* $\mathcal{A}_\sigma$ *by defining*

$$\mu(B) := \sup\{\mu(A) : \mathcal{A} \ni A \subset B\}. \tag{8.2}$$

This function $\mu : \mathcal{A}_\sigma \to [0,\infty]$ *then satisfies;*

1. ***(Monotonicity)*** *If* $A, B \in \mathcal{A}_\sigma$ *with* $A \subset B$ *then* $\mu(A) \le \mu(B)$.
2. ***(Continuity)*** *If* $A_n \in \mathcal{A}$ *and* $A_n \uparrow A \in \mathcal{A}_\sigma$, *then* $\mu(A_n) \uparrow \mu(A)$ *as* $n \to \infty$.
3. ***(Strong Additivity)*** *If* $A, B \in \mathcal{A}_\sigma$, *then*

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{8.3}$$

4. ***(Sub-Additivity on*** $\mathcal{A}_\sigma$***)*** *The function* μ *is sub-additive on* $\mathcal{A}_\sigma$, *i.e. if* $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$, *then*

$$\mu(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{8.4}$$

5. *(σ - Additivity* ***on*** $\mathcal{A}_\sigma$*)* *The function* μ *is countably additive on* $\mathcal{A}_\sigma$.

Proof. 1. and 2. Monotonicity follows directly from Eq. (8.2) which then implies $\mu(A_n) \le \mu(B)$ for all n. Therefore $M := \lim_{n\to\infty} \mu(A_n) \le \mu(B)$. To prove the reverse inequality, let $\mathcal{A} \ni A \subset B$. Then by the continuity of μ on $\mathcal{A}$ and the fact that $A_n \cap A \uparrow A$ we have $\mu(A_n \cap A) \uparrow \mu(A)$. As $\mu(A_n) \ge \mu(A_n \cap A)$ for all n it then follows that $M := \lim_{n\to\infty} \mu(A_n) \ge \mu(A)$. As $A \in \mathcal{A}$ with $A \subset B$ was arbitrary we may conclude,

$$\mu(B) = \sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \le M.$$

3. Suppose that $A, B \in \mathcal{A}_\sigma$ and $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $\mathcal{A}$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Then passing to the limit as $n \to \infty$ in the identity,

$$\mu(A_n \cup B_n) + \mu(A_n \cap B_n) = \mu(A_n) + \mu(B_n)$$

proves Eq. (8.3). In particular, it follows that μ is finitely additive on $\mathcal{A}_\sigma$.

4 and 5. Let $\{A_n\}_{n=1}^\infty$ be any sequence in $\mathcal{A}_\sigma$ and choose $\{A_{n,i}\}_{i=1}^\infty \subset \mathcal{A}$ such that $A_{n,i} \uparrow A_n$ as $i \to \infty$. Then we have,

$$\mu\left(\cup_{n=1}^N A_{n,N}\right) \le \sum_{n=1}^N \mu(A_{n,N}) \le \sum_{n=1}^N \mu(A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{8.5}$$

Since $\mathcal{A} \ni \cup_{n=1}^N A_{n,N} \uparrow \cup_{n=1}^\infty A_n \in \mathcal{A}_\sigma$, we may let $N \to \infty$ in Eq. (8.5) to conclude Eq. (8.4) holds. If we further assume that $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$ are pairwise disjoint, by the finite additivity and monotonicity of μ on $\mathcal{A}_\sigma$, we have

$$\sum_{n=1}^\infty \mu(A_n) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n) = \lim_{N\to\infty} \mu\left(\cup_{n=1}^N A_n\right) \le \mu(\cup_{n=1}^\infty A_n).$$

This inequality along with Eq. (8.4) shows that μ is $\sigma -$ additive on $\mathcal{A}_\sigma$. ■

Suppose μ is a **finite** premeasure on an algebra, $\mathcal{A} \subset 2^\Omega$, and $A \in \mathcal{A}_\delta \cap \mathcal{A}_\sigma$. Since $A, A^c \in \mathcal{A}_\sigma$ and $\Omega = A \cup A^c$, it follows that $\mu(\Omega) = \mu(A) + \mu(A^c)$. From this observation we may extend μ to a function on $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ by defining

$$\mu(A) := \mu(\Omega) - \mu(A^c) \text{ for all } A \in \mathcal{A}_\delta. \tag{8.6}$$

Lemma 8.23. *Suppose* μ *is a finite premeasure on an algebra,* $\mathcal{A} \subset 2^\Omega$, *and* μ *has been extended to* $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ *as described in Proposition 8.22 and Eq. (8.6) above.*

1. *If* $A \in \mathcal{A}_\delta$ *then* $\mu(A) = \inf\{\mu(B) : A \subset B \in \mathcal{A}\}$.
2. *If* $A \in \mathcal{A}_\delta$ *and* $A_n \in \mathcal{A}$ *such that* $A_n \downarrow A$, *then* $\mu(A) =\downarrow \lim_{n\to\infty} \mu(A_n)$.
3. μ *is strongly additive when restricted to* $\mathcal{A}_\delta$.
4. *If* $A \in \mathcal{A}_\delta$ *and* $C \in \mathcal{A}_\sigma$ *such that* $A \subset C$, *then* $\mu(C \setminus A) = \mu(C) - \mu(A)$.

Proof.

1. Since $\mu(B) = \mu(\Omega) - \mu(B^c)$ and $A \subset B$ iff $B^c \subset A^c$, it follows that

$$\begin{aligned} \inf\{\mu(B) : A \subset B \in \mathcal{A}\} &= \inf\{\mu(\Omega) - \mu(B^c) : \mathcal{A} \ni B^c \subset A^c\} \\ &= \mu(\Omega) - \sup\{\mu(B) : \mathcal{A} \ni B \subset A^c\} \\ &= \mu(\Omega) - \mu(A^c) = \mu(A). \end{aligned}$$

2. Similarly, since $A_n^c \uparrow A^c \in \mathcal{A}_\sigma$, by the definition of $\mu(A)$ and Proposition 8.22 it follows that

$$\begin{aligned}\mu(A) &= \mu(\Omega) - \mu(A^c) = \mu(\Omega) - \uparrow \lim_{n\to\infty} \mu(A_n^c) \\ &= \downarrow \lim_{n\to\infty} [\mu(\Omega) - \mu(A_n^c)] = \downarrow \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

3. Suppose $A, B \in \mathcal{A}_\delta$ and $A_n, B_n \in \mathcal{A}$ such that $A_n \downarrow A$ and $B_n \downarrow B$, then $A_n \cup B_n \downarrow A \cup B$ and $A_n \cap B_n \downarrow A \cap B$ and therefore,

$$\begin{aligned}\mu(A \cup B) + \mu(A \cap B) &= \lim_{n\to\infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)] \\ &= \lim_{n\to\infty} [\mu(A_n) + \mu(B_n)] = \mu(A) + \mu(B).\end{aligned}$$

All we really need is the finite additivity of μ which can be proved as follows. Suppose that $A, B \in \mathcal{A}_\delta$ are disjoint, then $A \cap B = \emptyset$ implies $A^c \cup B^c = \Omega$. So by the strong additivity of μ on $\mathcal{A}_\sigma$ it follows that

$$\mu(\Omega) + \mu(A^c \cap B^c) = \mu(A^c) + \mu(B^c)$$

from which it follows that

$$\begin{aligned}\mu(A \cup B) &= \mu(\Omega) - \mu(A^c \cap B^c) \\ &= \mu(\Omega) - [\mu(A^c) + \mu(B^c) - \mu(\Omega)] \\ &= \mu(A) + \mu(B).\end{aligned}$$

4. Since $A^c, C \in \mathcal{A}_\sigma$ we may use the strong additivity of μ on $\mathcal{A}_\sigma$ to conclude,

$$\mu(A^c \cup C) + \mu(A^c \cap C) = \mu(A^c) + \mu(C).$$

Because $\Omega = A^c \cup C$, and $\mu(A^c) = \mu(\Omega) - \mu(A)$, the above equation may be written as

$$\mu(\Omega) + \mu(C \setminus A) = \mu(\Omega) - \mu(A) + \mu(C)$$

which finishes the proof.

■

Notation 8.24 (Inner and outer measures) *Let $\mu : \mathcal{A} \to [0, \infty)$ be a finite premeasure extended to $\mathcal{A}_\sigma \cup \mathcal{A}_\delta$ as above. The for **any** $B \subset \Omega$ let*

$$\begin{aligned}\mu_*(B) &:= \sup\{\mu(A) : \mathcal{A}_\delta \ni A \subset B\} \text{ and} \\ \mu^*(B) &:= \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\}.\end{aligned}$$

*We refer to $\mu_*(B)$ and $\mu^*(B)$ as the **inner and outer** content of B respectively.*

If $B \subset \Omega$ has the same inner and outer content it is reasonable to define the measure of B as this common value. As we will see in Theorem 8.27 below, this extension becomes a σ – additive measure on a σ – algebra of subsets of Ω.

Definition 8.25 (Measurable Sets). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^\Omega$. We say that $B \subset \Omega$ is **measurable** if $\mu_*(B) = \mu^*(B)$. We will denote the collection of measurable subsets of Ω by $\mathcal{B} = \mathcal{B}(\mu)$ and define $\bar{\mu} : \mathcal{B} \to [0, \mu(\Omega)]$ by*

$$\bar{\mu}(B) := \mu_*(B) = \mu^*(B) \text{ for all } B \in \mathcal{B}. \tag{8.7}$$

Remark 8.26. Observe that $\mu_*(B) = \mu^*(B)$ iff for all $\varepsilon > 0$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu(C) - \mu(A) < \varepsilon, \tag{8.8}$$

wherein we have used Lemma 8.23 for the first equality. Moreover we will use below that if $B \in \mathcal{B}$ and $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$, then

$$\mu(A) \le \mu_*(B) = \bar{\mu}(B) = \mu^*(B) \le \mu(C). \tag{8.9}$$

Theorem 8.27 (Finite Premeasure Extension Theorem). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^\Omega$ and $\bar{\mu} : \mathcal{B} := \mathcal{B}(\mu) \to [0, \mu(\Omega)]$ be as in Definition 8.25. Then $\mathcal{B}$ is a σ – algebra on Ω which contains $\mathcal{A}$ and $\bar{\mu}$ is a σ – additive measure on $\mathcal{B}$. Moreover, $\bar{\mu}$ is the unique measure on $\mathcal{B}$ such that $\bar{\mu}|_{\mathcal{A}} = \mu$.*

Proof. 1. **$\mathcal{B}$ is an algebra.** It is clear that $\mathcal{A} \subset \mathcal{B}$ and that $\mathcal{B}$ is closed under complementation – see Eq. (8.8) and use the fact that $A^c \setminus C^c = C \setminus A$. Now suppose that $B_i \in \mathcal{B}$ for $i = 1, 2$ and $\varepsilon > 0$ is given. We may then choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon$ for $i = 1, 2$. Then with $A = A_1 \cup A_2$, $B = B_1 \cup B_2$ and $C = C_1 \cup C_2$, we have $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$. Since

$$C \setminus A = (C_1 \setminus A) \cup (C_2 \setminus A) \subset (C_1 \setminus A_1) \cup (C_2 \setminus A_2),$$

it follows from the sub-additivity of μ that with

$$\mu(C \setminus A) \le \mu(C_1 \setminus A_1) + \mu(C_2 \setminus A_2) < 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have shown that $B \in \mathcal{B}$ which completes the proof that $\mathcal{B}$ is an algebra.

2. **$\mathcal{B}$ is a σ–algebra.** As we know $\mathcal{B}$ is an algebra, to show $\mathcal{B}$ is a σ – algebra it suffices to show that $B = \sum_{n=1}^\infty B_n \in \mathcal{B}$ whenever $\{B_n\}_{n=1}^\infty$ is a disjoint sequence in $\mathcal{B}$. To this end, let $\varepsilon > 0$ be given and choose $A_i \subset B_i \subset C_i$ such

that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon 2^{-i}$ for all i. Let $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$ and for $n \in \mathbb{N}$ let $A^n := \sum_{i=1}^n A_i \in \mathcal{A}_\delta$. Since the $\{A_i\}_{i=1}^\infty$ are pairwise disjoint we may use Lemma 8.23 to show,

$$\begin{aligned}\sum_{i=1}^n \mu(C_i) &= \sum_{i=1}^n (\mu(A_i) + \mu(C_i \setminus A_i)) \\ &= \mu(A^n) + \sum_{i=1}^n \mu(C_i \setminus A_i) \le \mu(\Omega) + \sum_{i=1}^n \varepsilon 2^{-i}\end{aligned}$$

which on letting $n \to \infty$ implies

$$\sum_{i=1}^\infty \mu(C_i) \le \mu(\Omega) + \varepsilon < \infty. \tag{8.10}$$

Using

$$C \setminus A^n = \cup_{i=1}^\infty (C_i \setminus A^n) \subset [\cup_{i=1}^n (C_i \setminus A_i)] \cup \left[\cup_{i=n+1}^\infty C_i\right] \in \mathcal{A}_\sigma,$$

and the sub-additivity of μ on $\mathcal{A}_\sigma$ it follows that

$$\begin{aligned}\mu(C \setminus A^n) &\le \sum_{i=1}^n \mu(C_i \setminus A_i) + \sum_{i=n+1}^\infty \mu(C_i) \le \varepsilon \sum_{i=1}^n 2^{-i} + \sum_{i=n+1}^\infty \mu(C_i) \\ &\le \varepsilon + \sum_{i=n+1}^\infty \mu(C_i) \to \varepsilon \text{ as } n \to \infty,\end{aligned}$$

wherein we have used Eq. (8.10) in computing the limit. In summary, $B = \cup_{i=1}^\infty B_i$, $\mathcal{A}_\delta \ni A^n \subset B \subset C \in \mathcal{A}_\sigma$, $C \setminus A^n \in \mathcal{A}_\sigma$ with $\mu(C \backslash A^n) \le 2\varepsilon$ for all n sufficiently large. Since $\varepsilon > 0$ is arbitrary, it follows that $B \in \mathcal{B}$.

3. $\bar{\mu}$ is a measure. Continuing the notation in step 2, we have

$$\sum_{i=1}^\infty \mu(A_i) \overset{n\to\infty}{\longleftarrow} \sum_{i=1}^n \mu(A_i) = \mu(A^n) \le \bar{\mu}(B) \le \mu(C) \le \sum_{i=1}^\infty \mu(C_i). \tag{8.11}$$

On the other hand, since $A_i \subset B_i \subset C_i$, it follows (see Eq. (8.9)) that $\mu(A_i) \le \bar{\mu}(B_i) \le \mu(C_i)$ and therefore that

$$\sum_{i=1}^\infty \mu(A_i) \le \sum_{i=1}^\infty \bar{\mu}(B_i) \le \sum_{i=1}^\infty \mu(C_i). \tag{8.12}$$

Equations (8.11) and (8.12) show that $\bar{\mu}(B)$ and $\sum_{i=1}^\infty \bar{\mu}(B_i)$ are both between $\sum_{i=1}^\infty \mu(A_i)$ and $\sum_{i=1}^\infty \mu(C_i)$ and so

$$\left|\bar{\mu}(B) - \sum_{i=1}^\infty \bar{\mu}(B_i)\right| \le \sum_{i=1}^\infty \mu(C_i) - \sum_{i=1}^\infty \mu(A_i) = \sum_{i=1}^\infty \mu(C_i \setminus A_i) \le \sum_{i=1}^\infty \varepsilon 2^{-i} = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have shown $\bar{\mu}(B) = \sum_{i=1}^\infty \bar{\mu}(B_i)$, i.e. $\bar{\mu}$ is a measure on $\mathcal{B}$.

Since we really had no choice as to how to extend μ, it is to be expected that the extension is unique. You are asked to supply the details in Exercise 8.3 below. ■

Exercise 8.3. Let μ, $\bar{\mu}$, $\mathcal{A}$, and $\mathcal{B} := \mathcal{B}(\mu)$ be as in Theorem 8.27. Further suppose that $\mathcal{B}_0 \subset 2^\Omega$ is a σ – algebra such that $\mathcal{A} \subset \mathcal{B}_0 \subset \mathcal{B}$ and $\nu : \mathcal{B}_0 \to [0, \mu(\Omega)]$ is a σ – additive measure on $\mathcal{B}_0$ such that $\nu = \mu$ on $\mathcal{A}$. Show that $\nu = \bar{\mu}$ on $\mathcal{B}_0$ as well. (When $\mathcal{B}_0 = \sigma(\mathcal{A})$ this exercise is of course a consequence of Proposition 8.15. It is not necessary to use this information to complete the exercise.)

Corollary 8.28. *Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mu : \mathcal{B}_0 := \sigma(\mathcal{A}) \to [0, \mu(\Omega)]$ is a σ – additive finite measure. Then for every $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$;*

1. *there exists $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$ and $\varepsilon > 0$ such that $\mu(C \setminus A) < \varepsilon$ and*
2. *there exists $A \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$.*

Exercise 8.4. Prove corollary 8.28 by considering $\bar{\nu}$ where $\nu := \mu|_{\mathcal{A}}$. **Hint:** you may find Exercise 7.3 useful here.

Theorem 8.29 (σ - Finite Premeasure Extension Theorem). *Suppose that μ is a σ – finite premeasure on an algebra $\mathcal{A}$. Then*

$$\bar{\mu}(B) := \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\} \ \forall\ B \in \sigma(\mathcal{A}) \tag{8.13}$$

defines a measure on $\sigma(\mathcal{A})$ and this measure is the unique extension of μ on $\mathcal{A}$ to a measure on $\sigma(\mathcal{A})$. Recall that

$$\mu(C) = \sup\{\mu(A) : \mathcal{A} \ni A \subset C\} \text{ for all } C \in \mathcal{A}_\sigma.$$

Proof. Let $\{\Omega_n\}_{n=1}^\infty \subset \mathcal{A}$ be chosen so that $\mu(\Omega_n) < \infty$ for all n and $\Omega_n \uparrow \Omega$ as $n \to \infty$ and let

$$\mu_n(A) := \mu_n(A \cap \Omega_n) \text{ for all } A \in \mathcal{A}.$$

Each μ_n is a premeasure (as is easily verified) on $\mathcal{A}$ and hence by Theorem 8.27 each μ_n has an extension, $\bar{\mu}_n$, to a measure on $\sigma(\mathcal{A})$. Since the measure $\bar{\mu}_n$ are increasing, $\bar{\mu} := \lim_{n\to\infty} \bar{\mu}_n$ is a measure which extends μ.

The proof will be completed by verifying that Eq. (8.13) holds. Let $B \in \sigma(\mathcal{A})$, $B_m = \Omega_m \cap B$ and $\varepsilon > 0$ be given. By Theorem 8.27, there exists

$C_m \in \mathcal{A}_\sigma$ such that $B_m \subset C_m \subset \Omega_m$ and $\bar{\mu}(C_m \setminus B_m) = \bar{\mu}_m(C_m \setminus B_m) < \varepsilon 2^{-n}$. Then $C := \cup_{m=1}^\infty C_m \in \mathcal{A}_\sigma$ and

$$\bar{\mu}(C \setminus B) \le \bar{\mu}\left(\bigcup_{m=1}^\infty (C_m \setminus B)\right) \le \sum_{m=1}^\infty \bar{\mu}(C_m \setminus B) \le \sum_{m=1}^\infty \bar{\mu}(C_m \setminus B_m) < \varepsilon.$$

Thus

$$\bar{\mu}(B) \le \bar{\mu}(C) = \bar{\mu}(B) + \bar{\mu}(C \setminus B) \le \bar{\mu}(B) + \varepsilon$$

which, since $\varepsilon > 0$ is arbitrary, shows $\bar{\mu}$ satisfies Eq. (8.13). The uniqueness of the extension $\bar{\mu}$ is the subject of Exercise 8.11. ■

The following slight reformulation of Theorem 8.29 can be useful.

Corollary 8.30. *Let $\mathcal{A}$ be an algebra of sets, $\{\Omega_m\}_{m=1}^\infty \subset \mathcal{A}$ is a given sequence of sets such that $\Omega_m \uparrow \Omega$ as $m \to \infty$. Let*

$$\mathcal{A}_f := \{A \in \mathcal{A} : A \subset \Omega_m \text{ for some } m \in \mathbb{N}\}.$$

Notice that $\mathcal{A}_f$ is a ring, i.e. closed under differences, intersections and unions and contains the empty set. Further suppose that $\mu : \mathcal{A}_f \to [0,\infty)$ is an additive set function such that $\mu(A_n) \downarrow 0$ for any sequence, $\{A_n\} \subset \mathcal{A}_f$ such that $A_n \downarrow \emptyset$ as $n \to \infty$. Then μ extends uniquely to a σ – finite measure on $\mathcal{A}$.

Proof. Existence. By assumption, $\mu_m := \mu|_{\mathcal{A}_{\Omega_m}} : \mathcal{A}_{\Omega_m} \to [0,\infty)$ is a premeasure on $(\Omega_m, \mathcal{A}_{\Omega_m})$ and hence by Theorem 8.29 extends to a measure μ_m' on $(\Omega_m, \sigma(\mathcal{A}_{\Omega_m}) = \mathcal{B}_{\Omega_m})$. Let $\bar{\mu}_m(B) := \mu_m'(B \cap \Omega_m)$ for all $B \in \mathcal{B}$. Then $\{\bar{\mu}_m\}_{m=1}^\infty$ is an increasing sequence of measure on $(\Omega, \mathcal{B})$ and hence $\bar{\mu} := \lim_{m\to\infty} \bar{\mu}_m$ defines a measure on $(\Omega, \mathcal{B})$ such that $\bar{\mu}|_{\mathcal{A}_f} = \mu$.

Uniqueness. If μ_1 and μ_2 are two such extensions, then $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{A}$ and therefore by Proposition 8.15 or Exercise 8.11 we know that $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{B}$. We may now let $m \to \infty$ to see that in fact $\mu_1(B) = \mu_2(B)$ for all $B \in \mathcal{B}$, i.e. $\mu_1 = \mu_2$. ■

8.4 Radon Measures on $\mathbb{R}$

We say that a measure, μ, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is a **Radon measure** if $\mu([a,b]) < \infty$ for all $-\infty < a < b < \infty$. In this section we will give a characterization of all Radon measures on $\mathbb{R}$. We first need the following general result characterizing premeasures on an algebra generated by a semi-algebra.

Proposition 8.31. *Suppose that $\mathcal{S} \subset 2^\Omega$ is a semi-algebra, $\mathcal{A} = \mathcal{A}(\mathcal{S})$ and $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure. Then μ is a premeasure on $\mathcal{A}$ iff μ is countably sub-additive on $\mathcal{S}$.*

Proof. Clearly if μ is a premeasure on $\mathcal{A}$ then μ is σ - additive and hence sub-additive on $\mathcal{S}$. Because of Proposition 7.2, to prove the converse it suffices to show that the sub-additivity of μ on $\mathcal{S}$ implies the sub-additivity of μ on $\mathcal{A}$.

So suppose $A = \sum_{n=1}^\infty A_n \in \mathcal{A}$ with each $A_n \in \mathcal{A}$. By Proposition 6.25 we may write $A = \sum_{j=1}^k E_j$ and $A_n = \sum_{i=1}^{N_n} E_{n,i}$ with $E_j, E_{n,i} \in \mathcal{S}$. Intersecting the identity, $A = \sum_{n=1}^\infty A_n$, with E_j implies

$$E_j = A \cap E_j = \sum_{n=1}^\infty A_n \cap E_j = \sum_{n=1}^\infty \sum_{i=1}^{N_n} E_{n,i} \cap E_j.$$

By the assumed sub-additivity of μ on $\mathcal{S}$,

$$\mu(E_j) \le \sum_{n=1}^\infty \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j).$$

Summing this equation on j and using the finite additivity of μ shows

$$\begin{aligned}\mu(A) = \sum_{j=1}^k \mu(E_j) &\le \sum_{j=1}^k \sum_{n=1}^\infty \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j) \\ &= \sum_{n=1}^\infty \sum_{i=1}^{N_n} \sum_{j=1}^k \mu(E_{n,i} \cap E_j) = \sum_{n=1}^\infty \sum_{i=1}^{N_n} \mu(E_{n,i}) = \sum_{n=1}^\infty \mu(A_n).\end{aligned}$$

■

Suppose now that μ is a Radon measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that

$$\mu((a,b]) = F(b) - F(a) \text{ for all } -\infty < a \le b < \infty. \tag{8.14}$$

For example if $\mu(\mathbb{R}) < \infty$ we can take $F(x) = \mu((-\infty, x])$ while if $\mu(\mathbb{R}) = \infty$ we might take

$$F(x) = \begin{cases} \mu((0,x]) & \text{if } x \ge 0 \\ -\mu((x,0]) & \text{if } x \le 0 \end{cases}.$$

The function F is uniquely determined modulo translation by a constant.

Lemma 8.32. *If μ is a Radon measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that $\mu((a,b]) = F(b) - F(a)$, then F is increasing and right continuous.*

Proof. The function F is increasing by the monotonicity of μ. To see that F is right continuous, let $b \in \mathbb{R}$ and choose $a \in (-\infty, b)$ and any sequence $\{b_n\}_{n=1}^\infty \subset (b, \infty)$ such that $b_n \downarrow b$ as $n \to \infty$. Since $\mu((a, b_1]) < \infty$ and $(a, b_n] \downarrow (a, b]$ as $n \to \infty$, it follows that

$$F(b_n) - F(a) = \mu((a, b_n]) \downarrow \mu((a,b]) = F(b) - F(a).$$

Since $\{b_n\}_{n=1}^\infty$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $\lim_{y\downarrow b} F(y) = F(b)$. ■

The key result of this section is the converse to this lemma.

Theorem 8.33. *Suppose $F : \mathbb{R} \to \mathbb{R}$ is a right continuous increasing function. Then there exists a unique Radon measure, $\mu = \mu_F$, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that Eq. (8.14) holds.*

Proof. Let $\mathcal{S} := \{(a,b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$, and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ consists of those sets, $A \subset \mathbb{R}$ which may be written as finite disjoint unions of sets from $\mathcal{S}$ as in Example 6.26. Recall that $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{A}) = \sigma(\mathcal{S})$. Further define $F(\pm\infty) := \lim_{x\to\pm\infty} F(x)$ and let $\mu = \mu_F$ be the finitely additive measure on $(\mathbb{R}, \mathcal{A})$ described in Proposition 7.8 and Remark 7.9. To finish the proof it suffices by Theorem 8.29 to show that μ is a premeasure on $\mathcal{A} = \mathcal{A}(\mathcal{S})$ where $\mathcal{S} := \{(a,b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$. So in light of Proposition 8.31, to finish the proof it suffices to show μ is sub-additive on $\mathcal{S}$, i.e. we must show

$$\mu(J) \le \sum_{n=1}^\infty \mu(J_n). \tag{8.15}$$

where $J = \sum_{n=1}^\infty J_n$ with $J = (a,b] \cap \mathbb{R}$ and $J_n = (a_n, b_n] \cap \mathbb{R}$. Recall from Proposition 7.2 that the finite additivity of μ implies

$$\sum_{n=1}^\infty \mu(J_n) \le \mu(J). \tag{8.16}$$

We begin with the special case where $-\infty < a < b < \infty$. Our proof will be by "continuous induction." The strategy is to show $a \in \Lambda$ where

$$\Lambda := \left\{\alpha \in [a,b] : \mu(J \cap (\alpha, b]) \le \sum_{n=1}^\infty \mu(J_n \cap (\alpha, b])\right\}. \tag{8.17}$$

As $b \in J$, there exists an k such that $b \in J_k$ and hence $(a_k, b_k] = (a_k, b]$ for this k. It now easily follows that $J_k \subset \Lambda$ so that Λ is not empty. To finish the proof we are going to show $\bar{a} := \inf \Lambda \in \Lambda$ and that $\bar{a} = a$.

- Let $\alpha_m \in \Lambda$ such that $\alpha_m \downarrow \bar{a}$, i.e.

$$\mu(J \cap (\alpha_m, b]) \le \sum_{n=1}^\infty \mu(J_n \cap (\alpha_m, b]). \tag{8.18}$$

The right continuity of F implies $\alpha \to \mu(J_n \cap (\alpha, b])$ is right continuous. So by the dominated convergence theorem[2] for sums,

[2] DCT applies as $\mu(J_n \cap (\alpha_m, b]) \le \mu(J_n)$ and $\sum_{n=1}^\infty \mu(J_n) \le \mu(J) < \infty$ by Eq. (8.18).

$$\begin{aligned}
\mu(J \cap (\bar{a}, b]) &= \lim_{m\to\infty} \mu(J \cap (\alpha_m, b]) \\
&\le \lim_{m\to\infty} \sum_{n=1}^\infty \mu(J_n \cap (\alpha_m, b]) \\
&= \sum_{n=1}^\infty \lim_{m\to\infty} \mu(J_n \cap (\alpha_m, b]) = \sum_{n=1}^\infty \mu(J_n \cap (\bar{a}, b]),
\end{aligned}$$

i.e. $\bar{a} \in \Lambda$.

- If $\bar{a} > a$, then $\bar{a} \in J_l = (a_l, b_l]$ for some l. Letting $\alpha = a_l < \bar{a}$, we have,

$$\begin{aligned}
\mu(J \cap (\alpha, b]) &= \mu(J \cap (\alpha, \bar{a}]) + \mu(J \cap (\bar{a}, b]) \\
&\le \mu(J_l \cap (\alpha, \bar{a}]) + \sum_{n=1}^\infty \mu(J_n \cap (\bar{a}, b]) \\
&= \mu(J_l \cap (\alpha, \bar{a}]) + \mu(J_l \cap (\bar{a}, b]) + \sum_{n \ne l} \mu(J_n \cap (\bar{a}, b]) \\
&= \mu(J_l \cap (\alpha, b]) + \sum_{n\ne l} \mu(J_n \cap (\bar{a}, b]) \\
&\le \sum_{n=1}^\infty \mu(J_n \cap (\alpha, b]).
\end{aligned}$$

This shows $\alpha \in \Lambda$ and $\alpha < \bar{a}$ which violates the definition of $\bar{a}$. Thus we must conclude that $\bar{a} = a$.

The hard work is now done but we still have to check the cases where $a = -\infty$ or $b = \infty$. For example, suppose that $b = \infty$ so that

$$J = (a, \infty) = \sum_{n=1}^\infty J_n$$

with $J_n = (a_n, b_n] \cap \mathbb{R}$. Then

$$I_M := (a, M] = J \cap I_M = \sum_{n=1}^\infty J_n \cap I_M$$

and so by what we have already proved,

$$F(M) - F(a) = \mu(I_M) \le \sum_{n=1}^\infty \mu(J_n \cap I_M) \le \sum_{n=1}^\infty \mu(J_n).$$

Now let $M \to \infty$ in this last inequality to find that

$$\mu((a,\infty)) = F(\infty) - F(a) \le \sum_{n=1}^{\infty} \mu(J_n).$$

The other cases where $a = -\infty$ and $b \in \mathbb{R}$ and $a = -\infty$ and $b = \infty$ are handled similarly. ■

8.4.1 Lebesgue Measure

If $F(x) = x$ for all $x \in \mathbb{R}$, we denote μ_F by m and call m Lebesgue measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Theorem 8.34. *Lebesgue measure m is invariant under translations, i.e. for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$,*

$$m(x+B) = m(B). \tag{8.19}$$

Lebesgue measure, m, is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0,1]) = 1$ and Eq. (8.19) holds for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$. Moreover, m has the scaling property

$$m(\lambda B) = |\lambda|\, m(B) \tag{8.20}$$

where $\lambda \in \mathbb{R}$, $B \in \mathcal{B}_{\mathbb{R}}$ and $\lambda B := \{\lambda x : x \in B\}$.

Proof. Let $m_x(B) := m(x+B)$, then one easily shows that m_x is a measure on $\mathcal{B}_{\mathbb{R}}$ such that $m_x((a,b]) = b-a$ for all $a < b$. Therefore, $m_x = m$ by the uniqueness assertion in Exercise 8.11. For the converse, suppose that m is translation invariant and $m((0,1]) = 1$. Given $n \in \mathbb{N}$, we have

$$(0,1] = \cup_{k=1}^{n}(\frac{k-1}{n}, \frac{k}{n}] = \cup_{k=1}^{n}\left(\frac{k-1}{n} + (0, \frac{1}{n}]\right).$$

Therefore,

$$1 = m((0,1]) = \sum_{k=1}^{n} m\left(\frac{k-1}{n} + (0,\frac{1}{n}]\right) = \sum_{k=1}^{n} m((0,\frac{1}{n}]) = n \cdot m((0,\frac{1}{n}]).$$

That is to say

$$m((0,\frac{1}{n}]) = 1/n.$$

Similarly, $m((0,\frac{l}{n}]) = l/n$ for all $l, n \in \mathbb{N}$ and therefore by the translation invariance of m,

$$m((a,b]) = b-a \text{ for all } a,b \in \mathbb{Q} \text{ with } a < b.$$

Finally for $a, b \in \mathbb{R}$ such that $a < b$, choose $a_n, b_n \in \mathbb{Q}$ such that $b_n \downarrow b$ and $a_n \uparrow a$, then $(a_n, b_n] \downarrow (a,b]$ and thus

$$m((a,b]) = \lim_{n\to\infty} m((a_n,b_n]) = \lim_{n\to\infty}(b_n - a_n) = b-a,$$

i.e. m is Lebesgue measure. To prove Eq. (8.20) we may assume that $\lambda \neq 0$ since this case is trivial to prove. Now let $m_\lambda(B) := |\lambda|^{-1} m(\lambda B)$. It is easily checked that m_λ is again a measure on $\mathcal{B}_{\mathbb{R}}$ which satisfies

$$m_\lambda((a,b]) = \lambda^{-1} m((\lambda a, \lambda b]) = \lambda^{-1}(\lambda b - \lambda a) = b-a$$

if $\lambda > 0$ and

$$m_\lambda((a,b]) = |\lambda|^{-1} m([\lambda b, \lambda a)) = -|\lambda|^{-1}(\lambda b - \lambda a) = b - a$$

if $\lambda < 0$. Hence $m_\lambda = m$. ■

8.5 A Discrete Kolmogorov's Extension Theorem

For this section, let S be a finite or countable set (we refer to S as **state space**), $\Omega := S^\infty := S^{\mathbb{N}}$ (think of $\mathbb{N}$ as time and Ω as **path space**)

$$\mathcal{A}_n := \{B \times \Omega : B \subset S^n\} \text{ for all } n \in \mathbb{N},$$

$\mathcal{A} := \cup_{n=1}^{\infty}\mathcal{A}_n$, and $\mathcal{B} := \sigma(\mathcal{A})$. We call the elements, $A \subset \Omega$, the **cylinder subsets of** Ω. Notice that $A \subset \Omega$ is a cylinder set iff there exists $n \in \mathbb{N}$ and $B \subset S^n$ such that

$$A = B \times \Omega := \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}.$$

Also observe that we may write A as $A = B' \times \Omega$ where $B' = B \times S^k \subset S^{n+k}$ for any $k \geq 0$.

Exercise 8.5. Show;

1. $\mathcal{A}_n$ is a σ – algebra for each $n \in \mathbb{N}$,
2. $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ for all n, and
3. $\mathcal{A} \subset 2^\Omega$ is an algebra of subsets of Ω. (In fact, you might show that $\mathcal{A} = \cup_{n=1}^{\infty}\mathcal{A}_n$ is an algebra whenever $\{\mathcal{A}_n\}_{n=1}^{\infty}$ is an increasing sequence of algebras.)

Lemma 8.35 (Baby Tychonov Theorem). *Suppose $\{C_n\}_{n=1}^{\infty} \subset \mathcal{A}$ is a decreasing sequence of* ***non-empty*** *cylinder sets. Further assume there exists $N_n \in \mathbb{N}$ and $B_n \subset\subset S^{N_n}$ such that $C_n = B_n \times \Omega$. (This last assumption is vacuous when S is a finite set. Recall that we write $\Lambda \subset\subset A$ to indicate that Λ is a finite subset of A.) Then $\cap_{n=1}^{\infty} C_n \neq \emptyset$.*

Proof. Since $C_{n+1} \subset C_n$, if $N_n > N_{n+1}$, we would have $B_{n+1} \times S^{N_{n+1}-N_n} \subset B_n$. If S is an infinite set this would imply B_n is an infinite set and hence we must have $N_{n+1} \geq N_n$ for all n when $\#(S) = \infty$. On the other hand, if S is a finite set, we can always replace B_{n+1} by $B_{n+1} \times S^k$ for some appropriate k and arrange it so that $N_{n+1} \geq N_n$ for all n. So from now we assume that $N_{n+1} \geq N_n$.

Case 1. $\lim_{n\to\infty} N_n < \infty$ in which case there exists some $N \in \mathbb{N}$ such that $N_n = N$ for all large n. Thus for large N, $C_n = B_n \times \Omega$ with $B_n \subset\subset S^N$ and $B_{n+1} \subset B_n$ and hence $\#(B_n) \downarrow$ as $n \to \infty$. By assumption, $\lim_{n\to\infty} \#(B_n) \neq 0$ and therefore $\#(B_n) = k > 0$ for all n large. It then follows that there exists $n_0 \in \mathbb{N}$ such that $B_n = B_{n_0}$ for all $n \geq n_0$. Therefore $\cap_{n=1}^\infty C_n = B_{n_0} \times \Omega \neq \emptyset$.

Case 2. $\lim_{n\to\infty} N_n = \infty$. By assumption, there exists $\omega(n) = (\omega_1(n), \omega_2(n), \dots) \in \Omega$ such that $\omega(n) \in C_n$ for all n. Moreover, since $\omega(n) \in C_n \subset C_k$ for all $k \leq n$, it follows that

$$(\omega_1(n), \omega_2(n), \dots, \omega_{N_k}(n)) \in B_k \text{ for all } n \geq k \tag{8.21}$$

and as B_k is a finite set $\{\omega_i(n)\}_{n=1}^\infty$ must be a finite set for all $1 \leq i \leq N_k$. As $N_k \to \infty$ as $k \to \infty$ it follows that $\{\omega_i(n)\}_{n=1}^\infty$ is a finite set for all $i \in \mathbb{N}$. Using this observation, we may find, $s_1 \in S$ and an infinite subset, $\Gamma_1 \subset \mathbb{N}$ such that $\omega_1(n) = s_1$ for all $n \in \Gamma_1$. Similarly, there exists $s_2 \in S$ and an infinite set, $\Gamma_2 \subset \Gamma_1$, such that $\omega_2(n) = s_2$ for all $n \in \Gamma_2$. Continuing this procedure inductively, there exists (for all $j \in \mathbb{N}$) infinite subsets, $\Gamma_j \subset \mathbb{N}$ and points $s_j \in S$ such that $\Gamma_1 \supset \Gamma_2 \supset \Gamma_3 \supset \dots$ and $\omega_j(n) = s_j$ for all $n \in \Gamma_j$.

We are now going to complete the proof by showing $s := (s_1, s_2, \dots) \in \cap_{n=1}^\infty C_n$. By the construction above, for all $N \in \mathbb{N}$ we have

$$(\omega_1(n), \dots, \omega_N(n)) = (s_1, \dots, s_N) \text{ for all } n \in \Gamma_N.$$

Taking $N = N_k$ and $n \in \Gamma_{N_k}$ with $n \geq k$, we learn from Eq. (8.21) that

$$(s_1, \dots, s_{N_k}) = (\omega_1(n), \dots, \omega_{N_k}(n)) \in B_k.$$

But this is equivalent to showing $s \in C_k$. Since $k \in \mathbb{N}$ was arbitrary it follows that $s \in \cap_{n=1}^\infty C_n$. ∎

Let $\bar{S} := S$ is S is a finite set and $\bar{S} = S \cup \{\infty\}$ if S is an infinite set. Here, ∞, is simply another point not in S which we call infinity Let $\{x_n\}_{n=1}^\infty \subset \bar{S}$ be a sequence, then we way $\lim_{n\to\infty} x_n = \infty$ if for every $A \subset\subset S$, $x_n \notin A$ for almost all n and we say that $\lim_{n\to\infty} x_n = s \in S$ if $x_n = s$ for almost all n. For example this is the usual notion of convergence for $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ and $\bar{S} = S \cup \{0\} \subset [0,1]$, where 0 is playing the role of infinity here. Observe that either $\lim_{n\to\infty} x_n = \infty$ or there exists a finite subset $F \subset S$ such that $x_n \in F$ infinitely often. Moreover, there must be some point, $s \in F$ such that $x_n = s$ infinitely often. Thus if we let $\{n_1 < n_2 < \dots\} \subset \mathbb{N}$ be chosen such that $x_{n_k} = s$ for all k, then $\lim_{k\to\infty} x_{n_k} = s$. Thus we have shown that every sequence in $\bar{S}$ has a convergent subsequence.

Lemma 8.36 (Baby Tychonov Theorem I.). *Let $\bar{\Omega} := \bar{S}^{\mathbb{N}}$ and $\{\omega(n)\}_{n=1}^\infty$ be a sequence in $\bar{\Omega}$. Then there is a subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$ such that $\lim_{k\to\infty} \omega(n_k)$ exists in $\bar{\Omega}$ by which we mean, $\lim_{k\to\infty} \omega_i(n_k)$ exists in $\bar{S}$ for all $i \in \mathbb{N}$.*

Proof. This follows by the usual cantor's diagonalization argument. Indeed, let $\{n_k^1\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ be chosen so that $\lim_{k\to\infty} \omega_1(n_k^1) = s_1 \in \bar{S}$ exists. Then choose $\{n_k^2\}_{k=1}^\infty \subset \{n_k^1\}_{k=1}^\infty$ so that $\lim_{k\to\infty} \omega_2(n_k^2) = s_2 \in \bar{S}$ exists. Continue on this way to inductively choose

$$\{n_k^1\}_{k=1}^\infty \supset \{n_k^2\}_{k=1}^\infty \supset \dots \supset \{n_k^l\}_{k=1}^\infty \supset \dots$$

such that $\lim_{k\to\infty} \omega_l(n_k^l) = s_l \in \bar{S}$. The subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$, may now be defined by, $n_k = n_k^k$. ∎

Corollary 8.37 (Baby Tychonov Theorem II.). *Suppose that $\{F_n\}_{n=1}^\infty \subset \bar{\Omega}$ is decreasing sequence of non-empty sets which are closed under taking sequential limits, then $\cap_{n=1}^\infty F_n \neq \emptyset$.*

Proof. Since $F_n \neq \emptyset$ there exists $\omega(n) \in F_n$ for all n. Using Lemma 8.36, there exists $\{n_k\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ such that $\omega := \lim_{k\to\infty} \omega(n_k)$ exits in $\bar{\Omega}$. Since $\omega(n_k) \in F_n$ for all $k \geq n$, it follows that $\omega \in F_n$ for all n, i.e. $\omega \in \cap_{n=1}^\infty F_n$ and hence $\cap_{n=1}^\infty F_n \neq \emptyset$. ∎

Example 8.38. Suppose that $1 \leq N_1 < N_2 < N_3 < \dots$, $F_n = K_n \times \Omega$ with $K_n \subset\subset S^{N_n}$ such that $\{F_n\}_{n=1}^\infty \subset \Omega$ is a decreasing sequence of non-empty sets. Then $\cap_{n=1}^\infty F_n \neq \emptyset$. To prove this, let $\bar{F}_n := K_n \times \bar{\Omega}$ in which case $\bar{F}_n$ are non – empty sets closed under taking limits. Therefore by Corollary 8.37, $\cap_n \bar{F}_n \neq \emptyset$. This completes the proof since it is easy to check that $\cap_{n=1}^\infty F_n = \cap_n \bar{F}_n \neq \emptyset$.

Corollary 8.39. *If S is a finite set and $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \neq \emptyset$.*

Proof. This follows directly from Example 8.38 since necessarily, $A_n = K_n \times \Omega$, for some $K_n \subset\subset S^{N_n}$. ∎

Theorem 8.40 (Kolmogorov's Extension Theorem I.). *Let us continue the notation above with the further assumption that S is a finite set. Then **every** finitely additive probability measure, $P : \mathcal{A} \to [0,1]$, has a unique extension to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

Proof. From Theorem 8.27, it suffices to show $\lim_{n\to\infty} P(A_n) = 0$ whenever $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ with $A_n \downarrow \emptyset$. However, by Lemma 8.35 with $C_n = A_n$, $A_n \in \mathcal{A}$ and $A_n \downarrow \emptyset$, we must have that $A_n = \emptyset$ for a.a. n and in particular $P(A_n) = 0$ for a.a. n. This certainly implies $\lim_{n\to\infty} P(A_n) = 0$. ∎

For the next three exercises, suppose that S is a finite set and continue the notation from above. Further suppose that $P : \sigma(\mathcal{A}) \to [0,1]$ is a probability measure and for $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, let

$$p_n(s_1, \dots, s_n) := P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}). \tag{8.22}$$

Exercise 8.6 (Consistency Conditions). If p_n is defined as above, show:

1. $\sum_{s\in S} p_1(s) = 1$ and
2. for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$,

$$p_n(s_1, \dots, s_n) = \sum_{s\in S} p_{n+1}(s_1, \dots, s_n, s).$$

Exercise 8.7 (Converse to 8.6). Suppose for each $n \in \mathbb{N}$ we are given functions, $p_n : S^n \to [0,1]$ such that the consistency conditions in Exercise 8.6 hold. Then there exists a unique probability measure, P on $\sigma(\mathcal{A})$ such that Eq. (8.22) holds for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$.

Example 8.41 (Existence of iid simple R.V.s). Suppose now that $q : S \to [0,1]$ is a function such that $\sum_{s\in S} q(s) = 1$. Then there exists a unique probability measure P on $\sigma(\mathcal{A})$ such that, for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, we have

$$P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}) = q(s_1) \dots q(s_n).$$

This is a special case of Exercise 8.7 with $p_n(s_1, \dots, s_n) := q(s_1) \dots q(s_n)$.

Theorem 8.42 (Kolmogorov's Extension Theorem II). *Suppose now that S is countably infinite set and $P : \mathcal{A} \to [0,1]$ is a finitely additive measure such that $P|_{\mathcal{A}_n}$ is a σ – additive measure for each $n \in \mathbb{N}$. Then P extends uniquely to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

Proof. From Theorem 8.27 it suffice to show; if $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$, then $\cap_{m=1}^\infty A_m \neq \emptyset$. You are asked to verify this property of P in the next couple of exercises. ∎

For the next couple of exercises the hypothesis of Theorem 8.42 are to be assumed.

Exercise 8.8. Show for each $n \in \mathbb{N}$, $A \in \mathcal{A}_n$, and $\varepsilon > 0$ are given. Show there exists $F \in \mathcal{A}_n$ such that $F \subset A$, $F = K \times \Omega$ with $K \subset\subset S^n$, and $P(A \setminus F) < \varepsilon$.

Exercise 8.9. Let $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ be a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$. Using Exercise 8.8, choose $F_m = K_m \times \Omega \subset A_m$ with $K_m \subset\subset S^{N_n}$ and $P(A_m \setminus F_m) \le \varepsilon/2^{m+1}$. Further define $C_m := F_1 \cap \dots \cap F_m$ for each m. Show;

1. Show $A_m \setminus C_m \subset (A_1 \setminus F_1) \cup (A_2 \setminus F_2) \cup \dots \cup (A_m \setminus F_m)$ and use this to conclude that $P(A_m \setminus C_m) \le \varepsilon/2$.
2. Conclude C_m is not empty for m.
3. Use Lemma 8.35 to conclude that $\emptyset \neq \cap_{m=1}^\infty C_m \subset \cap_{m=1}^\infty A_m$.

Exercise 8.10. Convince yourself that the results of Exercise 8.6 and 8.7 are valid when S is a countable set. (See Example 7.6.)

In summary, the main result of this section states, to any sequence of functions, $p_n : S^n \to [0,1]$, such that $\sum_{\lambda\in S^n} p_n(\lambda) = 1$ and $\sum_{s\in S} p_{n+1}(\lambda, s) = p_n(\lambda)$ for all n and $\lambda \in S^n$, there exists a unique probability measure, P, on $\mathcal{B} := \sigma(\mathcal{A})$ such that

$$P(B \times \Omega) = \sum_{\lambda\in B} p_n(\lambda) \ \forall\ B \subset S^n \text{ and } n \in \mathbb{N}.$$

Example 8.43 (Markov Chain Probabilities). Let S be a finite or at most countable state space and $p : S \times S \to [0,1]$ be a **Markov kernel,** i.e.

$$\sum_{y\in S} p(x,y) = 1 \text{ for all } x \in S. \tag{8.23}$$

Also let $\pi : S \to [0,1]$ be a probability function, i.e. $\sum_{x\in S} \pi(x) = 1$. We now take

$$\Omega := S^{\mathbb{N}_0} = \{\omega = (s_0, s_1, \dots) : s_j \in S\}$$

and let $X_n : \Omega \to S$ be given by

$$X_n(s_0, s_1, \dots) = s_n \text{ for all } n \in \mathbb{N}_0.$$

Then there exists a unique probability measure, P_π, on $\sigma(\mathcal{A})$ such that

$$P_\pi(X_0 = x_0, \dots, X_n = x_n) = \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

for all $n \in \mathbb{N}_0$ and $x_0, x_1, \dots, x_n \in S$. To see such a measure exists, we need only verify that

$$p_n(x_0, \dots, x_n) := \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

verifies the hypothesis of Exercise 8.6 taking into account a shift of the n – index.

8.6 Appendix: Regularity and Uniqueness Results*

The goal of this appendix it to approximating measurable sets from inside and outside by classes of sets which are relatively easy to understand. Our first few results are already contained in Carathoédory's existence of measures proof. Nevertheless, we state these results again and give another somewhat independent proof.

Theorem 8.44 (Finite Regularity Result). *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} = \sigma(\mathcal{A})$ and $\mu : \mathcal{B} \to [0,\infty)$ is a finite measure, i.e. $\mu(\Omega) < \infty$. Then for every $\varepsilon > 0$ and $B \in \mathcal{B}$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$.*

Proof. Let $\mathcal{B}_0$ denote the collection of $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there here exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$. It is now clear that $\mathcal{A} \subset \mathcal{B}_0$ and that $\mathcal{B}_0$ is closed under complementation. Now suppose that $B_i \in \mathcal{B}_0$ for $i = 1, 2, \dots$ and $\varepsilon > 0$ is given. By assumption there exists $A_i \in \mathcal{A}_\delta$ and $C_i \in \mathcal{A}_\sigma$ such that $A_i \subset B_i \subset C_i$ and $\mu(C_i \setminus A_i) < 2^{-i}\varepsilon$.

Let $A := \cup_{i=1}^\infty A_i$, $A^N := \cup_{i=1}^N A_i \in \mathcal{A}_\delta$, $B := \cup_{i=1}^\infty B_i$, and $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$. Then $A^N \subset A \subset B \subset C$ and

$$C \setminus A = [\cup_{i=1}^\infty C_i] \setminus A = \cup_{i=1}^\infty [C_i \setminus A] \subset \cup_{i=1}^\infty [C_i \setminus A_i].$$

Therefore,

$$\mu(C \setminus A) = \mu(\cup_{i=1}^\infty [C_i \setminus A]) \le \sum_{i=1}^\infty \mu(C_i \setminus A) \le \sum_{i=1}^\infty \mu(C_i \setminus A_i) < \varepsilon.$$

Since $C \setminus A^N \downarrow C \setminus A$, it also follows that $\mu\left(C \setminus A^N\right) < \varepsilon$ for sufficiently large N and this shows $B = \cup_{i=1}^\infty B_i \in \mathcal{B}_0$. Hence $\mathcal{B}_0$ is a sub-σ-algebra of $\mathcal{B} = \sigma(\mathcal{A})$ which contains $\mathcal{A}$ which shows $\mathcal{B}_0 = \mathcal{B}$. ■

Many theorems in the sequel will require some control on the size of a measure μ. The relevant notion for our purposes (and most purposes) is that of a σ – finite measure defined next.

Definition 8.45. *Suppose Ω is a set, $\mathcal{E} \subset \mathcal{B} \subset 2^\Omega$ and $\mu : \mathcal{B} \to [0,\infty]$ is a function. The function μ is σ – finite on $\mathcal{E}$ if there exists $E_n \in \mathcal{E}$ such that $\mu(E_n) < \infty$ and $\Omega = \cup_{n=1}^\infty E_n$. If $\mathcal{B}$ is a σ – algebra and μ is a measure on $\mathcal{B}$ which is σ – finite on $\mathcal{B}$ we will say $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space.*

The reader should check that if μ is a finitely additive measure on an algebra, $\mathcal{B}$, then μ is σ – finite on $\mathcal{B}$ iff there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$.

Corollary 8.46 (σ – Finite Regularity Result). *Theorem 8.44 continues to hold under the weaker assumption that $\mu : \mathcal{B} \to [0,\infty]$ is a measure which is σ – finite on $\mathcal{A}$.*

Proof. Let $\Omega_n \in \mathcal{A}$ such that $\cup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all n.Since $A \in \mathcal{B} \to \mu_n(A) := \mu(\Omega_n \cap A)$ is a finite measure on $A \in \mathcal{B}$ for each n, by Theorem 8.44, for every $B \in \mathcal{B}$ there exists $C_n \in \mathcal{A}_\sigma$ such that $B \subset C_n$ and $\mu(\Omega_n \cap [C_n \setminus B]) = \mu_n(C_n \setminus B) < 2^{-n}\varepsilon$. Now let $C := \cup_{n=1}^\infty [\Omega_n \cap C_n] \in \mathcal{A}_\sigma$ and observe that $B \subset C$ and

$$\mu(C \setminus B) = \mu(\cup_{n=1}^\infty ([\Omega_n \cap C_n] \setminus B)) \le \sum_{n=1}^\infty \mu([\Omega_n \cap C_n] \setminus B) = \sum_{n=1}^\infty \mu(\Omega_n \cap [C_n \setminus B]) < \varepsilon.$$

Applying this result to B^c shows there exists $D \in \mathcal{A}_\sigma$ such that $B^c \subset D$ and

$$\mu(B \setminus D^c) = \mu(D \setminus B^c) < \varepsilon.$$

So if we let $A := D^c \in \mathcal{A}_\delta$, then $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu([B \setminus A] \cup [(C \setminus B) \setminus A]) \le \mu(B \setminus A) + \mu(C \setminus B) < 2\varepsilon$$

and the result is proved. ■

Exercise 8.11. Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and μ and ν are two measures on $\mathcal{B} = \sigma(\mathcal{A})$.

a. Suppose that μ and ν are finite measures such that $\mu = \nu$ on $\mathcal{A}$. Show $\mu = \nu$.
b. Generalize the previous assertion to the case where you only assume that μ and ν are σ – finite on $\mathcal{A}$.

Corollary 8.47. *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mu : \mathcal{B} = \sigma(\mathcal{A}) \to [0,\infty]$ is a measure which is σ – finite on $\mathcal{A}$. Then for all $B \in \mathcal{B}$, there exists $A \in \mathcal{A}_{\delta\sigma}$ and $C \in \mathcal{A}_{\sigma\delta}$ such that $A \subset B \subset C$ and $\mu(C \setminus A) = 0$.*

Proof. By Theorem 8.44, given $B \in \mathcal{B}$, we may choose $A_n \in \mathcal{A}_\delta$ and $C_n \in \mathcal{A}_\sigma$ such that $A_n \subset B \subset C_n$ and $\mu(C_n \setminus B) \le 1/n$ and $\mu(B \setminus A_n) \le 1/n$. By replacing A_N by $\cup_{n=1}^N A_n$ and C_N by $\cap_{n=1}^N C_n$, we may assume that $A_n \uparrow$ and $C_n \downarrow$ as n increases. Let $A = \cup A_n \in \mathcal{A}_{\delta\sigma}$ and $C = \cap C_n \in \mathcal{A}_{\sigma\delta}$, then $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu(C \setminus B) + \mu(B \setminus A) \le \mu(C_n \setminus B) + \mu(B \setminus A_n) \le 2/n \to 0 \text{ as } n \to \infty.$$

■

Exercise 8.12. Let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n} = \sigma(\{\text{open subsets of } \mathbb{R}^n\})$ be the Borel σ – algebra on $\mathbb{R}^n$ and μ be a probability measure on $\mathcal{B}$. Further, let $\mathcal{B}_0$ denote those sets $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there exists $F \subset B \subset V$ such that F is closed, V is open, and $\mu(V \setminus F) < \varepsilon$. Show:

1. $\mathcal{B}_0$ contains all closed subsets of $\mathcal{B}$. **Hint:** given a closed subset, $F \subset \mathbb{R}^n$ and $k \in \mathbb{N}$, let $V_k := \cup_{x \in F} B(x, 1/k)$, where $B(x,\delta) := \{y \in \mathbb{R}^n : |y - x| < \delta\}$. Show, $V_k \downarrow F$ as $k \to \infty$.
2. Show $\mathcal{B}_0$ is a σ – algebra and use this along with the first part of this exercise to conclude $\mathcal{B} = \mathcal{B}_0$. **Hint:** follow closely the method used in the first step of the proof of Theorem 8.44.
3. Show for every $\varepsilon > 0$ and $B \in \mathcal{B}$, there exist a compact subset, $K \subset \mathbb{R}^n$, such that $K \subset B$ and $\mu(B \setminus K) < \varepsilon$. **Hint:** take $K := F \cap \{x \in \mathbb{R}^n : |x| \le n\}$ for some sufficiently large n.

8.7 Appendix: Completions of Measure Spaces*

Definition 8.48. *A set $E \subset \Omega$ is a **null** set if $E \in \mathcal{B}$ and $\mu(E) = 0$. If P is some "property" which is either true or false for each $x \in \Omega$, we will use the terminology P a.e. (to be read P almost everywhere) to mean*

$$E := \{x \in \Omega : P \text{ is false for } x\}$$

is a null set. For example if f and g are two measurable functions on $(\Omega, \mathcal{B}, \mu)$, $f = g$ a.e. means that $\mu(f \neq g) = 0$.

Definition 8.49. *A measure space $(\Omega, \mathcal{B}, \mu)$ is **complete** if every subset of a null set is in $\mathcal{B}$, i.e. for all $F \subset \Omega$ such that $F \subset E \in \mathcal{B}$ with $\mu(E) = 0$ implies that $F \in \mathcal{B}$.*

Proposition 8.50 (Completion of a Measure). *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. Set*

$$\begin{aligned} \mathcal{N} = \mathcal{N}^\mu &:= \{N \subset \Omega : \exists\, F \in \mathcal{B} \text{ such that } N \subset F \text{ and } \mu(F) = 0\}, \\ \mathcal{B} = \bar{\mathcal{B}}^\mu &:= \{A \cup N : A \in \mathcal{B} \text{ and } N \in \mathcal{N}\} \text{ and} \\ \bar{\mu}(A \cup N) &:= \mu(A) \text{ for } A \in \mathcal{B} \text{ and } N \in \mathcal{N}, \end{aligned}$$

*see Fig. 8.2. Then $\bar{\mathcal{B}}$ is a σ – algebra, $\bar{\mu}$ is a well defined measure on $\bar{\mathcal{B}}$, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{B}}$ which extends μ on $\mathcal{B}$, and $(\Omega, \bar{\mathcal{B}}, \bar{\mu})$ is complete measure space. The σ-algebra, $\bar{\mathcal{B}}$, is called the **completion** of $\mathcal{B}$ relative to μ and $\bar{\mu}$, is called the **completion of** μ.*

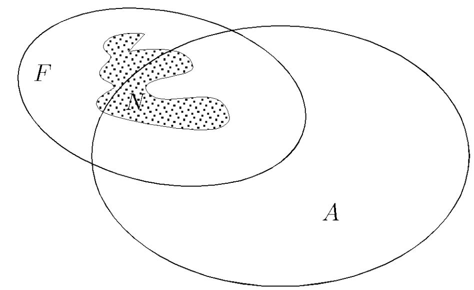


Fig. 8.2. Completing a σ – algebra.

Proof. Clearly $\Omega, \emptyset \in \bar{\mathcal{B}}$. Let $A \in \mathcal{B}$ and $N \in \mathcal{N}$ and choose $F \in \mathcal{B}$ such that $N \subset F$ and $\mu(F) = 0$. Since $N^c = (F \setminus N) \cup F^c$,

$$\begin{aligned} (A \cup N)^c &= A^c \cap N^c = A^c \cap (F \setminus N \cup F^c) \\ &= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c] \end{aligned}$$

where $[A^c \cap (F \setminus N)] \in \mathcal{N}$ and $[A^c \cap F^c] \in \mathcal{B}$. Thus $\bar{\mathcal{B}}$ is closed under complements. If $A_i \in \mathcal{B}$ and $N_i \subset F_i \in \mathcal{B}$ such that $\mu(F_i) = 0$ then $\cup(A_i \cup N_i) = (\cup A_i) \cup (\cup N_i) \in \bar{\mathcal{B}}$ since $\cup A_i \in \mathcal{B}$ and $\cup N_i \subset \cup F_i$ and $\mu(\cup F_i) \le \sum \mu(F_i) = 0$. Therefore, $\bar{\mathcal{B}}$ is a σ – algebra. Suppose $A \cup N_1 = B \cup N_2$ with $A, B \in \mathcal{B}$ and $N_1, N_2, \in \mathcal{N}$. Then $A \subset A \cup N_1 \subset A \cup N_1 \cup F_2 = B \cup F_2$ which shows that

$$\mu(A) \le \mu(B) + \mu(F_2) = \mu(B).$$

Similarly, we show that $\mu(B) \le \mu(A)$ so that $\mu(A) = \mu(B)$ and hence $\bar{\mu}(A \cup N) := \mu(A)$ is well defined. It is left as an exercise to show $\bar{\mu}$ is a measure, i.e. that it is countable additive. ∎

8.8 Appendix Monotone Class Theorems*

This appendix may be safely skipped!

Definition 8.51 (Montone Class). *$\mathcal{C} \subset 2^\Omega$ is a **monotone class** if it is closed under countable increasing unions and countable decreasing intersections.*

Lemma 8.52 (Monotone Class Theorem*). *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$. Then $\mathcal{C} = \sigma(\mathcal{A})$.*

Proof. For $C \in \mathcal{C}$ let

$$\mathcal{C}(C) = \{B \in \mathcal{C} : C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}\},$$

then $\mathcal{C}(C)$ is a monotone class. Indeed, if $B_n \in \mathcal{C}(C)$ and $B_n \uparrow B$, then $B_n^c \downarrow B^c$ and so

$$\begin{aligned}&\mathcal{C} \ni C \cap B_n \uparrow C \cap B\\&\mathcal{C} \ni C \cap B_n^c \downarrow C \cap B^c \text{ and}\\&\mathcal{C} \ni B_n \cap C^c \uparrow B \cap C^c.\end{aligned}$$

Since $\mathcal{C}$ is a monotone class, it follows that $C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}$, i.e. $B \in \mathcal{C}(C)$. This shows that $\mathcal{C}(C)$ is closed under increasing limits and a similar argument shows that $\mathcal{C}(C)$ is closed under decreasing limits. Thus we have shown that $\mathcal{C}(C)$ is a monotone class for all $C \in \mathcal{C}$. If $A \in \mathcal{A} \subset \mathcal{C}$, then $A \cap B, A \cap B^c, B \cap A^c \in \mathcal{A} \subset \mathcal{C}$ for all $B \in \mathcal{A}$ and hence it follows that $\mathcal{A} \subset \mathcal{C}(A) \subset \mathcal{C}$. Since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(A)$ is a monotone class containing $\mathcal{A}$, we conclude that $\mathcal{C}(A) = \mathcal{C}$ for any $A \in \mathcal{A}$. Let $B \in \mathcal{C}$ and notice that $A \in \mathcal{C}(B)$ happens iff $B \in \mathcal{C}(A)$. This observation and the fact that $\mathcal{C}(A) = \mathcal{C}$ for all $A \in \mathcal{A}$ implies $\mathcal{A} \subset \mathcal{C}(B) \subset \mathcal{C}$ for all $B \in \mathcal{C}$. Again since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(B)$ is a monotone class we conclude that $\mathcal{C}(B) = \mathcal{C}$ for all $B \in \mathcal{C}$. That is to say, if $A, B \in \mathcal{C}$ then $A \in \mathcal{C} = \mathcal{C}(B)$ and hence $A \cap B$, $A \cap B^c$, $A^c \cap B \in \mathcal{C}$. So $\mathcal{C}$ is closed under complements (since $\Omega \in \mathcal{A} \subset \mathcal{C}$) and finite intersections and increasing unions from which it easily follows that $\mathcal{C}$ is a σ – algebra. ∎

Measurable Functions (Random Variables)

Notation 9.1 *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E)|E \in \mathcal{E}\}.$$

If $\mathcal{G} \subset 2^X$, let

$$f_*\mathcal{G} := \{A \in 2^Y | f^{-1}(A) \in \mathcal{G}\}.$$

Definition 9.2. *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the* ***inclusion map*** *($i_A(x) = x$ for all $x \in A$) and*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

The following results will be used frequently (often without further reference) in the sequel.

Lemma 9.3 (A key measurability lemma). *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$, then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E})). \tag{9.1}$$

In particular, if $A \subset Y$ then

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A), \tag{9.2}$$

(Similar assertion hold with $\sigma(\cdot)$ being replaced by $\mathcal{A}(\cdot)$.)

Proof. Since $\mathcal{E} \subset \sigma(\mathcal{E})$, it follows that $f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E}))$. Moreover, by Exercise 9.1 below, $f^{-1}(\sigma(\mathcal{E}))$ is a σ – algebra and therefore,

$$\sigma(f^{-1}(\mathcal{E})) \subset f^{-1}(\sigma(\mathcal{E})).$$

To finish the proof we must show $f^{-1}(\sigma(\mathcal{E})) \subset \sigma(f^{-1}(\mathcal{E}))$, i.e. that $f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))$ for all $B \in \sigma(\mathcal{E})$. To do this we follow the usual measure theoretic mantra, namely let

$$\mathcal{M} := \left\{B \subset Y : f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))\right\} = f_*\sigma(f^{-1}(\mathcal{E})).$$

We will now finish the proof by showing $\sigma(\mathcal{E}) \subset \mathcal{M}$. This is easily achieved by observing that $\mathcal{M}$ is a σ – algebra (see Exercise 9.1) which contains $\mathcal{E}$ and therefore $\sigma(\mathcal{E}) \subset \mathcal{M}$.

Equation (9.2) is a special case of Eq. (9.1). Indeed, $f = i_A : A \to X$ we have

$$(\sigma(\mathcal{E}))_A = i_A^{-1}(\sigma(\mathcal{E})) = \sigma(i_A^{-1}(\mathcal{E})) = \sigma(\mathcal{E}_A).$$

■

Exercise 9.1. If $f : X \to Y$ is a function and $\mathcal{F} \subset 2^Y$ and $\mathcal{B} \subset 2^X$ are σ – algebras (algebras), then $f^{-1}\mathcal{F}$ and $f_*\mathcal{B}$ are σ – algebras (algebras).

Example 9.4. Let $\mathcal{E} = \{(a,b] : -\infty < a < b < \infty\}$ and $\mathcal{B} = \sigma(\mathcal{E})$ be the Borel σ – field on $\mathbb{R}$. Then

$$\mathcal{E}_{(0,1]} = \{(a,b] : 0 \le a < b \le 1\}$$

and we have

$$\mathcal{B}_{(0,1]} = \sigma\left(\mathcal{E}_{(0,1]}\right).$$

In particular, if $A \in \mathcal{B}$ such that $A \subset (0,1]$, then $A \in \sigma\left(\mathcal{E}_{(0,1]}\right)$.

9.1 Measurable Functions

Definition 9.5. *A* ***measurable space*** *is a pair $(X, \mathcal{M})$, where X is a set and $\mathcal{M}$ is a σ – algebra on X.*

To motivate the notion of a measurable function, suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{R}_+$ is a function. Roughly speaking, we are going to define $\int_X f d\mu$ as a certain limit of sums of the form,

$$\sum_{0<a_1<a_2<a_3<\dots}^{\infty} a_i \mu(f^{-1}(a_i, a_{i+1}]).$$

For this to make sense we will need to require $f^{-1}((a,b]) \in \mathcal{M}$ for all $a < b$. Because of Corollary 9.11 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$.

Definition 9.6. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is* ***measurable*** *of more precisely, $\mathcal{M}/\mathcal{F}$ – measurable or $(\mathcal{M}, \mathcal{F})$ – measurable, if $f^{-1}(\mathcal{F}) \subset \mathcal{M}$, i.e. if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{F}$.*

Remark 9.7. Let $f : X \to Y$ be a function. Given a σ – algebra $\mathcal{F} \subset 2^Y$, the σ – algebra $\mathcal{M} := f^{-1}(\mathcal{F})$ is the smallest σ – algebra on X such that f is $(\mathcal{M}, \mathcal{F})$ - measurable . Similarly, if $\mathcal{M}$ is a σ - algebra on X then

$$\mathcal{F} = f_*\mathcal{M} = \{A \in 2^Y | f^{-1}(A) \in \mathcal{M}\}$$

is the largest σ – algebra on Y such that f is $(\mathcal{M}, \mathcal{F})$ - measurable.

Example 9.8 (Indicator Functions). Let $(X, \mathcal{M})$ be a measurable space and $A \subset X$. Then 1_A is $(\mathcal{M}, \mathcal{B}_\mathbb{R})$ – measurable iff $A \in \mathcal{M}$. Indeed, $1_A^{-1}(W)$ is either $\emptyset$, X, A or A^c for any $W \subset \mathbb{R}$ with $1_A^{-1}(\{1\}) = A$.

Example 9.9. Suppose $f : X \to Y$ with Y being a finite or countable set and $\mathcal{F} = 2^Y$. Then f is measurable iff $f^{-1}(\{y\}) \in \mathcal{M}$ for all $y \in Y$.

Proposition 9.10. *Suppose that $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ are measurable spaces and further assume $\mathcal{E} \subset \mathcal{F}$ generates $\mathcal{F}$, i.e. $\mathcal{F} = \sigma(\mathcal{E})$. Then a map, $f : X \to Y$ is measurable iff $f^{-1}(\mathcal{E}) \subset \mathcal{M}$.*

Proof. If f is $\mathcal{M}/\mathcal{F}$ measurable, then $f^{-1}(\mathcal{E}) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}$. Conversely if $f^{-1}(\mathcal{E}) \subset \mathcal{M}$ then $\sigma(f^{-1}(\mathcal{E})) \subset \mathcal{M}$ and so making use of Lemma 9.3,

$$f^{-1}(\mathcal{F}) = f^{-1}(\sigma(\mathcal{E})) = \sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}.$$

■

Corollary 9.11. *Suppose that $(X, \mathcal{M})$ is a measurable space. Then the following conditions on a function $f : X \to \mathbb{R}$ are equivalent:*

1. *f is $(\mathcal{M}, \mathcal{B}_\mathbb{R})$ – measurable,*
2. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{Q}$,*
4. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

Exercise 9.2. Prove Corollary 9.11. **Hint:** See Exercise 6.7.

Exercise 9.3. If $\mathcal{M}$ is the σ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the σ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

Exercise 9.4. Let $(X, \mathcal{M})$ be a measure space and $f_n : X \to \mathbb{R}$ be a sequence of measurable functions on X. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \mathbf{in}\ \mathbb{R}\} \in \mathcal{M}$. Similarly show the same holds if $\mathbb{R}$ is replaced by $\mathbb{C}$.

Exercise 9.5. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_\mathbb{R}, \mathcal{B}_\mathbb{R})$ – measurable.

Definition 9.12. *Given measurable spaces $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ and a subset $A \subset X$. We say a function $f : A \to Y$ is measurable iff f is $\mathcal{M}_A/\mathcal{F}$ – measurable.*

Proposition 9.13 (Localizing Measurability). *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces and $f : X \to Y$ be a function.*

1. *If f is measurable and $A \subset X$ then $f|_A : A \to Y$ is $\mathcal{M}_A/\mathcal{F}$ – measurable.*
2. *Suppose there exist $A_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty A_n$ and $f|A_n$ is $\mathcal{M}_{A_n}/\mathcal{F}$ – measurable for all n, then f is $\mathcal{M}$ – measurable.*

Proof. 1. If $f : X \to Y$ is measurable, $f^{-1}(B) \in \mathcal{M}$ for all $B \in \mathcal{F}$ and therefore

$$f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathcal{M}_A \text{ for all } B \in \mathcal{F}.$$

2. If $B \in \mathcal{F}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty \left(f^{-1}(B) \cap A_n\right) = \cup_{n=1}^\infty f|_{A_n}^{-1}(B).$$

Since each $A_n \in \mathcal{M}$, $\mathcal{M}_{A_n} \subset \mathcal{M}$ and so the previous displayed equation shows $f^{-1}(B) \in \mathcal{M}$. ■

Lemma 9.14 (Composing Measurable Functions). *Suppose that $(X, \mathcal{M})$, $(Y, \mathcal{F})$ and $(Z, \mathcal{G})$ are measurable spaces. If $f : (X, \mathcal{M}) \to (Y, \mathcal{F})$ and $g : (Y, \mathcal{F}) \to (Z, \mathcal{G})$ are measurable functions then $g \circ f : (X, \mathcal{M}) \to (Z, \mathcal{G})$ is measurable as well.*

Proof. By assumption $g^{-1}(\mathcal{G}) \subset \mathcal{F}$ and $f^{-1}(\mathcal{F}) \subset \mathcal{M}$ so that

$$(g \circ f)^{-1}(\mathcal{G}) = f^{-1}\left(g^{-1}(\mathcal{G})\right) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}.$$

■

Definition 9.15 (σ – Algebras Generated by Functions). *Let X be a set and suppose there is a collection of measurable spaces $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$. Let $\sigma(f_\alpha : \alpha \in I)$ denote the smallest σ – algebra on X such that each f_α is measurable, i.e.*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(\cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)).$$

Example 9.16. Suppose that Y is a finite set, $\mathcal{F} = 2^Y$, and $X = Y^N$ for some $N \in \mathbb{N}$. Let $\pi_i : Y^N \to Y$ be the projection maps, $\pi_i(y_1, \ldots, y_N) = y_i$. Then, as the reader should check,

$$\sigma(\pi_1, \ldots, \pi_n) = \left\{A \times \Lambda^{N-n} : A \subset \Lambda^n\right\}.$$

Proposition 9.17. *Assuming the notation in Definition 9.15 (so $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$) and additionally let $(Z, \mathcal{M})$ be a measurable space. Then $g : Z \to X$ is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ g \left(Z \xrightarrow{g} X \xrightarrow{f_\alpha} Y_\alpha\right)$ is $(\mathcal{M}, \mathcal{F}_\alpha)$–measurable for all $\alpha \in I$.*

Proof. $(\Rightarrow)$ If g is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in I))$ – measurable, then the composition $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable by Lemma 9.14.

$(\Leftarrow)$ Since $\sigma(f_\alpha : \alpha \in I) = \sigma(\mathcal{E})$ where $\mathcal{E} := \cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)$, according to Proposition 9.10, it suffices to show $g^{-1}(A) \in \mathcal{M}$ for $A \in f_\alpha^{-1}(\mathcal{F}_\alpha)$. But this is true since if $A = f_\alpha^{-1}(B)$ for some $B \in \mathcal{F}_\alpha$, then $g^{-1}(A) = g^{-1}\left(f_\alpha^{-1}(B)\right) = (f_\alpha \circ g)^{-1}(B) \in \mathcal{M}$ because $f_\alpha \circ g : Z \to Y_\alpha$ is assumed to be measurable. ■

Definition 9.18. *If $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, then the product measure space, $(Y, \mathcal{F})$, is $Y := \prod_{\alpha \in I} Y_\alpha$, $\mathcal{F} := \sigma(\pi_\alpha : \alpha \in I)$ where $\pi_\alpha : Y \to Y_\alpha$ is the α – component projection. We call $\mathcal{F}$ the product σ – algebra and denote it by, $\mathcal{F} = \otimes_{\alpha \in I} \mathcal{F}_\alpha$.*

Let us record an important special case of Proposition 9.17.

Corollary 9.19. *If $(Z, \mathcal{M})$ is a measure space, then $g : Z \to Y = \prod_{\alpha \in I} Y_\alpha$ is $(\mathcal{M}, \mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha)$ – measurable iff $\pi_\alpha \circ g : Z \to Y_\alpha$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$.*

As a special case of the above corollary, if $A = \{1, 2, \ldots, n\}$, then $Y = Y_1 \times \cdots \times Y_n$ and $g = (g_1, \ldots, g_n) : Z \to Y$ is measurable iff each component, $g_i : Z \to Y_i$, is measurable. Here is another closely related result.

Proposition 9.20. *Suppose X is a set, $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, and we are given maps, $f_\alpha : X \to Y_\alpha$, for all $\alpha \in I$. If $f : X \to Y := \prod_{\alpha \in I} Y_\alpha$ is the unique map, such that $\pi_\alpha \circ f = f_\alpha$, then*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(f) = f^{-1}(\mathcal{F})$$

where $\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha$.

Proof. Since $\pi_\alpha \circ f = f_\alpha$ is $\sigma(f_\alpha : \alpha \in I) / \mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ it follows from Corollary 9.19 that $f : X \to Y$ is $\sigma(f_\alpha : \alpha \in I) / \mathcal{F}$ – measurable. Since $\sigma(f)$ is the smallest σ – algebra on X such that f is measurable we may conclude that $\sigma(f) \subset \sigma(f_\alpha : \alpha \in I)$.

Conversely, for each $\alpha \in I$, $f_\alpha = \pi_\alpha \circ f$ is $\sigma(f) / \mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ being the composition of two measurable functions. Since $\sigma(f_\alpha : \alpha \in I)$ is the smallest σ – algebra on X such that each $f_\alpha : X \to Y_\alpha$ is measurable, we learn that $\sigma(f_\alpha : \alpha \in I) \subset \sigma(f)$. ■

Exercise 9.6. Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\mathcal{E}_i$ is a subset of $\mathcal{F}_i$ such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for $i = 1$ and 2. Show $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{E})$ where $\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{E}_i \text{ for } i = 1, 2\}$. **Hints:**

1. First show that if Y is a set and $\mathcal{S}_1$ and $\mathcal{S}_2$ are two non-empty subsets of 2^Y, then $\sigma(\sigma(\mathcal{S}_1) \cup \sigma(\mathcal{S}_2)) = \sigma(\mathcal{S}_1 \cup \mathcal{S}_2)$. (In fact, one has that $\sigma(\cup_{\alpha \in I} \sigma(\mathcal{S}_\alpha)) = \sigma(\cup_{\alpha \in I} \mathcal{S}_\alpha)$ for any collection of non-empty subsets, $\{\mathcal{S}_\alpha\}_{\alpha \in I} \subset 2^Y$.)
2. After this you might start your proof as follows;

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\left(\pi_1^{-1}(\mathcal{F}_1) \cup \pi_2^{-1}(\mathcal{F}_2)\right) = \sigma\left(\pi_1^{-1}(\sigma(\mathcal{E}_2)) \cup \pi_2^{-1}(\sigma(\mathcal{E}_2))\right) = \ldots.$$

Remark 9.21. The reader should convince herself that Exercise 9.6 admits the following extension. If I is any finite or countable index set, $\{(Y_i, \mathcal{F}_i)\}_{i \in I}$ are measurable spaces and $\mathcal{E}_i \subset \mathcal{F}_i$ are such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for all $i \in I$, then

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{E}_j \text{ for all } j \in I\right\}\right)$$

and in particular,

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{F}_j \text{ for all } j \in I\right\}\right).$$

The last fact is easily verified directly without the aid of Exercise 9.6.

Exercise 9.7. Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\emptyset \neq B_i \subset Y_i$ for $i = 1, 2$. Show

$$[\mathcal{F}_1 \otimes \mathcal{F}_2]_{B_1 \times B_2} = [\mathcal{F}_1]_{B_1} \otimes [\mathcal{F}_2]_{B_2}.$$

Hint: you may find it useful to use the result of Exercise 9.6 with

$$\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{F}_i \text{ for } i = 1, 2\}.$$

Definition 9.22. *A function $f : X \to Y$ between two topological spaces is* ***Borel measurable*** *if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 9.23. *Let X and Y be two topological spaces and $f : X \to Y$ be a continuous function. Then f is Borel measurable.*

Proof. Using Lemma 9.3 and $\mathcal{B}_Y = \sigma(\tau_Y)$,

$$f^{-1}(\mathcal{B}_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = \mathcal{B}_X.$$

■

Example 9.24. For $i = 1, 2, \ldots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x) = x_i$. Then each π_i is continuous and therefore $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable.

Lemma 9.25. *Let $\mathcal{E}$ denote the collection of open rectangle in $\mathbb{R}^n$, then $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E})$. We also have that $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \ldots, \pi_n) = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ and in particular, $A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$ whenever $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the σ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$. (Also see Remark 9.21.)*

Proof. Assertion 1. Since $\mathcal{E} \subset \mathcal{B}_{\mathbb{R}^n}$, it follows that $\sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}$. Let

$$\mathcal{E}_0 := \{(a, b) : a, b \in \mathbb{Q}^n \ni a < b\},$$

where, for $a, b \in \mathbb{R}^n$, we write $a < b$ iff $a_i < b_i$ for $i = 1, 2, \ldots, n$ and let

$$(a, b) = (a_1, b_1) \times \cdots \times (a_n, b_n). \tag{9.3}$$

Since every open set, $V \subset \mathbb{R}^n$, may be written as a (necessarily) countable union of elements from $\mathcal{E}_0$, we have

$$V \in \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}),$$

i.e. $\sigma(\mathcal{E}_0)$ and hence $\sigma(\mathcal{E})$ contains all open subsets of $\mathbb{R}^n$. Hence we may conclude that

$$\mathcal{B}_{\mathbb{R}^n} = \sigma(\text{open sets}) \subset \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}.$$

Assertion 2. Since each $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is continuous, it is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable and therefore, $\sigma(\pi_1, \ldots, \pi_n) \subset \mathcal{B}_{\mathbb{R}^n}$. Moreover, if (a, b) is as in Eq. (9.3), then

$$(a, b) = \cap_{i=1}^n \pi_i^{-1}((a_i, b_i)) \in \sigma(\pi_1, \ldots, \pi_n).$$

Therefore, $\mathcal{E} \subset \sigma(\pi_1, \ldots, \pi_n)$ and $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E}) \subset \sigma(\pi_1, \ldots, \pi_n)$.

Assertion 3. If $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$, then

$$A_1 \times \cdots \times A_n = \cap_{i=1}^n \pi_i^{-1}(A_i) \in \sigma(\pi_1, \ldots, \pi_n) = \mathcal{B}_{\mathbb{R}^n}.$$

■

Corollary 9.26. *If $(X, \mathcal{M})$ is a measurable space, then*

$$f = (f_1, f_2, \ldots, f_n) : X \to \mathbb{R}^n$$

is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}^n})$ – measurable iff $f_i : X \to \mathbb{R}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable for each i. In particular, a function $f : X \to \mathbb{C}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable.

Proof. This is an application of Lemma 9.25 and Corollary 9.19 with $Y_i = \mathbb{R}$ for each i. ■

Corollary 9.27. *Let $(X, \mathcal{M})$ be a measurable space and $f, g : X \to \mathbb{C}$ be $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_\pm : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ by $F(x) = (f(x), g(x))$, $A_\pm(w, z) = w \pm z$ and $M(w, z) = wz$. Then $A_\pm$ and M are continuous and hence $(\mathcal{B}_{\mathbb{C}^2}, \mathcal{B}_{\mathbb{C}})$ – measurable. Also F is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}^2})$ – measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable. Therefore $A_\pm \circ F = f \pm g$ and $M \circ F = f \cdot g$, being the composition of measurable functions, are also measurable. ■

Lemma 9.28. *Let $\alpha \in \mathbb{C}$, $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{C}$ be a $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable function. Then*

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if} \quad f(x) \neq 0 \\ \alpha & \text{if} \quad f(x) = 0 \end{cases}$$

is measurable.

Proof. Define $i : \mathbb{C} \to \mathbb{C}$ by

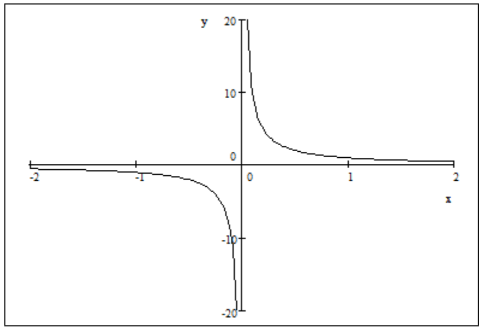
$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \neq 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

For any open set $V \subset \mathbb{C}$ we have

$$i^{-1}(V) = i^{-1}(V \setminus \{0\}) \cup i^{-1}(V \cap \{0\})$$

Because i is continuous except at $z = 0$, $i^{-1}(V \setminus \{0\})$ is an open set and hence in $\mathcal{B}_{\mathbb{C}}$. Moreover, $i^{-1}(V \cap \{0\}) \in \mathcal{B}_{\mathbb{C}}$ since $i^{-1}(V \cap \{0\})$ is either the empty set or the one point set $\{0\}$. Therefore $i^{-1}(\tau_{\mathbb{C}}) \subset \mathcal{B}_{\mathbb{C}}$ and hence $i^{-1}(\mathcal{B}_{\mathbb{C}}) = i^{-1}(\sigma(\tau_{\mathbb{C}})) = \sigma(i^{-1}(\tau_{\mathbb{C}})) \subset \mathcal{B}_{\mathbb{C}}$ which shows that i is Borel measurable. Since $F = i \circ f$ is the composition of measurable functions, F is also measurable. ■

Remark 9.29. For the real case of Lemma 9.28, define i as above but now take z to real. From the plot of i, Figure 9.29, the reader may easily verify that $i^{-1}((-\infty, a])$ is an infinite half interval for all a and therefore i is measurable. See Example 9.34 for another proof of this fact.



We will often deal with functions $f : X \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. When talking about measurability in this context we will refer to the σ – algebra on $\bar{\mathbb{R}}$ defined by

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma\left(\{[a,\infty] : a \in \mathbb{R}\}\right). \tag{9.4}$$

Proposition 9.30 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$). *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{9.5}$$

In particular $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ and $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}}$.

Proof. Let us first observe that

$$\{-\infty\} = \cap_{n=1}^{\infty}[-\infty,-n) = \cap_{n=1}^{\infty}[-n,\infty]^c \in \mathcal{B}_{\bar{\mathbb{R}}},$$
$$\{\infty\} = \cap_{n=1}^{\infty}[n,\infty] \in \mathcal{B}_{\bar{\mathbb{R}}} \text{ and } \mathbb{R} = \bar{\mathbb{R}} \setminus \{\pm\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}.$$

Letting $i : \mathbb{R} \to \bar{\mathbb{R}}$ be the inclusion map,

$$\begin{aligned} i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) &= \sigma\left(i^{-1}\left(\left\{[a,\infty] : a \in \bar{\mathbb{R}}\right\}\right)\right) = \sigma\left(\left\{i^{-1}\left([a,\infty]\right) : a \in \bar{\mathbb{R}}\right\}\right) \\ &= \sigma\left(\left\{[a,\infty] \cap \mathbb{R} : a \in \bar{\mathbb{R}}\right\}\right) = \sigma\left(\{[a,\infty) : a \in \mathbb{R}\}\right) = \mathcal{B}_{\mathbb{R}}. \end{aligned}$$

Thus we have shown

$$\mathcal{B}_{\mathbb{R}} = i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) = \{A \cap \mathbb{R} : A \in \mathcal{B}_{\bar{\mathbb{R}}}\}.$$

This implies:

1. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \Longrightarrow A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and
2. if $A \subset \bar{\mathbb{R}}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\bar{\mathbb{R}}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \Delta B \subset \{\pm\infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ we may conclude that $A \in \mathcal{B}_{\bar{\mathbb{R}}}$ as well.

This proves Eq. (9.5). ■

The proofs of the next two corollaries are left to the reader, see Exercises 9.8 and 9.9.

Corollary 9.31. *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \bar{\mathbb{R}}$ be a function. Then the following are equivalent*

1. *f is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ - measurable,*
2. *$f^{-1}((a,\infty]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((-\infty,a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
4. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$ and $f^0 : X \to \mathbb{R}$ defined by*

$$f^0(x) := \begin{cases} f(x) & \text{if } \ f(x) \in \mathbb{R} \\ 0 & \text{if } f(x) \in \{\pm\infty\} \end{cases}$$

is measurable.

Corollary 9.32. *Let $(X, \mathcal{M})$ be a measurable space, $f, g : X \to \bar{\mathbb{R}}$ be functions and define $f \cdot g : X \to \bar{\mathbb{R}}$ and $(f+g) : X \to \bar{\mathbb{R}}$ using the conventions, $0 \cdot \infty = 0$ and $(f+g)(x) = 0$ if $f(x) = \infty$ and $g(x) = -\infty$ or $f(x) = -\infty$ and $g(x) = \infty$. Then $f \cdot g$ and $f + g$ are measurable functions on X if both f and g are measurable.*

Exercise 9.8. Prove Corollary 9.31 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 9.11. Use Proposition 9.30 to handle item 4.

Exercise 9.9. Prove Corollary 9.32.

Proposition 9.33 (Closure under sups, infs and limits). *Suppose that $(X, \mathcal{M})$ is a measurable space and $f_j : (X, \mathcal{M}) \to \overline{\mathbb{R}}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. Then*

$$\sup{}_j f_j, \quad \inf{}_j f_j, \quad \limsup_{j\to\infty} \ f_j \text{ and } \liminf_{j\to\infty} \ f_j$$

are all $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. (Note that this result is in generally false when $(X, \mathcal{M})$ is a topological space and measurable is replaced by continuous in the statement.)

Proof. Define $g_+(x) := \sup{}_j \ f_j(x)$, then

$$\begin{aligned} \{x : g_+(x) \le a\} &= \{x : f_j(x) \le a \ \forall j\} \\ &= \cap_j \{x : f_j(x) \le a\} \in \mathcal{M} \end{aligned}$$

so that g_+ is measurable. Similarly if $g_-(x) = \inf_j f_j(x)$ then

$$\{x : g_-(x) \ge a\} = \cap_j \{x : f_j(x) \ge a\} \in \mathcal{M}.$$

Since

$$\limsup_{j\to\infty} \ f_j = \inf_n \sup\{f_j : j \ge n\} \ \text{ and}$$
$$\liminf_{j\to\infty} \ f_j = \sup_n \inf\{f_j : j \ge n\}$$

we are done by what we have already proved. ■

Example 9.34. As we saw in Remark 9.29, $i : \mathbb{R} \to \mathbb{R}$ defined by

$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \neq 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

is measurable by a simple direct argument. For an alternative argument, let

$$i_n(z) := \frac{z}{z^2 + \frac{1}{n}} \text{ for all } n \in \mathbb{N}.$$

Then i_n is continuous and $\lim_{n\to\infty} i_n(z) = i(z)$ for all $z \in \mathbb{R}$ from which it follows that i is Borel measurable.

Example 9.35. Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q} \cap [0,1]$ and define

$$f(x) = \sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x - r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x - r_n|}} = 5 \text{ if } x = r_n.$$

Then $f : \mathbb{R} \to \bar{\mathbb{R}}$ is measurable. Indeed, if

$$g_n(x) = \begin{cases} \frac{1}{\sqrt{|x - r_n|}} & \text{if } x \neq r_n \\ 0 & \text{if } x = r_n \end{cases}$$

then $g_n(x) = \sqrt{|i(x - r_n)|}$ is measurable as the composition of measurable is measurable. Therefore $g_n + 5 \cdot 1_{\{r_n\}}$ is measurable as well. Finally,

$$f(x) = \lim_{N\to\infty} \sum_{n=1}^N 2^{-n} \frac{1}{\sqrt{|x - r_n|}}$$

is measurable since sums of measurable functions are measurable and limits of measurable functions are measurable. **Moral:** if you can explicitly write a function $f : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ down then it is going to be measurable.

Definition 9.36. *Given a function $f : X \to \bar{\mathbb{R}}$ let $f_+(x) := \max\{f(x), 0\}$ and $f_-(x) := \max(-f(x), 0) = -\min(f(x), 0)$. Notice that $f = f_+ - f_-$.*

Corollary 9.37. *Suppose $(X, \mathcal{M})$ is a measurable space and $f : X \to \bar{\mathbb{R}}$ is a function. Then f is measurable iff $f_\pm$ are measurable.*

Proof. If f is measurable, then Proposition 9.33 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f = f_+ - f_-$. ■

Definition 9.38. *Let $(X, \mathcal{M})$ be a measurable space. A function $\varphi : X \to \mathbb{F}$ ($\mathbb{F}$ denotes either $\mathbb{R}$, $\mathbb{C}$ or $[0,\infty] \subset \bar{\mathbb{R}}$) is a **simple function** if φ is $\mathcal{M}$ – $\mathcal{B}_\mathbb{F}$ measurable and $\varphi(X)$ contains only finitely many elements.*

Any such simple functions can be written as

$$\varphi = \sum_{i=1}^n \lambda_i 1_{A_i} \text{ with } A_i \in \mathcal{M} \text{ and } \lambda_i \in \mathbb{F}. \tag{9.6}$$

Indeed, take $\lambda_1, \lambda_2, \ldots, \lambda_n$ to be an enumeration of the range of φ and $A_i = \varphi^{-1}(\{\lambda_i\})$. Note that this argument shows that any simple function may be written intrinsically as

$$\varphi = \sum_{y\in\mathbb{F}} y 1_{\varphi^{-1}(\{y\})}. \tag{9.7}$$

The next theorem shows that simple functions are "pointwise dense" in the space of measurable functions.

Theorem 9.39 (Approximation Theorem). *Let $f : X \to [0,\infty]$ be measurable and define, see Figure 9.1,*

$$\begin{aligned}\varphi_n(x) &:= \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right)}(x) + 2^n 1_{f^{-1}((2^n,\infty])}(x) \\ &= \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{\left\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\right\}}(x) + 2^n 1_{\{f > 2^n\}}(x)\end{aligned}$$

then $\varphi_n \le f$ for all n, $\varphi_n(x) \uparrow f(x)$ for all $x \in X$ and $\varphi_n \uparrow f$ uniformly on the sets $X_M := \{x \in X : f(x) \le M\}$ with $M < \infty$.

Moreover, if $f : X \to \mathbb{C}$ is a measurable function, then there exists simple functions φ_n such that $\lim_{n\to\infty} \varphi_n(x) = f(x)$ for all x and $|\varphi_n| \uparrow |f|$ as $n \to \infty$.

Proof. Since $f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right)$ and $f^{-1}((2^n,\infty])$ are in $\mathcal{M}$ as f is measurable, φ_n is a measurable simple function for each n. Because

$$\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] = \left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \cup \left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right],$$

if $x \in f^{-1}\left(\left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right]\right)$ then $\varphi_n(x) = \varphi_{n+1}(x) = \frac{2k}{2^{n+1}}$ and if $x \in f^{-1}\left(\left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right]\right)$ then $\varphi_n(x) = \frac{2k}{2^{n+1}} < \frac{2k+1}{2^{n+1}} = \varphi_{n+1}(x)$. Similarly

$$(2^n,\infty] = (2^n, 2^{n+1}] \cup (2^{n+1},\infty],$$

and so for $x \in f^{-1}((2^{n+1},\infty])$, $\varphi_n(x) = 2^n < 2^{n+1} = \varphi_{n+1}(x)$ and for $x \in f^{-1}((2^n, 2^{n+1}])$, $\varphi_{n+1}(x) \ge 2^n = \varphi_n(x)$. Therefore $\varphi_n \le \varphi_{n+1}$ for all n. It is

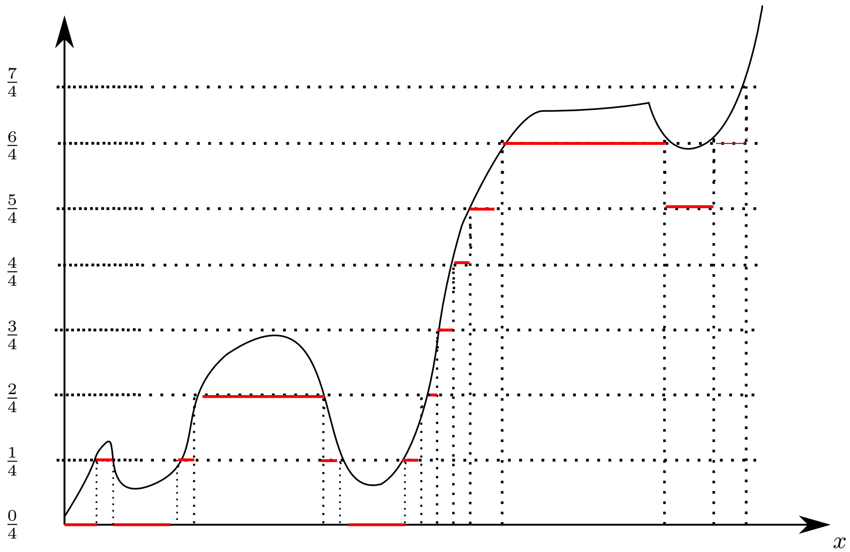


Fig. 9.1. Constructing the simple function, φ_2, approximating a function, $f : X \to [0, \infty]$. The graph of φ_2 is in red.

clear by construction that $0 \le \varphi_n(x) \le f(x)$ for all x and that $0 \le f(x) - \varphi_n(x) \le 2^{-n}$ if $x \in X_{2^n} = \{f \le 2^n\}$. Hence we have shown that $\varphi_n(x) \uparrow f(x)$ for all $x \in X$ and $\varphi_n \uparrow f$ uniformly on bounded sets.

For the second assertion, first assume that $f : X \to \mathbb{R}$ is a measurable function and choose $\varphi_n^{\pm}$ to be non-negative simple functions such that $\varphi_n^{\pm} \uparrow f_{\pm}$ as $n \to \infty$ and define $\varphi_n = \varphi_n^+ - \varphi_n^-$. Then (using $\varphi_n^+ \cdot \varphi_n^- \le f_+ \cdot f_- = 0$)

$$|\varphi_n| = \varphi_n^+ + \varphi_n^- \le \varphi_{n+1}^+ + \varphi_{n+1}^- = |\varphi_{n+1}|$$

and clearly $|\varphi_n| = \varphi_n^+ + \varphi_n^- \uparrow f_+ + f_- = |f|$ and $\varphi_n = \varphi_n^+ - \varphi_n^- \to f_+ - f_- = f$ as $n \to \infty$. Now suppose that $f : X \to \mathbb{C}$ is measurable. We may now choose simple function u_n and v_n such that $|u_n| \uparrow |\mathrm{Re}\, f|$, $|v_n| \uparrow |\mathrm{Im}\, f|$, $u_n \to \mathrm{Re}\, f$ and $v_n \to \mathrm{Im}\, f$ as $n \to \infty$. Let $\varphi_n = u_n + iv_n$, then

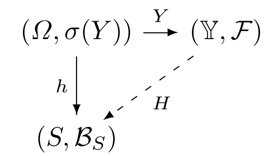
$$|\varphi_n|^2 = u_n^2 + v_n^2 \uparrow |\mathrm{Re}\, f|^2 + |\mathrm{Im}\, f|^2 = |f|^2$$

and $\varphi_n = u_n + iv_n \to \mathrm{Re}\, f + i\,\mathrm{Im}\, f = f$ as $n \to \infty$. ∎

9.2 Factoring Random Variables

Lemma 9.40. *Suppose that $(\mathbb{Y}, \mathcal{F})$ is a measurable space and $Y : \Omega \to \mathbb{Y}$ is a map. Then to every $(\sigma(Y), \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $h : \Omega \to \bar{\mathbb{R}}$, there is a*

$(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $H : \mathbb{Y} \to \bar{\mathbb{R}}$ such that $h = H \circ Y$. More generally, $\bar{\mathbb{R}}$ may be replaced by any "standard Borel space,"[1] *i.e. a space, $(S, \mathcal{B}_S)$ which is measure theoretic isomorphic to a Borel subset of $\mathbb{R}$.*



Proof. First suppose that $h = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $H = 1_B$. More generally if $h = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $h = H \circ Y$ with $H := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, h, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions h_n converging to h. Let $H_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $h_n = H_n \circ Y$. Then it follows that

$$h = \lim_{n\to\infty} h_n = \limsup_{n\to\infty} h_n = \limsup_{n\to\infty} H_n \circ Y = H \circ Y$$

where $H := \limsup_{n\to\infty} H_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$.

For the last assertion we may assume that $S \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_S = (\mathcal{B}_{\mathbb{R}})_S = \{A \cap S : A \in \mathcal{B}_{\mathbb{R}}\}$. Since $i_S : S \to \mathbb{R}$ is measurable, what we have just proved shows there exists, $H : \mathbb{Y} \to \bar{\mathbb{R}}$ which is $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable such that $h = i_S \circ h = H \circ Y$. The only problems with H is that $H(\mathbb{Y})$ may not be contained in S. To fix this, let

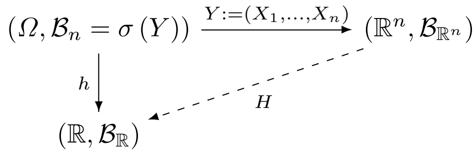
$$H_S = \begin{cases} H|_{H^{-1}(S)} & \text{on } H^{-1}(S) \\ * & \text{on } \mathbb{Y} \setminus H^{-1}(S) \end{cases}$$

where $*$ is some fixed arbitrary point in S. It follows from Proposition 9.13 that $H_S : \mathbb{Y} \to S$ is $(\mathcal{F}, \mathcal{B}_S)$ – measurable and we still have $h = H_S \circ Y$ as the range of Y must necessarily be in $H^{-1}(S)$. ∎

Here is how this lemma will often be used in these notes.

Corollary 9.41. *Suppose that $(\Omega, \mathcal{B})$ is a measurable space, $X_n : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_{\mathbb{R}}$ – measurable functions, and $\mathcal{B}_n := \sigma(X_1, \dots, X_n) \subset \mathcal{B}$ for each $n \in \mathbb{N}$. Then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff there exists $H : \mathbb{R}^n \to \mathbb{R}$ which is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable such that $h = H(X_1, \dots, X_n)$.*

[1] Standard Borel spaces include almost any measurable space that we will consider in these notes. For example they include all complete seperable metric spaces equipped with the Borel σ – algebra, see Section **??**.



Proof. By Lemma 9.25 and Corollary 9.19, the map, $Y := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ is $(\mathcal{B}, \mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}})$ – measurable and by Proposition 9.20, $\mathcal{B}_n = \sigma(X_1, \dots, X_n) = \sigma(Y)$. Thus we may apply Lemma 9.40 to see that there exists a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$, such that $h = H \circ Y = H(X_1, \dots, X_n)$. ■

9.3 Summary of Measurability Statements

It may be worthwhile to gather the statements of the main measurability results of Sections 9.1 and 9.2 in one place. To do this let $(\Omega, \mathcal{B})$, $(X, \mathcal{M})$, and $\{(Y_\alpha, \mathcal{F}_\alpha)\}_{\alpha \in I}$ be measurable spaces and $f_\alpha : \Omega \to Y_\alpha$ be given maps for all $\alpha \in I$. Also let $\pi_\alpha : Y \to Y_\alpha$ be the α – projection map,

$$\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha := \sigma(\pi_\alpha : \alpha \in I)$$

be the product σ – algebra on Y, and $f : \Omega \to Y$ be the unique map determined by $\pi_\alpha \circ f = f_\alpha$ for all $\alpha \in I$. Then the following measurability results hold;

1. For $A \subset \Omega$, the indicator function, 1_A, is $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$ – measurable iff $A \in \mathcal{B}$. (Example 9.8).
2. If $\mathcal{E} \subset \mathcal{M}$ generates $\mathcal{M}$ (i.e. $\mathcal{M} = \sigma(\mathcal{E})$), then a map, $g : \Omega \to X$ is $(\mathcal{B}, \mathcal{M})$ – measurable iff $g^{-1}(\mathcal{E}) \subset \mathcal{B}$ (Lemma 9.3 and Proposition 9.10).
3. The notion of measurability may be localized (Proposition 9.13).
4. Composition of measurable functions are measurable (Lemma 9.14).
5. Continuous functions between two topological spaces are also Borel measurable (Proposition 9.23).
6. $\sigma(f) = \sigma(f_\alpha : \alpha \in I)$ (Proposition 9.20).
7. A map, $h : X \to \Omega$ is $(\mathcal{M}, \sigma(f) = \sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Proposition 9.17).
8. A map, $h : X \to Y$ is $(\mathcal{M}, \mathcal{F})$ – measurable iff $\pi_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Corollary 9.19).
9. If $I = \{1, 2, \dots, n\}$, then
$$\otimes_{\alpha \in I} \mathcal{F}_\alpha = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma(\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{F}_i \text{ for } i \in I\}),$$
this is a special case of Remark 9.21.
10. $\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ (n - times) for all $n \in \mathbb{N}$, i.e. the Borel σ – algebra on $\mathbb{R}^n$ is the same as the product σ – algebra. (Lemma 9.25).
11. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ is closed under the usual pointwise algebraic operations (Corollary 9.32). They are also closed under the countable supremums, infimums, and limits (Proposition 9.33).
12. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ is closed under the usual pointwise algebraic operations and countable limits. (Corollary 9.27 and Proposition 9.33). The limiting assertion follows by considering the real and imaginary parts of all functions involved.
13. The class of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ and from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ may be well approximated by measurable simple functions (Theorem 9.39).

14. If $X_i : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_{\mathbb{R}}$ – measurable maps and $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$, then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff $h = H(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$ (Corollary 9.41).
15. We also have the more general factorization Lemma 9.40.

For the most part most of our future measurability issues can be resolved by one or more of the items on this list.

9.4 Distributions / Laws of Random Vectors

The proof of the following proposition is routine and will be left to the reader.

Proposition 9.42. *Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Define a function $\nu : \mathcal{F} \to [0, \infty]$ by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$. Then ν is a measure on $(Y, \mathcal{F})$. (In the future we will denote ν by $f_*\mu$ or $\mu \circ f^{-1}$ or $\mathrm{Law}_\mu(f)$ and call $f_*\mu$ the **push-forward of** μ **by** f or the **law of** f **under** μ.*

Definition 9.43. *Suppose that $\{X_i\}_{i=1}^n$ is a sequence of random variables on a probability space, $(\Omega, \mathcal{B}, P)$. The probability measure,*

$$\mu = (X_1, \dots, X_n)_* P = P \circ (X_1, \dots, X_n)^{-1} \text{ on } \mathcal{B}_{\mathbb{R}}$$

*(see Proposition 9.42) is called the **joint distribution** (or **law**) of $(X_1, \dots, X_n)$. To be more explicit,*

$$\mu(B) := P((X_1, \dots, X_n) \in B) := P(\{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\})$$

for all $B \in \mathcal{B}_{\mathbb{R}^n}$.

Corollary 9.44. *The joint distribution, μ is uniquely determined from the knowledge of*

$$P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) \text{ for all } A_i \in \mathcal{B}_{\mathbb{R}}$$

or from the knowledge of

$$P(X_1 \le x_1, \dots, X_n \le x_n) \text{ for all } A_i \in \mathcal{B}_{\mathbb{R}}$$

for all $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Proof. Apply Proposition 8.15 with $\mathcal{P}$ being the π – systems defined by

$$\mathcal{P} := \{A_1 \times \dots \times A_n \in \mathcal{B}_{\mathbb{R}^n} : A_i \in \mathcal{B}_{\mathbb{R}}\}$$

for the first case and

$$\mathcal{P} := \{(-\infty, x_1] \times \dots \times (-\infty, x_n] \in \mathcal{B}_{\mathbb{R}^n} : x_i \in \mathbb{R}\}$$

for the second case. ∎

Definition 9.45. *Suppose that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are two finite sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ respectively. We write $(X_1, \dots, X_n) \overset{d}{=} (Y_1, \dots, Y_n)$ if $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ have the **same distribution / law**, i.e. if*

$$P((X_1,\dots,X_n)\in B)=P'((Y_1,\dots,Y_n)\in B) \text{ for all } B\in\mathcal{B}_{\mathbb{R}^n}.$$

More generally, if $\{X_i\}_{i=1}^\infty$ *and* $\{Y_i\}_{i=1}^\infty$ *are two sequences of random variables on two probability spaces,* $(\Omega,\mathcal{B},P)$ *and* $(\Omega',\mathcal{B}',P')$ *we write* $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ *iff* $(X_1,\dots,X_n)\overset{d}{=}(Y_1,\dots,Y_n)$ *for all* $n\in\mathbb{N}$.

Proposition 9.46. *Let us continue using the notation in Definition 9.45. Further let*

$$X=(X_1,X_2,\dots):\Omega\to\mathbb{R}^{\mathbb{N}} \text{ and } Y:=(Y_1,Y_2,\dots):\Omega'\to\mathbb{R}^{\mathbb{N}}$$

and let $\mathcal{F}:=\otimes_{n\in\mathbb{N}}\mathcal{B}_{\mathbb{R}}$ *– be the product* σ *– algebra on* $\mathbb{R}^{\mathbb{N}}$*. Then* $\{X_i\}_{i=1}^\infty\overset{d}{=}\{Y_i\}_{i=1}^\infty$ *iff* $X_*P=Y_*P'$ *as measures on* $(\mathbb{R}^{\mathbb{N}},\mathcal{F})$.

Proof. Let

$$\mathcal{P}:=\cup_{n=1}^\infty\left\{A_1\times A_2\times\cdots\times A_n\times\mathbb{R}^{\mathbb{N}}:A_i\in\mathcal{B}_{\mathbb{R}} \text{ for } 1\le i\le n\right\}.$$

Notice that $\mathcal{P}$ is a π – system and it is easy to show $\sigma(\mathcal{P})=\mathcal{F}$ (see Exercise 9.6). Therefore by Proposition 8.15, $X_*P=Y_*P'$ iff $X_*P=Y_*P'$ on $\mathcal{P}$. Now for $A_1\times A_2\times\cdots\times A_n\times\mathbb{R}^{\mathbb{N}}\in\mathcal{P}$ we have,

$$X_*P\left(A_1\times A_2\times\cdots\times A_n\times\mathbb{R}^{\mathbb{N}}\right)=P((X_1,\dots,X_n)\in A_1\times A_2\times\cdots\times A_n)$$

and hence the condition becomes,

$$P((X_1,\dots,X_n)\in A_1\times A_2\times\cdots\times A_n)=P'((Y_1,\dots,Y_n)\in A_1\times A_2\times\cdots\times A_n)$$

for all $n\in\mathbb{N}$ and $A_i\in\mathcal{B}_{\mathbb{R}}$. Another application of Proposition 8.15 or using Corollary 9.44 allows us to conclude that shows that $X_*P=Y_*P'$ iff $(X_1,\dots,X_n)\overset{d}{=}(Y_1,\dots,Y_n)$ for all $n\in\mathbb{N}$. ∎

Corollary 9.47. *Continue the notation above and assume that* $\{X_i\}_{i=1}^\infty\overset{d}{=}\{Y_i\}_{i=1}^\infty$*. Further let*

$$X_\pm=\begin{cases}\limsup_{n\to\infty}X_n & \text{if } +\\ \liminf_{n\to\infty}X_n & \text{if } -\end{cases}$$

and define $Y_\pm$ *similarly. Then* $(X_-,X_+)\overset{d}{=}(Y_-,Y_+)$ *as random variables into* $\left(\bar{\mathbb{R}}^2,\mathcal{B}_{\bar{\mathbb{R}}}\otimes\mathcal{B}_{\bar{\mathbb{R}}}\right)$*. In particular,*

$$P\left(\lim_{n\to\infty}X_n \text{ exists in } \mathbb{R}\right)=P'\left(\lim_{n\to\infty}Y \text{ exists in } \mathbb{R}\right). \tag{9.8}$$

Proof. First suppose that $(\Omega',\mathcal{B}',P')=\left(\mathbb{R}^{\mathbb{N}},\mathcal{F},P':=X_*P\right)$ where $Y_i(a_1,a_2,\dots):=a_i=\pi_i(a_1,a_2,\dots)$. Then for $C\in\mathcal{B}_{\bar{\mathbb{R}}}\otimes\mathcal{B}_{\bar{\mathbb{R}}}$ we have,

$$X^{-1}(\{(Y_-,Y_+)\in C\})=\{(Y_-\circ X,Y_+\circ X)\in C\}=\{(X_-,X_+)\in C\},$$

since, for example,

$$Y_-\circ X=\liminf_{n\to\infty}Y_n\circ X=\liminf_{n\to\infty}X_n=X_-.$$

Therefore it follows that

$$P((X_-,X_+)\in C)=P\circ X^{-1}(\{(Y_-,Y_+)\in C\})=P'(\{(Y_-,Y_+)\in C\}). \tag{9.9}$$

The general result now follows by two applications of this special case.

For the last assertion, take

$$C=\{(x,x):x\in\mathbb{R}\}\in\mathcal{B}_{\mathbb{R}^2}=\mathcal{B}_{\mathbb{R}}\otimes\mathcal{B}_{\mathbb{R}}\subset\mathcal{B}_{\bar{\mathbb{R}}}\otimes\mathcal{B}_{\bar{\mathbb{R}}}.$$

Then $(X_-,X_+)\in C$ iff $X_-=X_+\in\mathbb{R}$ which happens iff $\lim_{n\to\infty}X_n$ exists in $\mathbb{R}$. Similarly, $(Y_-,Y_+)\in C$ iff $\lim_{n\to\infty}Y_n$ exists in $\mathbb{R}$ and therefore Eq. (9.8) holds as a consequence of Eq. (9.9). ∎

Exercise 9.10. Let $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ be two sequences of random variables such that $\{X_i\}_{i=1}^\infty\overset{\mathrm{d}}{=}\{Y_i\}_{i=1}^\infty$. Let $\{S_n\}_{n=1}^\infty$ and $\{T_n\}_{n=1}^\infty$ be defined by, $S_n:=X_1+\cdots+X_n$ and $T_n:=Y_1+\cdots+Y_n$. Prove the following assertions.

1. Suppose that $f:\mathbb{R}^n\to\mathbb{R}^k$ is a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}^k}$ – measurable function, then $f(X_1,\dots,X_n)\overset{\mathrm{d}}{=}f(Y_1,\dots,Y_n)$.
2. Use your result in item 1. to show $\{S_n\}_{n=1}^\infty\overset{\mathrm{d}}{=}\{T_n\}_{n=1}^\infty$.
   Hint: Apply item 1. with $k=n$ after making a judicious choice for $f:\mathbb{R}^n\to\mathbb{R}^n$.

9.5 Generating All Distributions from the Uniform Distribution

Theorem 9.48. *Given a distribution function,* $F:\mathbb{R}\to[0,1]$ *let* $G:(0,1)\to\mathbb{R}$ *be defined (see Figure 9.2) by,*

$$G(y):=\inf\{x:F(x)\ge y\}.$$

Then $G:(0,1)\to\mathbb{R}$ *is Borel measurable and* $G_*m=\mu_F$ *where* μ_F *is the unique measure on* $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ *such that* $\mu_F((a,b])=F(b)-F(a)$ *for all* $-\infty<a<b<\infty$.

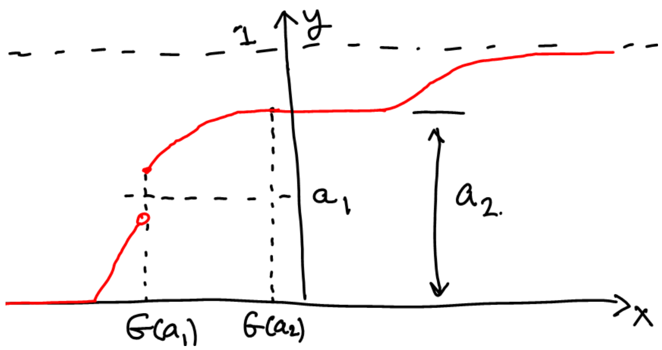


Fig. 9.2. A pictorial definition of G.

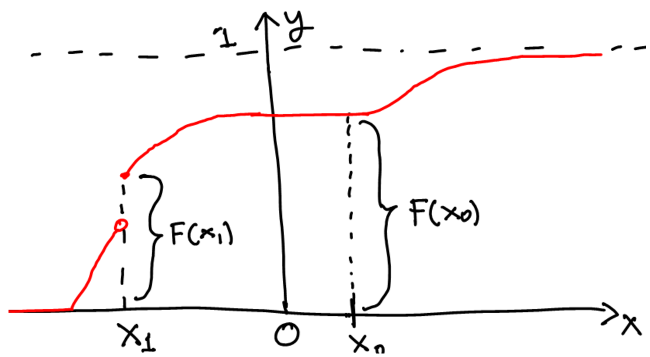


Fig. 9.3. As can be seen from this picture, $G(y) \le x_0$ iff $y \le F(x_0)$ and similarly, $G(y) \le x_1$ iff $y \le x_1$.

Proof. Since $G : (0,1) \to \mathbb{R}$ is a non-decreasing function, G is measurable. We also claim that, for all $x_0 \in \mathbb{R}$, that

$$G^{-1}((0, x_0]) = \{y : G(y) \le x_0\} = (0, F(x_0)] \cap \mathbb{R}, \tag{9.10}$$

see Figure 9.3.

To give a formal proof of Eq. (9.10), $G(y) = \inf\{x : F(x) \ge y\} \le x_0$, there exists $x_n \ge x_0$ with $x_n \downarrow x_0$ such that $F(x_n) \ge y$. By the right continuity of F, it follows that $F(x_0) \ge y$. Thus we have shown

$$\{G \le x_0\} \subset (0, F(x_0)] \cap (0,1).$$

For the converse, if $y \le F(x_0)$ then $G(y) = \inf\{x : F(x) \ge y\} \le x_0$, i.e. $y \in \{G \le x_0\}$. Indeed, $y \in G^{-1}((-\infty, x_0])$ iff $G(y) \le x_0$. Observe that

$$G(F(x_0)) = \inf\{x : F(x) \ge F(x_0)\} \le x_0$$

and hence $G(y) \le x_0$ whenever $y \le F(x_0)$. This shows that

$$(0, F(x_0)] \cap (0,1) \subset G^{-1}((0, x_0]).$$

As a consequence we have $G_* m = \mu_F$. Indeed,

$$\begin{aligned}(G_* m)((-\infty, x]) &= m\left(G^{-1}((-\infty, x])\right) = m(\{y \in (0,1) : G(y) \le x\}) \\ &= m((0, F(x)] \cap (0,1)) = F(x).\end{aligned}$$

See section 2.5.2 on p. 61 of Resnick for more details. ■

Theorem 9.49 (Durret's Version). *Given a distribution function, $F : \mathbb{R} \to [0,1]$ let $Y : (0,1) \to \mathbb{R}$ be defined (see Figure 9.4) by,*

$$Y(x) := \sup\{y : F(y) < x\}.$$

Then $Y : (0,1) \to \mathbb{R}$ is Borel measurable and $Y_* m = \mu_F$ where μ_F is the unique measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.

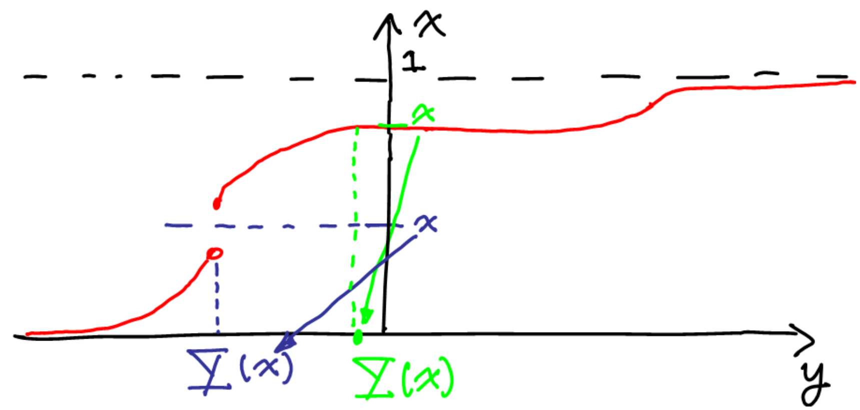


Fig. 9.4. A pictorial definition of $Y(x)$.

Proof. Since $Y : (0,1) \to \mathbb{R}$ is a non-decreasing function, Y is measurable. Also observe, if $y < Y(x)$, then $F(y) < x$ and hence,

$$F(Y(x)-) = \lim_{y \uparrow Y(x)} F(y) \le x.$$

For $y > Y(x)$, we have $F(y) \geq x$ and therefore,

$$F(Y(x)) = F(Y(x)+) = \lim_{y \downarrow Y(x)} F(y) \geq x$$

and so we have shown

$$F(Y(x)-) \leq x \leq F(Y(x)).$$

We will now show

$$\{x \in (0,1) : Y(x) \leq y_0\} = (0, F(y_0)] \cap (0,1). \tag{9.11}$$

For the inclusion "$\subset$," if $x \in (0,1)$ and $Y(x) \leq y_0$, then $x \leq F(Y(x)) \leq F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely if $x \in (0,1)$ and $x \leq F(y_0)$ then (by definition of $Y(x)$) $y_0 \geq Y(x)$.

From the identity in Eq. (9.11), it follows that Y is measurable and

$$(Y_* m)((-\infty, y_0)) = m\left(Y^{-1}(-\infty, y_0)\right) = m((0, F(y_0)] \cap (0,1)) = F(y_0).$$

Therefore, $Law(Y) = \mu_F$ as desired. ■

10

Integration Theory

In this chapter, we will greatly extend the "simple" integral or expectation which was developed in Section 7.3 above. Recall there that if $(\Omega, \mathcal{B}, \mu)$ was measurable space and $\varphi : \Omega \to [0, \infty)$ was a measurable simple function, then we let

$$\mathbb{E}_\mu \varphi := \sum_{\lambda \in [0,\infty)} \lambda \mu\left(\varphi = \lambda\right).$$

The conventions being use here is that $0 \cdot \mu\left(\varphi = 0\right) = 0$ even when $\mu\left(\varphi = 0\right) = \infty$. This convention is necessary in order to make the integral linear – at a minimum we will want $\mathbb{E}_\mu\left[0\right] = 0$. Please be careful not blindly apply the $0 \cdot \infty = 0$ convention in other circumstances.

10.1 Integrals of positive functions

Definition 10.1. *Let* $L^+ = L^+\left(\mathcal{B}\right) = \{f : \Omega \to [0, \infty] : f \text{ is measurable}\}$. *Define*

$$\int_\Omega f\left(\omega\right) d\mu\left(\omega\right) = \int_\Omega f d\mu := \sup\left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le f\right\}.$$

We say the $f \in L^+$ *is* ***integrable*** *if* $\int_\Omega f d\mu < \infty$. *If* $A \in \mathcal{B}$, *let*

$$\int_A f\left(\omega\right) d\mu\left(\omega\right) = \int_A f d\mu := \int_\Omega 1_A f \; d\mu.$$

We also use the notation,

$$\mathbb{E}f = \int_\Omega f d\mu \text{ and } \mathbb{E}\left[f : A\right] := \int_A f d\mu.$$

Remark 10.2. Because of item 3. of Proposition 7.19, if φ is a non-negative simple function, $\int_\Omega \varphi d\mu = \mathbb{E}_\mu \varphi$ so that $\int_\Omega$ is an extension of $\mathbb{E}_\mu$.

Lemma 10.3. *Let* $f, g \in L^+\left(\mathcal{B}\right)$. *Then:*

1. *if* $\lambda \ge 0$, *then*
$$\int_\Omega \lambda f d\mu = \lambda \int_\Omega f d\mu$$
wherein $\lambda \int_\Omega f d\mu \equiv 0$ *if* $\lambda = 0$, *even if* $\int_\Omega f d\mu = \infty$.
2. *if* $0 \le f \le g$, *then*
$$\int_\Omega f d\mu \le \int_\Omega g d\mu. \tag{10.1}$$
3. *For all* $\varepsilon > 0$ *and* $p > 0$,
$$\mu(f \ge \varepsilon) \le \frac{1}{\varepsilon^p} \int_\Omega f^p 1_{\{f \ge \varepsilon\}} d\mu \le \frac{1}{\varepsilon^p} \int_\Omega f^p d\mu. \tag{10.2}$$
The inequality in Eq. (10.2) is called Chebyshev's Inequality for $p = 1$ *and Markov's inequality for* $p = 2$.
4. *If* $\int_\Omega f d\mu < \infty$ *then* $\mu(f = \infty) = 0$ *(i.e.* $f < \infty$ *a.e.) and the set* $\{f > 0\}$ *is* σ *– finite.*

Proof. 1. We may assume $\lambda > 0$ in which case,

$$\begin{aligned}
\int_\Omega \lambda f d\mu &= \sup\left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le \lambda f\right\} \\
&= \sup\left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \lambda^{-1}\varphi \le f\right\} \\
&= \sup\left\{\mathbb{E}_\mu\left[\lambda \psi\right] : \psi \text{ is simple and } \psi \le f\right\} \\
&= \sup\left\{\lambda \mathbb{E}_\mu\left[\psi\right] : \psi \text{ is simple and } \psi \le f\right\} \\
&= \lambda \int_\Omega f d\mu.
\end{aligned}$$

2. Since

$$\left\{\varphi \text{ is simple and } \varphi \le f\right\} \subset \left\{\varphi \text{ is simple and } \varphi \le g\right\},$$

Eq. (10.1) follows from the definition of the integral.

3. Since $1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \frac{1}{\varepsilon} f \le \frac{1}{\varepsilon} f$ we have

$$1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \left(\frac{1}{\varepsilon} f\right)^p \le \left(\frac{1}{\varepsilon} f\right)^p$$

and by monotonicity and the multiplicative property of the integral,

$$\mu(f \ge \varepsilon) = \int_\Omega 1_{\{f \ge \varepsilon\}} d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega 1_{\{f \ge \varepsilon\}} f^p d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega f^p d\mu.$$

4. If $\mu(f=\infty)>0$, then $\varphi_n := n1_{\{f=\infty\}}$ is a simple function such that $\varphi_n \le f$ for all n and hence

$$n\mu(f=\infty)=\mathbb{E}_\mu(\varphi_n)\le\int_\Omega f d\mu$$

for all n. Letting $n\to\infty$ shows $\int_\Omega f d\mu=\infty$. Thus if $\int_\Omega f d\mu<\infty$ then $\mu(f=\infty)=0$.

Moreover,

$$\{f>0\}=\cup_{n=1}^\infty\{f>1/n\}$$

with $\mu(f>1/n)\le n\int_\Omega f d\mu<\infty$ for each n. ∎

Theorem 10.4 (Monotone Convergence Theorem). *Suppose $f_n\in L^+$ is a sequence of functions such that $f_n\uparrow f$ (f is necessarily in L^+) then*

$$\int f_n\uparrow\int f \text{ as } n\to\infty.$$

Proof. Since $f_n\le f_m\le f$, for all $n\le m<\infty$,

$$\int f_n\le\int f_m\le\int f$$

from which if follows $\int f_n$ is increasing in n and

$$\lim_{n\to\infty}\int f_n\le\int f. \tag{10.3}$$

For the opposite inequality, let $\varphi:\Omega\to[0,\infty)$ be a simple function such that $0\le\varphi\le f$, $\alpha\in(0,1)$ and $\Omega_n:=\{f_n\ge\alpha\varphi\}$. Notice that $\Omega_n\uparrow\Omega$ and $f_n\ge\alpha 1_{\Omega_n}\varphi$ and so by definition of $\int f_n$,

$$\int f_n\ge\mathbb{E}_\mu[\alpha 1_{\Omega_n}\varphi]=\alpha\mathbb{E}_\mu[1_{\Omega_n}\varphi]. \tag{10.4}$$

Then using the identity

$$1_{\Omega_n}\varphi=1_{\Omega_n}\sum_{y>0}y1_{\{\varphi=y\}}=\sum_{y>0}y1_{\{\varphi=y\}\cap\Omega_n},$$

and the linearity of $\mathbb{E}_\mu$ we have,

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}_\mu[1_{\Omega_n}\varphi]&=\lim_{n\to\infty}\sum_{y>0}y\cdot\mu(\Omega_n\cap\{\varphi=y\})\\&=\sum_{y>0}y\lim_{n\to\infty}\mu(\Omega_n\cap\{\varphi=y\})\text{ (finite sum)}\\&=\sum_{y>0}y\mu(\{\varphi=y\})=\mathbb{E}_\mu[\varphi],\end{aligned}$$

wherein we have used the continuity of μ under increasing unions for the third equality. This identity allows us to let $n\to\infty$ in Eq. (10.4) to conclude $\lim_{n\to\infty}\int f_n\ge\alpha\mathbb{E}_\mu[\varphi]$ and since $\alpha\in(0,1)$ was arbitrary we may further conclude, $\mathbb{E}_\mu[\varphi]\le\lim_{n\to\infty}\int f_n$. The latter inequality being true for all simple functions φ with $\varphi\le f$ then implies that

$$\int f=\sup_{0\le\varphi\le f}\mathbb{E}_\mu[\varphi]\le\lim_{n\to\infty}\int f_n,$$

which combined with Eq. (10.3) proves the theorem. ∎

Remark 10.5 ("Explicit" Integral Formula). Given $f:\Omega\to[0,\infty]$ measurable, we know from the approximation Theorem 9.39 $\varphi_n\uparrow f$ where

$$\varphi_n:=\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{\{\frac{k}{2^n}<f\le\frac{k+1}{2^n}\}}+2^n1_{\{f>2^n\}}.$$

Therefore by the monotone convergence theorem,

$$\begin{aligned}\int_\Omega f d\mu&=\lim_{n\to\infty}\int_\Omega\varphi_n d\mu\\&=\lim_{n\to\infty}\left[\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}\mu\left(\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right)+2^n\mu(f>2^n)\right].\end{aligned}$$

Corollary 10.6. *If $f_n\in L^+$ is a sequence of functions then*

$$\int\sum_{n=1}^\infty f_n=\sum_{n=1}^\infty\int f_n.$$

In particular, if $\sum_{n=1}^\infty\int f_n<\infty$ then $\sum_{n=1}^\infty f_n<\infty$ a.e.

Proof. First off we show that

$$\int(f_1+f_2)=\int f_1+\int f_2$$

by choosing non-negative simple function φ_n and ψ_n such that $\varphi_n\uparrow f_1$ and $\psi_n\uparrow f_2$. Then $(\varphi_n+\psi_n)$ is simple as well and $(\varphi_n+\psi_n)\uparrow(f_1+f_2)$ so by the monotone convergence theorem,

$$\begin{aligned}\int(f_1+f_2)&=\lim_{n\to\infty}\int(\varphi_n+\psi_n)=\lim_{n\to\infty}\left(\int\varphi_n+\int\psi_n\right)\\&=\lim_{n\to\infty}\int\varphi_n+\lim_{n\to\infty}\int\psi_n=\int f_1+\int f_2.\end{aligned}$$

Now to the general case. Let $g_N := \sum_{n=1}^{N} f_n$ and $g = \sum_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

$$\sum_{n=1}^{\infty} \int f_n := \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n$$
$$= \lim_{N\to\infty} \int g_N = \int g =: \int \sum_{n=1}^{\infty} f_n.$$

■

Remark 10.7. It is in the proof of Corollary 10.6 (i.e. the linearity of the integral) that we really make use of the assumption that all of our functions are measurable. In fact the definition $\int f d\mu$ makes sense for **all** functions $f : \Omega \to [0,\infty]$ not just measurable functions. Moreover the monotone convergence theorem holds in this generality with no change in the proof. However, in the proof of Corollary 10.6, we use the approximation Theorem 9.39 which relies heavily on the measurability of the functions to be approximated.

Example 10.8 (Sums as Integrals I). Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, $\mu(A) = \#(A)$ for $A \subset \Omega$ is the counting measure on $\mathcal{B}$, and $f : \mathbb{N} \to [0,\infty]$ is a function. Since

$$f = \sum_{n=1}^{\infty} f(n)\, 1_{\{n\}},$$

it follows from Corollary 10.6 that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty} \int_{\mathbb{N}} f(n)\, 1_{\{n\}} d\mu = \sum_{n=1}^{\infty} f(n)\, \mu(\{n\}) = \sum_{n=1}^{\infty} f(n).$$

Thus the integral relative to counting measure is simply the infinite sum.

Lemma 10.9 (Sums as Integrals II*). *Let Ω be a set and $\rho : \Omega \to [0,\infty]$ be a function, let $\mu = \sum_{\omega\in\Omega} \rho(\omega)\delta_\omega$ on $\mathcal{B} = 2^{\Omega}$, i.e.*

$$\mu(A) = \sum_{\omega\in A} \rho(\omega).$$

If $f : \Omega \to [0,\infty]$ is a function (which is necessarily measurable), then

$$\int_\Omega f d\mu = \sum_\Omega f\rho.$$

Proof. Suppose that $\varphi : \Omega \to [0,\infty)$ is a simple function, then $\varphi = \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}$ and

$$\sum_\Omega \varphi\rho = \sum_{\omega\in\Omega} \rho(\omega) \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}(\omega) = \sum_{z\in[0,\infty)} z \sum_{\omega\in\Omega} \rho(\omega) 1_{\{\varphi=z\}}(\omega)$$
$$= \sum_{z\in[0,\infty)} z\mu(\{\varphi = z\}) = \int_\Omega \varphi d\mu.$$

So if $\varphi : \Omega \to [0,\infty)$ is a simple function such that $\varphi \le f$, then

$$\int_\Omega \varphi d\mu = \sum_\Omega \varphi\rho \le \sum_\Omega f\rho.$$

Taking the sup over φ in this last equation then shows that

$$\int_\Omega f d\mu \le \sum_\Omega f\rho.$$

For the reverse inequality, let $\Lambda \subset\subset \Omega$ be a finite set and $N \in (0,\infty)$. Set $f^N(\omega) = \min\{N, f(\omega)\}$ and let $\varphi_{N,\Lambda}$ be the simple function given by $\varphi_{N,\Lambda}(\omega) := 1_\Lambda(\omega) f^N(\omega)$. Because $\varphi_{N,\Lambda}(\omega) \le f(\omega)$,

$$\sum_\Lambda f^N \rho = \sum_\Omega \varphi_{N,\Lambda}\rho = \int_\Omega \varphi_{N,\Lambda} d\mu \le \int_\Omega f d\mu.$$

Since $f^N \uparrow f$ as $N \to \infty$, we may let $N \to \infty$ in this last equation to concluded

$$\sum_\Lambda f\rho \le \int_\Omega f d\mu.$$

Since Λ is arbitrary, this implies

$$\sum_\Omega f\rho \le \int_\Omega f d\mu.$$

■

Exercise 10.1. Suppose that $\mu_n : \mathcal{B} \to [0,\infty]$ are measures on $\mathcal{B}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{B}$. Prove that $\mu : \mathcal{B} \to [0,\infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Proposition 10.10. *Suppose that $f \ge 0$ is a measurable function. Then $\int_\Omega f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \ge 0$ are measurable functions such that $f \le g$ a.e. then $\int f d\mu \le \int g d\mu$. In particular if $f = g$ a.e. then $\int f d\mu = \int g d\mu$.*

Proof. If $f = 0$ a.e. and $\varphi \le f$ is a simple function then $\varphi = 0$ a.e. This implies that $\mu(\varphi^{-1}(\{y\})) = 0$ for all $y > 0$ and hence $\int_\Omega \varphi d\mu = 0$ and therefore $\int_\Omega f d\mu = 0$. Conversely, if $\int f d\mu = 0$, then by (Lemma 10.3),

$$\mu(f \ge 1/n) \le n \int f d\mu = 0 \text{ for all } n.$$

Therefore, $\mu(f > 0) \le \sum_{n=1}^\infty \mu(f \ge 1/n) = 0$, i.e. $f = 0$ a.e.

For the second assertion let E be the exceptional set where $f > g$, i.e.

$$E := \{\omega \in \Omega : f(\omega) > g(\omega)\}.$$

By assumption E is a null set and $1_{E^c} f \le 1_{E^c} g$ everywhere. Because $g = 1_{E^c} g + 1_E g$ and $1_E g = 0$ a.e.,

$$\int g d\mu = \int 1_{E^c} g d\mu + \int 1_E g d\mu = \int 1_{E^c} g d\mu$$

and similarly $\int f d\mu = \int 1_{E^c} f d\mu$. Since $1_{E^c} f \le 1_{E^c} g$ everywhere,

$$\int f d\mu = \int 1_{E^c} f d\mu \le \int 1_{E^c} g d\mu = \int g d\mu.$$

■

Corollary 10.11. *Suppose that $\{f_n\}$ is a sequence of non-negative measurable functions and f is a measurable function such that $f_n \uparrow f$ off a null set, then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Let $E \subset \Omega$ be a null set such that $f_n 1_{E^c} \uparrow f 1_{E^c}$ as $n \to \infty$. Then by the monotone convergence theorem and Proposition 10.10,

$$\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.$$

■

Lemma 10.12 (Fatou's Lemma). *If $f_n : \Omega \to [0, \infty]$ is a sequence of measurable functions then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n$$

Proof. Define $g_k := \inf_{n \ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \le f_n$ for all $k \le n$,

$$\int g_k \le \int f_n \text{ for all } n \ge k$$

and therefore

$$\int g_k \le \liminf_{n\to\infty} \int f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\int \liminf_{n\to\infty} f_n = \int \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \int g_k \le \liminf_{n\to\infty} \int f_n.$$

■

The following Corollary and the next lemma are simple applications of Corollary 10.6.

Corollary 10.13. *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{B}$ is a collection of sets such that $\mu(A_i \cap A_j) = 0$ for all $i \ne j$, then*

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n).$$

Proof. Since

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \int_\Omega 1_{\cup_{n=1}^\infty A_n} d\mu \text{ and}$$
$$\sum_{n=1}^\infty \mu(A_n) = \int_\Omega \sum_{n=1}^\infty 1_{A_n} d\mu$$

it suffices to show

$$\sum_{n=1}^\infty 1_{A_n} = 1_{\cup_{n=1}^\infty A_n} \ \mu - \text{a.e.} \tag{10.5}$$

Now $\sum_{n=1}^\infty 1_{A_n} \ge 1_{\cup_{n=1}^\infty A_n}$ and $\sum_{n=1}^\infty 1_{A_n}(\omega) \ne 1_{\cup_{n=1}^\infty A_n}(\omega)$ iff $\omega \in A_i \cap A_j$ for some $i \ne j$, that is

$$\left\{\omega : \sum_{n=1}^\infty 1_{A_n}(\omega) \ne 1_{\cup_{n=1}^\infty A_n}(\omega)\right\} = \cup_{i<j} A_i \cap A_j$$

and the latter set has measure 0 being the countable union of sets of measure zero. This proves Eq. (10.5) and hence the corollary. ■

Lemma 10.14 (The First Borell – Cantelli Lemma). *Let $(\Omega,\mathcal{B},\mu)$ be a measure space, $A_n \in \mathcal{B}$, and set*

$$\{A_n \text{ i.o.}\} = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\text{'s}\} = \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} A_n.$$

If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then $\mu(\{A_n \text{ i.o.}\}) = 0$.

Proof. (First Proof.) Let us first observe that

$$\{A_n \text{ i.o.}\} = \left\{\omega \in \Omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty\right\}.$$

Hence if $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then

$$\infty > \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \int_{\Omega} 1_{A_n}\, d\mu = \int_{\Omega} \sum_{n=1}^{\infty} 1_{A_n}\, d\mu$$

implies that $\sum_{n=1}^{\infty} 1_{A_n}(\omega) < \infty$ for μ-a.e. ω. That is to say $\mu(\{A_n \text{ i.o.}\}) = 0$.

(Second Proof.) Of course we may give a strictly measure theoretic proof of this fact:

$$\begin{aligned}\mu(A_n \text{ i.o.}) &= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} A_n\right)\\ &\leq \lim_{N\to\infty} \sum_{n\geq N} \mu(A_n)\end{aligned}$$

and the last limit is zero since $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. ■

Example 10.15. Suppose that $(\Omega,\mathcal{B},P)$ is a probability space (i.e. $P(\Omega) = 1$) and $X_n : \Omega \to \{0,1\}$ are Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. If $\sum_{n=1}^{\infty} p_n < \infty$, then $P(X_n = 1 \text{ i.o.}) = 0$ and hence $P(X_n = 0 \text{ a.a.}) = 1$. In particular, $P(\lim_{n\to\infty} X_n = 0) = 1$.

10.2 Integrals of Complex Valued Functions

Definition 10.16. *A measurable function $f : \Omega \to \bar{\mathbb{R}}$ is **integrable** if $f_+ := f 1_{\{f\geq 0\}}$ and $f_- = -f\ 1_{\{f\leq 0\}}$ are **integrable**. We write $\mathrm{L}^1(\mu;\mathbb{R})$ for the space of real valued integrable functions. For $f \in \mathrm{L}^1(\mu;\mathbb{R})$, let*

$$\int_{\Omega} f d\mu = \int_{\Omega} f_+ d\mu - \int_{\Omega} f_- d\mu.$$

To shorten notation in this chapter we may simply write $\int f d\mu$ or even $\int f$ for $\int_{\Omega} f d\mu$.

Convention: If $f,g : \Omega \to \bar{\mathbb{R}}$ are two measurable functions, let $f+g$ denote the collection of measurable functions $h : \Omega \to \bar{\mathbb{R}}$ such that $h(\omega) = f(\omega) + g(\omega)$ whenever $f(\omega) + g(\omega)$ is well defined, i.e. is not of the form $\infty - \infty$ or $-\infty + \infty$. We use a similar convention for $f - g$. Notice that if $f,g \in \mathrm{L}^1(\mu;\mathbb{R})$ and $h_1, h_2 \in f + g$, then $h_1 = h_2$ a.e. because $|f| < \infty$ and $|g| < \infty$ a.e.

Notation 10.17 (Abuse of notation) *We will sometimes denote the integral $\int_{\Omega} f d\mu$ by $\mu(f)$. With this notation we have $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{B}$.*

Remark 10.18. Since

$$f_{\pm} \leq |f| \leq f_+ + f_-,$$

a measurable function f is **integrable** iff $\int |f|\ d\mu < \infty$. Hence

$$\mathrm{L}^1(\mu;\mathbb{R}) := \left\{f : \Omega \to \bar{\mathbb{R}} :\ f \text{ is measurable and } \int_{\Omega} |f|\ d\mu < \infty\right\}.$$

If $f,g \in \mathrm{L}^1(\mu;\mathbb{R})$ and $f = g$ a.e. then $f_{\pm} = g_{\pm}$ a.e. and so it follows from Proposition 10.10 that $\int f d\mu = \int g d\mu$. In particular if $f,g \in \mathrm{L}^1(\mu;\mathbb{R})$ we may define

$$\int_{\Omega} (f+g)\, d\mu = \int_{\Omega} h d\mu$$

where h is any element of $f+g$.

Proposition 10.19. *The map*

$$f \in \mathrm{L}^1(\mu;\mathbb{R}) \to \int_{\Omega} f d\mu \in \mathbb{R}$$

is linear and has the monotonicity property: $\int f d\mu \leq \int g d\mu$ for all $f,g \in \mathrm{L}^1(\mu;\mathbb{R})$ such that $f \leq g$ a.e.

Proof. Let $f,g \in \mathrm{L}^1(\mu;\mathbb{R})$ and $a,b \in \mathbb{R}$. By modifying f and g on a null set, we may assume that f,g are real valued functions. We have $af + bg \in \mathrm{L}^1(\mu;\mathbb{R})$ because

$$|af + bg| \leq |a|\,|f| + |b|\,|g| \in \mathrm{L}^1(\mu;\mathbb{R}).$$

If $a < 0$, then

$$(af)_+ = -af_- \text{ and } (af)_- = -af_+$$

so that

$$\int af = -a\int f_- + a\int f_+ = a(\int f_+ - \int f_-) = a\int f.$$

A similar calculation works for $a > 0$ and the case $a = 0$ is trivial so we have shown that

$$\int af = a\int f.$$

Now set $h = f + g$. Since $h = h_+ - h_-$,

$$h_+ - h_- = f_+ - f_- + g_+ - g_-$$

or

$$h_+ + f_- + g_- = h_- + f_+ + g_+.$$

Therefore,

$$\int h_+ + \int f_- + \int g_- = \int h_- + \int f_+ + \int g_+$$

and hence

$$\int h = \int h_+ - \int h_- = \int f_+ + \int g_+ - \int f_- - \int g_- = \int f + \int g.$$

Finally if $f_+ - f_- = f \le g = g_+ - g_-$ then $f_+ + g_- \le g_+ + f_-$ which implies that

$$\int f_+ + \int g_- \le \int g_+ + \int f_-$$

or equivalently that

$$\int f = \int f_+ - \int f_- \le \int g_+ - \int g_- = \int g.$$

The monotonicity property is also a consequence of the linearity of the integral, the fact that $f \le g$ a.e. implies $0 \le g - f$ a.e. and Proposition 10.10. ■

Definition 10.20. *A measurable function $f : \Omega \to \mathbb{C}$ is **integrable** if $\int_\Omega |f|\ d\mu < \infty$. Analogously to the real case, let*

$$\mathrm{L}^1(\mu;\mathbb{C}) := \left\{ f : \Omega \to \mathbb{C} : \ f \text{ is measurable and } \int_\Omega |f|\ d\mu < \infty \right\}.$$

denote the complex valued integrable functions. Because, $\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|) \le |f| \le \sqrt{2}\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|)$, $\int |f|\ d\mu < \infty$ *iff*

$$\int |\mathrm{Re}\, f|\, d\mu + \int |\mathrm{Im}\, f|\, d\mu < \infty.$$

For $f \in \mathrm{L}^1(\mu;\mathbb{C})$ define

$$\int f\ d\mu = \int \mathrm{Re}\, f\ d\mu + i\int \mathrm{Im}\, f\ d\mu.$$

It is routine to show the integral is still linear on $\mathrm{L}^1(\mu;\mathbb{C})$ (prove!). In the remainder of this section, let $\mathrm{L}^1(\mu)$ be either $\mathrm{L}^1(\mu;\mathbb{C})$ or $\mathrm{L}^1(\mu;\mathbb{R})$. If $A \in \mathcal{B}$ and $f \in \mathrm{L}^1(\mu;\mathbb{C})$ or $f : \Omega \to [0,\infty]$ is a measurable function, let

$$\int_A f d\mu := \int_\Omega 1_A f d\mu.$$

Proposition 10.21. *Suppose that $f \in \mathrm{L}^1(\mu;\mathbb{C})$, then*

$$\left|\int_\Omega f d\mu\right| \le \int_\Omega |f|\, d\mu. \tag{10.6}$$

Proof. Start by writing $\int_\Omega f\ d\mu = Re^{i\theta}$ with $R \ge 0$. We may assume that $R = \left|\int_\Omega f d\mu\right| > 0$ since otherwise there is nothing to prove. Since

$$R = e^{-i\theta}\int_\Omega f\ d\mu = \int_\Omega e^{-i\theta} f\ d\mu = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu + i\int_\Omega \mathrm{Im}\left(e^{-i\theta} f\right) d\mu,$$

it must be that $\int_\Omega \mathrm{Im}\left[e^{-i\theta} f\right] d\mu = 0$. Using the monotonicity in Proposition 10.10,

$$\left|\int_\Omega f d\mu\right| = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_\Omega \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_\Omega |f|\, d\mu.$$

■

Proposition 10.22. *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \ne 0\}$ is σ – finite, in fact $\{|f| \ge \frac{1}{n}\} \uparrow \{f \ne 0\}$ and $\mu(|f| \ge \frac{1}{n}) < \infty$ for all n.*
2. *The following are equivalent*
   a) $\int_E f = \int_E g$ for all $E \in \mathcal{B}$
   b) $\int_\Omega |f - g| = 0$
   c) $f = g$ a.e.

Proof. 1. By Chebyshev's inequality, Lemma 10.3,

$$\mu(|f| \ge \frac{1}{n}) \le n\int_\Omega |f|\, d\mu < \infty$$

for all n.

2. (a) $\Longrightarrow$ (c) Notice that

$$\int_E f = \int_E g \Leftrightarrow \int_E (f - g) = 0$$

for all $E \in \mathcal{B}$. Taking $E = \{\operatorname{Re}(f-g) > 0\}$ and using $1_E \operatorname{Re}(f-g) \geq 0$, we learn that

$$0 = \operatorname{Re} \int_E (f-g) d\mu = \int 1_E \operatorname{Re}(f-g) \Longrightarrow 1_E \operatorname{Re}(f-g) = 0 \text{ a.e.}$$

This implies that $1_E = 0$ a.e. which happens iff

$$\mu\left(\{\operatorname{Re}(f-g) > 0\}\right) = \mu(E) = 0.$$

Similar $\mu(\operatorname{Re}(f-g) < 0) = 0$ so that $\operatorname{Re}(f-g) = 0$ a.e. Similarly, $\operatorname{Im}(f-g) = 0$ a.e and hence $f - g = 0$ a.e., i.e. $f = g$ a.e.

(c) $\Longrightarrow$ (b) is clear and so is (b) $\Longrightarrow$ (a) since

$$\left| \int_E f - \int_E g \right| \leq \int |f-g| = 0.$$

■

Lemma 10.23 (Integral Comparison I). *Suppose that $h \in L^1(\mu)$ satisfies*

$$\int_A h d\mu \geq 0 \text{ for all } A \in \mathcal{B}, \tag{10.7}$$

then $h \geq 0$ a.e.

Proof. Since by assumption,

$$0 = \operatorname{Im} \int_A h d\mu = \int_A \operatorname{Im} h d\mu \text{ for all } A \in \mathcal{B},$$

we may apply Proposition 10.22 to conclude that $\operatorname{Im} h = 0$ a.e. Thus we may now assume that h is real valued. Taking $A = \{h < 0\}$ in Eq. (10.7) implies

$$\int_\Omega 1_A |h| \, d\mu = \int_\Omega -1_A h d\mu = -\int_A h d\mu \leq 0.$$

However $1_A |h| \geq 0$ and therefore it follows that $\int_\Omega 1_A |h| \, d\mu = 0$ and so Proposition 10.22 implies $1_A |h| = 0$ a.e. which then implies $0 = \mu(A) = \mu(h < 0) = 0$. ■

Lemma 10.24 (Integral Comparison II). *Suppose $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space (i.e. there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$ for all n) and $f, g : \Omega \to [0, \infty]$ are $\mathcal{B}$ – measurable functions. Then $f \geq g$ a.e. iff*

$$\int_A f d\mu \geq \int_A g d\mu \text{ for all } A \in \mathcal{B}. \tag{10.8}$$

In particular $f = g$ a.e. iff equality holds in Eq. (10.8).

Proof. It was already shown in Proposition 10.10 that $f \geq g$ a.e. implies Eq. (10.8). For the converse assertion, let $B_n := \{f \leq n 1_{\Omega_n}\}$. Then from Eq. (10.8),

$$\infty > n\mu(\Omega_n) \geq \int f 1_{B_n} d\mu \geq \int g 1_{B_n} d\mu$$

from which it follows that both $f1_{B_n}$ and $g1_{B_n}$ are in $L^1(\mu)$ and hence $h := f1_{B_n} - g1_{B_n} \in L^1(\mu)$. Using Eq. (10.8) again we know that

$$\int_A h = \int f 1_{B_n \cap A} - \int g 1_{B_n \cap A} \geq 0 \text{ for all } A \in \mathcal{B}.$$

An application of Lemma 10.23 implies $h \geq 0$ a.e., i.e. $f1_{B_n} \geq g1_{B_n}$ a.e. Since $B_n \uparrow \{f < \infty\}$, we may conclude that

$$f 1_{\{f<\infty\}} = \lim_{n\to\infty} f 1_{B_n} \geq \lim_{n\to\infty} g 1_{B_n} = g 1_{\{f<\infty\}} \text{ a.e.}$$

Since $f \geq g$ whenever $f = \infty$, we have shown $f \geq g$ a.e.

If equality holds in Eq. (10.8), then we know that $g \leq f$ and $f \leq g$ a.e., i.e. $f = g$ a.e. ■

Notice that we can not drop the σ – finiteness assumption in Lemma 10.24. For example, let μ be the measure on $\mathcal{B}$ such that $\mu(A) = \infty$ when $A \neq \emptyset$, $g = 3$, and $f = 2$. Then equality holds (both sides are infinite unless $A = \emptyset$ when they are both zero) in Eq. (10.8) holds even though $f < g$ **everywhere.**

Definition 10.25. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $L^1(\mu) = L^1(\Omega, \mathcal{B}, \mu)$ denote the set of $\mathrm{L}^1(\mu)$ functions modulo the equivalence relation; $f \sim g$ iff $f = g$ a.e. We make this into a normed space using the norm*

$$\|f - g\|_{L^1} = \int |f - g| \, d\mu$$

and into a metric space using $\rho_1(f, g) = \|f - g\|_{L^1}$.

Warning: in the future we will often not make much of a distinction between $L^1(\mu)$ and $\mathrm{L}^1(\mu)$. On occasion this can be dangerous and this danger will be pointed out when necessary.

Remark 10.26. More generally we may define $L^p(\mu) = L^p(\Omega, \mathcal{B}, \mu)$ for $p \in [1, \infty)$ as the set of measurable functions f such that

$$\int_\Omega |f|^p \, d\mu < \infty$$

modulo the equivalence relation; $f \sim g$ iff $f = g$ a.e.

We will see in later that

$$\|f\|_{L^p} = \left(\int |f|^p \, d\mu\right)^{1/p} \quad \text{for } f \in L^p(\mu)$$

is a norm and $(L^p(\mu), \|\cdot\|_{L^p})$ is a Banach space in this norm and in particular,

$$\|f+g\|_p \le \|f\|_p + \|g\|_p \text{ for all } f,g \in L^p(\mu).$$

Theorem 10.27 (Dominated Convergence Theorem). *Suppose $f_n, g_n, g \in \mathrm{L}^1(\mu)$, $f_n \to f$ a.e., $|f_n| \le g_n \in \mathrm{L}^1(\mu)$, $g_n \to g$ a.e. and $\int_\Omega g_n d\mu \to \int_\Omega g d\mu$. Then $f \in \mathrm{L}^1(\mu)$ and*

$$\int_\Omega f d\mu = \lim_{h\to\infty} \int_\Omega f_n d\mu.$$

(In most typical applications of this theorem $g_n = g \in \mathrm{L}^1(\mu)$ for all n.)

Proof. Notice that $|f| = \lim_{n\to\infty} |f_n| \le \lim_{n\to\infty} |g_n| \le g$ a.e. so that $f \in \mathrm{L}^1(\mu)$. By considering the real and imaginary parts of f separately, it suffices to prove the theorem in the case where f is real. By Fatou's Lemma,

$$\begin{aligned}\int_\Omega (g \pm f) d\mu &= \int_\Omega \liminf_{n\to\infty} (g_n \pm f_n)\, d\mu \le \liminf_{n\to\infty} \int_\Omega (g_n \pm f_n)\, d\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu\right) \\ &= \int_\Omega g d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu\right)\end{aligned}$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\int_\Omega g d\mu \pm \int_\Omega f d\mu \le \int_\Omega g d\mu + \begin{cases} \liminf_{n\to\infty} \int_\Omega f_n d\mu \\ -\limsup_{n\to\infty} \int_\Omega f_n d\mu \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \le \int_\Omega f d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

This shows that $\lim_{n\to\infty} \int_\Omega f_n d\mu$ exists and is equal to $\int_\Omega f d\mu$. ∎

Exercise 10.2. Give another proof of Proposition 10.21 by first proving Eq. (10.6) with f being a simple function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 9.39 along with the dominated convergence Theorem 10.27 to handle the general case.

Corollary 10.28. *Let $\{f_n\}_{n=1}^\infty \subset \mathrm{L}^1(\mu)$ be a sequence such that $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$, then $\sum_{n=1}^\infty f_n$ is convergent a.e. and*

$$\int_\Omega \left(\sum_{n=1}^\infty f_n\right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

Proof. The condition $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$ is equivalent to $\sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu)$. Hence $\sum_{n=1}^\infty f_n$ is almost everywhere convergent and if $S_N := \sum_{n=1}^N f_n$, then

$$|S_N| \le \sum_{n=1}^N |f_n| \le \sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu).$$

So by the dominated convergence theorem,

$$\begin{aligned}\int_\Omega \left(\sum_{n=1}^\infty f_n\right) d\mu &= \int_\Omega \lim_{N\to\infty} S_N d\mu = \lim_{N\to\infty} \int_\Omega S_N d\mu \\ &= \lim_{N\to\infty} \sum_{n=1}^N \int_\Omega f_n d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.\end{aligned}$$

∎

Example 10.29 (Sums as integrals). Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, μ is counting measure on $\mathcal{B}$ (see Example 10.8), and $f : \mathbb{N} \to \mathbb{C}$ is a function. From Example 10.8 we have $f \in L^1(\mu)$ iff $\sum_{n=1}^\infty |f(n)| < \infty$, i.e. iff the sum, $\sum_{n=1}^\infty f(n)$ is absolutely convergent. Moreover, if $f \in L^1(\mu)$, we may again write

$$f = \sum_{n=1}^\infty f(n)\, 1_{\{n\}}$$

and then use Corollary 10.28 to conclude that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^\infty \int_{\mathbb{N}} f(n)\, 1_{\{n\}} d\mu = \sum_{n=1}^\infty f(n)\, \mu(\{n\}) = \sum_{n=1}^\infty f(n).$$

So again the integral relative to counting measure is simply the infinite sum **provided** the sum is absolutely convergent.

However if $f(n) = (-1)^n \frac{1}{n}$, then

$$\sum_{n=1}^\infty f(n) := \lim_{N\to\infty} \sum_{n=1}^N f(n)$$

is perfectly well defined while $\int_{\mathbb{N}} f d\mu$ is **not**. In fact in this case we have,

$$\int_{\mathbb{N}} f_{\pm} d\mu = \infty.$$

The point is that when we write $\sum_{n=1}^{\infty} f(n)$ the ordering of the terms in the sum may matter. On the other hand, $\int_{\mathbb{N}} f d\mu$ knows nothing about the integer ordering.

The following corollary will be routinely be used in the sequel – often without explicit mention.

Corollary 10.30 (Differentiation Under the Integral). *Suppose that $J \subset \mathbb{R}$ is an open interval and $f : J \times \Omega \to \mathbb{C}$ is a function such that*

1. *$\omega \to f(t, \omega)$ is measurable for each $t \in J$.*
2. *$f(t_0, \cdot) \in L^1(\mu)$ for some $t_0 \in J$.*
3. *$\frac{\partial f}{\partial t}(t, \omega)$ exists for all (t, ω).*
4. *There is a function $g \in \mathrm{L}^1(\mu)$ such that $\left|\frac{\partial f}{\partial t}(t, \cdot)\right| \leq g$ for each $t \in J$.*

Then $f(t, \cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$ (i.e. $\int_\Omega |f(t, \omega)| d\mu(\omega) < \infty$), $t \to \int_\Omega f(t, \omega) d\mu(\omega)$ is a differentiable function on J, and

$$\frac{d}{dt} \int_\Omega f(t, \omega) d\mu(\omega) = \int_\Omega \frac{\partial f}{\partial t}(t, \omega) d\mu(\omega).$$

Proof. By considering the real and imaginary parts of f separately, we may assume that f is real. Also notice that

$$\frac{\partial f}{\partial t}(t, \omega) = \lim_{n \to \infty} n(f(t + n^{-1}, \omega) - f(t, \omega))$$

and therefore, for $\omega \to \frac{\partial f}{\partial t}(t, \omega)$ is a sequential limit of measurable functions and hence is measurable for all $t \in J$. By the mean value theorem,

$$|f(t, \omega) - f(t_0, \omega)| \leq g(\omega) |t - t_0| \text{ for all } t \in J \tag{10.9}$$

and hence

$$|f(t, \omega)| \leq |f(t, \omega) - f(t_0, \omega)| + |f(t_0, \omega)| \leq g(\omega) |t - t_0| + |f(t_0, \omega)|.$$

This shows $f(t, \cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$. Let $G(t) := \int_\Omega f(t, \omega) d\mu(\omega)$, then

$$\frac{G(t) - G(t_0)}{t - t_0} = \int_\Omega \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} d\mu(\omega).$$

By assumption,

$$\lim_{t \to t_0} \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} = \frac{\partial f}{\partial t}(t, \omega) \text{ for all } \omega \in \Omega$$

and by Eq. (10.9),

$$\left|\frac{f(t, \omega) - f(t_0, \omega)}{t - t_0}\right| \leq g(\omega) \text{ for all } t \in J \text{ and } \omega \in \Omega.$$

Therefore, we may apply the dominated convergence theorem to conclude

$$\begin{aligned}
\lim_{n \to \infty} \frac{G(t_n) - G(t_0)}{t_n - t_0} &= \lim_{n \to \infty} \int_\Omega \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0} d\mu(\omega) \\
&= \int_\Omega \lim_{n \to \infty} \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0} d\mu(\omega) \\
&= \int_\Omega \frac{\partial f}{\partial t}(t_0, \omega) d\mu(\omega)
\end{aligned}$$

for **all** sequences $t_n \in J \setminus \{t_0\}$ such that $t_n \to t_0$. Therefore, $\dot{G}(t_0) = \lim_{t \to t_0} \frac{G(t) - G(t_0)}{t - t_0}$ exists and

$$\dot{G}(t_0) = \int_\Omega \frac{\partial f}{\partial t}(t_0, \omega) d\mu(\omega).$$

■

Corollary 10.31. *Suppose that $\{a_n\}_{n=0}^\infty \subset \mathbb{C}$ is a sequence of complex numbers such that series*

$$f(z) := \sum_{n=0}^\infty a_n (z - z_0)^n$$

is convergent for $|z - z_0| < R$, where R is some positive number. Then $f : D(z_0, R) \to \mathbb{C}$ is complex differentiable on $D(z_0, R)$ and

$$f'(z) = \sum_{n=0}^\infty n a_n (z - z_0)^{n-1} = \sum_{n=1}^\infty n a_n (z - z_0)^{n-1}. \tag{10.10}$$

By induction it follows that $f^{(k)}$ exists for all k and that

$$f^{(k)}(z) = \sum_{n=0}^\infty n(n-1) \dots (n-k+1) a_n (z - z_0)^{n-k}.$$

Proof. Let $\rho < R$ be given and choose $r \in (\rho, R)$. Since $z = z_0 + r \in D(z_0, R)$, by assumption the series $\sum_{n=0}^\infty a_n r^n$ is convergent and in particular

$M := \sup_n |a_n r^n| < \infty$. We now apply Corollary 10.30 with $X = \mathbb{N}\cup\{0\}$, μ being counting measure, $\Omega = D(z_0,\rho)$ and $g(z,n) := a_n(z-z_0)^n$. Since

$$\begin{aligned}|g'(z,n)| &= |na_n(z-z_0)^{n-1}| \le n\,|a_n|\,\rho^{n-1} \\ &\le \frac{1}{r}n\left(\frac{\rho}{r}\right)^{n-1}|a_n|\,r^n \le \frac{1}{r}n\left(\frac{\rho}{r}\right)^{n-1}M\end{aligned}$$

and the function $G(n) := \frac{M}{r}n\left(\frac{\rho}{r}\right)^{n-1}$ is summable (by the Ratio test for example), we may use G as our dominating function. It then follows from Corollary 10.30

$$f(z) = \int_X g(z,n)d\mu(n) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$$

is complex differentiable with the differential given as in Eq. (10.10). ■

Definition 10.32 (Moment Generating Function). *Let $(\Omega,\mathcal{B},P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. The* ***moment generating function*** *of X is $M_X : \mathbb{R} \to [0,\infty]$ defined by*

$$M_X(t) := \mathbb{E}\left[e^{tX}\right].$$

Proposition 10.33. *Suppose there exists $\varepsilon > 0$ such that $\mathbb{E}\left[e^{\varepsilon|X|}\right] < \infty$, then $M_X(t)$ is a smooth function of $t \in (-\varepsilon,\varepsilon)$ and*

$$M_X(t) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}X^n \text{ if } |t| \le \varepsilon. \tag{10.11}$$

In particular,

$$\mathbb{E}X^n = \left(\frac{d}{dt}\right)^n|_{t=0}M_X(t) \text{ for all } n \in \mathbb{N}_0. \tag{10.12}$$

Proof. If $|t| \le \varepsilon$, then

$$\mathbb{E}\left[\sum_{n=0}^{\infty}\frac{|t|^n}{n!}|X|^n\right] \le \mathbb{E}\left[\sum_{n=0}^{\infty}\frac{\varepsilon^n}{n!}|X|^n\right] = \mathbb{E}\left[e^{\varepsilon|X|}\right] < \infty.$$

it $e^{tX} \le e^{\varepsilon|X|}$ for all $|t| \le \varepsilon$. Hence it follows from Corollary 10.28 that, for $|t| \le \varepsilon$,

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty}\frac{t^n}{n!}X^n\right] = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}X^n.$$

Equation (10.12) now is a consequence of Corollary 10.31. ■

Exercise 10.3. Let $d \in \mathbb{N}$, $\Omega = \mathbb{N}_0^d$, $\mathcal{B} = 2^{\Omega}$, $\mu : \mathcal{B} \to \mathbb{N}_0 \cup \{\infty\}$ be counting measure on Ω, and for $x \in \mathbb{R}^d$ and $\omega \in \Omega$, let $x^{\omega} := x_1^{\omega_1}\dots x_n^{\omega_n}$. Further suppose that $f : \Omega \to \mathbb{C}$ is function and $r_i > 0$ for $1 \le i \le d$ such that

$$\sum_{\omega\in\Omega}|f(\omega)|\,r^{\omega} = \int_{\Omega}|f(\omega)|\,r^{\omega}d\mu(\omega) < \infty,$$

where $r := (r_1,\dots,r_d)$. Show;

1. There is a constant, $C < \infty$ such that $|f(\omega)| \le \frac{C}{r^{\omega}}$ for all $\omega \in \Omega$.
2. Let
$$U := \left\{x \in \mathbb{R}^d : |x_i| < r_i \ \forall\ i\right\} \text{ and } \bar{U} = \left\{x \in \mathbb{R}^d : |x_i| \le r_i \ \forall\ i\right\}$$
Show $\sum_{\omega\in\Omega}|f(\omega)\,x^{\omega}| < \infty$ for all $x \in \bar{U}$ and the function, $F : U \to \mathbb{R}$ defined by
$$F(x) = \sum_{\omega\in\Omega} f(\omega)\,x^{\omega} \text{ is continuous on } \bar{U}.$$
3. Show, for all $x \in U$ and $1 \le i \le d$, that
$$\frac{\partial}{\partial x_i}F(x) = \sum_{\omega\in\Omega}\omega_i f(\omega)\,x^{\omega-e_i}$$
where $e_i = (0,\dots,0,1,0,\dots,0)$ is the i^{th} – standard basis vector on $\mathbb{R}^d$.
4. For any $\alpha \in \Omega$, let $\partial^{\alpha} := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\dots\left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$ and $\alpha! := \prod_{i=1}^d \alpha_i!$ Explain why we may now conclude that
$$\partial^{\alpha}F(x) = \sum_{\omega\in\Omega}\alpha! f(\omega)\,x^{\omega-\alpha} \text{ for all } x \in U. \tag{10.13}$$
5. Conclude that $f(\alpha) = \frac{(\partial\alpha F)(0)}{\alpha!}$ for all $\alpha \in \Omega$.
6. If $g : \Omega \to \mathbb{C}$ is another function such that $\sum_{\omega\in\Omega} g(\omega)\,x^{\omega} = \sum_{\omega\in\Omega} f(\omega)\,x^{\omega}$ for x in a neighborhood of $0 \in \mathbb{R}^d$, then $g(\omega) = f(\omega)$ for all $\omega \in \Omega$.

10.2.1 Square Integrable Random Variables and Correlations

Suppose that $(\Omega,\mathcal{B},P)$ is a probability space. We say that $X : \Omega \to \mathbb{R}$ is **integrable** if $X \in L^1(P)$ and **square integrable** if $X \in L^2(P)$. When X is integrable we let $a_X := \mathbb{E}X$ be the **mean** of X.

Now suppose that $X, Y : \Omega \to \mathbb{R}$ are two square integrable random variables. Since

$$0 \le |X-Y|^2 = |X|^2 + |Y|^2 - 2\,|X|\,|Y|,$$

it follows that

$$|XY| \le \frac{1}{2}|X|^2 + \frac{1}{2}|Y|^2 \in L^1(P).$$

In particular by taking $Y = 1$, we learn that $|X| \le \frac{1}{2}\left(1 + |X^2|\right)$ which shows that every square integrable random variable is also integrable.

Definition 10.34. *The* ***covariance,*** $\mathrm{Cov}(X,Y)$, *of two square integrable random variables, X and Y, is defined by*

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - a_X)(Y - a_Y)\right] = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

where $a_X := \mathbb{E}X$ and $a_Y := \mathbb{E}Y$. The ***variance*** *of X,*

$$\mathrm{Var}(X) := \mathrm{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \tag{10.14}$$

We say that X and Y are ***uncorrelated*** *if $\mathrm{Cov}(X,Y) = 0$, i.e. $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$. More generally we say $\{X_k\}_{k=1}^n \subset L^2(P)$ are* ***uncorrelated*** *iff $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \ne j$.*

It follows from Eq. (10.14) that

$$\mathrm{Var}(X) \le \mathbb{E}\left[X^2\right] \text{ for all } X \in L^2(P). \tag{10.15}$$

Lemma 10.35. *The covariance function, $\mathrm{Cov}(X,Y)$ is bilinear in X and Y and $\mathrm{Cov}(X,Y) = 0$ if either X or Y is constant. For any constant k, $\mathrm{Var}(X+k) = \mathrm{Var}(X)$ and $\mathrm{Var}(kX) = k^2\,\mathrm{Var}(X)$. If $\{X_k\}_{k=1}^n$ are uncorrelated $L^2(P)$ – random variables, then*

$$\mathrm{Var}(S_n) = \sum_{k=1}^n \mathrm{Var}(X_k).$$

Proof. We leave most of this simple proof to the reader. As an example of the type of argument involved, let us prove $\mathrm{Var}(X+k) = \mathrm{Var}(X)$;

$$\begin{aligned}\mathrm{Var}(X+k) &= \mathrm{Cov}(X+k, X+k) = \mathrm{Cov}(X+k, X) + \mathrm{Cov}(X+k, k) \\ &= \mathrm{Cov}(X+k, X) = \mathrm{Cov}(X,X) + \mathrm{Cov}(k, X) \\ &= \mathrm{Cov}(X,X) = \mathrm{Var}(X),\end{aligned}$$

wherein we have used the bilinearity of $\mathrm{Cov}(\cdot,\cdot)$ and the property that $\mathrm{Cov}(Y,k) = 0$ whenever k is a constant. ■

Exercise 10.4 (A Weak Law of Large Numbers). Assume $\{X_n\}_{n=1}^\infty$ is a sequence if uncorrelated square integrable random variables which are identically distributed, i.e. $X_n \overset{d}{=} X_m$ for all $m, n \in \mathbb{N}$. Let $S_n := \sum_{k=1}^n X_k$, $\mu := \mathbb{E}X_k$ and $\sigma^2 := \mathrm{Var}(X_k)$ (these are independent of k). Show;

$$\begin{aligned}\mathbb{E}\left[\frac{S_n}{n}\right] &= \mu, \\ \mathbb{E}\left(\frac{S_n}{n} - \mu\right)^2 &= \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}, \text{ and} \\ P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) &\le \frac{\sigma^2}{n\varepsilon^2}\end{aligned}$$

for all $\varepsilon > 0$ and $n \in \mathbb{N}$. (Compare this with Exercise 7.13.)

10.2.2 Some Discrete Distributions

Definition 10.36 (Generating Function). *Suppose that $N : \Omega \to \mathbb{N}_0$ is an integer valued random variable on a probability space, $(\Omega, \mathcal{B}, P)$. The generating function associated to N is defined by*

$$G_N(z) := \mathbb{E}\left[z^N\right] = \sum_{n=0}^\infty P(N = n)\, z^n \text{ for } |z| \le 1. \tag{10.16}$$

By Corollary 10.31, it follows that $P(N = n) = \frac{1}{n!} G_N^{(n)}(0)$ so that G_N can be used to completely recover the distribution of N.

Proposition 10.37 (Generating Functions). *The generating function satisfies,*

$$G_N^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\, z^{N-k}\right] \text{ for } |z| < 1$$

and

$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\right],$$

where it is possible that one and hence both sides of this equation are infinite. In particular, $G'(1) := \lim_{z\uparrow 1} G'(z) = \mathbb{E}N$ and if $\mathbb{E}N^2 < \infty$,

$$\mathrm{Var}(N) = G''(1) + G'(1) - \left[G'(1)\right]^2. \tag{10.17}$$

Proof. By Corollary 10.31 for $|z| < 1$,

$$\begin{aligned}G_N^{(k)}(z) &= \sum_{n=0}^\infty P(N = n) \cdot n(n-1)\dots(n-k+1)\, z^{n-k} \\ &= \mathbb{E}\left[N(N-1)\dots(N-k+1)\, z^{N-k}\right].\end{aligned} \tag{10.18}$$

Since, for $z \in (0,1)$,

$$0 \le N(N-1)\dots(N-k+1)\, z^{N-k} \uparrow N(N-1)\dots(N-k+1) \text{ as } z \uparrow 1,$$

we may apply the MCT to pass to the limit as $z \uparrow 1$ in Eq. (10.18) to find,

$$G^{(k)}(1) = \lim_{z \uparrow 1} G^{(k)}(z) = \mathbb{E}[N(N-1)\dots(N-k+1)].$$

■

Exercise 10.5 (Some Discrete Distributions). Let $p \in (0,1]$ and $\lambda > 0$. In the four parts below, the distribution of N will be described. You should work out the generating function, $G_N(z)$, in each case and use it to verify the given formulas for $\mathbb{E}N$ and $\operatorname{Var}(N)$.

1. Bernoulli(p) : $P(N=1) = p$ and $P(N=0) = 1-p$. You should find $\mathbb{E}N = p$ and $\operatorname{Var}(N) = p - p^2$.
2. Binomial(n,p) : $P(N=k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0,1,\dots,n$. ($P(N=k)$ is the probability of k successes in a sequence of n independent yes/no experiments with probability of success being p.) You should find $\mathbb{E}N = np$ and $\operatorname{Var}(N) = n(p-p^2)$.
3. Geometric(p) : $P(N=k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}$. ($P(N=k)$ is the probability that the k^{th} – trial is the first time of success out a sequence of independent trials with probability of success being p.) You should find $\mathbb{E}N = 1/p$ and $\operatorname{Var}(N) = \frac{1-p}{p^2}$.
4. Poisson(λ) : $P(N=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for all $k \in \mathbb{N}_0$. You should find $\mathbb{E}N = \lambda = \operatorname{Var}(N)$.

Exercise 10.6. Let $S_{n,p} \overset{d}{=} \text{Binomial}(n,p)$, $k \in \mathbb{N}$, $p_n = \lambda_n / n$ where $\lambda_n \to \lambda > 0$ as $n \to \infty$. Show that

$$\lim_{n\to\infty} P(S_{n,p_n} = k) = \frac{\lambda^k}{k!} e^{-\lambda} = P(\text{Poisson}(\lambda) = k).$$

Thus we see that for $p = O(1/n)$ and k not too large relative to n that for large n,

$$P(\text{Binomial}(n,p) = k) \cong P(\text{Poisson}(pn) = k) = \frac{(pn)^k}{k!} e^{-pn}.$$

(We will come back to the Poisson distribution and the related Poisson process later on.)

10.3 Integration on $\mathbb{R}$

Notation 10.38 *If m is Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$, f is a non-negative Borel measurable function and $a < b$ with $a, b \in \bar{\mathbb{R}}$, we will often write $\int_a^b f(x)\,dx$ or $\int_a^b f dm$ for $\int_{(a,b]\cap\mathbb{R}} f dm$.*

Example 10.39. Suppose $-\infty < a < b < \infty$, $f \in C([a,b],\mathbb{R})$ and m be Lebesgue measure on $\mathbb{R}$. Given a partition,

$$\pi = \{a = a_0 < a_1 < \cdots < a_n = b\},$$

let

$$\operatorname{mesh}(\pi) := \max\{|a_j - a_{j-1}| : j = 1,\dots,n\}$$

and

$$f_\pi(x) := \sum_{l=0}^{n-1} f(a_l) 1_{(a_l, a_{l+1}]}(x).$$

Then

$$\int_a^b f_\pi \, dm = \sum_{l=0}^{n-1} f(a_l) m((a_l, a_{l+1}]) = \sum_{l=0}^{n-1} f(a_l)(a_{l+1} - a_l)$$

is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k\to\infty} \operatorname{mesh}(\pi_k) = 0$, we know that

$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b f(x)\,dx \tag{10.19}$$

where the latter integral is the Riemann integral. Using the (uniform) continuity of f on $[a,b]$, it easily follows that $\lim_{k\to\infty} f_{\pi_k}(x) = f(x)$ and that $|f_{\pi_k}(x)| \le g(x) := M 1_{(a,b]}(x)$ for all $x \in (a,b]$ where $M := \max_{x\in[a,b]} |f(x)| < \infty$. Since $\int_{\mathbb{R}} g dm = M(b-a) < \infty$, we may apply D.C.T. to conclude,

$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b \lim_{k\to\infty} f_{\pi_k} \, dm = \int_a^b f \, dm.$$

This equation with Eq. (10.19) shows

$$\int_a^b f \, dm = \int_a^b f(x)\,dx$$

whenever $f \in C([a,b],\mathbb{R})$, i.e. the Lebesgue and the Riemann integral agree on continuous functions. See Theorem 10.67 below for a more general statement along these lines.

Theorem 10.40 (The Fundamental Theorem of Calculus). *Suppose $-\infty < a < b < \infty$, $f \in C((a,b),\mathbb{R}) \cap L^1((a,b),m)$ and $F(x) := \int_a^x f(y) dm(y)$. Then*

1. *$F \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$.*
2. *$F'(x) = f(x)$ for all $x \in (a,b)$.*

3. If $G \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$ is an anti-derivative of f on (a,b) (i.e. $f = G'|_{(a,b)}$) then

$$\int_a^b f(x)dm(x) = G(b) - G(a).$$

Proof. Since $F(x) := \int_{\mathbb{R}} 1_{(a,x)}(y)f(y)dm(y)$, $\lim_{x\to z} 1_{(a,x)}(y) = 1_{(a,z)}(y)$ for m – a.e. y and $\left|1_{(a,x)}(y)f(y)\right| \le 1_{(a,b)}(y)\left|f(y)\right|$ is an L^1 – function, it follows from the dominated convergence Theorem 10.27 that F is continuous on $[a,b]$. Simple manipulations show,

$$\begin{aligned}\left|\frac{F(x+h)-F(x)}{h} - f(x)\right| &= \frac{1}{|h|}\begin{cases}\left|\int_x^{x+h}[f(y)-f(x)]\,dm(y)\right| & \text{if } h>0\\ \left|\int_{x+h}^{x}[f(y)-f(x)]\,dm(y)\right| & \text{if } h<0\end{cases}\\ &\le \frac{1}{|h|}\begin{cases}\int_x^{x+h}|f(y)-f(x)|\,dm(y) & \text{if } h>0\\ \int_{x+h}^{x}|f(y)-f(x)|\,dm(y) & \text{if } h<0\end{cases}\\ &\le \sup\{|f(y)-f(x)| : y\in[x-|h|,x+|h|]\}\end{aligned}$$

and the latter expression, by the continuity of f, goes to zero as $h \to 0$. This shows $F' = f$ on (a,b).

For the converse direction, we have by assumption that $G'(x) = F'(x)$ for $x \in (a,b)$. Therefore by the mean value theorem, $F - G = C$ for some constant C. Hence

$$\begin{aligned}\int_a^b f(x)dm(x) &= F(b) = F(b) - F(a)\\ &= (G(b)+C) - (G(a)+C) = G(b) - G(a).\end{aligned}$$

■

We can use the above results to integrate some non-Riemann integrable functions:

Example 10.41. For all $\lambda > 0$,

$$\int_0^\infty e^{-\lambda x}dm(x) = \lambda^{-1} \text{ and } \int_{\mathbb{R}}\frac{1}{1+x^2}dm(x) = \pi.$$

The proof of these identities are similar. By the monotone convergence theorem, Example 10.39 and the fundamental theorem of calculus for Riemann integrals (or Theorem 10.40 below),

$$\begin{aligned}\int_0^\infty e^{-\lambda x}dm(x) &= \lim_{N\to\infty}\int_0^N e^{-\lambda x}dm(x) = \lim_{N\to\infty}\int_0^N e^{-\lambda x}dx\\ &= -\lim_{N\to\infty}\frac{1}{\lambda}e^{-\lambda x}|_0^N = \lambda^{-1}\end{aligned}$$

and

$$\begin{aligned}\int_{\mathbb{R}}\frac{1}{1+x^2}dm(x) &= \lim_{N\to\infty}\int_{-N}^N\frac{1}{1+x^2}dm(x) = \lim_{N\to\infty}\int_{-N}^N\frac{1}{1+x^2}dx\\ &= \lim_{N\to\infty}\left[\tan^{-1}(N) - \tan^{-1}(-N)\right] = \pi.\end{aligned}$$

Let us also consider the functions x^{-p}. Using the MCT and the fundamental theorem of calculus,

$$\begin{aligned}\int_{(0,1]}\frac{1}{x^p}\,dm(x) &= \lim_{n\to\infty}\int_0^1 1_{(\frac{1}{n},1]}(x)\frac{1}{x^p}dm(x)\\ &= \lim_{n\to\infty}\int_{\frac{1}{n}}^1\frac{1}{x^p}dx = \lim_{n\to\infty}\left.\frac{x^{-p+1}}{1-p}\right|_{1/n}^1\\ &= \begin{cases}\frac{1}{1-p} & \text{if } p<1\\ \infty & \text{if } p>1\end{cases}\end{aligned}$$

If $p = 1$ we find

$$\int_{(0,1]}\frac{1}{x^p}\,dm(x) = \lim_{n\to\infty}\int_{\frac{1}{n}}^1\frac{1}{x}dx = \lim_{n\to\infty}\ln(x)|_{1/n}^1 = \infty.$$

Exercise 10.7. Show

$$\int_1^\infty\frac{1}{x^p}dm(x) = \begin{cases}\infty & \text{if } p\le 1\\ \frac{1}{p-1} & \text{if } p>1\end{cases}.$$

Example 10.42 (Integration of Power Series). Suppose $R > 0$ and $\{a_n\}_{n=0}^\infty$ is a sequence of complex numbers such that $\sum_{n=0}^\infty |a_n| r^n < \infty$ for all $r \in (0,R)$. Then

$$\int_\alpha^\beta\left(\sum_{n=0}^\infty a_n x^n\right)dm(x) = \sum_{n=0}^\infty a_n\int_\alpha^\beta x^n dm(x) = \sum_{n=0}^\infty a_n\frac{\beta^{n+1}-\alpha^{n+1}}{n+1}$$

for all $-R < \alpha < \beta < R$. Indeed this follows from Corollary 10.28 since

$$\begin{aligned}\sum_{n=0}^\infty\int_\alpha^\beta |a_n||x|^n\,dm(x) &\le \sum_{n=0}^\infty\left(\int_0^{|\beta|}|a_n||x|^n\,dm(x) + \int_0^{|\alpha|}|a_n||x|^n\,dm(x)\right)\\ &\le \sum_{n=0}^\infty |a_n|\frac{|\beta|^{n+1}+|\alpha|^{n+1}}{n+1} \le 2r\sum_{n=0}^\infty |a_n| r^n < \infty\end{aligned}$$

where $r = \max(|\beta|, |\alpha|)$.

Example 10.43. Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q} \cap [0,1]$ and define

$$f(x) = \sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x - r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x - r_n|}} = 5 \text{ if } x = r_n.$$

Since, By Theorem 10.40,

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{|x - r_n|}}\, dx &= \int_{r_n}^1 \frac{1}{\sqrt{x - r_n}} dx + \int_0^{r_n} \frac{1}{\sqrt{r_n - x}} dx \\ &= 2\sqrt{x - r_n}|_{r_n}^1 - 2\sqrt{r_n - x}|_0^{r_n} = 2\left(\sqrt{1 - r_n} - \sqrt{r_n}\right) \\ &\le 4,\end{aligned}$$

we find

$$\int_{[0,1]} f(x) dm(x) = \sum_{n=1}^\infty 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x - r_n|}} dx \le \sum_{n=1}^\infty 2^{-n} 4 = 4 < \infty.$$

In particular, $m(f = \infty) = 0$, i.e. that $f < \infty$ for almost every $x \in [0,1]$ and this implies that

$$\sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x - r_n|}} < \infty \text{ for a.e. } x \in [0,1].$$

This result is somewhat surprising since the singularities of the summands form a dense subset of $[0,1]$.

Example 10.44. The following limit holds,

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n dm(x) = 1. \tag{10.20}$$

DCT Proof. To verify this, let $f_n(x) := \left(1 - \frac{x}{n}\right)^n 1_{[0,n]}(x)$. Then $\lim_{n\to\infty} f_n(x) = e^{-x}$ for all $x \ge 0$. Moreover by simple calculus[1]

$$1 - x \le e^{-x} \text{ for all } x \in \mathbb{R}.$$

Therefore, for $x < n$, we have

$$0 \le 1 - \frac{x}{n} \le e^{-x/n} \implies \left(1 - \frac{x}{n}\right)^n \le \left[e^{-x/n}\right]^n = e^{-x},$$

from which it follows that

$$0 \le f_n(x) \le e^{-x} \text{ for all } x \ge 0.$$

From Example 10.41, we know

$$\int_0^\infty e^{-x} dm(x) = 1 < \infty,$$

so that e^{-x} is an integrable function on $[0, \infty)$. Hence by the dominated convergence theorem,

$$\begin{aligned}\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n dm(x) &= \lim_{n\to\infty} \int_0^\infty f_n(x) dm(x) \\ &= \int_0^\infty \lim_{n\to\infty} f_n(x) dm(x) = \int_0^\infty e^{-x} dm(x) = 1.\end{aligned}$$

MCT Proof. The limit in Eq. (10.20) may also be computed using the monotone convergence theorem. To do this we must show that $n \to f_n(x)$ is increasing in n for each x and for this it suffices to consider $n > x$. But for $n > x$,

$$\begin{aligned}\frac{d}{dn} \ln f_n(x) &= \frac{d}{dn}\left[n \ln\left(1 - \frac{x}{n}\right)\right] = \ln\left(1 - \frac{x}{n}\right) + \frac{n}{1 - \frac{x}{n}} \frac{x}{n^2} \\ &= \ln\left(1 - \frac{x}{n}\right) + \frac{\frac{x}{n}}{1 - \frac{x}{n}} = h(x/n)\end{aligned}$$

where, for $0 \le y < 1$,

$$h(y) := \ln(1 - y) + \frac{y}{1 - y}.$$

Since $h(0) = 0$ and

$$h'(y) = -\frac{1}{1 - y} + \frac{1}{1 - y} + \frac{y}{(1 - y)^2} > 0$$

it follows that $h \ge 0$. Thus we have shown, $f_n(x) \uparrow e^{-x}$ as $n \to \infty$ as claimed.

Example 10.45. Suppose that $f_n(x) := n 1_{(0, \frac{1}{n}]}(x)$ for $n \in \mathbb{N}$. Then $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$ while

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\, dx = \lim_{n\to\infty} 1 = 1 \ne 0 = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)\, dx.$$

[1] Since $y = 1 - x$ is the tangent line to $y = e^{-x}$ at $x = 0$ and e^{-x} is convex up, it follows that $1 - x \le e^{-x}$ for all $x \in \mathbb{R}$.

The problem is that the best dominating function we can take is

$$g(x) = \sup_n f_n(x) = \sum_{n=1}^{\infty} n \cdot 1_{\left(\frac{1}{n+1}, \frac{1}{n}\right]}(x).$$

Notice that

$$\int_{\mathbb{R}} g(x)\,dx = \sum_{n=1}^{\infty} n \cdot \left(\frac{1}{n} - \frac{1}{n+1}\right) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty.$$

Example 10.46 (Jordan's Lemma). In this example, let us consider the limit;

$$\lim_{n\to\infty} \int_0^{\pi} \cos\left(\sin\frac{\theta}{n}\right) e^{-n\sin(\theta)} d\theta.$$

Let

$$f_n(\theta) := 1_{(0,\pi]}(\theta) \cos\left(\sin\frac{\theta}{n}\right) e^{-n\sin(\theta)}.$$

Then

$$|f_n| \le 1_{(0,\pi]} \in L^1(m)$$

and

$$\lim_{n\to\infty} f_n(\theta) = 1_{(0,\pi]}(\theta)\, 1_{\{\pi\}}(\theta) = 1_{\{\pi\}}(\theta).$$

Therefore by the D.C.T.,

$$\lim_{n\to\infty} \int_0^{\pi} \cos\left(\sin\frac{\theta}{n}\right) e^{-n\sin(\theta)} d\theta = \int_{\mathbb{R}} 1_{\{\pi\}}(\theta)\, dm(\theta) = m(\{\pi\}) = 0.$$

Example 10.47. Recall from Example 10.41 that

$$\lambda^{-1} = \int_{[0,\infty)} e^{-\lambda x} dm(x) \text{ for all } \lambda > 0.$$

Let $\varepsilon > 0$. For $\lambda \ge 2\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $C_n(\varepsilon) < \infty$ such that

$$0 \le \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} = x^n e^{-\lambda x} \le C_n(\varepsilon) e^{-\varepsilon x}.$$

Using this fact, Corollary 10.30 and induction gives

$$\begin{aligned} n!\lambda^{-n-1} &= \left(-\frac{d}{d\lambda}\right)^n \lambda^{-1} = \int_{[0,\infty)} \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} dm(x) \\ &= \int_{[0,\infty)} x^n e^{-\lambda x} dm(x). \end{aligned}$$

That is

$$n! = \lambda^n \int_{[0,\infty)} x^n e^{-\lambda x} dm(x). \tag{10.21}$$

Remark 10.48. Corollary 10.30 may be generalized by allowing the hypothesis to hold for $x \in X \setminus E$ where $E \in \mathcal{B}$ is a **fixed** null set, i.e. E must be independent of t. Consider what happens if we formally apply Corollary 10.30 to $g(t) := \int_0^\infty 1_{x \le t} dm(x)$,

$$\dot{g}(t) = \frac{d}{dt}\int_0^\infty 1_{x\le t}\, dm(x) \overset{?}{=} \int_0^\infty \frac{\partial}{\partial t} 1_{x \le t}\, dm(x).$$

The last integral is zero since $\frac{\partial}{\partial t} 1_{x\le t} = 0$ unless $t = x$ in which case it is not defined. On the other hand $g(t) = t$ so that $\dot{g}(t) = 1$. (The reader should decide which hypothesis of Corollary 10.30 has been violated in this example.)

Exercise 10.8 (Folland 2.28 on p. 60.). Compute the following limits and justify your calculations:

1. $\lim_{n\to\infty} \int_0^\infty \frac{\sin(\frac{x}{n})}{(1+\frac{x}{n})^n}\, dx.$
2. $\lim_{n\to\infty} \int_0^1 \frac{1+nx^2}{(1+x^2)^n}\, dx$
3. $\lim_{n\to\infty} \int_0^\infty \frac{n\sin(x/n)}{x(1+x^2)}\, dx$
4. For all $a \in \mathbb{R}$ compute,

$$f(a) := \lim_{n\to\infty} \int_a^\infty n(1+n^2x^2)^{-1} dx.$$

[Hints: for parts 1. and 2. you might use the binomial expansion to estimate the denominators.]

Exercise 10.9 (Integration by Parts). Suppose that $f, g : \mathbb{R} \to \mathbb{R}$ are two continuously differentiable functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $\mathbb{R}$. Prove the following integration by parts formula;

$$\int_{\mathbb{R}} f'(x) \cdot g(x)\, dx = -\int_{\mathbb{R}} f(x) \cdot g'(x)\, dx. \tag{10.22}$$

Similarly show that; if $f, g : [0,\infty) \to [0,\infty)$ are continuously differentiable functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $[0,\infty)$, then

$$\int_0^\infty f'(x) \cdot g(x)\, dx = -f(0)\, g(0) - \int_0^\infty f(x) \cdot g'(x)\, dx. \tag{10.23}$$

Outline: 1. First notice that Eq. (10.22) holds if $f(x) = 0$ for $|x| \ge N$ for some $N < \infty$ by undergraduate calculus.

2. Let $\psi : \mathbb{R} \to [0,1]$ be a continuously differentiable function such that $\psi(x) = 1$ if $|x| \le 1$ and $\psi(x) = 0$ if $|x| \ge 2$. For any $\varepsilon > 0$ let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ Write out the identity in Eq. (10.22) with $f(x)$ being replaced by $f(x)\,\psi_\varepsilon(x)$.

3. Now use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in the identity you found in step 2.
4. A similar outline works to prove Eq. (10.23).

Definition 10.49 (Gamma Function). *The **Gamma function**, $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by*

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du \tag{10.24}$$

(The reader should check that $\Gamma(x) < \infty$ for all $x > 0$.)

Here are some of the more basic properties of this function.

Example 10.50 (Γ – function properties). Let Γ be the gamma function, then;

1. $\Gamma(1) = 1$ as is easily verified.
2. $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$ as follows by integration by parts;
$$\Gamma(x+1) = \int_0^\infty e^{-u}\ u^{x+1}\ \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\ e^{-u}\right) du$$
$$= x \int_0^\infty u^{x-1}\ e^{-u}\ du = x\ \Gamma(x).$$
In particular, it follows from items 1. and 2. and induction that
$$\Gamma(n+1) = n! \text{ for all } n \in \mathbb{N}. \tag{10.25}$$
(Equation 10.25was also proved in Eq. (10.21).)
3. $\Gamma(1/2) = \sqrt{\pi}$. This last assertion is a bit trickier. One proof is to make use of the fact (proved below in Lemma **??**) that
$$\int_{-\infty}^\infty e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a > 0. \tag{10.26}$$
Taking $a = 1$ and making the change of variables, $u = r^2$ below implies,
$$\sqrt{\pi} = \int_{-\infty}^\infty e^{-r^2} dr = 2\int_0^\infty u^{-1/2} e^{-u} du = \Gamma(1/2).$$
$$\Gamma(1/2) = 2\int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr$$
$$= I_1(1) = \sqrt{\pi}.$$
4. A simple induction argument using items 2. and 3. now shows that
$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi}$$
where $(-1)!! := 1$ and $(2n-1)!! = (2n-1)(2n-3)\dots 3 \cdot 1$ for $n \in \mathbb{N}$.

10.4 Densities and Change of Variables Theorems

Exercise 10.10 (Measures and Densities). Let $(X, \mathcal{M}, \mu)$ be a measure space and $\rho : X \to [0,\infty]$ be a measurable function. For $A \in \mathcal{M}$, set $\nu(A) := \int_A \rho d\mu$.

1. Show $\nu : \mathcal{M} \to [0,\infty]$ is a measure.
2. Let $f : X \to [0,\infty]$ be a measurable function, show
$$\int_X f d\nu = \int_X f \rho d\mu. \tag{10.27}$$
Hint: first prove the relationship for characteristic functions, then for simple functions, and then for general positive measurable functions.
3. Show that a measurable function $f : X \to \mathbb{C}$ is in $L^1(\nu)$ iff $|f|\rho \in L^1(\mu)$ and if $f \in L^1(\nu)$ then Eq. (10.27) still holds.

Notation 10.51 *It is customary to informally describe ν defined in Exercise 10.10 by writing $d\nu = \rho d\mu$.*

Exercise 10.11 (Abstract Change of Variables Formula). Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Recall that $\nu = f_*\mu : \mathcal{F} \to [0,\infty]$ defined by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$ is a measure on $\mathcal{F}$.

1. Show
$$\int_Y g d\nu = \int_X (g \circ f)\, d\mu \tag{10.28}$$
for all measurable functions $g : Y \to [0,\infty]$. **Hint:** see the hint from Exercise 10.10.
2. Show a measurable function $g : Y \to \mathbb{C}$ is in $L^1(\nu)$ iff $g \circ f \in L^1(\mu)$ and that Eq. (10.28) holds for all $g \in \mathrm{L}^1(\nu)$.

Example 10.52. Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\{X_i\}_{i=1}^n$ are random variables on Ω with $\nu := \mathrm{Law}_P(X_1, \dots, X_n)$, then

$$\mathbb{E}\left[g(X_1, \dots, X_n)\right] = \int_{\mathbb{R}^n} g\ d\nu$$

for all $g : \mathbb{R}^n \to \mathbb{R}$ which are Borel measurable and either bounded or nonnegative. This follows directly from Exercise 10.11 with $f := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ and $\mu = P$.

Remark 10.53. As a special case of Example 10.52, suppose that X is a random variable on a probability space, $(\Omega, \mathcal{B}, P)$, and $F(x) := P(X \le x)$. Then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\, dF(x) \tag{10.29}$$

where $dF(x)$ is shorthand for $d\mu_F(x)$ and μ_F is the unique probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((-\infty, x]) = F(x)$ for all $x \in \mathbb{R}$. Moreover if $F : \mathbb{R} \to [0,1]$ happens to be C^1-function, then

$$d\mu_F(x) = F'(x)\, dm(x) \tag{10.30}$$

and Eq. (10.29) may be written as

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) F'(x)\, dm(x). \tag{10.31}$$

To verify Eq. (10.30) it suffices to observe, by the fundamental theorem of calculus, that

$$\mu_F((a,b]) = F(b) - F(a) = \int_a^b F'(x)\, dx = \int_{(a,b]} F' dm.$$

From this equation we may deduce that $\mu_F(A) = \int_A F' dm$ for all $A \in \mathcal{B}_{\mathbb{R}}$. Equation 10.31 now follows from Exercise 10.10.

Exercise 10.12. Let $F : \mathbb{R} \to \mathbb{R}$ be a C^1-function such that $F'(x) > 0$ for all $x \in \mathbb{R}$ and $\lim_{x\to\pm\infty} F(x) = \pm\infty$. (Notice that F is strictly increasing so that $F^{-1} : \mathbb{R} \to \mathbb{R}$ exists and moreover, by the inverse function theorem that F^{-1} is a C^1 – function.) Let m be Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$ and

$$\nu(A) = m(F(A)) = m(\left(F^{-1}\right)^{-1}(A)) = \left(F_*^{-1} m\right)(A)$$

for all $A \in \mathcal{B}_{\mathbb{R}}$. Show $d\nu = F' dm$. Use this result to prove the change of variable formula,

$$\int_{\mathbb{R}} h \circ F \cdot F' dm = \int_{\mathbb{R}} h dm \tag{10.32}$$

which is valid for all Borel measurable functions $h : \mathbb{R} \to [0, \infty]$.

Hint: Start by showing $d\nu = F' dm$ on sets of the form $A = (a, b]$ with $a, b \in \mathbb{R}$ and $a < b$. Then use the uniqueness assertions in Exercise 8.11 to conclude $d\nu = F' dm$ on all of $\mathcal{B}_{\mathbb{R}}$. To prove Eq. (10.32) apply Exercise 10.11 with $g = h \circ F$ and $f = F^{-1}$.

10.5 Normal (Gaussian) Random Variables

Definition 10.54 (Normal / Gaussian Random Variables). *A random variable, Y,* ***is normal with mean*** *μ* ***standard deviation*** *σ^2 iff*

$$P(Y \in B) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ for all } B \in \mathcal{B}_{\mathbb{R}}. \tag{10.33}$$

We will abbreviate this by writing $Y \stackrel{d}{=} N(\mu, \sigma^2)$. When $\mu = 0$ and $\sigma^2 = 1$ we will simply write N for $N(0,1)$ and if $Y \stackrel{d}{=} N$, we will say Y is a ***standard normal*** *random variable.*

Observe that Eq. (10.33) is equivalent to writing

$$\mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy$$

for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$. Also observe that $Y \stackrel{d}{=} N(\mu, \sigma^2)$ is equivalent to $Y \stackrel{d}{=} \sigma N + \mu$. Indeed, by making the change of variable, $y = \sigma x + \mu$, we find

$$\begin{aligned}\mathbb{E}[f(\sigma N + \mu)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\sigma x + \mu) e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \frac{dy}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy.\end{aligned}$$

Lastly the constant, $\left(2\pi\sigma^2\right)^{-1/2}$ is chosen so that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}y^2} dy = 1,$$

see Example 10.50 and Lemma **??**.

Exercise 10.13. Suppose that $X \stackrel{d}{=} N(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ is a C^1 – function such that $Xf(X)$, $f'(X)$ and $f(X)$ are all integrable random variables. Show

$$\begin{aligned}\mathbb{E}[Xf(X)] &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \frac{d}{dx} e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(x) e^{-\frac{1}{2}x^2} dx = \mathbb{E}[f'(X)].\end{aligned}$$

Example 10.55. Suppose that $X \stackrel{d}{=} N(0,1)$ and define $\alpha_k := \mathbb{E}\left[X^{2k}\right]$ for all $k \in \mathbb{N}_0$. By Exercise 10.13,

$$\alpha_{k+1} = \mathbb{E}\left[X^{2k+1} \cdot X\right] = (2k+1)\,\alpha_k \text{ with } \alpha_0 = 1.$$

Hence it follows that

$$\alpha_1 = \alpha_0 = 1,\ \alpha_2 = 3\alpha_1 = 3,\ \alpha_3 = 5 \cdot 3$$

and by a simple induction argument,

$$\mathbb{E}X^{2k} = \alpha_k = (2k-1)!!, \tag{10.34}$$

where $(-1)!! := 0$. Actually we can use the Γ – function to say more. Namely for any $\beta > -1$,

$$\mathbb{E}\,|X|^\beta = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^\beta\, e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x^\beta e^{-\frac{1}{2}x^2} dx.$$

Now make the change of variables, $y = x^2/2$ (i.e. $x = \sqrt{2y}$ and $dx = \frac{1}{\sqrt{2}} y^{-1/2} dy$) to learn,

$$\begin{aligned}\mathbb{E}\,|X|^\beta &= \frac{1}{\sqrt{\pi}} \int_0^\infty (2y)^{\beta/2}\, e^{-y} y^{-1/2} dy \\ &= \frac{1}{\sqrt{\pi}} 2^{\beta/2} \int_0^\infty y^{(\beta+1)/2} e^{-y} y^{-1} dy = \frac{1}{\sqrt{\pi}} 2^{\beta/2} \Gamma\left(\frac{\beta+1}{2}\right).\end{aligned} \tag{10.35}$$

Exercise 10.14. Suppose that $X \overset{d}{=} N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right] = \exp\left(-\lambda^2/2\right). \tag{10.36}$$

Hint: Use Corollary 10.30 to show, $f'(\lambda) = i\mathbb{E}\left[X e^{i\lambda X}\right]$ and then use Exercise 10.13 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

Exercise 10.15. Suppose that $X \overset{d}{=} N(0,1)$ and $t \in \mathbb{R}$. Show $\mathbb{E}\left[e^{tX}\right] = \exp\left(t^2/2\right)$. (You could follow the hint in Exercise 10.14 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

Exercise 10.16. Use Exercise 10.15 and Proposition 10.33 to give another proof that $\mathbb{E}X^{2k} = (2k-1)!!$ when $X \overset{d}{=} N(0,1)$.

Exercise 10.17. Let $X \overset{d}{=} N(0,1)$ and $\alpha \in \mathbb{R}$, find $\rho : \mathbb{R}_+ \to \mathbb{R}_+ := (0,\infty)$ such that

$$\mathbb{E}\left[f\left(|X|^\alpha\right)\right] = \int_{\mathbb{R}_+} f(x)\,\rho(x)\,dx$$

for all continuous functions, $f : \mathbb{R}_+ \to \mathbb{R}$ with compact support in $\mathbb{R}_+$.

Lemma 10.56 (Gaussian tail estimates). *Suppose that X is a standard normal random variable, i.e.*

$$P(X \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx \text{ for all } A \in \mathcal{B}_{\mathbb{R}},$$

then for all $x \geq 0$,

$$P(X \geq x) \leq \min\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right) \leq \frac{1}{2} e^{-x^2/2}. \tag{10.37}$$

Moreover (see [18, Lemma 2.5]),

$$P(X \geq x) \geq \max\left(1 - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) \tag{10.38}$$

which combined with Eq. (10.37) proves Mill's ratio (see [7]);

$$\lim_{x\to\infty} \frac{P(X \geq x)}{\frac{1}{\sqrt{2\pi}x} e^{-x^2/2}} = 1. \tag{10.39}$$

Proof. See Figure 10.1 where; the green curve is the plot of $P(X \geq x)$, the black is the plot of

$$\min\left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right),$$

the red is the plot of $\frac{1}{2} e^{-x^2/2}$, and the blue is the plot of

$$\max\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right).$$

The formal proof of these estimates for the reader who is not convinced by Figure 10.1 is given below.

We begin by observing that

$$\begin{aligned}P(X \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \leq \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2} dy \\ &\leq -\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-y^2/2}|_x^{-\infty} = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.\end{aligned} \tag{10.40}$$

If we only want to prove Mill's ratio (10.39), we could proceed as follows. Let $\alpha > 1$, then for $x > 0$,

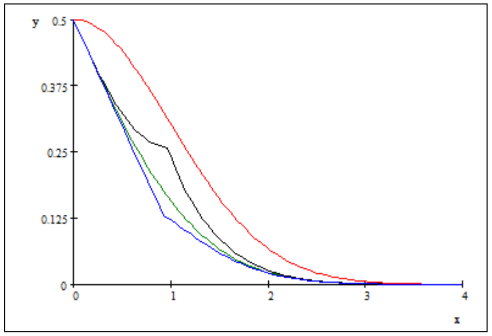


Fig. 10.1. Plots of $P(X \geq x)$ and its estimates.

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \\ &\geq \frac{1}{\sqrt{2\pi}} \int_x^{\alpha x} \frac{y}{\alpha x} e^{-y^2/2} dy = -\frac{1}{\sqrt{2\pi}} \frac{1}{\alpha x} e^{-y^2/2}|_{y=x}^{y=\alpha x} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\alpha x} e^{-x^2/2} \left[1 - e^{-\alpha^2 x^2/2}\right] \end{aligned}$$

from which it follows,

$$\liminf_{x\to\infty} \left[\sqrt{2\pi} x e^{x^2/2} \cdot P(X \geq x)\right] \geq 1/\alpha \uparrow 1 \text{ as } \alpha \downarrow 1.$$

The estimate in Eq. (10.40) shows $\limsup_{x\to\infty} \left[\sqrt{2\pi} x e^{x^2/2} \cdot P(X \geq x)\right] \leq 1$.

To get more precise estimates, we begin by observing,

$$\begin{aligned} P(X \geq x) &= \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy \qquad (10.41) \\ &\leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-x^2/2} dy \leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x. \end{aligned}$$

This equation along with Eq. (10.40) gives the first equality in Eq. (10.37). To prove the second equality observe that $\sqrt{2\pi} > 2$, so

$$\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} \leq \frac{1}{2} e^{-x^2/2} \text{ if } x \geq 1.$$

For $x \leq 1$ we must show,

$$\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2} \leq \frac{1}{2} e^{-x^2/2}$$

or equivalently that $f(x) := e^{x^2/2} - \sqrt{\frac{2}{\pi}} x \leq 1$ for $0 \leq x \leq 1$. Since f is convex $\left(f''(x) = \left(x^2+1\right) e^{x^2/2} > 0\right)$, $f(0) = 1$ and $f(1) \cong 0.85 < 1$, it follows that $f \leq 1$ on $[0,1]$. This proves the second inequality in Eq. (10.37).

It follows from Eq. (10.41) that

$$\begin{aligned} P(X \geq x) &= \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy \\ &\geq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x 1 dy = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} x \text{ for all } x \geq 0. \end{aligned}$$

So to finish the proof of Eq. (10.38) we must show,

$$\begin{aligned} f(x) &:= \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} - \left(1 + x^2\right) P(X \geq x) \\ &= \frac{1}{\sqrt{2\pi}} \left[x e^{-x^2/2} - \left(1 + x^2\right) \int_x^\infty e^{-y^2/2} dy\right] \leq 0 \text{ for all } 0 \leq x < \infty. \end{aligned}$$

This follows by observing that $f(0) = -1/2 < 0$, $\lim_{x\uparrow\infty} f(x) = 0$ and

$$\begin{aligned} f'(x) &= \frac{1}{\sqrt{2\pi}} \left[e^{-x^2/2}\left(1 - x^2\right) - 2xP(X \geq x) + \left(1 + x^2\right) e^{-x^2/2}\right] \\ &= 2\left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2} - xP(X \geq y)\right) \geq 0, \end{aligned}$$

where the last inequality is a consequence Eq. (10.37). ∎

10.6 Stirling's Formula

On occasion one is faced with estimating an integral of the form, $\int_J e^{-G(t)} dt$, where $J = (a,b) \subset \mathbb{R}$ and $G(t)$ is a C^1 – function with a unique (for simplicity) global minimum at some point $t_0 \in J$. The idea is that the majority contribution of the integral will often come from some neighborhood, $(t_0 - \alpha, t_0 + \alpha)$, of t_0. Moreover, it may happen that $G(t)$ can be well approximated on this neighborhood by its Taylor expansion to order 2;

$$G(t) \cong G(t_0) + \frac{1}{2} \ddot{G}(t_0)(t - t_0)^2.$$

Notice that the linear term is zero since t_0 is a minimum and therefore $\dot{G}(t_0) = 0$. We will further assume that $\ddot{G}(t_0) \neq 0$ and hence $\ddot{G}(t_0) > 0$. Under these hypothesis we will have,

$$\int_J e^{-G(t)}dt \cong e^{-G(t_0)}\int_{|t-t_0|<\alpha}\exp\left(-\frac{1}{2}\ddot{G}(t_0)(t-t_0)^2\right)dt.$$

Making the change of variables, $s=\sqrt{\ddot{G}(t_0)}(t-t_0)$, in the above integral then gives,

$$\begin{aligned}\int_J e^{-G(t)}dt &\cong \frac{1}{\sqrt{\ddot{G}(t_0)}}e^{-G(t_0)}\int_{|s|<\sqrt{\ddot{G}(t_0)}\cdot\alpha}e^{-\frac{1}{2}s^2}ds\\ &= \frac{1}{\sqrt{\ddot{G}(t_0)}}e^{-G(t_0)}\left[\sqrt{2\pi}-\int_{\sqrt{\ddot{G}(t_0)}\cdot\alpha}^{\infty}e^{-\frac{1}{2}s^2}ds\right]\\ &= \frac{1}{\sqrt{\ddot{G}(t_0)}}e^{-G(t_0)}\left[\sqrt{2\pi}-O\left(\frac{1}{\sqrt{\ddot{G}(t_0)}\cdot\alpha}e^{-\frac{1}{2}\ddot{G}(t_0)\cdot\alpha^2}\right)\right].\end{aligned}$$

If α is sufficiently large, for example if $\sqrt{\ddot{G}(t_0)}\cdot\alpha=3$, then the error term is about 0.0037 and we should be able to conclude that

$$\int_J e^{-G(t)}dt \cong \sqrt{\frac{2\pi}{\ddot{G}(t_0)}}e^{-G(t_0)}. \tag{10.42}$$

The proof of the next theorem (Stirling's formula for the Gamma function) will illustrate these ideas and what one has to do to carry them out rigorously.

Theorem 10.57 (Stirling's formula). *The Gamma function (see Definition 10.49), satisfies Stirling's formula,*

$$\lim_{x\to\infty}\frac{\Gamma(x+1)}{\sqrt{2\pi}e^{-x}x^{x+1/2}}=1. \tag{10.43}$$

In particular, if $n\in\mathbb{N}$, we have

$$n! = \Gamma(n+1)\sim\sqrt{2\pi}e^{-n}n^{n+1/2}$$

where we write $a_n\sim b_n$ to mean, $\lim_{n\to\infty}\frac{a_n}{b_n}=1$. (See Example 10.62 below for a slightly cruder but more elementary estimate of $n!$)

Proof. (The following proof is an elaboration of the proof found on page 236-237 in Krantz's Real Analysis and Foundations.) We begin with the formula for $\Gamma(x+1)$;

$$\Gamma(x+1)=\int_0^\infty e^{-t}t^x dt=\int_0^\infty e^{-G_x(t)}dt, \tag{10.44}$$

where

$$G_x(t):=t-x\ln t.$$

Then $\dot{G}_x(t)=1-x/t$, $\ddot{G}_x(t)=x/t^2$, G_x has a global minimum (since $\ddot{G}_x>0$) at $t_0=x$ where

$$G_x(x)=x-x\ln x \text{ and } \ddot{G}_x(x)=1/x.$$

So if Eq. (10.42) is valid in this case we should expect,

$$\Gamma(x+1)\cong\sqrt{2\pi x}e^{-(x-x\ln x)}=\sqrt{2\pi}e^{-x}x^{x+1/2}$$

which would give Stirling's formula. The rest of the proof will be spent on rigorously justifying the approximations involved.

Let us begin by making the change of variables $s=\sqrt{\ddot{G}(t_0)}(t-t_0)=\frac{1}{\sqrt{x}}(t-x)$ as suggested above. Then

$$\begin{aligned}G_x(t)-G_x(x)&=(t-x)-x\ln(t/x)=\sqrt{x}s-x\ln\left(\frac{x+\sqrt{x}s}{x}\right)\\ &=x\left[\frac{s}{\sqrt{x}}-\ln\left(1+\frac{s}{\sqrt{x}}\right)\right]=s^2q\left(\frac{s}{\sqrt{x}}\right)\end{aligned}$$

where

$$q(u):=\frac{1}{u^2}[u-\ln(1+u)] \text{ for } u>-1 \text{ with } q(0):=\frac{1}{2}.$$

Setting $q(0)=1/2$ makes q a continuous and in fact smooth function on $(-1,\infty)$, see Figure 10.2. Using the power series expansion for $\ln(1+u)$ we find,

$$q(u)=\frac{1}{2}+\sum_{k=3}^\infty\frac{(-u)^{k-2}}{k} \text{ for } |u|<1. \tag{10.45}$$

Making the change of variables, $t=x+\sqrt{x}s$ in the second integral in Eq. (10.44) yields,

$$\Gamma(x+1)=e^{-(x-x\ln x)}\sqrt{x}\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds=x^{x+1/2}e^{-x}\cdot I(x),$$

where

$$I(x)=\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds=\int_{-\infty}^\infty 1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds. \tag{10.46}$$

From Eq. (10.45) it follows that $\lim_{u\to0}q(u)=1/2$ and therefore,

$$\int_{-\infty}^\infty\lim_{x\to\infty}\left[1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}\right]ds=\int_{-\infty}^\infty e^{-\frac{1}{2}s^2}ds=\sqrt{2\pi}. \tag{10.47}$$

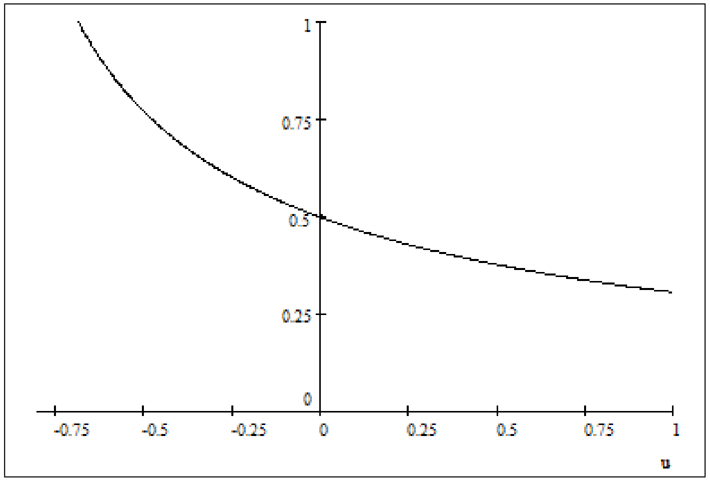


Fig. 10.2. Plot of $q(u)$.

So if there exists a dominating function, $F \in L^1(\mathbb{R}, m)$, such that

$$1_{s \geq -\sqrt{x}} \cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2} \leq F(s) \text{ for all } s \in \mathbb{R} \text{ and } x \geq 1,$$

we can apply the DCT to learn that $\lim_{x\to\infty} I(x) = \sqrt{2\pi}$ which will complete the proof of Stirling's formula.

We now construct the desired function F. From Eq. (10.45) it follows that $q(u) \geq 1/2$ for $-1 < u \leq 0$. Since $u - \ln(1+u) > 0$ for $u \neq 0$ ($u - \ln(1+u)$ is convex and has a minimum of 0 at $u = 0$) we may conclude that $q(u) > 0$ for all $u > -1$ therefore by compactness (on $[0, M]$), $\min_{-1<u\leq M} q(u) = \varepsilon(M) > 0$ for all $M \in (0, \infty)$, see Remark 10.58 for more explicit estimates. Lastly, since $\frac{1}{u}\ln(1+u) \to 0$ as $u \to \infty$, there exists $M < \infty$ ($M = 3$ would due) such that $\frac{1}{u}\ln(1+u) \leq \frac{1}{2}$ for $u \geq M$ and hence,

$$q(u) = \frac{1}{u}\left[1 - \frac{1}{u}\ln(1+u)\right] \geq \frac{1}{2u} \text{ for } u \geq M.$$

So there exists $\varepsilon > 0$ and $M < \infty$ such that (for all $x \geq 1$),

$$\begin{aligned} 1_{s \geq -\sqrt{x}} e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2} &\leq 1_{-\sqrt{x}<s\leq M} e^{-\varepsilon s^2} + 1_{s\geq M} e^{-\sqrt{x}s/2} \\ &\leq 1_{-\sqrt{x}<s\leq M} e^{-\varepsilon s^2} + 1_{s\geq M} e^{-s/2} \\ &\leq e^{-\varepsilon s^2} + e^{-|s|/2} =: F(s) \in L^1(\mathbb{R}, ds). \end{aligned}$$

■

We will sometimes use the following variant of Eq. (10.43);

$$\lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x} = 1 \tag{10.48}$$

To prove this let x go to $x - 1$ in Eq. (10.43) in order to find,

$$1 = \lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{2\pi}e^{-x} \cdot e \cdot (x-1)^{x-1/2}} = \lim_{x\to\infty} \frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x \cdot \sqrt{\frac{x}{x-1}} \cdot e \cdot \left(1 - \frac{1}{x}\right)^x}$$

which gives Eq. (10.48) since

$$\lim_{x\to\infty} \sqrt{\frac{x}{x-1}} \cdot e \cdot \left(1 - \frac{1}{x}\right)^x = 1.$$

Remark 10.58 (Estimating $q(u)$ by Taylor's Theorem). Another way to estimate $q(u)$ is to use Taylor's theorem with integral remainder. In general if h is C^2 – function on $[0, 1]$, then by the fundamental theorem of calculus and integration by parts,

$$\begin{aligned} h(1) - h(0) &= \int_0^1 \dot{h}(t)\,dt = -\int_0^1 \dot{h}(t)\,d(1-t) \\ &= -\dot{h}(t)(1-t)|_0^1 + \int_0^1 \ddot{h}(t)(1-t)\,dt \\ &= \dot{h}(0) + \frac{1}{2}\int_0^1 \ddot{h}(t)\,d\nu(t) \end{aligned} \tag{10.49}$$

where $d\nu(t) := 2(1-t)\,dt$ which is a probability measure on $[0, 1]$. Applying this to $h(t) = F(a + t(b-a))$ for a C^2 – function on an interval of points between a and b in $\mathbb{R}$ then implies,

$$F(b) - F(a) = (b-a)\dot{F}(a) + \frac{1}{2}(b-a)^2 \int_0^1 \ddot{F}(a + t(b-a))\,d\nu(t). \tag{10.50}$$

(Similar formulas hold to any order.) Applying this result with $F(x) = x - \ln(1+x)$, $a = 0$, and $b = u \in (-1, \infty)$ gives,

$$u - \ln(1+u) = \frac{1}{2}u^2 \int_0^1 \frac{1}{(1+tu)^2}\,d\nu(t),$$

i.e.

$$q(u) = \frac{1}{2}\int_0^1 \frac{1}{(1+tu)^2}\,d\nu(t).$$

From this expression for $q(u)$ it now easily follows that

$$q(u) \geq \frac{1}{2}\int_0^1 \frac{1}{(1+0)^2}d\nu(t) = \frac{1}{2} \text{ if } -1 < u \leq 0$$

and

$$q(u) \geq \frac{1}{2}\int_0^1 \frac{1}{(1+u)^2}d\nu(t) = \frac{1}{2(1+u)^2}.$$

So an explicit formula for $\varepsilon(M)$ is $\varepsilon(M) = (1+M)^{-2}/2.$

10.6.1 Two applications of Stirling's formula

In this subsection suppose $x \in (0,1)$ and $S_n \overset{d}{=}$Binomial(n,x) for all $n \in \mathbb{N}$, i.e.

$$P_x(S_n = k) = \binom{n}{k}x^k(1-x)^{n-k} \text{ for } 0 \leq k \leq n. \tag{10.51}$$

Recall that $\mathbb{E}S_n = nx$ and $\operatorname{Var}(S_n) = n\sigma^2$ where $\sigma^2 := x(1-x)$. The weak law of large numbers states (Exercise 7.13) that

$$P\left(\left|\frac{S_n}{n} - x\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\sigma^2$$

and therefore, $\frac{S_n}{n}$ is concentrating near its mean value, x, for n large, i.e. $S_n \cong nx$ for n large. The next central limit theorem describes the fluctuations of S_n about nx.

Theorem 10.59 (De Moivre-Laplace Central Limit Theorem). *For all* $-\infty < a < b < \infty$,

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - nx}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}dy$$
$$= P(a \leq N \leq b)$$

where $N \overset{d}{=} N(0,1)$. *Informally,* $\frac{S_n - nx}{\sigma\sqrt{n}} \overset{d}{\cong} N$ *or equivalently,* $S_n \overset{d}{\cong} nx + \sigma\sqrt{n}\cdot N$ *which if valid in a neighborhood of* nx *whose length is order* $\sqrt{n}$.

Proof. (We are not going to cover all the technical details in this proof as we will give much more general versions of this theorem later.) Starting with the definition of the Binomial distribution we have,

$$\begin{aligned} p_n &:= P\left(a \leq \frac{S_n - nx}{\sigma\sqrt{n}} \leq b\right) = P\left(S_n \in nx + \sigma\sqrt{n}[a,b]\right) \\ &= \sum_{k \in nx+\sigma\sqrt{n}[a,b]} P(S_n = k) \\ &= \sum_{k \in nx+\sigma\sqrt{n}[a,b]} \binom{n}{k}x^k(1-x)^{n-k}. \end{aligned}$$

Letting $k = nx + \sigma\sqrt{n}y_k$, i.e. $y_k = (k - nx)/\sigma\sqrt{n}$ we see that $\Delta y_k = y_{k+1} - y_k = 1/(\sigma\sqrt{n})$. Therefore we may write p_n as

$$p_n = \sum_{y_k \in [a,b]} \sigma\sqrt{n}\binom{n}{k}x^k(1-x)^{n-k}\Delta y_k. \tag{10.52}$$

So to finish the proof we need to show, for $k = O(\sqrt{n})$ ($y_k = O(1)$), that

$$\sigma\sqrt{n}\binom{n}{k}x^k(1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_k^2} \text{ as } n \to \infty \tag{10.53}$$

in which case the sum in Eq. (10.52) may be well approximated by the "Riemann sum;"

$$p_n \sim \sum_{y_k \in [a,b]} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_k^2}\Delta y_k \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}dy \text{ as } n \to \infty.$$

By Stirling's formula,

$$\begin{aligned} \sigma\sqrt{n}\binom{n}{k} &= \sigma\sqrt{n}\frac{1}{k!}\frac{n!}{(n-k)!} \sim \frac{\sigma\sqrt{n}}{\sqrt{2\pi}}\frac{n^{n+1/2}}{k^{k+1/2}(n-k)^{n-k+1/2}} \\ &= \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\left(\frac{k}{n}\right)^{k+1/2}\left(1-\frac{k}{n}\right)^{n-k+1/2}} \\ &= \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\left(x+\frac{\sigma}{\sqrt{n}}y_k\right)^{k+1/2}\left(1-x-\frac{\sigma}{\sqrt{n}}y_k\right)^{n-k+1/2}} \\ &\sim \frac{\sigma}{\sqrt{2\pi}}\frac{1}{\sqrt{x(1-x)}}\frac{1}{\left(x+\frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1-x-\frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\left(x+\frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1-x-\frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}}. \end{aligned}$$

In order to shorten the notation, let $z_k := \frac{\sigma}{\sqrt{n}}y_k = O\left(n^{-1/2}\right)$ so that $k = nx + nz_k = n(x+z_k)$. In this notation we have shown,

$$\sqrt{2\pi}\sigma\sqrt{n}\binom{n}{k}x^k(1-x)^{n-k} \sim \frac{x^k(1-x)^{n-k}}{(x+z_k)^k(1-x-z_k)^{n-k}}$$
$$= \frac{1}{\left(1+\frac{1}{x}z_k\right)^k\left(1-\frac{1}{1-x}z_k\right)^{n-k}}$$
$$= \frac{1}{\left(1+\frac{1}{x}z_k\right)^{n(x+z_k)}\left(1-\frac{1}{1-x}z_k\right)^{n(1-x-z_k)}} =: q(n,k)\,. \tag{10.54}$$

Taking logarithms and using Taylor's theorem we learn

$$n(x+z_k)\ln\left(1+\frac{1}{x}z_k\right)$$
$$= n(x+z_k)\left(\frac{1}{x}z_k - \frac{1}{2x^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= nz_k + \frac{n}{2x}z_k^2 + O\left(n^{-3/2}\right) \text{ and}$$
$$n(1-x-z_k)\ln\left(1-\frac{1}{1-x}z_k\right)$$
$$= n(1-x-z_k)\left(-\frac{1}{1-x}z_k - \frac{1}{2(1-x)^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= -nz_k + \frac{n}{2(1-x)}z_k^2 + O\left(n^{-3/2}\right).$$

and then adding these expressions shows,

$$-\ln q(n,k) = \frac{n}{2}z_k^2\left(\frac{1}{x}+\frac{1}{1-x}\right) + O\left(n^{-3/2}\right)$$
$$= \frac{n}{2\sigma^2}z_k^2 + O\left(n^{-3/2}\right) = \frac{1}{2}y_k^2 + O\left(n^{-3/2}\right).$$

Combining this with Eq. (10.54) shows,

$$\sigma\sqrt{n}\binom{n}{k}x^k(1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_k^2 + O\left(n^{-3/2}\right)\right)$$

which gives the desired estimate in Eq. (10.53). ■

The previous central limit theorem has shown that

$$\frac{S_n}{n} \overset{d}{\cong} x + \frac{\sigma}{\sqrt{n}}N$$

which implies the major fluctuations of S_n/n occur within intervals about x of length $O\left(\frac{1}{\sqrt{n}}\right)$. The next result aims to understand the rare events where S_n/n makes a "large" deviation from its mean value, x – in this case a large deviation is something of size $O(1)$ as $n \to \infty$.

Theorem 10.60 (Binomial Large Deviation Bounds). *Let us continue to use the notation in Theorem 10.59. Then for all $y \in (0,x)$,*

$$\lim_{n\to\infty}\frac{1}{n}\ln P_x\left(\frac{S_n}{n} \le y\right) = y\ln\frac{x}{y} + (1-y)\ln\frac{1-x}{1-y}.$$

Roughly speaking,

$$P_x\left(\frac{S_n}{n} \le y\right) \approx e^{-nI_x(y)}$$

where $I_x(y)$ is the "rate function,"

$$I_x(y) := y\ln\frac{y}{x} + (1-y)\ln\frac{1-y}{1-x},$$

see Figure 10.3 for the graph of $I_{1/2}$.

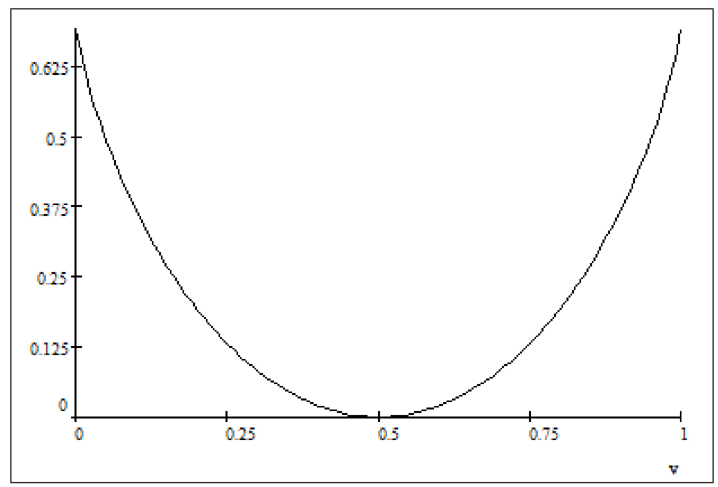


Fig. 10.3. A plot of the rate function, $I_{1/2}$.

Proof. By definition of the binomial distribution,

$$P_x\left(\frac{S_n}{n}\le y\right)=P_x\left(S_n\le ny\right)=\sum_{k\le ny}\binom{n}{k}x^k\left(1-x\right)^{n-k}.$$

If $a_k\ge 0$, then we have the following crude estimates on $\sum_{k=0}^{m-1}a_k$,

$$\max_{k<m}a_k\le\sum_{k=0}^{m-1}a_k\le m\cdot\max_{k<m}a_k. \tag{10.55}$$

In order to apply this with $a_k=\binom{n}{k}x^k\left(1-x\right)^{n-k}$ and $m=[ny]$, we need to find the maximum of the a_k for $0\le k\le ny$. This is easy to do since a_k is increasing for $0\le k\le ny$ as we now show. Consider,

$$\begin{aligned}\frac{a_{k+1}}{a_k}&=\frac{\binom{n}{k+1}x^{k+1}\left(1-x\right)^{n-k-1}}{\binom{n}{k}x^k\left(1-x\right)^{n-k}}\\&=\frac{k!\left(n-k\right)!\cdot x}{\left(k+1\right)!\cdot\left(n-k-1\right)!\cdot\left(1-x\right)}\\&=\frac{\left(n-k\right)\cdot x}{\left(k+1\right)\cdot\left(1-x\right)}.\end{aligned}$$

Therefore, where the latter expression is greater than or equal to 1 iff

$$\begin{aligned}\frac{a_{k+1}}{a_k}\ge 1&\iff\left(n-k\right)\cdot x\ge\left(k+1\right)\cdot\left(1-x\right)\\&\iff nx\ge k+1-x\iff k<\left(n-1\right)x-1.\end{aligned}$$

Thus for $k<\left(n-1\right)x-1$ we may conclude that $\binom{n}{k}x^k\left(1-x\right)^{n-k}$ is increasing in k.

Thus the crude bound in Eq. (10.55) implies,

$$\binom{n}{[ny]}x^{[ny]}\left(1-x\right)^{n-[ny]}\le P_x\left(\frac{S_n}{n}\le y\right)\le[ny]\binom{n}{[ny]}x^{[ny]}\left(1-x\right)^{n-[ny]}$$

or equivalently,

$$\begin{aligned}\frac{1}{n}\ln\left[\binom{n}{[ny]}x^{[ny]}\left(1-x\right)^{n-[ny]}\right]&\\&\le\frac{1}{n}\ln P_x\left(\frac{S_n}{n}\le y\right)\\&\le\frac{1}{n}\ln\left[\left(ny\right)\binom{n}{[ny]}x^{[ny]}\left(1-x\right)^{n-[ny]}\right].\end{aligned}$$

By Stirling's formula, for k such that k and $n-k$ is large we have,

$$\binom{n}{k}\sim\frac{1}{\sqrt{2\pi}}\frac{n^{n+1/2}}{k^{k+1/2}\cdot\left(n-k\right)^{n-k+1/2}}=\frac{\sqrt{n}}{\sqrt{2\pi}}\frac{1}{\left(\frac{k}{n}\right)^{k+1/2}\cdot\left(1-\frac{k}{n}\right)^{n-k+1/2}}$$

and therefore,

$$\frac{1}{n}\ln\binom{n}{k}\sim-\frac{k}{n}\ln\left(\frac{k}{n}\right)-\left(1-\frac{k}{n}\right)\ln\left(1-\frac{k}{n}\right).$$

So taking $k=[ny]$, we learn that

$$\lim_{n\to\infty}\frac{1}{n}\ln\binom{n}{[ny]}=-y\ln y-\left(1-y\right)\ln\left(1-y\right)$$

and therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\ln P_x\left(\frac{S_n}{n}\le y\right)&=-y\ln y-\left(1-y\right)\ln\left(1-y\right)+y\ln x+\left(1-y\right)\ln\left(1-x\right)\\&=y\ln\frac{x}{y}+\left(1-y\right)\ln\left(\frac{1-x}{1-y}\right).\end{aligned}$$

■

As a consistency check it is worth noting, by Jensen's inequality described below, that

$$-I_x\left(y\right)=y\ln\frac{x}{y}+\left(1-y\right)\ln\left(\frac{1-x}{1-y}\right)\le\ln\left(y\frac{x}{y}+\left(1-y\right)\frac{1-x}{1-y}\right)=\ln\left(1\right)=0.$$

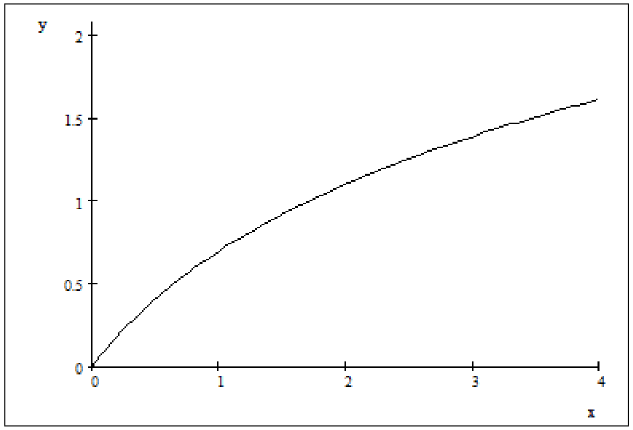
This must be the case since

$$-I_x\left(y\right)=\lim_{n\to\infty}\frac{1}{n}\ln P_x\left(\frac{S_n}{n}\le y\right)\le\lim_{n\to\infty}\frac{1}{n}\ln 1=0.$$

10.6.2 A primitive Stirling type approximation

Theorem 10.61. *Suppose that $f:(0,\infty)\to\mathbb{R}$ is an increasing concave down function (like $f\left(x\right)=\ln x$) and let $s_n:=\sum_{k=1}^{n}f\left(k\right)$, then*

$$\begin{aligned}s_n-\frac{1}{2}\left(f\left(n\right)+f\left(1\right)\right)&\le\int_1^n f\left(x\right)dx\\&\le s_n-\frac{1}{2}\left[f\left(n+1\right)+2f\left(1\right)\right]+\frac{1}{2}f\left(2\right)\\&\le s_n-\frac{1}{2}\left[f\left(n\right)+2f\left(1\right)\right]+\frac{1}{2}f\left(2\right).\end{aligned}$$



Proof. On the interval, $[k-1,k]$, we have that $f(x)$ is larger than the straight line segment joining $(k-1, f(k-1))$ and $(k, f(k))$ and thus

$$\frac{1}{2}(f(k)+f(k-1)) \le \int_{k-1}^{k} f(x)\,dx.$$

Summing this equation on $k=2,\dots,n$ shows,

$$\begin{aligned} s_n - \frac{1}{2}(f(n)+f(1)) &= \sum_{k=2}^{n} \frac{1}{2}(f(k)+f(k-1)) \\ &\le \sum_{k=2}^{n} \int_{k-1}^{k} f(x)\,dx = \int_1^n f(x)\,dx. \end{aligned}$$

For the upper bound on the integral we observe that $f(x) \le f(k) - f'(k)(x-k)$ for all x and therefore,

$$\int_{k-1}^{k} f(x)\,dx \le \int_{k-1}^{k} [f(k) - f'(k)(x-k)]\,dx = f(k) - \frac{1}{2} f'(k).$$

Summing this equation on $k=2,\dots,n$ then implies,

$$\int_1^n f(x)\,dx \le \sum_{k=2}^{n} f(k) - \frac{1}{2}\sum_{k=2}^{n} f'(k).$$

Since $f''(x) \le 0$, $f'(x)$ is decreasing and therefore $f'(x) \le f'(k-1)$ for $x \in [k-1,k]$ and integrating this equation over $[k-1,k]$ gives

$$f(k) - f(k-1) \le f'(k-1).$$

Summing the result on $k=3,\dots,n+1$ then shows,

$$f(n+1) - f(2) \le \sum_{k=2}^{n} f'(k)$$

and thus ti follows that

$$\begin{aligned} \int_1^n f(x)\,dx &\le \sum_{k=2}^{n} f(k) - \frac{1}{2}(f(n+1) - f(2)) \\ &= s_n - \frac{1}{2}[f(n+1) + 2f(1)] + \frac{1}{2} f(2) \\ &\le s_n - \frac{1}{2}[f(n) + 2f(1)] + \frac{1}{2} f(2) \end{aligned}$$

■

Example 10.62 (Approximating $n!$). Let us take $f(n) = \ln n$ and recall that

$$\int_1^n \ln x dx = n \ln n - n + 1.$$

Thus we may conlcud that

$$s_n - \frac{1}{2}\ln n \le n \ln n - n + 1 \le s_n - \frac{1}{2}\ln n + \frac{1}{2}\ln 2.$$

Thus it follows that

$$\left(n + \frac{1}{2}\right)\ln n - n + 1 - \ln\sqrt{2} \le s_n \le \left(n + \frac{1}{2}\right)\ln n - n + 1.$$

Exponentiating this identity then implies,

$$\frac{e}{\sqrt{2}} \cdot e^{-n} n^{n+1/2} \le n! \le e \cdot e^{-n} n^{n+1/2}$$

which compares well with Strirling's formula (Theorem 10.57) which states,

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

Observe that

$$\frac{e}{\sqrt{2}} \cong 1.922\,1 \le \sqrt{2\pi} \cong 2.506 \le e \cong 2.718\,3.$$

10.7 Comparison of the Lebesgue and the Riemann Integral*

For the rest of this chapter, let $-\infty < a < b < \infty$ and $f : [a,b] \to \mathbb{R}$ be a bounded function. A partition of $[a,b]$ is a finite subset $\pi \subset [a,b]$ containing $\{a,b\}$. To each partition

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{10.56}$$

of $[a, b]$ let

$$\text{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1, \ldots, n\},$$

$$M_j = \sup\{f(x) : t_j \le x \le t_{j-1}\}, \;\; m_j = \inf\{f(x) : t_j \le x \le t_{j-1}\}$$

$$G_\pi = f(a)1_{\{a\}} + \sum_1^n M_j 1_{(t_{j-1}, t_j]}, \;\; g_\pi = f(a)1_{\{a\}} + \sum_1^n m_j 1_{(t_{j-1}, t_j]} \text{ and}$$

$$S_\pi f = \sum M_j (t_j - t_{j-1}) \text{ and } s_\pi f = \sum m_j (t_j - t_{j-1}).$$

Notice that

$$S_\pi f = \int_a^b G_\pi dm \text{ and } s_\pi f = \int_a^b g_\pi dm.$$

The upper and lower Riemann integrals are defined respectively by

$$\overline{\int_a^b} f(x)dx = \inf_\pi S_\pi f \text{ and } \underline{\int_b^a} f(x)dx = \sup_\pi \;\; s_\pi f.$$

Definition 10.63. *The function f is **Riemann integrable** iff $\overline{\int_a^b} f = \underline{\int_a^b} f \in \mathbb{R}$ and which case the Riemann integral $\int_a^b f$ is defined to be the common value:*

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

The proof of the following Lemma is left to the reader as Exercise 10.28.

Lemma 10.64. *If π' and π are two partitions of $[a, b]$ and $\pi \subset \pi'$ then*

$$G_\pi \ge G_{\pi'} \ge f \ge g_{\pi'} \ge g_\pi \text{ and}$$
$$S_\pi f \ge S_{\pi'} f \ge s_{\pi'} f \ge s_\pi f.$$

There exists an increasing sequence of partitions $\{\pi_k\}_{k=1}^\infty$ such that $\text{mesh}(\pi_k) \downarrow 0$ and

$$S_{\pi_k} f \downarrow \overline{\int_a^b} f \text{ and } s_{\pi_k} f \uparrow \underline{\int_a^b} f \text{ as } k \to \infty.$$

If we let

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{10.57}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} gdm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{10.58}$$

and

$$\int_{[a,b]} Gdm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{10.59}$$

Notation 10.65 *For $x \in [a, b]$, let*

$$H(x) = \limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \;\sup\{f(y) : |y - x| \le \varepsilon, \;\; y \in [a, b]\} \text{ and}$$
$$h(x) = \liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \;\inf\;\{f(y) : |y - x| \le \varepsilon, \;\; y \in [a, b]\}.$$

Lemma 10.66. *The functions $H, h : [a, b] \to \mathbb{R}$ satisfy:*

1. *$h(x) \le f(x) \le H(x)$ for all $x \in [a, b]$ and $h(x) = H(x)$ iff f is continuous at x.*
2. *If $\{\pi_k\}_{k=1}^\infty$ is any increasing sequence of partitions such that $\text{mesh}(\pi_k) \downarrow 0$ and G and g are defined as in Eq. (10.57), then*

$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall \; x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{10.60}$$

(Note π is a countable set.)

3. *H and h are Borel measurable.*

Proof. Let $G_k := G_{\pi_k} \downarrow G$ and $g_k := g_{\pi_k} \uparrow g$.

1. It is clear that $h(x) \le f(x) \le H(x)$ for all x and $H(x) = h(x)$ iff $\lim_{y\to x} f(y)$ exists and is equal to $f(x)$. That is $H(x) = h(x)$ iff f is continuous at x.
2. For $x \notin \pi$,

$$G_k(x) \ge H(x) \ge f(x) \ge h(x) \ge g_k(x) \; \forall \; k$$

and letting $k \to \infty$ in this equation implies

$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x) \; \forall \; x \notin \pi. \tag{10.61}$$

Moreover, given $\varepsilon > 0$ and $x \notin \pi$,

$$\sup\{f(y) : |y - x| \le \varepsilon, \;\; y \in [a, b]\} \ge G_k(x)$$

for all k large enough, since eventually $G_k(x)$ is the supremum of $f(y)$ over some interval contained in $[x - \varepsilon, x + \varepsilon]$. Again letting $k \to \infty$ implies $\sup\limits_{|y-x|\le\varepsilon} f(y) \ge G(x)$ and therefore, that

$$H(x) = \limsup_{y\to x} f(y) \ge G(x)$$

for all $x \notin \pi$. Combining this equation with Eq. (10.61) then implies $H(x) = G(x)$ if $x \notin \pi$. A similar argument shows that $h(x) = g(x)$ if $x \notin \pi$ and hence Eq. (10.60) is proved.

3. The functions G and g are limits of measurable functions and hence measurable. Since $H = G$ and $h = g$ except possibly on the countable set π, both H and h are also Borel measurable. (You justify this statement.)

■

Theorem 10.67. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{10.62}$$

and the following statements are equivalent:

1. *$H(x) = h(x)$ for m -a.e. x,*
2. *the set*
$$E := \{x \in [a,b] : f \text{ is discontinuous at } x\}$$
is an $\bar{m}$ – null set.
3. *f is Riemann integrable.*

If f is Riemann integrable then f is Lebesgue measurable[2], i.e. f is $\mathcal{L}/\mathcal{B}$ – measurable where $\mathcal{L}$ is the Lebesgue σ – algebra and $\mathcal{B}$ is the Borel σ – algebra on $[a,b]$. Moreover if we let $\bar{m}$ denote the completion of m, then

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{10.63}$$

Proof. Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 10.64 and let G and g be defined as in Lemma 10.66. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (10.62) is a consequence of Eqs. (10.58) and (10.59). From Eq. (10.62), f is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

and because $h \le f \le H$ this happens iff $h(x) = H(x)$ for m - a.e. x. Since $E = \{x : H(x) \neq h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (10.60), the remaining assertions including Eq. (10.63) are now consequences of Lemma 10.70. ■

Notation 10.68 *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

[2] f need not be Borel measurable.

10.8 Measurability on Complete Measure Spaces*

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

Proposition 10.69. *Suppose that $(X, \mathcal{B}, \mu)$ is a complete measure space[3] and $f : X \to \mathbb{R}$ is measurable.*

1. *If $g : X \to \mathbb{R}$ is a function such that $f(x) = g(x)$ for μ – a.e. x, then g is measurable.*
2. *If $f_n : X \to \mathbb{R}$ are measurable and $f : X \to \mathbb{R}$ is a function such that $\lim_{n\to\infty} f_n = f$, μ-a.e., then f is measurable as well.*

Proof. 1. Let $E = \{x : f(x) \neq g(x)\}$ which is assumed to be in $\mathcal{B}$ and $\mu(E) = 0$. Then $g = 1_{E^c} f + 1_E g$ since $f = g$ on E^c. Now $1_{E^c} f$ is measurable so g will be measurable if we show $1_E g$ is measurable. For this consider,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{10.64}$$

Since $(1_E g)^{-1}(B) \subset E$ if $0 \notin B$ and $\mu(E) = 0$, it follow by completeness of $\mathcal{B}$ that $(1_E g)^{-1}(B) \in \mathcal{B}$ if $0 \notin B$. Therefore Eq. (10.64) shows that $1_E g$ is measurable. 2. Let $E = \{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, g is measurable. Because $f = g$ on E^c and $\mu(E) = 0$, $f = g$ a.e. so by part 1. f is also measurable. ■

The above results are in general false if $(X, \mathcal{B}, \mu)$ is not complete. For example, let $X = \{0, 1, 2\}$, $\mathcal{B} = \{\{0\}, \{1, 2\}, X, \varphi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet g is not measurable.

Lemma 10.70. *Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and $\bar{\mathcal{M}}$ is the completion of $\mathcal{M}$ relative to μ and $\bar{\mu}$ is the extension of μ to $\bar{\mathcal{M}}$. Then a function $f : X \to \mathbb{R}$ is $(\bar{\mathcal{M}}, \mathcal{B} = \mathcal{B}_\mathbb{R})$ – measurable iff there exists a function $g : X \to \mathbb{R}$ that is $(\mathcal{M}, \mathcal{B})$ – measurable such $E = \{x : f(x) \neq g(x)\} \in \bar{\mathcal{M}}$ and $\bar{\mu}(E) = 0$, i.e. $f(x) = g(x)$ for $\bar{\mu}$ – a.e. x. Moreover for such a pair f and g, $f \in L^1(\bar{\mu})$ iff $g \in L^1(\mu)$ and in which case*

$$\int_X f d\bar{\mu} = \int_X g d\mu.$$

Proof. Suppose first that such a function g exists so that $\bar{\mu}(E) = 0$. Since g is also $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, we see from Proposition 10.69 that f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may

[3] Recall this means that if $N \subset X$ is a set such that $N \subset A \in \mathcal{M}$ and $\mu(A) = 0$, then $N \in \mathcal{M}$ as well.

assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ as $n \to \infty$. Writing

$$\varphi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \setminus B_k) = 0$. Letting

$$\tilde{\varphi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\varphi}_n \geq 0$ such that $E_n := \{\varphi_n \neq \tilde{\varphi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \cdot \tilde{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (10.65) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{10.65}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (10.65) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 9.39 it holds for all $\mathcal{M}$ – measurable functions $g : X \to [0, \infty]$. The rest of the assertions follow in the standard way by considering $(\operatorname{Re} g)_\pm$ and $(\operatorname{Im} g)_\pm$. ■

10.9 More Exercises

Exercise 10.18. Let μ be a measure on an algebra $\mathcal{A} \subset 2^X$, then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for all $A, B \in \mathcal{A}$.

Exercise 10.19 (From problem 12 on p. 27 of Folland.). Let $(X, \mathcal{M}, \mu)$ be a finite measure space and for $A, B \in \mathcal{M}$ let $\rho(A, B) = \mu(A \Delta B)$ where $A \Delta B = (A \setminus B) \cup (B \setminus A)$. It is clear that $\rho(A, B) = \rho(B, A)$. Show:

1. ρ satisfies the triangle inequality:
$$\rho(A, C) \leq \rho(A, B) + \rho(B, C) \text{ for all } A, B, C \in \mathcal{M}.$$
2. Define $A \sim B$ iff $\mu(A \Delta B) = 0$ and notice that $\rho(A, B) = 0$ iff $A \sim B$. Show "$\sim$" is an equivalence relation.
3. Let $\mathcal{M}/\sim$ denote $\mathcal{M}$ modulo the equivalence relation, $\sim$, and let $[A] := \{B \in \mathcal{M} : B \sim A\}$. Show that $\bar{\rho}([A], [B]) := \rho(A, B)$ is gives a well defined metric on $\mathcal{M}/\sim$.
4. Similarly show $\tilde{\mu}([A]) = \mu(A)$ is a well defined function on $\mathcal{M}/\sim$ and show $\tilde{\mu} : (\mathcal{M}/\sim) \to \mathbb{R}_+$ is $\bar{\rho}$ – continuous.

Exercise 10.20. Suppose that $\mu_n : \mathcal{M} \to [0, \infty]$ are measures on $\mathcal{M}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{M}$. Prove that $\mu : \mathcal{M} \to [0, \infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Exercise 10.21. Now suppose that Λ is some index set and for each $\lambda \in \Lambda$, $\mu_\lambda : \mathcal{M} \to [0, \infty]$ is a measure on $\mathcal{M}$. Define $\mu : \mathcal{M} \to [0, \infty]$ by $\mu(A) = \sum_{\lambda \in \Lambda} \mu_\lambda(A)$ for each $A \in \mathcal{M}$. Show that μ is also a measure.

Exercise 10.22. Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, show

$$\mu(\{A_n \text{ a.a.}\}) \leq \liminf_{n\to\infty} \mu(A_n)$$

and if $\mu(\cup_{m\geq n} A_m) < \infty$ for some n, then

$$\mu(\{A_n \text{ i.o.}\}) \geq \limsup_{n\to\infty} \mu(A_n).$$

Exercise 10.23 (Folland 2.13 on p. 52.). Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of non-negative measurable functions such that $f_n \to f$ pointwise and

$$\lim_{n\to\infty} \int f_n = \int f < \infty.$$

Then

$$\int_E f = \lim_{n\to\infty} \int_E f_n$$

for all measurable sets $E \in \mathcal{M}$. The conclusion need not hold if $\lim_{n\to\infty} \int f_n = \int f$. **Hint:** "Fatou times two."

Exercise 10.24. Give examples of measurable functions $\{f_n\}$ on $\mathbb{R}$ such that f_n decreases to 0 uniformly yet $\int f_n dm = \infty$ for all n. Also give an example of a sequence of measurable functions $\{g_n\}$ on $[0, 1]$ such that $g_n \to 0$ while $\int g_n dm = 1$ for all n.

Exercise 10.25. Suppose $\{a_n\}_{n=-\infty}^\infty \subset \mathbb{C}$ is a summable sequence (i.e. $\sum_{n=-\infty}^\infty |a_n| < \infty$), then $f(\theta) := \sum_{n=-\infty}^\infty a_n e^{in\theta}$ is a continuous function for $\theta \in \mathbb{R}$ and

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta.$$

Exercise 10.26. For any function $f \in L^1(m)$, show $x \in \mathbb{R} \to \int_{(-\infty, x]} f(t)\, dm(t)$ is continuous in x. Also find a finite measure, μ, on $\mathcal{B}_\mathbb{R}$ such that $x \to \int_{(-\infty, x]} f(t)\, d\mu(t)$ is not continuous.

Exercise 10.27. Folland 2.31b and 2.31e on p. 60. (The answer in 2.13b is wrong by a factor of -1 and the sum is on $k = 1$ to ∞. In part (e), s should be taken to be a. You may also freely use the Taylor series expansion

$$(1-z)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{2^n n!} z^n = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} z^n \text{ for } |z| < 1.$$

Exercise 10.28. Prove Lemma 10.64.

Functional Forms of the $\pi - \lambda$ Theorem

In this chapter we will develop a very useful function analogue of the $\pi - \lambda$ theorem. The results in this section will be used often in the sequel.

11.1 Multiplicative System Theorems

Notation 11.1 *Let Ω be a set and $\mathbb{H}$ be a subset of the bounded real valued functions on Ω. We say that $\mathbb{H}$ is* ***closed under bounded convergence*** *if; for every sequence, $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $|f_n(\omega)| \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f(\omega) := \lim_{n\to\infty} f_n(\omega)$ exists for all $\omega \in \Omega$, then $f \in \mathbb{H}$.*

A subset, $\mathbb{M}$, of $\mathbb{H}$ is called a ***multiplicative system*** *if $\mathbb{M}$ is closed under finite products.*

The following result may be found in Dellacherie [2, p. 14]. The style of proof given here may be found in Janson [10, Appendix A., p. 309].

Theorem 11.2 (Dynkin's Multiplicative System Theorem). *Suppose that $\mathbb{H}$ is a vector subspace of bounded functions from Ω to $\mathbb{R}$ which contains the constant functions and is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions. In short, $\sigma(\mathbb{M})_b \subset \mathbb{H}$ where we are using $\sigma(\mathbb{M})_b$ denote the bounded real valued $\sigma(\mathbb{M})$ – measurable functions on Ω.*

Proof. In this proof, we may (and do) assume that $\mathbb{H}$ is the smallest subspace of bounded functions on Ω which contains the constant functions, contains $\mathbb{M}$, and is closed under bounded convergence. (As usual such a space exists by taking the intersection of all such spaces.) The remainder of the proof will be broken into four steps.

Step 1. ($\mathbb{H}$ is an algebra of functions.) For $f \in \mathbb{H}$, let $\mathbb{H}^f := \{g \in \mathbb{H} : gf \in \mathbb{H}\}$. The reader will now easily verify that $\mathbb{H}^f$ is a linear subspace of $\mathbb{H}$, $1 \in \mathbb{H}^f$, and $\mathbb{H}^f$ is closed under bounded convergence. Moreover if $f \in \mathbb{M}$, since $\mathbb{M}$ is a multiplicative system, $\mathbb{M} \subset \mathbb{H}^f$. Hence by the definition of $\mathbb{H}$, $\mathbb{H} = \mathbb{H}^f$, i.e. $fg \in \mathbb{H}$ for all $f \in \mathbb{M}$ and $g \in \mathbb{H}$. Having proved this it now follows for any $f \in \mathbb{H}$ that $\mathbb{M} \subset \mathbb{H}^f$ and therefore as before, $\mathbb{H}^f = \mathbb{H}$. Thus we may conclude that $fg \in \mathbb{H}$ whenever $f, g \in \mathbb{H}$, i.e. $\mathbb{H}$ is an algebra of functions.

Step 2. ($\mathcal{B} := \{A \subset \Omega : 1_A \in \mathbb{H}\}$ is a σ – algebra.) Using the fact that $\mathbb{H}$ is an algebra containing constants, the reader will easily verify that $\mathcal{B}$ is closed under complementation, finite intersections, and contains Ω, i.e. $\mathcal{B}$ is an algebra. Using the fact that $\mathbb{H}$ is closed under bounded convergence, it follows that $\mathcal{B}$ is closed under increasing unions and hence that $\mathcal{B}$ is σ – algebra.

Step 3. ($\mathcal{B}_b \subset \mathbb{H}$, i.e. $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.) Since $\mathbb{H}$ is a vector space and $\mathbb{H}$ contains 1_A for all $A \in \mathcal{B}$, $\mathbb{H}$ contains all $\mathcal{B}$ – measurable simple functions. Since every bounded $\mathcal{B}$ – measurable function may be written as a bounded limit of such simple functions (see Theorem 9.39), it follows that $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.

Step 4. ($\sigma(\mathbb{M}) \subset \mathcal{B}$.) Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 11.1 below) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f \in \mathbb{M}$ and $a \in \mathbb{R}$, let $F_n := \varphi_n(f - a)$ and $M := \sup_{\omega\in\Omega} |f(\omega) - a|$. By the Weierstrass approximation Theorem 7.36, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$. Since p_l is a polynomial and $\mathbb{H}$ is an algebra, $p_l(f - a) \in \mathbb{H}$ for all l. Moreover, $p_l \circ (f - a) \to F_n$ uniformly as $l \to \infty$, from with it follows that $F_n \in \mathbb{H}$ for all n. Since, $F_n \uparrow 1_{\{f>a\}}$ it follows that $1_{\{f>a\}} \in \mathbb{H}$, i.e. $\{f > a\} \in \mathcal{B}$. As the sets $\{f > a\}$ with $a \in \mathbb{R}$ and $f \in \mathbb{M}$ generate $\sigma(\mathbb{M})$, it follows that $\sigma(\mathbb{M}) \subset \mathcal{B}$.

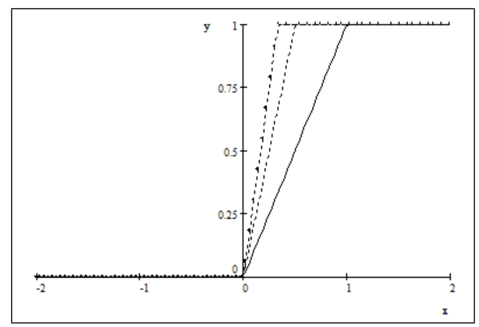


Fig. 11.1. Plots of φ_1, φ_2 and φ_3.

Step 5. From steps 3. and 4. we have, $\sigma(\mathbb{M})_b \subset \mathcal{B}_b \subset \mathbb{H}$ which completes the proof.

■

Corollary 11.3. *Suppose $\mathbb{H}$ is a subspace of bounded real valued functions such that $1 \in \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. If $\mathcal{P} \subset 2^{\Omega}$ is a multiplicative class such that $1_A \in \mathbb{H}$ for all $A \in \mathcal{P}$, then $\mathbb{H}$ contains all bounded $\sigma(\mathcal{P})$ – measurable functions.*

Proof. Let $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$. Then $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system and the proof is completed with an application of Theorem 11.2. ■

Example 11.4. Suppose μ and ν are two probability measure on $(\Omega, \mathcal{B})$ such that

$$\int_{\Omega} f d\mu = \int_{\Omega} f d\nu \tag{11.1}$$

for all f in a multiplicative subset, $\mathbb{M}$, of bounded measurable functions on Ω. Then $\mu = \nu$ on $\sigma(\mathbb{M})$. Indeed, apply Theorem 11.2 with $\mathbb{H}$ being the bounded measurable functions on Ω such that Eq. (11.1) holds. In particular if $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$ with $\mathcal{P}$ being a multiplicative class we learn that $\mu = \nu$ on $\sigma(\mathbb{M}) = \sigma(\mathcal{P})$.

Example 11.5. Suppose μ is a measure and $\rho \in L^1(\mu) = L^1(\Omega, \mathcal{B}, \mu)$ (ρ is a complex integrable function). Suppose that $\int_\Omega f \rho d\mu = 0$ for all f in a multiplicative subset, $\mathbb{M}$, of bounded real measurable functions on Ω and there exists $f_n \in \mathbb{M}$ such that $f_n \to 1$ boundedly. If $\mathcal{B} = \sigma(\mathbb{M})$, then by Theorem 11.2, $\int_\Omega f \rho d\mu = 0$ for all bounded real measurable functions on $(\Omega, \mathcal{B})$. Using the linearity of the integral it now follows that $\int_\Omega f \rho d\mu = 0$ for all bounded complex measurable functions on $(\Omega, \mathcal{B})$. Taking

$$f = \operatorname{sgn}(\rho) = \begin{cases} \frac{\bar{\rho}}{\rho} \text{ if } \rho \neq 0 \\ 1 \text{ if } \rho = 0 \end{cases}$$

implies,

$$0 = \int_\Omega |\rho| 1_{\rho \neq 0} d\mu = \int_\Omega |\rho| d\mu,$$

i.e. $\rho = 0$ a.e.

Example 11.6. Let us continue the notation of Example 11.5 but specialize to the case that $\Omega = \mathbb{R}$ and $\mathcal{B} = \mathcal{B}_{\mathbb{R}}$. In this case we might take $\mathbb{M} = \left\{1_{(a,b]} : -\infty < a < b < \infty\right\}$ or $\mathbb{M} = C_c(\mathbb{R}, \mathbb{R})$. In the first case we then conclude that $\int_{(a,b]} \rho d\mu = 0$ for all $-\infty < a < b < \infty$ implies $\rho = 0$ (μ – a.e.) and in the second if $\int_{\mathbb{R}} f \rho d\mu = 0$ for all $f \in C_c(\mathbb{R}, \mathbb{R})$, then $\rho = 0$ (μ – a.e.). See Corollary 11.9 below where it is shown that $\sigma(C_c(\mathbb{R}, \mathbb{R})) = \mathcal{B}_{\mathbb{R}}$.

Exercise 11.1. Let $\Omega := \{1, 2, 3, 4\}$ and $\mathbb{M} := \{1_A, 1_B\}$ where $A := \{1, 2\}$ and $B := \{2, 3\}$.

a) Show $\sigma(\mathbb{M}) = 2^{\Omega}$.

b) Find two distinct probability measures, μ and ν on 2^{Ω} such that $\mu(A) = \nu(A)$ and $\mu(B) = \nu(B)$, i.e.

$$\int_\Omega f d\mu = \int_\Omega f d\nu \tag{11.2}$$

holds for all $f \in \mathbb{M}$.

Moral: the assumption that $\mathbb{M}$ is multiplicative can not be dropped from multiplicative system theorem.

Here is a complex version of Theorem 11.2.

Theorem 11.7 (Complex Multiplicative System Theorem). *Suppose $\mathbb{H}$ is a complex linear subspace of the bounded complex functions on Ω, $1 \in \mathbb{H}$, $\mathbb{H}$ is closed under complex conjugation, and $\mathbb{H}$ is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system which is closed under conjugation, then $\mathbb{H}$ contains all bounded complex valued $\sigma(\mathbb{M})$-measurable functions.*

Proof. Let $\mathbb{M}_0 = \operatorname{span}_{\mathbb{C}}(\mathbb{M} \cup \{1\})$ be the complex span of $\mathbb{M}$. As the reader should verify, $\mathbb{M}_0$ is an algebra, $\mathbb{M}_0 \subset \mathbb{H}$, $\mathbb{M}_0$ is closed under complex conjugation and $\sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$. Let

$$\mathbb{H}^{\mathbb{R}} := \{f \in \mathbb{H} : f \text{ is real valued}\} \text{ and}$$
$$\mathbb{M}_0^{\mathbb{R}} := \{f \in \mathbb{M}_0 : f \text{ is real valued}\}.$$

Then $\mathbb{H}^{\mathbb{R}}$ is a real linear space of bounded real valued functions 1 which is closed under bounded convergence and $\mathbb{M}_0^{\mathbb{R}} \subset \mathbb{H}^{\mathbb{R}}$. Moreover, $\mathbb{M}_0^{\mathbb{R}}$ is a multiplicative system (as the reader should check) and therefore by Theorem 11.2, $\mathbb{H}^{\mathbb{R}}$ contains all bounded $\sigma\left(\mathbb{M}_0^{\mathbb{R}}\right)$ – measurable real valued functions. Since $\mathbb{H}$ and $\mathbb{M}_0$ are complex linear spaces closed under complex conjugation, for any $f \in \mathbb{H}$ or $f \in \mathbb{M}_0$, the functions $\operatorname{Re} f = \frac{1}{2}\left(f + \bar{f}\right)$ and $\operatorname{Im} f = \frac{1}{2i}\left(f - \bar{f}\right)$ are in $\mathbb{H}$ or $\mathbb{M}_0$ respectively. Therefore $\mathbb{M}_0 = \mathbb{M}_0^{\mathbb{R}} + i\mathbb{M}_0^{\mathbb{R}}$, $\sigma\left(\mathbb{M}_0^{\mathbb{R}}\right) = \sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$, and $\mathbb{H} = \mathbb{H}^{\mathbb{R}} + i\mathbb{H}^{\mathbb{R}}$. Hence if $f : \Omega \to \mathbb{C}$ is a bounded $\sigma(\mathbb{M})$ – measurable function, then $f = \operatorname{Re} f + i \operatorname{Im} f \in \mathbb{H}$ since $\operatorname{Re} f$ and $\operatorname{Im} f$ are in $\mathbb{H}^{\mathbb{R}}$. ■

Lemma 11.8. *Suppose that $-\infty < a < b < \infty$ and let Trig($\mathbb{R}$) $\subset C(\mathbb{R}, \mathbb{C})$ be the complex linear span of $\left\{x \to e^{i\lambda x} : \lambda \in \mathbb{R}\right\}$. Then there exists $f_n \in C_c(\mathbb{R}, [0,1])$ and $g_n \in$Trig($\mathbb{R}$) such that $\lim_{n\to\infty} f_n(x) = 1_{(a,b]}(x) = \lim_{n\to\infty} g_n(x)$ for all $x \in \mathbb{R}$.*

Proof. The assertion involving $f_n \in C_c(\mathbb{R}, [0,1])$ was the content of one of your homework assignments. For the assertion involving $g_n \in$Trig($\mathbb{R}$), it will suffice to show that any $f \in C_c(\mathbb{R})$ may be written as $f(x) = \lim_{n\to\infty} g_n(x)$

for some $\{g_n\} \subset \mathrm{Trig}(\mathbb{R})$ where the limit is uniform for x in compact subsets of $\mathbb{R}$.

So suppose that $f \in C_c(\mathbb{R})$ and $L > 0$ such that $f(x) = 0$ if $|x| \geq L/4$. Then

$$f_L(x) := \sum_{n=-\infty}^{\infty} f(x + nL)$$

is a continuous L – periodic function on $\mathbb{R}$, see Figure 11.2. If $\varepsilon > 0$ is given,

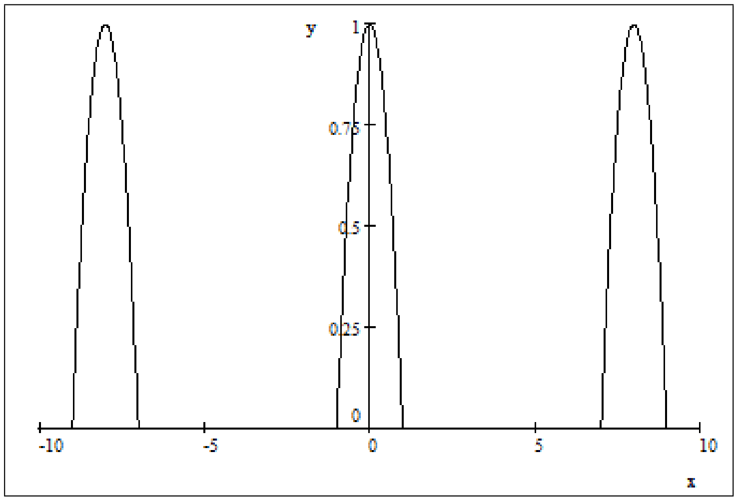


Fig. 11.2. This is plot of $f_8(x)$ where $f(x) = (1 - x^2)\, 1_{|x| \leq 1}$. The center hump by itself would be the plot of $f(x)$.

we may apply Theorem 7.42 to find $\Lambda \subset\subset \mathbb{Z}$ such that

$$\left| f_L\left(\frac{L}{2\pi}x\right) - \sum_{\alpha \in \Lambda} a_\lambda e^{i\alpha x} \right| \leq \varepsilon \text{ for all } x \in \mathbb{R},$$

wherein we have use the fact that $x \to f_L\left(\frac{L}{2\pi}x\right)$ is a 2π – periodic function of x. Equivalently we have,

$$\max_x \left| f_L(x) - \sum_{\alpha \in \Lambda} a_\lambda e^{i\frac{2\pi\alpha}{L}x} \right| \leq \varepsilon.$$

In particular it follows that $f_L(x)$ is a uniform limit of functions from $\mathrm{Trig}(\mathbb{R})$. Since $\lim_{L\to\infty} f_L(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$, it is easy to conclude there exists $g_n \in \mathrm{Trig}(\mathbb{R})$ such that $\lim_{n\to\infty} g_n(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$. ■

Corollary 11.9. *Each of the following σ – algebras on $\mathbb{R}^d$ are equal to $\mathcal{B}_{\mathbb{R}^d}$;*

1. $\mathcal{M}_1 := \sigma\left(\cup_{i=1}^n \{x \to f(x_i) : f \in C_c(\mathbb{R})\}\right)$,
2. $\mathcal{M}_2 := \sigma\left(x \to f_1(x_1) \dots f_d(x_d) : f_i \in C_c(\mathbb{R})\right)$
3. $\mathcal{M}_3 = \sigma\left(C_c\left(\mathbb{R}^d\right)\right)$, *and*
4. $\mathcal{M}_4 := \sigma\left(\left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}\right)$.

Proof. As the functions defining each $\mathcal{M}_i$ are continuous and hence Borel measurable, it follows that $\mathcal{M}_i \subset \mathcal{B}_{\mathbb{R}^d}$ for each i. So to finish the proof it suffices to show $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_i$ for each i.

$\mathcal{M}_1$ case. Let $a, b \in \mathbb{R}$ with $-\infty < a < b < \infty$. By Lemma 11.8, there exists $f_n \in C_c(\mathbb{R})$ such that $\lim_{n\to\infty} f_n = 1_{(a,b]}$. Therefore it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_1$ – measurable for each i. Moreover if $-\infty < a_i < b_i < \infty$ for each i, then we may conclude that

$$x \to \prod_{i=1}^d 1_{(a_i,b_i]}(x_i) = 1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x)$$

is $\mathcal{M}_1$ – measurable as well and hence $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_1$. As such sets generate $\mathcal{B}_{\mathbb{R}^d}$ we may conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_1$.

and therefore $\mathcal{M}_1 = \mathcal{B}_{\mathbb{R}^d}$.

$\mathcal{M}_2$ case. As above, we may find $f_{i,n} \to 1_{(a_i,b_i]}$ as $n \to \infty$ for each $1 \leq i \leq d$ and therefore,

$$1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x) = \lim_{n\to\infty} f_{1,n}(x_1) \dots f_{d,n}(x_d) \text{ for all } x \in \mathbb{R}^d.$$

This shows that $1_{(a_1,b_1]\times\dots\times(a_d,b_d]}$ is $\mathcal{M}_2$ – measurable and therefore $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_2$.

$\mathcal{M}_3$ case. This is easy since $\mathcal{B}_{\mathbb{R}^d} = \mathcal{M}_2 \subset \mathcal{M}_3$.

$\mathcal{M}_4$ case. By Lemma 11.8 here exists $g_n \in \mathrm{Trig}(\mathbb{R})$ such that $\lim_{n\to\infty} g_n = 1_{(a,b]}$. Since $x \to g_n(x_i)$ is in the span $\left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}$ for each n, it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_4$ – measurable for all $-\infty < a < b < \infty$. Therefore, just as in the proof of case 1., we may now conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_4$. ■

Corollary 11.10. *Suppose that $\mathbb{H}$ is a subspace of complex valued functions on $\mathbb{R}^d$ which is closed under complex conjugation and bounded convergence. If $\mathbb{H}$ contains any one of the following collection of functions;*

1. $\mathbb{M} := \{x \to f_1(x_1) \dots f_d(x_d) : f_i \in C_c(\mathbb{R})\}$
2. $\mathbb{M} := C_c\left(\mathbb{R}^d\right)$, *or*
3. $\mathbb{M} := \left\{x \to e^{i\lambda\cdot x} : \lambda \in \mathbb{R}^d\right\}$

then $\mathbb{H}$ *contains all bounded complex Borel measurable functions on* $\mathbb{R}^d$.

Proof. Observe that if $f \in C_c(\mathbb{R})$ such that $f(x) = 1$ in a neighborhood of 0, then $f_n(x) := f(x/n) \to 1$ as $n \to \infty$. Therefore in cases 1. and 2., $\mathbb{H}$ contains the constant function, 1, since

$$1 = \lim_{n\to\infty} f_n(x_1) \dots f_n(x_d).$$

In case 3, $1 \in \mathbb{M} \subset \mathbb{H}$ as well. The result now follows from Theorem 11.7 and Corollary 11.9. ■

Proposition 11.11 (Change of Variables Formula). *Suppose that* $-\infty < a < b < \infty$ *and* $u : [a,b] \to \mathbb{R}$ *is a continuously differentiable function. Let* $[c,d] = u([a,b])$ *where* $c = \min u([a,b])$ *and* $d = \max u([a,b])$. *(By the intermediate value theorem* $u([a,b])$ *is an interval.) Then for all bounded measurable functions,* $f : [c,d] \to \mathbb{R}$ *we have*

$$\int_{u(a)}^{u(b)} f(x)\,dx = \int_a^b f(u(t))\,\dot{u}(t)\,dt. \tag{11.3}$$

Moreover, Eq. (11.3) is also valid if $f : [c,d] \to \mathbb{R}$ *is measurable and*

$$\int_a^b |f(u(t))|\,|\dot{u}(t)|\,dt < \infty. \tag{11.4}$$

Proof. Let $\mathbb{H}$ denote the space of bounded measurable functions such that Eq. (11.3) holds. It is easily checked that $\mathbb{H}$ is a linear space closed under bounded convergence. Next we show that $\mathbb{M} = C([c,d], \mathbb{R}) \subset \mathbb{H}$ which coupled with Corollary 11.10 will show that $\mathbb{H}$ contains all bounded measurable functions from $[c,d]$ to $\mathbb{R}$.

If $f : [c,d] \to \mathbb{R}$ is a continuous function and let F be an anti-derivative of f. Then by the fundamental theorem of calculus,

$$\begin{aligned}\int_a^b f(u(t))\,\dot{u}(t)\,dt &= \int_a^b F'(u(t))\,\dot{u}(t)\,dt \\ &= \int_a^b \frac{d}{dt} F(u(t))\,dt = F(u(t))|_a^b \\ &= F(u(b)) - F(u(a)) = \int_{u(a)}^{u(b)} F'(x)\,dx = \int_{u(a)}^{u(b)} f(x)\,dx.\end{aligned}$$

Thus $\mathbb{M} \subset \mathbb{H}$ and the first assertion of the proposition is proved.

Now suppose that $f : [c,d] \to \mathbb{R}$ is measurable and Eq. (11.4) holds. For $M < \infty$, let $f_M(x) = f(x) \cdot 1_{|f(x)| \le M}$ – a bounded measurable function. Therefore applying Eq. (11.3) with f replaced by $|f_M|$ shows,

$$\left|\int_{u(a)}^{u(b)} |f_M(x)|\,dx\right| = \left|\int_a^b |f_M(u(t))|\,\dot{u}(t)\,dt\right| \le \int_a^b |f_M(u(t))|\,|\dot{u}(t)|\,dt.$$

Using the MCT, we may let $M \uparrow \infty$ in the previous inequality to learn

$$\left|\int_{u(a)}^{u(b)} |f(x)|\,dx\right| \le \int_a^b |f(u(t))|\,|\dot{u}(t)|\,dt < \infty.$$

Now apply Eq. (11.3) with f replaced by f_M to learn

$$\int_{u(a)}^{u(b)} f_M(x)\,dx = \int_a^b f_M(u(t))\,\dot{u}(t)\,dt.$$

Using the DCT we may now let $M \to \infty$ in this equation to show that Eq. (11.3) remains valid. ■

Exercise 11.2. Suppose that $u : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that $\dot{u}(t) \ge 0$ for all t and $\lim_{t\to\pm\infty} u(t) = \pm\infty$. Show that

$$\int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} f(u(t))\,\dot{u}(t)\,dt \tag{11.5}$$

for all measurable functions $f : \mathbb{R} \to [0,\infty]$. In particular applying this result to $u(t) = at + b$ where $a > 0$ implies,

$$\int_{\mathbb{R}} f(x)\,dx = a \int_{\mathbb{R}} f(at+b)\,dt.$$

Definition 11.12. *The* ***Fourier transform*** *or* ***characteristic function*** *of a finite measure,* μ, *on* $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, *is the function,* $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ *defined by*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^d} e^{i\lambda\cdot x}\,d\mu(x) \text{ for all } \lambda \in \mathbb{R}^d$$

Corollary 11.13. *Suppose that* μ *and* ν *are two probability measures on* $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. *Then any one of the next three conditions implies that* $\mu = \nu$;

1. $\int_{\mathbb{R}^d} f_1(x_1)\dots f_d(x_d)\,d\nu(x) = \int_{\mathbb{R}^d} f_1(x_1)\dots f_d(x_d)\,d\mu(x)$ *for all* $f_i \in C_c(\mathbb{R})$.
2. $\int_{\mathbb{R}^d} f(x)\,d\nu(x) = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$ *for all* $f \in C_c(\mathbb{R}^d)$.
3. $\hat{\nu} = \hat{\mu}$.

Item 3. asserts that the Fourier transform is injective.

Proof. Let $\mathbb{H}$ be the collection of bounded complex measurable functions from $\mathbb{R}^d$ to $\mathbb{C}$ such that

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\nu. \tag{11.6}$$

It is easily seen that $\mathbb{H}$ is a linear space closed under complex conjugation and bounded convergence (by the DCT). Since $\mathbb{H}$ contains one of the multiplicative systems appearing in Corollary 11.10, it contains all bounded Borel measurable functions form $\mathbb{R}^d \to \mathbb{C}$. Thus we may take $f = 1_A$ with $A \in \mathcal{B}_{\mathbb{R}^d}$ in Eq. (11.6) to learn, $\mu(A) = \nu(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$. ∎

In many cases we can replace the condition in item 3. of Corollary 11.13 by;

$$\int_{\mathbb{R}^d} e^{\lambda \cdot x} d\mu(x) = \int_{\mathbb{R}^d} e^{\lambda \cdot x} d\nu(x) \text{ for all } \lambda \in U, \tag{11.7}$$

where U is a neighborhood of $0 \in \mathbb{R}^d$. In order to do this, one must assume at least assume that the integrals involved are finite for all $\lambda \in U$. The idea is to show that Condition 11.7 implies $\hat{\nu} = \hat{\mu}$. You are asked to carry out this argument in Exercise 11.3 making use of the following lemma.

Lemma 11.14 (Analytic Continuation). *Let $\varepsilon > 0$ and $S_\varepsilon := \{x + iy \in \mathbb{C} : |x| < \varepsilon\}$ be an ε strip in $\mathbb{C}$ about the imaginary axis. Suppose that $h : S_\varepsilon \to \mathbb{C}$ is a function such that for each $b \in \mathbb{R}$, there exists $\{c_n(b)\}_{n=0}^\infty \subset \mathbb{C}$ such that*

$$h(z + ib) = \sum_{n=0}^\infty c_n(b) z^n \text{ for all } |z| < \varepsilon. \tag{11.8}$$

If $c_n(0) = 0$ for all $n \in \mathbb{N}_0$, then $h \equiv 0$.

Proof. It suffices to prove the following assertion; if for some $b \in \mathbb{R}$ we know that $c_n(b) = 0$ for all n, then $c_n(y) = 0$ for all n and $y \in (b - \varepsilon, b + \varepsilon)$. We now prove this assertion.

Let us assume that $b \in \mathbb{R}$ and $c_n(b) = 0$ for all $n \in \mathbb{N}_0$. It then follows from Eq. (11.8) that $h(z + ib) = 0$ for all $|z| < \varepsilon$. Thus if $|y - b| < \varepsilon$, we may conclude that $h(x + iy) = 0$ for x in a (possibly very small) neighborhood $(-\delta, \delta)$ of 0. Since

$$\sum_{n=0}^\infty c_n(y) x^n = h(x + iy) = 0 \text{ for all } |x| < \delta,$$

it follows that

$$0 = \frac{1}{n!} \frac{d^n}{dx^n} h(x + iy)|_{x=0} = c_n(y)$$

and the proof is complete. ∎

11.2 Exercises

Exercise 11.3. Suppose $\varepsilon > 0$ and X and Y are two random variables such that $\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{tY}\right] < \infty$ for all $|t| \le \varepsilon$. Show;

1. $\mathbb{E}\left[e^{\varepsilon|X|}\right]$ and $\mathbb{E}\left[e^{\varepsilon|Y|}\right]$ are finite.
2. $\mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[e^{itY}\right]$ for all $t \in \mathbb{R}$. **Hint:** Consider $h(z) := \mathbb{E}\left[e^{zX}\right] - \mathbb{E}\left[e^{zY}\right]$ for $z \in S_\varepsilon$. Now show for $|z| \le \varepsilon$ and $b \in \mathbb{R}$, that
$$h(z + ib) = \mathbb{E}\left[e^{ibX} e^{zX}\right] - \mathbb{E}\left[e^{ibY} e^{zY}\right] = \sum_{n=0}^\infty c_n(b) z^n \tag{11.9}$$
where
$$c_n(b) := \frac{1}{n!}\left(\mathbb{E}\left[e^{ibX} X^n\right] - \mathbb{E}\left[e^{ibY} Y^n\right]\right). \tag{11.10}$$
3. Conclude from item 2. that $X \overset{d}{=} Y$, i.e. that $\mathrm{Law}_P(X) = \mathrm{Law}_P(Y)$.

Exercise 11.4. Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ be a pair of random variables such that

$$\mathbb{E}[f(X) g(Y)] = \mathbb{E}[f(X) g(X)]$$

for every pair of bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$. Show $P(X = Y) = 1$. **Hint:** Let $\mathbb{H}$ denote the bounded Borel measurable functions, $h : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathbb{E}[h(X, Y)] = \mathbb{E}[h(X, X)].$$

Use the multiplicative systems Theorem 11.2 to show $\mathbb{H}$ is the vector space of all bounded Borel measurable functions. Then take $h(x, y) = 1_{\{x=y\}}$.

Exercise 11.5 (Density of $\mathcal{A}$ – simple functions). Let $(\Omega, \mathcal{B}, P)$ be a probability space and assume that $\mathcal{A}$ is a sub-algebra of $\mathcal{B}$ such that $\mathcal{B} = \sigma(\mathcal{A})$. Let $\mathbb{H}$ denote the bounded measurable functions $f : \Omega \to \mathbb{R}$ such that for every $\varepsilon > 0$ there exists an an $\mathcal{A}$ – simple function[1], $\varphi : \Omega \to \mathbb{R}$ such that $\mathbb{E}|f - \varphi| < \varepsilon$. Show $\mathbb{H}$ consists of all bounded measurable functions, $f : \Omega \to \mathbb{R}$. **Hint:** let $\mathbb{M}$ denote the collection of $\mathcal{A}$ – simple functions.

Corollary 11.15. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $\{X_n\}_{n=1}^\infty$ is a collection of random variables on Ω, and $\mathcal{B}_\infty := \sigma(X_1, X_2, X_3, \dots)$. Then for all $\varepsilon > 0$ and all bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, there exists an $n \in \mathbb{N}$ and a bounded $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}|f - G(X_1, \dots, X_n)| < \varepsilon$. Moreover we may assume that $\sup_{x \in \mathbb{R}^n} |G(x)| \le M := \sup_{\omega \in \Omega} |f(\omega)|$.*

[1] Recall from Definition 7.10 than f is an $\mathcal{A}$ – simple function if f is a simple function such that $f^{-1}(\{y\}) \in \mathcal{A}$ for all $y \in \mathbb{R}$.

Proof. Apply Exercise 11.5 with $\mathcal{A} := \cup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)$ in order to find an $\mathcal{A}$ – measurable simple function, φ, such that $\mathbb{E}|f - \varphi| < \varepsilon$. By the definition of $\mathcal{A}$ we know that φ is $\sigma(X_1, \dots, X_n)$ – measurable for some $n \in \mathbb{N}$. It now follows by the factorization Lemma 9.40 that $\varphi = G(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$. If necessary, replace G by $[G \wedge M] \vee (-M)$ in order to insure $\sup_{x \in \mathbb{R}^n} |G(x)| \leq M$. ■

Proposition 11.16 (Density of $\mathcal{A}$ in $\mathcal{B} = \sigma(\mathcal{A})$). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and assume that $\mathcal{A}$ is a sub-algebra of $\mathcal{B}$ such that $\mathcal{B} = \sigma(\mathcal{A})$. Then to each $B \in \mathcal{B}$ and $\varepsilon > 0$ there exists a $D \in \mathcal{A}$ such that*

$$P(B \Delta D) = \mathbb{E}|1_B - 1_D| < \varepsilon.$$

Proof. Let $f = 1_B$ and choose an $\mathcal{A}$ – simple function, $\varphi : \Omega \to \mathbb{R}$ such that $\mathbb{E}|f - \varphi| < \varepsilon$ by Exercise 11.5. Let $\lambda_0 = 0$ and $\{\lambda_i\}_{i=1}^n$ be an enumeration of $\varphi(\Omega) \setminus \{0\}$ so that $\varphi = \sum_{i=0}^n \lambda_i 1_{A_i}$ where $A_i := \{\varphi = \lambda_i\}$. Then

$$\begin{aligned}
\mathbb{E}|1_B - \varphi| &= \sum_{i=0}^n \mathbb{E}[1_{A_i}|1_B - \varphi|] = \sum_{i=0}^n \mathbb{E}[1_{A_i}|1_B - \lambda_i|] \\
&= \sum_{i=0}^n \mathbb{E}\left[1_{A_i \cap B}|1 - \lambda_i| + 1_{A_i \setminus B}|\lambda_i|\right] \\
&= P(A_0 \cap B) + \sum_{i=1}^n [|1 - \lambda_i| P(B \cap A_i) + |\lambda_i| P(A_i \setminus B)] \quad (11.11) \\
&\geq P(A_0 \cap B) + \sum_{i=1}^n [|1 - \lambda_i| + |\lambda_i|] \min\{P(B \cap A_i), P(A_i \setminus B)\} \\
&\geq P(A_0 \cap B) + \sum_{i=1}^n \min\{P(B \cap A_i), P(A_i \setminus B)\} \quad (11.12)
\end{aligned}$$

where the last equality is a consequence of the fact that $1 \leq |\lambda_i| + |1 - \lambda_i|$.

Now let $\psi = \sum_{i=0}^n \alpha_i 1_{A_i}$ where $\alpha_0 = 0$ and for $1 \leq i \leq n$,

$$\alpha_i = \begin{cases} 1 \text{ if } P(A_i \setminus B) \leq P(B \cap A_i) \\ 0 \text{ if } P(A_i \setminus B) > P(B \cap A_i) \end{cases}.$$

From Eq. (11.11) with φ replaced by ψ and λ_i by α_i for all i Then show that

$$\mathbb{E}|1_B - \psi| = P(A_0 \cap B) + \sum_{i=1}^n \min\{P(B \cap A_i), P(A_i \setminus B)\} \leq \mathbb{E}|1_B - \varphi|.$$

where the last equality is a conseqnece of Eq. (11.12). Since $\psi = 1_D$ where $D = \cup_{i:\alpha_i=1} A_i \in \mathcal{A}$ we have shown there exists a $D \in \mathcal{A}$ such that

$$P(B \Delta D) = \mathbb{E}|1_B - 1_D| < \varepsilon.$$

■

Proposition 11.17. *Suppose that $\{(X_i, \mathcal{B}_i)\}_{i=1}^n$ are measurable spaces and for each i, $\mathbb{M}_i$ is a multiplicative system of real bounded measurable functions on X_i such that $\sigma(\mathbb{M}_i) = \mathcal{B}_i$ and there exist $\chi_n \in \mathbb{M}_i$ such that $\chi_n \to 1$ boundedly as $n \to \infty$. Given $f_i : X_i \to \mathbb{R}$ let $f_1 \otimes \dots \otimes f_n : X_1 \times \dots \times X_n \to \mathbb{R}$ be defined by*

$$(f_1 \otimes \dots \otimes f_n)(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n).$$

Show

$$\mathbb{M}_1 \otimes \dots \otimes \mathbb{M}_n := \{f_1 \otimes \dots \otimes f_n : f_i \in \mathbb{M}_i \text{ for } 1 \leq i \leq n\}$$

is a multiplicative system of bounded measurable functions on $(X := X_1 \times \dots \times X_n, \mathcal{B} := \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_n)$ such that $\sigma(\mathbb{M}_1 \otimes \dots \otimes \mathbb{M}_n) = \mathcal{B}$.

Proof. I will give the proof in case that $n = 2$. The generalization to higher n is straight forward. Let $\pi_i : X \to X_i$ be the projection maps, $\pi_1(x_1, x_2) = x_1$ and $\pi_2(x_1, x_2) = x_2$. For $f_i \in \mathbb{M}_i$, $f_i \circ \pi_i : X \to \mathbb{R}$ is the composition of measurable functions and hence measurable. Therefore $f_1 \otimes f_2 = (f_1 \circ \pi_1) \cdot (f_2 \circ \pi_2)$ is a bounded $\mathcal{B}_1 \otimes \mathcal{B}_2$ – measurable function and therefore $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2) \subset \mathcal{B}_1 \otimes \mathcal{B}_2$. Since it is clear that $\mathbb{M}_1 \otimes \mathbb{M}_2$ is a multiplicative system, to finish the proof we must show $\mathcal{B}_1 \otimes \mathcal{B}_2 \subset \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ which we now do.

Let $g \in \mathbb{M}_2$ and let

$$\mathbb{H}_g := \{f \in (\mathcal{B}_1)_b : f \otimes g \text{ is } \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2) \text{ – measurable}\}.$$

You may easily check that $\mathbb{H}_g$ is closed under bounded convergence, $\mathbb{M}_1 \subset \mathbb{H}_g$, and $\mathbb{H}_g$ contains the constant functions. Since $\sigma(\mathbb{M}_1) = \mathcal{B}_1$ it now follows by Dynkin's multiplicative systems Theorem 11.2, that $\mathbb{H}_g = (\mathcal{B}_1)_b$. Thus we have shown that $(\mathcal{B}_1)_b \otimes \mathbb{M}_2$ consists of $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ – measurable functions. By the same logic we may now conclude that $(\mathcal{B}_1)_b \otimes (\mathcal{B}_2)_b$ consists of $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ – measurable functions as well. In particular this shows for any $A_i \in \mathcal{B}_i$ that $1_{A_1 \times A_2} = 1_{A_1} \otimes 1_{A_2}$ is $\sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ – measurable and therefore $A_1 \times A_2 \in \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$ for all $A_i \in \mathcal{B}_i$. As the set $\{A_1 \times A_2 : A_i \in \mathcal{B}_i\}$ generate $\mathcal{B}_1 \otimes \mathcal{B}_2$ we may conclude that $\mathcal{B}_1 \otimes \mathcal{B}_2 \subset \sigma(\mathbb{M}_1 \otimes \mathbb{M}_2)$. ■

11.3 A Strengthening of the Multiplicative System Theorem*

Notation 11.18 *We say that $\mathbb{H} \subset \ell^\infty(\Omega, \mathbb{R})$ is* ***closed under monotone convergence*** *if; for every sequence, $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$, satisfying:*

1. *there exists* $M < \infty$ *such that* $0 \le f_n(\omega) \le M$ *for all* $\omega \in \Omega$ *and* $n \in \mathbb{N}$,
2. $f_n(\omega)$ *is increasing in* n *for all* $\omega \in \Omega$, *then* $f := \lim_{n\to\infty} f_n \in \mathbb{H}$.

Clearly if $\mathbb{H}$ is closed under bounded convergence then it is also closed under monotone convergence. I learned the proof of the converse from Pat Fitzsimmons but this result appears in Sharpe [20, p. 365].

Proposition 11.19. **Let Ω be a set. Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions from Ω to $\mathbb{R}$ which is closed under monotone convergence. Then $\mathbb{H}$ is closed under uniform convergence as well, i.e. $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ with $\sup_{n\in\mathbb{N}}\sup_{\omega\in\Omega}|f_n(\omega)| < \infty$ and $f_n \to f$, then $f \in \mathbb{H}$.*

Proof. Let us first assume that $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ such that f_n converges uniformly to a bounded function, $f : \Omega \to \mathbb{R}$. Let $\|f\|_\infty := \sup_{\omega\in\Omega}|f(\omega)|$. Let $\varepsilon > 0$ be given. By passing to a subsequence if necessary, we may assume $\|f - f_n\|_\infty \le \varepsilon 2^{-(n+1)}$. Let

$$g_n := f_n - \delta_n + M$$

with δ_n and M constants to be determined shortly. We then have

$$g_{n+1} - g_n = f_{n+1} - f_n + \delta_n - \delta_{n+1} \ge -\varepsilon 2^{-(n+1)} + \delta_n - \delta_{n+1}.$$

Taking $\delta_n := \varepsilon 2^{-n}$, then $\delta_n - \delta_{n+1} = \varepsilon 2^{-n}(1 - 1/2) = \varepsilon 2^{-(n+1)}$ in which case $g_{n+1} - g_n \ge 0$ for all n. By choosing M sufficiently large, we will also have $g_n \ge 0$ for all n. Since $\mathbb{H}$ is a vector space containing the constant functions, $g_n \in \mathbb{H}$ and since $g_n \uparrow f + M$, it follows that $f = f + M - M \in \mathbb{H}$. So we have shown that $\mathbb{H}$ is closed under uniform convergence. ■

This proposition immediately leads to the following strengthening of Theorem 11.2.

Theorem 11.20. **Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions on Ω which contains the constant functions and is closed under monotone convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. Proposition 11.19 reduces this theorem to Theorem 11.2. ■

11.4 The Bounded Approximation Theorem*

This section should be skipped until needed (if ever!).

Notation 11.21 *Given a collection of bounded functions, $\mathbb{M}$, from a set, Ω, to $\mathbb{R}$, let $\mathbb{M}_\uparrow$ $(\mathbb{M}_\downarrow)$ denote the the bounded monotone increasing (decreasing) limits of functions from $\mathbb{M}$. More explicitly a bounded function, $f : \Omega \to \mathbb{R}$ is in $\mathbb{M}_\uparrow$ respectively $\mathbb{M}_\downarrow$ iff there exists $f_n \in \mathbb{M}$ such that $f_n \uparrow f$ respectively $f_n \downarrow f$.*

Theorem 11.22 (Bounded Approximation Theorem*). *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that:*

1. $\sigma(\mathbb{M}) = \mathcal{B}$,
2. $1 \in \mathbb{M}$, *and*
3. $|f| \in \mathbb{M}$ *for all* $f \in \mathbb{M}$.

Then for every bounded $\sigma(\mathbb{M})$ measurable function, $g : \Omega \to \mathbb{R}$, and every $\varepsilon > 0$, there exists $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$.[2]

Proof. Let us begin with a few simple observations.

1. $\mathbb{M}$ is a "lattice" – if $f, g \in \mathbb{M}$ then
$$f \vee g = \frac{1}{2}(f + g + |f - g|) \in \mathbb{M}$$
and
$$f \wedge g = \frac{1}{2}(f + g - |f - g|) \in \mathbb{M}.$$
2. If $f, g \in \mathbb{M}_\uparrow$ or $f, g \in \mathbb{M}_\downarrow$ then $f + g \in \mathbb{M}_\uparrow$ or $f + g \in \mathbb{M}_\downarrow$ respectively.
3. If $\lambda \ge 0$ and $f \in \mathbb{M}_\uparrow$ $(f \in \mathbb{M}_\downarrow)$, then $\lambda f \in \mathbb{M}_\uparrow$ $(\lambda f \in \mathbb{M}_\downarrow)$.
4. If $f \in \mathbb{M}_\uparrow$ then $-f \in \mathbb{M}_\downarrow$ and visa versa.
5. If $f_n \in \mathbb{M}_\uparrow$ and $f_n \uparrow f$ where $f : \Omega \to \mathbb{R}$ is a bounded function, then $f \in \mathbb{M}_\uparrow$. Indeed, by assumption there exists $f_{n,i} \in \mathbb{M}$ such that $f_{n,i} \uparrow f_n$ as $i \to \infty$. By observation (1), $g_n := \max\{f_{ij} : i, j \le n\} \in \mathbb{M}$. Moreover it is clear that $g_n \le \max\{f_k : k \le n\} = f_n \le f$ and hence $g_n \uparrow g := \lim_{n\to\infty} g_n \le f$. Since $f_{ij} \le g$ for all i, j, it follows that $f_n = \lim_{j\to\infty} f_{nj} \le g$ and consequently that $f = \lim_{n\to\infty} f_n \le g \le f$. So we have shown that $g_n \uparrow f \in \mathbb{M}_\uparrow$.

Now let $\mathbb{H}$ denote the collection of bounded measurable functions which satisfy the assertion of the theorem. Clearly, $\mathbb{M} \subset \mathbb{H}$ and in fact it is also easy to see that $\mathbb{M}_\uparrow$ and $\mathbb{M}_\downarrow$ are contained in $\mathbb{H}$ as well. For example, if $f \in \mathbb{M}_\uparrow$, by definition, there exists $f_n \in \mathbb{M} \subset \mathbb{M}_\downarrow$ such that $f_n \uparrow f$. Since $\mathbb{M}_\downarrow \ni f_n \le f \le f \in \mathbb{M}_\uparrow$ and $\mu(f - f_n) \to 0$ by the dominated convergence theorem, it follows that $f \in \mathbb{H}$. As similar argument shows $\mathbb{M}_\downarrow \subset \mathbb{H}$. We will now show $\mathbb{H}$ is a vector sub-space of the bounded $\mathcal{B} = \sigma(\mathbb{M})$ – measurable functions.

$\mathbb{H}$ is closed under addition. If $g_i \in \mathbb{H}$ for $i = 1, 2$, and $\varepsilon > 0$ is given, we may find $f_i \in \mathbb{M}_\downarrow$ and $h_i \in \mathbb{M}_\uparrow$ such that $f_i \le g_i \le h_i$ and $\mu(h_i - f_i) < \varepsilon/2$ for $i = 1, 2$. Since $h = h_1 + h_2 \in \mathbb{M}_\uparrow$, $f := f_1 + f_2 \in \mathbb{M}_\downarrow$, $f \le g_1 + g_2 \le h$, and

$$\mu(h - f) = \mu(h_1 - f_1) + \mu(h_2 - f_2) < \varepsilon,$$

[2] Bruce: rework the Daniel integral section in the Analysis notes to stick to latticies of bounded functions.

it follows that $g_1+g_2\in\mathbb{H}$.

$\mathbb{H}$ **is closed under scalar multiplication.** If $g\in\mathbb{H}$ then $\lambda g\in\mathbb{H}$ for all $\lambda\in\mathbb{R}$. Indeed suppose that $\varepsilon>0$ is given and $f\in\mathbb{M}_\downarrow$ and $h\in\mathbb{M}_\uparrow$ such that $f\le g\le h$ and $\mu(h-f)<\varepsilon$. Then for $\lambda\ge 0$, $\mathbb{M}_\downarrow\ni\lambda f\le\lambda g\le\lambda h\in\mathbb{M}_\uparrow$ and

$$\mu(\lambda h-\lambda f)=\lambda\mu(h-f)<\lambda\varepsilon.$$

Since $\varepsilon>0$ was arbitrary, if follows that $\lambda g\in\mathbb{H}$ for $\lambda\ge 0$. Similarly, $\mathbb{M}_\downarrow\ni -h\le -g\le -f\in\mathbb{M}_\uparrow$ and

$$\mu(-f-(-h))=\mu(h-f)<\varepsilon.$$

which shows $-g\in\mathbb{H}$ as well.

Because of Theorem 11.20, to complete this proof, it suffices to show $\mathbb{H}$ is closed under monotone convergence. So suppose that $g_n\in\mathbb{H}$ and $g_n\uparrow g$, where $g:\Omega\to\mathbb{R}$ is a bounded function. Since $\mathbb{H}$ is a vector space, it follows that $0\le\delta_n:=g_{n+1}-g_n\in\mathbb{H}$ for all $n\in\mathbb{N}$. So if $\varepsilon>0$ is given, we can find, $\mathbb{M}_\downarrow\ni u_n\le\delta_n\le v_n\in\mathbb{M}_\uparrow$ such that $\mu(v_n-u_n)\le 2^{-n}\varepsilon$ for all n. By replacing u_n by $u_n\vee 0\in\mathbb{M}_\downarrow$ (by observation 1.), we may further assume that $u_n\ge 0$. Let

$$v:=\sum_{n=1}^\infty v_n=\uparrow\lim_{N\to\infty}\sum_{n=1}^N v_n\in\mathbb{M}_\uparrow \text{ (using observations 2. and 5.)}$$

and for $N\in\mathbb{N}$, let

$$u^N:=\sum_{n=1}^N u_n\in\mathbb{M}_\downarrow \text{ (using observation 2).}$$

Then

$$\sum_{n=1}^\infty\delta_n=\lim_{N\to\infty}\sum_{n=1}^N\delta_n=\lim_{N\to\infty}(g_{N+1}-g_1)=g-g_1$$

and $u^N\le g-g_1\le v$. Moreover,

$$\mu\left(v-u^N\right)=\sum_{n=1}^N\mu(v_n-u_n)+\sum_{n=N+1}^\infty\mu(v_n)\le\sum_{n=1}^N\varepsilon 2^{-n}+\sum_{n=N+1}^\infty\mu(v_n)$$
$$\le\varepsilon+\sum_{n=N+1}^\infty\mu(v_n).$$

However, since

$$\sum_{n=1}^\infty\mu(v_n)\le\sum_{n=1}^\infty\mu\left(\delta_n+\varepsilon 2^{-n}\right)=\sum_{n=1}^\infty\mu(\delta_n)+\varepsilon\mu(\Omega)$$
$$=\sum_{n=1}^\infty\mu(g-g_1)+\varepsilon\mu(\Omega)<\infty,$$

it follows that for $N\in\mathbb{N}$ sufficiently large that $\sum_{n=N+1}^\infty\mu(v_n)<\varepsilon$. Therefore, for this N, we have $\mu\left(v-u^N\right)<2\varepsilon$ and since $\varepsilon>0$ is arbitrary, if follows that $g-g_1\in\mathbb{H}$. Since $g_1\in\mathbb{H}$ and $\mathbb{H}$ is a vector space, we may conclude that $g=(g-g_1)+g_1\in\mathbb{H}$. ■

11.5 Exercises

Exercise 11.6 (Density of $\mathcal{A}$ in $\mathcal{B}=\sigma(\mathcal{A})$, same as Proposition 11.16). Keeping[3] the same notation as in Exercise 11.5 but now take $f=1_B$ for some $B\in\mathcal{B}$ and given $\varepsilon>0$, write $\varphi=\sum_{i=0}^n\lambda_i 1_{A_i}$ where $\lambda_0=0$, $\{\lambda_i\}_{i=1}^n$ is an enumeration of $\varphi(\Omega)\setminus\{0\}$, and $A_i:=\{\varphi=\lambda_i\}$. Show; 1.

$$\mathbb{E}|1_B-\varphi|=P(A_0\cap B)+\sum_{i=1}^n\left[|1-\lambda_i|P(B\cap A_i)+|\lambda_i|P(A_i\setminus B)\right] \quad (11.13)$$
$$\ge P(A_0\cap B)+\sum_{i=1}^n\min\{P(B\cap A_i),P(A_i\setminus B)\}. \quad (11.14)$$

2. Now let $\psi=\sum_{i=0}^n\alpha_i 1_{A_i}$ with

$$\alpha_i=\begin{cases}1 \text{ if } P(A_i\setminus B)\le P(B\cap A_i)\\ 0 \text{ if } P(A_i\setminus B)>P(B\cap A_i)\end{cases}.$$

Then show that

$$\mathbb{E}|1_B-\psi|=P(A_0\cap B)+\sum_{i=1}^n\min\{P(B\cap A_i),P(A_i\setminus B)\}\le\mathbb{E}|1_B-\varphi|.$$

Observe that $\psi=1_D$ where $D=\cup_{i:\alpha_{i=1}}A_i\in\mathcal{A}$ and so you have shown; for every $\varepsilon>0$ there exists a $D\in\mathcal{A}$ such that

$$P(B\Delta D)=\mathbb{E}|1_B-1_D|<\varepsilon.$$

Exercise 11.7 (This is Proposition 11.17). Suppose that $\{(X_i,\mathcal{B}_i)\}_{i=1}^n$ are measurable spaces and for each i, $\mathbb{M}_i$ is a multiplicative system of real bounded measurable functions on X_i such that $\sigma(\mathbb{M}_i)=\mathcal{B}_i$ and there exist $\chi_n\in\mathbb{M}_i$ such that $\chi_n\to 1$ boundedly as $n\to\infty$. Given $f_i:X_i\to\mathbb{R}$ let $f_1\otimes\cdots\otimes f_n:X_1\times\cdots\times X_n\to\mathbb{R}$ be defined by

$$(f_1\otimes\cdots\otimes f_n)(x_1,\dots,x_n)=f_1(x_1)\dots f_n(x_n).$$

Show

[3] This is already done in Proposition 11.16.

$$\mathbb{M}_1 \otimes \cdots \otimes \mathbb{M}_n := \{f_1 \otimes \cdots \otimes f_n : f_i \in \mathbb{M}_i \text{ for } 1 \leq i \leq n\}$$

is a multiplicative system of bounded measurable functions on

$$(X := X_1 \times \cdots \times X_n, \mathcal{B} := \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n)$$

such that $\sigma(\mathbb{M}_1 \otimes \cdots \otimes \mathbb{M}_n) = \mathcal{B}$. (It is enough to write your solution in the special case where $n = 2$.)

11.6 σ – Function Algebras (Older Version of above notes from 280 Notes!!)

In this subsection, we are going to relate σ – algebras of subsets of a set X to certain algebras of functions on X.

Example 11.23. Suppose $\mathcal{M}$ is a σ – algebra on X, then

$$\ell^\infty(\mathcal{M}, \mathbb{R}) := \{f \in \ell^\infty(X, \mathbb{R}) : f \text{ is } \mathcal{M}/\mathcal{B}_\mathbb{R} \text{ – measurable}\} \tag{11.15}$$

is a σ – function algebra. The next theorem will show that these are the only example of σ – function algebras. (See Exercise **??** above for examples of function algebras on X.)

The next theorem is the σ – algebra analogue of Exercise **??**.

Theorem 11.24. *Let $\mathcal{H}$ be a σ – function algebra on a set X. Then*

1. *$\mathcal{M}(\mathcal{H})$ is a σ – algebra on X.*
2. *$\mathcal{H} = \ell^\infty(\mathcal{M}(\mathcal{H}), \mathbb{R})$.*
3. *The map*

$$\mathcal{M} \in \{\sigma \text{ – algebras on } X\} \to \ell^\infty(\mathcal{M}, \mathbb{R}) \in \{\sigma \text{ – function algebras on } X\} \tag{11.16}$$

is bijective with inverse given by $\mathcal{H} \to \mathcal{M}(\mathcal{H})$.

Proof. Let $\mathcal{M} := \mathcal{M}(\mathcal{H})$.

1. Since $0, 1 \in \mathcal{H}$, $\emptyset, X \in \mathcal{M}$. If $A \in \mathcal{M}$ then, since $\mathcal{H}$ is a linear subspace of $\ell^\infty(X, \mathbb{R})$, $1_{A^c} = 1 - 1_A \in \mathcal{H}$ which shows $A^c \in \mathcal{M}$. If $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, then since $\mathcal{H}$ is an algebra,
$$1_{\cap_{n=1}^N A_n} = \prod_{n=1}^N 1_{A_n} =: f_N \in \mathcal{H}$$
for all $N \in \mathbb{N}$. Because $\mathcal{H}$ is closed under bounded convergence it follows that
$$1_{\cap_{n=1}^\infty A_n} = \lim_{N\to\infty} f_N \in \mathcal{H}$$
and this implies $\cap_{n=1}^\infty A_n \in \mathcal{M}$. Hence we have shown $\mathcal{M}$ is a σ – algebra.

2. Since $\mathcal{H}$ is an algebra, $p(f) \in \mathcal{H}$ for any $f \in \mathcal{H}$ and any polynomial p on $\mathbb{R}$. The Weierstrass approximation Theorem 32.39, asserts that polynomials on $\mathbb{R}$ are uniformly dense in the space of continuous functions on any compact subinterval of $\mathbb{R}$. Hence if $f \in \mathcal{H}$ and $\varphi \in C(\mathbb{R})$, there exists polynomials p_n on $\mathbb{R}$ such that $p_n \circ f(x)$ converges to $\varphi \circ f(x)$ uniformly (and hence boundedly) in $x \in X$ as $n \to \infty$. Therefore $\varphi \circ f \in \mathcal{H}$ for all $f \in \mathcal{H}$ and $\varphi \in C(\mathbb{R})$ and in particular $|f| \in \mathcal{H}$ and $f_\pm := \frac{|f| \pm f}{2} \in \mathcal{H}$ if $f \in \mathcal{H}$. Fix an $\alpha \in \mathbb{R}$ and for $n \in \mathbb{N}$ let $\varphi_n(t) := (t-\alpha)_+^{1/n}$, where $(t-\alpha)_+ := \max\{t-\alpha, 0\}$. Then $\varphi_n \in C(\mathbb{R})$ and $\varphi_n(t) \to 1_{t>\alpha}$ as $n \to \infty$ and the convergence is bounded when t is restricted to any compact subset of $\mathbb{R}$. Hence if $f \in \mathcal{H}$ it follows that $1_{f>\alpha} = \lim_{n\to\infty} \varphi_n(f) \in \mathcal{H}$ for all $\alpha \in \mathbb{R}$, i.e. $\{f > \alpha\} \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$. Therefore if $f \in \mathcal{H}$ then $f \in \ell^\infty(\mathcal{M}, \mathbb{R})$ and we have shown $\mathcal{H} \subset \ell^\infty(\mathcal{M}, \mathbb{R})$.
Conversely if $f \in \ell^\infty(\mathcal{M}, \mathbb{R})$, then for any $\alpha < \beta$, $\{\alpha < f \leq \beta\} \in \mathcal{M} = \mathcal{M}(\mathcal{H})$ and so by assumption $1_{\{\alpha < f \leq \beta\}} \in \mathcal{H}$. Combining this remark with the approximation Theorem **??** and the fact that $\mathcal{H}$ is closed under bounded convergence shows that $f \in \mathcal{H}$. Hence we have shown $\ell^\infty(\mathcal{M}, \mathbb{R}) \subset \mathcal{H}$ which combined with $\mathcal{H} \subset \ell^\infty(\mathcal{M}, \mathbb{R})$ already proved shows $\mathcal{H} = \ell^\infty(\mathcal{M}(\mathcal{H}), \mathbb{R})$.
$2'$ (BRUCE: it suffices to use the results of Exercise 4.9 here.) Exercise 4.9, there exists polynomials $p_m(x)$ such that $\sqrt{x} = \lim_{n\to\infty} p_n(x)$ uniformly in $x \in [0, M]$ for any $M < \infty$. Therefore for any $\alpha \in \mathbb{R}$ and M chosen sufficiently large we have $p_n\left((f-\alpha)^2\right) \to |f - \alpha|$ boundedly as $n \to \infty$ and hence $|f - \alpha| \in \mathcal{H}$. Since $(f-\alpha)_+ = \frac{|f-\alpha| + (f-\alpha)}{2}$, it follows that $(f-\alpha)_+ \in \mathcal{H}$. Similarly, we have $(f-\alpha)_+^{1/2} = \lim_{n\to\infty} p_n\left((f-\alpha)_+\right) \in \mathcal{H}$ and inductively it follows that $(f-\alpha)_+^{1/2^n} \in \mathcal{H}$ for all n. Since $(f-\alpha)_+^{1/2^n}$ converges boundedly to $1_{\{f>\alpha\}}$, it follows that $1_{\{f>\alpha\}} \in \mathcal{H}$, i.e. that $\{f > \alpha\} \in \mathcal{M}$. Hence if $f \in \mathcal{H}$, $\{f > \alpha\} \in \mathcal{M}$ for all $\alpha \in \mathbb{R}$. Therefore if $f \in \mathcal{H}$ then $f \in \ell^\infty(\mathcal{M}, \mathbb{R})$ and we have shown $\mathcal{H} \subset \ell^\infty(\mathcal{M}, \mathbb{R})$.
Conversely if $f \in \ell^\infty(\mathcal{M}, \mathbb{R})$, then for any $\alpha < \beta$, $\{\alpha < f \leq \beta\} \in \mathcal{M} = \mathcal{M}(\mathcal{H})$ and so by assumption $1_{\{\alpha < f \leq \beta\}} \in \mathcal{H}$. Combining this remark with the approximation Theorem **??** and the fact that $\mathcal{H}$ is closed under bounded convergence shows that $f \in \mathcal{H}$. Hence we have shown $\ell^\infty(\mathcal{M}, \mathbb{R}) \subset \mathcal{H}$ which combined with $\mathcal{H} \subset \ell^\infty(\mathcal{M}, \mathbb{R})$ already proved shows $\mathcal{H} = \ell^\infty(\mathcal{M}(\mathcal{H}), \mathbb{R})$.
3. Items 1. and 2. shows the map in Eq. (11.16) is surjective. To see the map is injective suppose $\mathcal{M}$ and $\mathcal{F}$ are two σ – algebras on X. If $\ell^\infty(\mathcal{M}, \mathbb{R}) = \ell^\infty(\mathcal{F}, \mathbb{R})$, then
$$\begin{aligned}\mathcal{M} &= \{A \subset X : 1_A \in \ell^\infty(\mathcal{M}, \mathbb{R})\} \\ &= \{A \subset X : 1_A \in \ell^\infty(\mathcal{F}, \mathbb{R})\} = \mathcal{F}\end{aligned}$$
and the proof is complete.

■

Notation 11.25 *Suppose M is a subset of $\ell^\infty(X,\mathbb{R})$.*

1. *Let $\mathcal{H}(M)$ denote the smallest subspace of $\ell^\infty(X,\mathbb{R})$ which contains M, the constant functions, and is closed under bounded convergence.*
2. *Let $\mathcal{H}_\sigma(M)$ denote the smallest σ – function algebra containing M.*

Theorem 11.26. *Suppose M is a subset of $\ell^\infty(X,\mathbb{R})$, then $\mathcal{H}_\sigma(M) = \ell^\infty(\sigma(M),\mathbb{R})$ or in other words the following diagram commutes:*

$$\begin{array}{ccccccc} & M & \longrightarrow & \sigma(M) & \\ & \pitchfork & & \pitchfork & \\ M\in & \{\text{Subsets of } \ell^\infty(X,\mathbb{R})\} & \longrightarrow & \{\sigma - \text{algebras on } X\} & \ni \mathcal{M} \\ \downarrow & \downarrow & & \downarrow & \downarrow \\ \mathcal{H}_\sigma(M)\in & \left\{\begin{matrix}\sigma-\text{function algebras}\\ \text{on } X\end{matrix}\right\} & = & \left\{\begin{matrix}\sigma-\text{function algebras}\\ \text{on } X\end{matrix}\right\} & \ni \ell^\infty(\mathcal{M},\mathbb{R}). \end{array}$$

Proof. Since $\ell^\infty(\sigma(M),\mathbb{R})$ is σ – function algebra which contains M it follows that

$$\mathcal{H}_\sigma(M) \subset \ell^\infty(\sigma(M),\mathbb{R}).$$

For the opposite inclusion, let

$$\mathcal{M} = \mathcal{M}(\mathcal{H}_\sigma(M)) := \{A \subset X : 1_A \in \mathcal{H}_\sigma(M)\}.$$

By Theorem 11.24, $M \subset \mathcal{H}_\sigma(M) = \ell^\infty(\mathcal{M},\mathbb{R})$ which implies that every $f \in M$ is $\mathcal{M}$ – measurable. This then implies $\sigma(M) \subset \mathcal{M}$ and therefore

$$\ell^\infty(\sigma(M),\mathbb{R}) \subset \ell^\infty(\mathcal{M},\mathbb{R}) = \mathcal{H}_\sigma(M).$$

■

Definition 11.27 (Multiplicative System). *A collection of bounded real or complex valued functions, M, on a set X is called a **multiplicative system** if $f \cdot g \in M$ whenever f and g are in M.*

Theorem 11.28 (Dynkin's Multiplicative System Theorem). *Suppose $M \subset \ell^\infty(X,\mathbb{R})$ is a multiplicative system, then*

$$\mathcal{H}(M) = \mathcal{H}_\sigma(M) = \ell^\infty(\sigma(M),\mathbb{R}). \tag{11.17}$$

This can also be stated as follows.

Suppose $\mathcal{H}$ is a linear subspace of $\ell^\infty(X,\mathbb{R})$ such that: $\mathbf{1} \in \mathcal{H}$, $\mathcal{H}$ is closed under bounded convergence, and $M \subset \mathcal{H}$. Then $\mathcal{H}$ contains all bounded real valued $\sigma(M)$-measurable functions, i.e. $\ell^\infty(\sigma(M),\mathbb{R}) \subset \mathcal{H}$.

(In words, the smallest subspace of bounded real valued functions on X which contains M that is closed under bounded convergence is the same as the space of bounded real valued $\sigma(M)$ – measurable functions on X.)

Proof. The assertion that $\mathcal{H}_\sigma(M) = \ell^\infty(\sigma(M),\mathbb{R})$ has already been proved (without the assumption that M is multiplicative) in Theorem 11.26. Since any σ – function algebra containing M is also a subspace of $\ell^\infty(X,\mathbb{R})$ which contains the constant functions and is closed under bounded convergence (compare with Exercise 11.13), it follows that $\mathcal{H}(M) \subset \mathcal{H}_\sigma(M)$. To complete the proof it suffices to show the inclusion, $\mathcal{H}(M) \subset \mathcal{H}_\sigma(M)$, is an equality. We will accomplish this below by showing $\mathcal{H}(M)$ is also a σ – function algebra.

For any $f \in \mathcal{H} := \mathcal{H}(M)$ let

$$\mathcal{H}_f := \{g \in \mathcal{H} : fg \in \mathcal{H}\} \subset \mathcal{H}$$

and notice that $\mathcal{H}_f$ is a linear subspace of $\ell^\infty(X,\mathbb{R})$ which is closed under bounded convergence. Moreover if $f \in M$, $M \subset \mathcal{H}_f$ since M is multiplicative. Therefore $\mathcal{H}_f = \mathcal{H}$ and we have shown that $fg \in \mathcal{H}$ whenever $f \in M$ and $g \in \mathcal{H}$. Given this it now follows that $M \subset \mathcal{H}_f$ for any $f \in \mathcal{H}$ and by the same reasoning just used, $\mathcal{H}_f = \mathcal{H}$. Since $f \in \mathcal{H}$ is arbitrary, we have shown $fg \in \mathcal{H}$ for all $f, g \in \mathcal{H}$, i.e. $\mathcal{H}$ is an algebra, which by the definition of $\mathcal{H}(M)$ in Notation 11.25 contains the constant functions, i.e. $\mathcal{H}(M)$ is a σ – function algebra. ■

Theorem 11.29 (Complex Multiplicative System Theorem). *Suppose $\mathcal{H}$ is a complex linear subspace of $\ell^\infty(X,\mathbb{C})$ such that: $1 \in \mathcal{H}$, $\mathcal{H}$ is closed under complex conjugation, and $\mathcal{H}$ is closed under bounded convergence. If $M \subset \mathcal{H}$ is multiplicative system which is closed under conjugation, then $\mathcal{H}$ contains all bounded complex valued $\sigma(M)$-measurable functions, i.e. $\ell^\infty(\sigma(M),\mathbb{C}) \subset \mathcal{H}$.*

Proof. Let $M_0 = \operatorname{span}_{\mathbb{C}}(M \cup \{1\})$ be the complex span of M. As the reader should verify, M_0 is an algebra, $M_0 \subset \mathcal{H}$, M_0 is closed under complex conjugation and that $\sigma(M_0) = \sigma(M)$. Let $\mathcal{H}^{\mathbb{R}} := \mathcal{H} \cap \ell^\infty(X,\mathbb{R})$ and $M_0^{\mathbb{R}} = M \cap \ell^\infty(X,\mathbb{R})$. Then (you verify) $M_0^{\mathbb{R}}$ is a multiplicative system, $M_0^{\mathbb{R}} \subset \mathcal{H}^{\mathbb{R}}$ and $\mathcal{H}^{\mathbb{R}}$ is a linear space containing 1 which is closed under bounded convergence. Therefore by Theorem 11.28, $\ell^\infty\left(\sigma\left(M_0^{\mathbb{R}}\right),\mathbb{R}\right) \subset \mathcal{H}^{\mathbb{R}}$. Since $\mathcal{H}$ and M_0 are complex linear spaces closed under complex conjugation, for any $f \in \mathcal{H}$ or $f \in M_0$, the functions $\operatorname{Re} f = \frac{1}{2}\left(f + \bar{f}\right)$ and $\operatorname{Im} f = \frac{1}{2i}\left(f - \bar{f}\right)$ are in $\mathcal{H}$ (M_0) or M_0 respectively. Therefore $\mathcal{H} = \mathcal{H}^{\mathbb{R}} + i\mathcal{H}^{\mathbb{R}}$, $M_0 = M_0^{\mathbb{R}} + iM_0^{\mathbb{R}}$, $\sigma\left(M_0^{\mathbb{R}}\right) = \sigma(M_0) = \sigma(M)$ and

$$\begin{aligned}\ell^\infty(\sigma(M),\mathbb{C}) &= \ell^\infty\left(\sigma\left(M_0^{\mathbb{R}}\right),\mathbb{R}\right) + i\ell^\infty\left(\sigma\left(M_0^{\mathbb{R}}\right),\mathbb{R}\right) \\ &\subset \mathcal{H}^{\mathbb{R}} + i\mathcal{H}^{\mathbb{R}} = \mathcal{H}.\end{aligned}$$

■

Definition 11.30. *A collection of subsets, $\mathcal{C}$, of X is a **multiplicative class**(or a π – **class**) if $\mathcal{C}$ is closed under finite intersections.*

Corollary 11.31. *Suppose $\mathcal{H}$ is a subspace of $\ell^\infty(X,\mathbb{R})$ which is closed under bounded convergence and $1 \in \mathcal{H}$. If $\mathcal{C} \subset 2^X$ is a multiplicative class such that $1_A \in \mathcal{H}$ for all $A \in \mathcal{C}$, then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{C})$ – measurable functions.*

Proof. Let $M = \{1\} \cup \{1_A : A \in \mathcal{C}\}$. Then $M \subset \mathcal{H}$ is a multiplicative system and the proof is completed with an application of Theorem 11.28. ■

Corollary 11.32. *Suppose that (X,d) is a metric space and $\mathcal{B}_X = \sigma(\tau_d)$ is the Borel σ – algebra on X and $\mathcal{H}$ is a subspace of $\ell^\infty(X,\mathbb{R})$ such that $BC(X,\mathbb{R}) \subset \mathcal{H}$ and $\mathcal{H}$ is closed under bounded convergence[4]. Then $\mathcal{H}$ contains all bounded $\mathcal{B}_X$ – measurable real valued functions on X. (This may be stated as follows: the smallest vector space of bounded functions which is closed under bounded convergence and contains $BC(X,\mathbb{R})$ is the space of bounded $\mathcal{B}_X$ – measurable real valued functions on X.)*

Proof. Let $V \in \tau_d$ be an open subset of X and for $n \in \mathbb{N}$ let

$$f_n(x) := \min(n \cdot d_{V^c}(x), 1) \text{ for all } x \in X.$$

Notice that $f_n = \varphi_n \circ d_{V^c}$ where $\varphi_n(t) = \min(nt, 1)$ (see Figure 11.3) which is continuous and hence $f_n \in BC(X,\mathbb{R})$ for all n. Furthermore, f_n converges boundedly to $1_{d_{V^c}>0} = 1_V$ as $n \to \infty$ and therefore $1_V \in \mathcal{H}$ for all $V \in \tau$. Since τ is a π – class, the result now follows by an application of Corollary 11.31.

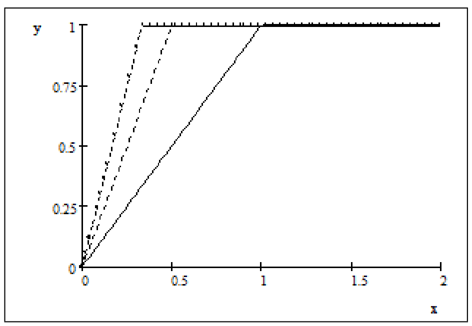


Fig. 11.3. Plots of ϕ_1, ϕ_2 and ϕ_3.

■

Here are some more variants of Corollary 11.32.

Proposition 11.33. *Let (X,d) be a metric space, $\mathcal{B}_X = \sigma(\tau_d)$ be the Borel σ – algebra and assume there exists compact sets $K_k \subset X$ such that $K_k^o \uparrow X$. Suppose that $\mathcal{H}$ is a subspace of $\ell^\infty(X,\mathbb{R})$ such that $C_c(X,\mathbb{R}) \subset \mathcal{H}$ ($C_c(X,\mathbb{R})$ is the space of continuous functions with compact support) and $\mathcal{H}$ is closed under bounded convergence. Then $\mathcal{H}$ contains all bounded $\mathcal{B}_X$ – measurable real valued functions on X.*

Proof. Let k and n be positive integers and set $\psi_{n,k}(x) = \min(1, n \cdot d_{(K_k^o)^c}(x))$. Then $\psi_{n,k} \in C_c(X,\mathbb{R})$ and $\{\psi_{n,k} \neq 0\} \subset K_k^o$. Let $\mathcal{H}_{n,k}$ denote those bounded $\mathcal{B}_X$ – measurable functions, $f : X \to \mathbb{R}$, such that $\psi_{n,k} f \in \mathcal{H}$. It is easily seen that $\mathcal{H}_{n,k}$ is closed under bounded convergence and that $\mathcal{H}_{n,k}$ contains $BC(X,\mathbb{R})$ and therefore by Corollary 11.32, $\psi_{n,k} f \in \mathcal{H}$ for all bounded measurable functions $f : X \to \mathbb{R}$. Since $\psi_{n,k} f \to 1_{K_k^o} f$ boundedly as $n \to \infty$, $1_{K_k^o} f \in \mathcal{H}$ for all k and similarly $1_{K_k^o} f \to f$ boundedly as $k \to \infty$ and therefore $f \in \mathcal{H}$. ■

Lemma 11.34. *Suppose that (X,τ) is a locally compact second countable Hausdorff space.[5] Then:*

1. *every open subset $U \subset X$ is σ – compact. In fact U is still a locally compact second countable Hausdorff space.*
2. *If $F \subset X$ is a closed set, there exist open sets $V_n \subset X$ such that $V_n \downarrow F$ as $n \to \infty$.*
3. *To each open set $U \subset X$ there exists $f_n \prec U$ (i.e. $f_n \in C_c(U,[0,1])$) such that $\lim_{n\to\infty} f_n = 1_U$.*
4. *$\mathcal{B}_X = \sigma(C_c(X,\mathbb{R}))$, i.e. the σ – algebra generated by $C_c(X)$ is the Borel σ – algebra on X.*

Proof.

1. Let U be an open subset of X, $\mathcal{V}$ be a countable base for τ and

$$\mathcal{V}^U := \{W \in \mathcal{V} : \bar{W} \subset U \text{ and } \bar{W} \text{ is compact}\}.$$

For each $x \in U$, by Proposition 37.7, there exists an open neighborhood V of x such that $\bar{V} \subset U$ and $\bar{V}$ is compact. Since $\mathcal{V}$ is a base for the topology τ, there exists $W \in \mathcal{V}$ such that $x \in W \subset V$. Because $\bar{W} \subset \bar{V}$, it follows that $\bar{W}$ is compact and hence $W \in \mathcal{V}^U$. As $x \in U$ was arbitrary, $U = \cup \mathcal{V}^U$. This shows $\mathcal{V}^U$ is a countable basis for the topology on U and that U is still locally compact.

Let $\{W_n\}_{n=1}^\infty$ be an enumeration of $\mathcal{V}^U$ and set $K_n := \cup_{k=1}^n \bar{W}_k$. Then $K_n \uparrow U$ as $n \to \infty$ and K_n is compact for each n. This shows U is σ – compact. (See Exercise 35.21.)

[4] Recall that $BC(X,\mathbb{R})$ are the bounded continuous functions on X.

[5] For example any separable locally compact metric space and in particular any open subset of $\mathbb{R}^n$.

2. Let $\{K_n\}_{n=1}^\infty$ be compact subsets of F^c such that $K_n \uparrow F^c$ as $n \to \infty$ and set $V_n := K_n^c = X \setminus K_n$. Then $V_n \downarrow F$ and by Proposition 37.5, V_n is open for each n.
3. Let $U \subset X$ be an open set and $\{K_n\}_{n=1}^\infty$ be compact subsets of U such that $K_n \uparrow U$. By Urysohn's Lemma 37.8, there exist $f_n \prec U$ such that $f_n = 1$ on K_n. These functions satisfy, $1_U = \lim_{n\to\infty} f_n$.
4. By item 3., 1_U is $\sigma(C_c(X,\mathbb{R}))$ – measurable for all $U \in \tau$ and hence $\tau \subset \sigma(C_c(X,\mathbb{R}))$. Therefore $\mathcal{B}_X = \sigma(\tau) \subset \sigma(C_c(X,\mathbb{R}))$. The converse inclusion holds because continuous functions are always Borel measurable.

■

Here is a variant of Corollary 11.32.

Corollary 11.35. *Suppose that (X,τ) is a second countable locally compact Hausdorff space and $\mathcal{B}_X = \sigma(\tau)$ is the Borel σ – algebra on X. If $\mathcal{H}$ is a subspace of $\ell^\infty(X,\mathbb{R})$ which is closed under bounded convergence and contains $C_c(X,\mathbb{R})$, then $\mathcal{H}$ contains all bounded $\mathcal{B}_X$ – measurable real valued functions on X.*

Proof. By Item 3. of Lemma 11.34, for every $U \in \tau$ the characteristic function, 1_U, may be written as a bounded pointwise limit of functions from $C_c(X,\mathbb{R})$. Therefore $1_U \in \mathcal{H}$ for all $U \in \tau$. Since τ is a π – class, the proof is finished with an application of Corollary 11.31 ■

11.6.1 Another (Better) Multiplicative System Theorem

Notation 11.36 *Let Ω be a set and $\mathbb{H}$ be a subset of the bounded real valued functions on Ω. We say that $\mathbb{H}$ is **closed under bounded convergence** if; for every sequence, $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $|f_n(\omega)| \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f(\omega) := \lim_{n\to\infty} f_n(\omega)$ exists for all $\omega \in \Omega$, then $f \in \mathbb{H}$.*
   *Similarly we say that $\mathbb{H}$ is **closed under monotone convergence** if; for every sequence, $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$, satisfying:*
3. *there exists $M < \infty$ such that $0 \le f_n(\omega) \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
4. *$f_n(\omega)$ is increasing in n for all $\omega \in \Omega$,*

then $f := \lim_{n\to\infty} f_n \in \mathbb{H}$.

Clearly if $\mathbb{H}$ is closed under bounded convergence then it is also closed under monotone convergence. I learned the following converse result from Pat Fitzsimmons.

Proposition 11.37. *Let Ω be a set. Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions from Ω to $\mathbb{R}$ which is closed under monotone convergence. Then $\mathbb{H}$ is closed under uniform convergence as well, i.e. $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ with $\sup_{n\in\mathbb{N}} \sup_{\omega\in\Omega} |f_n(\omega)| < \infty$ and $f_n \to f$, then $f \in \mathbb{H}$.*

Proof. Let us first assume that $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ such that f_n converges uniformly to a bounded function, $f : \Omega \to \mathbb{R}$. Let $\|f\|_\infty := \sup_{\omega\in\Omega} |f(\omega)|$. Let $\varepsilon > 0$ be given. By passing to a subsequence if necessary, we may assume $\|f - f_n\|_\infty \le \varepsilon 2^{-(n+1)}$. Let

$$g_n := f_n - \delta_n + M$$

with δ_n and M constants to be determined shortly. We then have

$$g_{n+1} - g_n = f_{n+1} - f_n + \delta_n - \delta_{n+1} \ge -\varepsilon 2^{-(n+1)} + \delta_n - \delta_{n+1}.$$

Taking $\delta_n := \varepsilon 2^{-n}$, then $\delta_n - \delta_{n+1} = \varepsilon 2^{-n}(1 - 1/2) = \varepsilon 2^{-(n+1)}$ in which case $g_{n+1} - g_n \ge 0$ for all n. By choosing M sufficiently large, we will also have $g_n \ge 0$ for all n. Since $\mathbb{H}$ is a vector space containing the constant functions, $g_n \in \mathbb{H}$ and since $g_n \uparrow f + M$, it follows that $f = f + M - M \in \mathbb{H}$. So we have shown that $\mathbb{H}$ is closed under uniform convergence. ■

Theorem 11.38 (Dynkin's Multiplicative System Theorem (Old Proof)). *Suppose that $\mathbb{H}$ is a vector subspace of bounded functions from Ω to $\mathbb{R}$ which contains the constant functions and is closed under monotone convergence. If $\mathbb{M}$ is **multiplicative system** (i.e. $\mathbb{M}$ is a subset of $\mathbb{H}$ which is closed under pointwise multiplication), then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. Let

$$\mathcal{L} := \{A \subset \Omega : 1_A \in \mathbb{H}\}.$$

We then have $\Omega \in \mathcal{L}$ since $1_\Omega = 1 \in \mathbb{H}$, if $A, B \in \mathcal{L}$ with $A \subset B$ then $B \setminus A \in \mathcal{L}$ since $1_{B\setminus A} = 1_B - 1_A \in \mathbb{H}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $A \in \mathcal{L}$ because $1_{A_n} \in \mathbb{H}$ and $1_{A_n} \uparrow 1_A \in \mathbb{H}$. Therefore $\mathcal{L}$ is λ – system.

Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 11.4 below) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f_1, f_2, \dots, f_k \in \mathbb{M}$ and $a_1, \dots, a_k \in \mathbb{R}$, let

$$F_n := \prod_{i=1}^k \varphi_n(f_i - a_i)$$

and let

$$M := \sup_{i=1,\dots,k} \sup_\omega |f_i(\omega) - a_i|.$$

By the Weierstrass approximation Theorem **??**, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$ .Since p_l is a polynomial it is easily seen that $\prod_{i=1}^k p_l \circ (f_i - a_i) \in \mathbb{H}$. Moreover,

$$\prod_{i=1}^k p_l \circ (f_i - a_i) \to F_n \text{ uniformly as } l \to \infty,$$

from with it follows that $F_n \in \mathbb{H}$ for all n. Since,

$$F_n \uparrow \prod_{i=1}^{k} 1_{\{f_i > a_i\}} = 1_{\cap_{i=1}^{k}\{f_i > a_i\}}$$

it follows that $1_{\cap_{i=1}^{k}\{f_i > a_i\}} \in \mathbb{H}$ or equivalently that $\cap_{i=1}^{k} \{f_i > a_i\} \in \mathcal{L}$. Therefore $\mathcal{L}$ contains the π – system, $\mathcal{P}$, consisting of finite intersections of sets of the form, $\{f > a\}$ with $f \in \mathbb{M}$ and $a \in \mathbb{R}$.

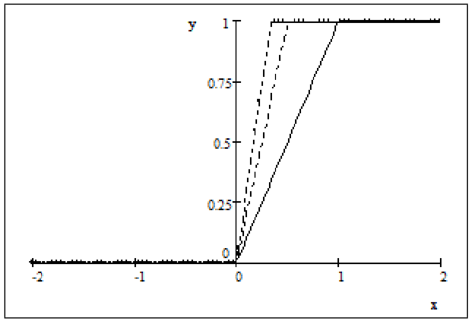


Fig. 11.4. Plots of φ_1, φ_2 and φ_3.

As a consequence of the above paragraphs and the $\pi - \lambda$ Theorem 10.4, $\mathcal{L}$ contains $\sigma(\mathcal{P}) = \sigma(\mathbb{M})$. In particular it follows that $1_A \in \mathbb{H}$ for all $A \in \sigma(\mathbb{M})$. Since any positive $\sigma(\mathbb{M})$ – measurable function may be written as a increasing limit of simple functions, it follows that $\mathbb{H}$ contains all non-negative bounded $\sigma(\mathbb{M})$ – measurable functions. Finally, since any bounded $\sigma(\mathbb{M})$ – measurable functions may be written as the difference of two such non-negative simple functions, it follows that $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions. ■

The proof of the following standard result is taken from Janson [10, Appendix A., p. 309.].

Theorem 11.39 (Multiplicative System Theorem (New Proof)). *Suppose $\mathbb{H}$ is a vector subspace of $\ell^\infty(X,\mathbb{R})$ such that: 1. $1 \in \mathbb{H}$, and 2. if $0 \le f_n \in \mathbb{H}$ such that $f_n \uparrow f \in \ell^\infty(X,\mathbb{R})$, then $f \in \mathbb{H}$. If $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system, then $\sigma(\mathbb{M})_b \subset \mathbb{H}$. (In words, the smallest subspace of bounded real valued functions on X which contains $\mathbb{M}$ that is closed under monotone bounded convergence is the same as the space of bounded real valued $\sigma(\mathbb{M})$ – measurable functions on X.)*

Proof. Let $\mathbb{M}'$ be the subsapce of $\mathbb{H}$ spanned by $\mathbb{M} \cup \{1\}$. As $\mathbb{M}'$ is a vector space closed under multiplication, it is an algebra. Let $\mathbb{M}'_+ := \{f \in \mathbb{M}' : f \ge 0\}$.

Let $\mathbb{K}$ be the smallest linear subspace of $\ell^\infty(X,\mathbb{R})$ which contains $\mathbb{M}'$ and is closed under bounded monotone convergence. (Such an $\mathbb{K}$ is found by intersecting all such subspace together noting that $\mathbb{H}$ is a such a subspace.) Our first goal is to show that $\mathbb{K}$ is a algebra.

For any $f \in \mathbb{K}$, let $\mathbb{K}_f := \{g \in \mathbb{K} : fg \in \mathbb{K}\}$. Then $\mathbb{K}_f$ is a subspace of $\mathbb{K}$ and if $f \ge 0$, then $\mathbb{K}_f$ is closed under bounded monotone convergence.

If $f \in \mathbb{M}'_+ \subset \mathbb{M}' \subset \mathbb{K}$, we have $\mathbb{M}' \subset \mathbb{K}_f$ and $\mathbb{K}_f$ is closed under bounded monotone convergence and therefore, $\mathbb{K} \subset \mathbb{K}_f$, i.e. $\mathbb{K}_f = \mathbb{K}$. Thus we have shown $fg \in \mathbb{K}$ if $f \in \mathbb{M}'_+$ and $g \in \mathbb{K}$. Moreover if $f \in \mathbb{M}'$ and $m := \max f(\Omega)$, then $f + m \in \mathbb{M}'_+$ and hence $fg = (f+m)g - mg \in \mathbb{K}$. Therefore, $fg \in \mathbb{K}$ if $f \in \mathbb{M}'$ and $g \in \mathbb{K}$.

Similarly, if $f \in \mathbb{K}_+$, then $g \in \mathbb{K}_f$ for all $g \in \mathbb{M}'$ and therefore $\mathbb{K}_f \subset \mathbb{K}$ again. Thus we have shown that $gf \in \mathbb{K}$ whenever $f \in \mathbb{K}_+$ and $g \in \mathbb{K}$. So if $f \in \mathbb{K}$ and $m := \max f(\Omega)$, then $f + m \in \mathbb{M}'_+$ and hence $fg = (f+m)g - mg \in \mathbb{K}$ for all $g \in \mathbb{K}$. This completes the proof that $\mathbb{K}$ is an algebra.

Next we are going to show

$$\mathcal{M} := \{A \subset \Omega : 1_A \in \mathbb{K}\} \tag{11.18}$$

is a σ – algebra. As 0 and 1 are in $\mathbb{K}$, $\emptyset, \Omega \in \mathcal{M}$. If $A \in \mathcal{M}$, then $A^c \in \mathcal{M}$ since $1_{A^c} = 1 - 1_A \in \mathbb{K}$. If we further suppose that $B \in \mathcal{M}$, then $1_{A\cap B} = 1_A \cdot 1_B \in \mathbb{K}$ which shows that $A \cap B \in \mathcal{M}$. Thuse we have shown that $\mathcal{M}$ is an algebra. Finally if $A_n \in \mathcal{M}$ and $A_n \uparrow A \subset \Omega$, then $1_{A_n} \uparrow 1_A$ showing that $1_A \in \mathbb{K}$, i.e. $A \in \mathcal{M}$. This implies that $\mathcal{M}$ is a σ – algebra. STOP BRUCE – look up to see how Janson finishes the proof here.

The Weierstrass approximation Theorem 32.39, asserts that polynomials on $\mathbb{R}$ are uniformly dense in the space of continuous functions on any compact subinterval of $\mathbb{R}$. Hence if $f \in \mathbb{K}$ and $\varphi \in C(\mathbb{R})$, there exists polynomials p_n on $\mathbb{R}$ such that $p_n \circ f(x)$ converges to $\varphi \circ f(x)$ uniformly (and hence boundedly) in $x \in X$ as $n \to \infty$. Hence by Propostion 11.37, it follows that $\varphi \circ f \in \mathbb{K}$ for all $f \in \mathbb{K}$ and $\varphi \in C(\mathbb{R})$ and in particular $|f| \in \mathbb{K}$ and $f_\pm := \frac{|f| \pm f}{2} \in \mathbb{K}$ if $f \in \mathbb{K}$. Fix an $\alpha \in \mathbb{R}$ and for $n \in \mathbb{N}$ let $\varphi_n(t) := (t-\alpha)_+^{1/n}$, where $(t-\alpha)_+ := \max\{t-\alpha, 0\}$. Then $\varphi_n \in C(\mathbb{R})$ and $\varphi_n(t) \to 1_{t>\alpha}$ as $n \to \infty$ and the convergence is bounded when t is restricted to any compact subset of $\mathbb{R}$. Hence, again using Proposition 11.37, if $f \in \mathbb{K}$ it follows that $1_{f>\alpha} = \lim_{n\to\infty} \varphi_n(f) \in \mathbb{K}$ for all $\alpha \in \mathbb{R}$, i.e. $\{f > \alpha\} \in \mathbb{M}$ for all $\alpha \in \mathbb{R}$. Therefore if $f \in \mathbb{K}$ then $f \in \ell^\infty(\mathcal{M},\mathbb{R})$ and we have shown $\mathbb{K} \subset \ell^\infty(\mathcal{M},\mathbb{R})$.

Conversely if $f \in \ell^\infty(\mathcal{M},\mathbb{R})$, then for any $\alpha < \beta$, $\{\alpha < f \le \beta\} \in \mathcal{M} = \mathcal{M}(\mathbb{K})$ and so by assumption $1_{\{\alpha < f \le \beta\}} \in \mathbb{K}$. Combining this remark with the approximation Theorem **??** and the fact that $\mathbb{K}$ is closed under bounded convergence shows that $f \in \mathbb{K}$. Hence we have shown $\ell^\infty(\mathcal{M},\mathbb{R}) \subset \mathbb{K}$ which combined with $\mathbb{K} \subset \ell^\infty(\mathcal{M},\mathbb{R})$ already proved shows $\ell^\infty(\mathcal{M}(\mathbb{K}),\mathbb{R}) = \mathbb{K} \subset \mathbb{H}$. ■

11.7 Exercises

Exercise 11.8. Prove Corollary **??**. **Hint:** See Exercise **??**.

Exercise 11.9. If $\mathcal{M}$ is the σ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the σ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

Exercise 11.10. Let $(X, \mathcal{M})$ be a measure space and $f_n : X \to \mathbb{F}$ be a sequence of measurable functions on X. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \mathbf{in}\ \mathbb{F}\} \in \mathcal{M}$.

Exercise 11.11. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_\mathbb{R}, \mathcal{B}_\mathbb{R})$ – measurable.

Exercise 11.12. Show by example that the supremum of an uncountable family of measurable functions need not be measurable. (Folland problem 2.6 on p. 48.)

Exercise 11.13. Let $X = \{1, 2, 3, 4\}$, $A = \{1, 2\}$, $B = \{2, 3\}$ and $M := \{1_A, 1_B\}$. Show $\mathcal{H}_\sigma(M) \neq \mathcal{H}(M)$ in this case.

12

Multiple and Iterated Integrals

12.1 Iterated Integrals

Notation 12.1 (Iterated Integrals) *If $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are two measure spaces and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function, the* ***iterated integrals*** *of f (when they make sense) are:*

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) := \int_X \left[\int_Y f(x,y) d\nu(y) \right] d\mu(x)$$

and

$$\int_Y d\nu(y) \int_X d\mu(x) f(x,y) := \int_Y \left[\int_X f(x,y) d\mu(x) \right] d\nu(y).$$

Notation 12.2 *Suppose that $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ are functions, let $f \otimes g$ denote the function on $X \times Y$ given by*

$$f \otimes g(x,y) = f(x)g(y).$$

Notice that if f, g are measurable, then $f \otimes g$ is $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{C}})$ – measurable. To prove this let $F(x,y) = f(x)$ and $G(x,y) = g(y)$ so that $f \otimes g = F \cdot G$ will be measurable provided that F and G are measurable. Now $F = f \circ \pi_1$ where $\pi_1 : X \times Y \to X$ is the projection map. This shows that F is the composition of measurable functions and hence measurable. Similarly one shows that G is measurable.

12.2 Tonelli's Theorem and Product Measure

Theorem 12.3. *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces and f is a nonnegative $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable function, then for each $y \in Y$,*

$$x \to f(x,y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{12.1}$$

for each $x \in X$,

$$y \to f(x,y) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{12.2}$$

$$x \to \int_Y f(x,y) d\nu(y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{12.3}$$

$$y \to \int_X f(x,y) d\mu(x) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{12.4}$$

and

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \int_Y d\nu(y) \int_X d\mu(x) f(x,y). \tag{12.5}$$

Proof. Suppose that $E = A \times B \in \mathcal{E} := \mathcal{M} \times \mathcal{N}$ and $f = 1_E$. Then

$$f(x,y) = 1_{A \times B}(x,y) = 1_A(x) 1_B(y)$$

and one sees that Eqs. (12.1) and (12.2) hold. Moreover

$$\int_Y f(x,y) d\nu(y) = \int_Y 1_A(x) 1_B(y) d\nu(y) = 1_A(x) \nu(B),$$

so that Eq. (12.3) holds and we have

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \nu(B) \mu(A). \tag{12.6}$$

Similarly,

$$\int_X f(x,y) d\mu(x) = \mu(A) 1_B(y) \text{ and}$$
$$\int_Y d\nu(y) \int_X d\mu(x) f(x,y) = \nu(B) \mu(A)$$

from which it follows that Eqs. (12.4) and (12.5) hold in this case as well.

For the moment let us now further assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ and let $\mathbb{H}$ be the collection of all bounded $(\mathcal{M} \otimes \mathcal{N}, \mathcal{B}_{\mathbb{R}})$ – measurable functions on $X \times Y$ such that Eqs. (12.1) – (12.5) hold. Using the fact that measurable functions are closed under pointwise limits and the dominated convergence theorem (the dominating function always being a constant), one easily shows that $\mathbb{H}$ closed under bounded convergence. Since we have just verified that $1_E \in \mathbb{H}$ for all E in the π – class, $\mathcal{E}$, it follows by Corollary 11.3 that $\mathbb{H}$ is the space of all bounded $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable functions on $X\times Y$. Moreover, if $f: X\times Y\to[0,\infty]$ is a $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, let $f_M = M\wedge f$ so that $f_M\uparrow f$ as $M\to\infty$. Then Eqs. (12.1) – (12.5) hold with f replaced by f_M for all $M\in\mathbb{N}$. Repeated use of the monotone convergence theorem allows us to pass to the limit $M\to\infty$ in these equations to deduce the theorem in the case μ and ν are finite measures.

For the σ – finite case, choose $X_n\in\mathcal{M}$, $Y_n\in\mathcal{N}$ such that $X_n\uparrow X$, $Y_n\uparrow Y$, $\mu(X_n)<\infty$ and $\nu(Y_n)<\infty$ for all $m,n\in\mathbb{N}$. Then define $\mu_m(A)=\mu(X_m\cap A)$ and $\nu_n(B)=\nu(Y_n\cap B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ or equivalently $d\mu_m = 1_{X_m}d\mu$ and $d\nu_n = 1_{Y_n}d\nu$. By what we have just proved Eqs. (12.1) – (12.5) with μ replaced by μ_m and ν by ν_n for all $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable functions, $f: X\times Y\to[0,\infty]$. The validity of Eqs. (12.1) – (12.5) then follows by passing to the limits $m\to\infty$ and then $n\to\infty$ making use of the monotone convergence theorem in the following context. For all $u\in L^+(X,\mathcal{M})$,

$$\int_X u d\mu_m = \int_X u1_{X_m}d\mu \uparrow \int_X u d\mu \text{ as } m\to\infty,$$

and for all and $v\in L^+(Y,\mathcal{N})$,

$$\int_Y v d\mu_n = \int_Y v1_{Y_n}d\mu \uparrow \int_Y v d\mu \text{ as } n\to\infty.$$

■

Corollary 12.4. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ – finite measure spaces. Then there exists a unique measure π on $\mathcal{M}\otimes\mathcal{N}$ such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$. Moreover π is given by*

$$\pi(E)=\int_X d\mu(x)\int_Y d\nu(y)1_E(x,y)=\int_Y d\nu(y)\int_X d\mu(x)1_E(x,y) \tag{12.7}$$

for all $E\in\mathcal{M}\otimes\mathcal{N}$ and π is σ – finite.

Proof. Notice that any measure π such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ is necessarily σ – finite. Indeed, let $X_n\in\mathcal{M}$ and $Y_n\in\mathcal{N}$ be chosen so that $\mu(X_n)<\infty$, $\nu(Y_n)<\infty$, $X_n\uparrow X$ and $Y_n\uparrow Y$, then $X_n\times Y_n\in\mathcal{M}\otimes\mathcal{N}$, $X_n\times Y_n\uparrow X\times Y$ and $\pi(X_n\times Y_n)<\infty$ for all n. The uniqueness assertion is a consequence of the combination of Exercises 6.10 and 8.11 Proposition 6.25 with $\mathcal{E}=\mathcal{M}\times\mathcal{N}$. For the existence, it suffices to observe, using the monotone convergence theorem, that π defined in Eq. (12.7) is a measure on $\mathcal{M}\otimes\mathcal{N}$. Moreover this measure satisfies $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ from Eq. (12.6). ■

Notation 12.5 *The measure π is called the product measure of μ and ν and will be denoted by $\mu\otimes\nu$.*

Theorem 12.6 (Tonelli's Theorem). *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ – finite measure spaces and $\pi=\mu\otimes\nu$ is the product measure on $\mathcal{M}\otimes\mathcal{N}$. If $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$, then $f(\cdot,y)\in L^+(X,\mathcal{M})$ for all $y\in Y$, $f(x,\cdot)\in L^+(Y,\mathcal{N})$ for all $x\in X$,*

$$\int_Y f(\cdot,y)d\nu(y)\in L^+(X,\mathcal{M}),\ \int_X f(x,\cdot)d\mu(x)\in L^+(Y,\mathcal{N})$$

and

$$\int_{X\times Y} f\,d\pi = \int_X d\mu(x)\int_Y d\nu(y)f(x,y) \tag{12.8}$$
$$= \int_Y d\nu(y)\int_X d\mu(x)f(x,y). \tag{12.9}$$

Proof. By Theorem 12.3 and Corollary 12.4, the theorem holds when $f=1_E$ with $E\in\mathcal{M}\otimes\mathcal{N}$. Using the linearity of all of the statements, the theorem is also true for non-negative simple functions. Then using the monotone convergence theorem repeatedly along with the approximation Theorem 9.39, one deduces the theorem for general $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$. ■

Example 12.7. In this example we are going to show, $I:=\int_{\mathbb{R}} e^{-x^2/2}dm(x)=\sqrt{2\pi}$. To this end we observe, using Tonelli's theorem, that

$$I^2 = \left[\int_{\mathbb{R}} e^{-x^2/2}dm(x)\right]^2 = \int_{\mathbb{R}} e^{-y^2/2}\left[\int_{\mathbb{R}} e^{-x^2/2}dm(x)\right]dm(y)$$
$$= \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}dm^2(x,y)$$

where $m^2 = m\otimes m$ is "Lebesgue measure" on $\left(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2}=\mathcal{B}_{\mathbb{R}}\otimes\mathcal{B}_{\mathbb{R}}\right)$. From the monotone convergence theorem,

$$I^2 = \lim_{R\to\infty}\int_{D_R} e^{-(x^2+y^2)/2}dm^2(x,y)$$

where $D_R=\left\{(x,y): x^2+y^2<R^2\right\}$. Using the change of variables theorem described in Section **??** below,[1] we find

$$\int_{D_R} e^{-(x^2+y^2)/2}d\pi(x,y) = \int_{(0,R)\times(0,2\pi)} e^{-r^2/2}r dr d\theta$$
$$= 2\pi\int_0^R e^{-r^2/2}r dr = 2\pi\left(1-e^{-R^2/2}\right).$$

[1] Alternatively, you can easily show that the integral $\int_{D_R} f dm^2$ agrees with the multiple integral in undergraduate analysis when f is continuous. Then use the change of variables theorem from undergraduate analysis.

From this we learn that

$$I^2 = \lim_{R\to\infty} 2\pi \left(1 - e^{-R^2/2}\right) = 2\pi$$

as desired.

12.3 Fubini's Theorem

Notation 12.8 *If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is any measurable function, let*

$$\bar{\int}_X f d\mu := \begin{cases} \int_X f d\mu & \text{if} \quad \int_X |f| \, d\mu < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 12.9 (Fubini's Theorem). *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces, $\pi = \mu \otimes \nu$ is the product measure on $\mathcal{M} \otimes \mathcal{N}$ and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Then the following three conditions are equivalent:*

$$\int_{X\times Y} |f| \, d\pi < \infty, \text{ i.e. } f \in L^1(\pi), \tag{12.10}$$

$$\int_X \left(\int_Y |f(x, y)| \, d\nu(y)\right) d\mu(x) < \infty \text{ and} \tag{12.11}$$

$$\int_Y \left(\int_X |f(x, y)| \, d\mu(x)\right) d\nu(y) < \infty. \tag{12.12}$$

If any one (and hence all) of these condition hold, then $f(x, \cdot) \in L^1(\nu)$ for μ-a.e. x, $f(\cdot, y) \in L^1(\mu)$ for ν-a.e. y, $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$, $\bar{\int}_X f(x, \cdot)d\mu(x) \in L^1(\nu)$ and Eqs. (12.8) and (12.9) are still valid after putting a bar over the integral symbols.

Proof. The equivalence of Eqs. (12.10) – (12.12) is a direct consequence of Tonelli's Theorem 12.6. Now suppose $f \in L^1(\pi)$ is a real valued function and let

$$E := \left\{x \in X : \int_Y |f(x, y)| \, d\nu(y) = \infty\right\}. \tag{12.13}$$

Then by Tonelli's theorem, $x \to \int_Y |f(x, y)| \, d\nu(y)$ is measurable and hence $E \in \mathcal{M}$. Moreover Tonelli's theorem implies

$$\int_X \left[\int_Y |f(x, y)| \, d\nu(y)\right] d\mu(x) = \int_{X\times Y} |f| \, d\pi < \infty$$

which implies that $\mu(E) = 0$. Let $f_\pm$ be the positive and negative parts of f, then

$$\begin{aligned}\bar{\int}_Y f(x, y) \, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y) \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) [f_+(x, y) - f_-(x, y)] \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) f_+(x, y) \, d\nu(y) - \int_Y 1_{E^c}(x) f_-(x, y) \, d\nu(y).\end{aligned} \tag{12.14}$$

Noting that $1_{E^c}(x) f_\pm(x, y) = (1_{E^c} \otimes 1_Y \cdot f_\pm)(x, y)$ is a positive $\mathcal{M} \otimes \mathcal{N}$ – measurable function, it follows from another application of Tonelli's theorem that $x \to \bar{\int}_Y f(x, y) \, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

$$\int_X \left|\bar{\int}_Y f(x, y) \, d\nu(y)\right| d\mu(x) \le \int_X \left[\int_Y |f(x, y)| \, d\nu(y)\right] d\mu(x) < \infty,$$

which shows $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$. Integrating Eq. (12.14) on x and using Tonelli's theorem repeatedly implies,

$$\begin{aligned}&\int_X \left[\bar{\int}_Y f(x, y) \, d\nu(y)\right] d\mu(x) \\ &= \int_X d\mu(x) \int_Y d\nu(y) \, 1_{E^c}(x) f_+(x, y) - \int_X d\mu(x) \int_Y d\nu(y) \, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) \, 1_{E^c}(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) \, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) f_-(x, y) \\ &= \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-) \, d\pi = \int_{X\times Y} f d\pi \end{aligned} \tag{12.15}$$

which proves Eq. (12.8) holds.

Now suppose that $f = u + iv$ is complex valued and again let E be as in Eq. (12.13). Just as above we still have $E \in \mathcal{M}$ and $\mu(E) = 0$ and

$$\begin{aligned}\bar{\int}_Y f(x, y) \, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y) \, d\nu(y) = \int_Y 1_{E^c}(x) [u(x, y) + iv(x, y)] \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) u(x, y) \, d\nu(y) + i \int_Y 1_{E^c}(x) v(x, y) \, d\nu(y).\end{aligned}$$

The last line is a measurable in x as we have just proved. Similarly one shows $\bar{\int}_Y f(\cdot, y)\, d\nu(y) \in L^1(\mu)$ and Eq. (12.8) still holds by a computation similar to that done in Eq. (12.15). The assertions pertaining to Eq. (12.9) may be proved in the same way. ■

The previous theorems generalize to products of any finite number of σ – finite measure spaces.

Theorem 12.10. *Suppose $\{(X_i, \mathcal{M}_i, \mu_i)\}_{i=1}^n$ are σ – finite measure spaces and $X := X_1 \times \cdots \times X_n$. Then there exists a unique measure (π) on $(X, \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n)$ such that*

$$\pi(A_1 \times \cdots \times A_n) = \mu_1(A_1) \ldots \mu_n(A_n) \text{ for all } A_i \in \mathcal{M}_i. \tag{12.16}$$

(This measure and its completion will be denoted by $\mu_1 \otimes \cdots \otimes \mu_n$.) If $f : X \to [0, \infty]$ is a $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ – measurable function then

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) \; f(x_1, \ldots, x_n) \tag{12.17}$$

where σ is any permutation of $\{1, 2, \ldots, n\}$. In particular $f \in L^1(\pi)$, iff

$$\int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) \; |f(x_1, \ldots, x_n)| < \infty$$

for some (and hence all) permutations, σ. Furthermore, if $f \in L^1(\pi)$, then

$$\int_X f d\pi = \bar{\int}_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \bar{\int}_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) \; f(x_1, \ldots, x_n) \tag{12.18}$$

for all permutations σ.

Proof. (* I would consider skipping this tedious proof.) The proof will be by induction on n with the case $n = 2$ being covered in Theorems 12.6 and 12.9. So let $n \geq 3$ and assume the theorem is valid for $n - 1$ factors or less. To simplify notation, for $1 \leq i \leq n$, let $X^i = \prod_{j \neq i} X_j$, $\mathcal{M}^i := \otimes_{j \neq i} \mathcal{M}_i$, and $\mu^i := \otimes_{j \neq i} \mu_j$ be the product measure on $(X^i, \mathcal{M}^i)$ which is assumed to exist by the induction hypothesis. Also let $\mathcal{M} := \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ and for $x = (x_1, \ldots, x_i, \ldots, x_n) \in X$ let

$$x^i := (x_1, \ldots, \hat{x}_i, \ldots, x_n) := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Here is an outline of the argument with some details being left to the reader.

1. If $f : X \to [0, \infty]$ is $\mathcal{M}$ -measurable, then

$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} f(x_1, \ldots, x_i, \ldots, x_n)\, d\mu_i(x_i)$$

is $\mathcal{M}^i$ -measurable. Thus by the induction hypothesis, the right side of Eq. (12.17) is well defined.

2. If $\sigma \in S_n$ (the permutations of $\{1, 2, \ldots, n\}$) we may define a measure π on $(X, \mathcal{M})$ by;

$$\pi(A) := \int_{X_{\sigma 1}} d\mu_{\sigma 1}(x_{\sigma 1}) \ldots \int_{X_{\sigma n}} d\mu_{\sigma n}(x_{\sigma n})\, 1_A(x_1, \ldots, x_n). \tag{12.19}$$

It is easy to check that π is a measure which satisfies Eq. (12.16). Using the σ – finiteness assumptions and the fact that

$$\mathcal{P} := \{A_1 \times \cdots \times A_n : A_i \in \mathcal{M}_i \text{ for } 1 \leq i \leq n\}$$

is a π – system such that $\sigma(\mathcal{P}) = \mathcal{M}$, it follows from Exercise 8.1 that there is only one such measure satisfying Eq. (12.16). Thus the formula for π in Eq. (12.19) is independent of $\sigma \in S_n$.

3. From Eq. (12.19) and the usual simple function approximation arguments we may conclude that Eq. (12.17) is valid.
Now suppose that $f \in L^1(X, \mathcal{M}, \pi)$.

4. Using step 1 it is easy to check that

$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \bar{\int}_{X_i} f(x_1, \ldots, x_i, \ldots, x_n)\, d\mu_i(x_i)$$

is $\mathcal{M}^i$ – measurable. Indeed,

$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)|\, d\mu_i(x_i)$$

is $\mathcal{M}^i$ – measurable and therefore

$$E := \left\{(x_1, \ldots, \hat{x}_i, \ldots, x_n) : \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)|\, d\mu_i(x_i) < \infty\right\} \in \mathcal{M}^i.$$

Now let $u := \operatorname{Re} f$ and $v := \operatorname{Im} f$ and $u_\pm$ and $v_\pm$ are the positive and negative parts of u and v respectively, then

$$\begin{aligned} \bar{\int}_{X_i} f(x)\, d\mu_i(x_i) &= \int_{X_i} 1_E(x^i)\, f(x)\, d\mu_i(x_i) \\ &= \int_{X_i} 1_E(x^i)\, u(x)\, d\mu_i(x_i) + i \int_{X_i} 1_E(x^i)\, v(x)\, d\mu_i(x_i). \end{aligned}$$

Both of these later terms are $\mathcal{M}^i$ – measurable since, for example,

$$\int_{X_i} 1_E(x^i)\, u(x)\, d\mu_i(x_i) = \int_{X_i} 1_E(x^i)\, u_+(x)\, d\mu_i(x_i) - \int_{X_i} 1_E(x^i)\, u_-(x)\, d\mu_i(x_i)$$

which is $\mathcal{M}^i$ – measurable by step 1.

5. It now follows by induction that the right side of Eq. (12.18) is well defined.
6. Let $i := \sigma n$ and $T : X \to X_i \times X^i$ be the obvious identification;

$$T\left(x_i, (x_1, \ldots, \hat{x}_i, \ldots, x_n)\right) = (x_1, \ldots, x_n).$$

One easily verifies T is $\mathcal{M}/\mathcal{M}_i \otimes \mathcal{M}^i$ – measurable (use Corollary 9.19 repeatedly) and that $\pi \circ T^{-1} = \mu_i \otimes \mu^i$ (see Exercise 8.1).
7. Let $f \in L^1(\pi)$. Combining step 6. with the abstract change of variables Theorem (Exercise 10.11) implies

$$\int_X f d\pi = \int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right). \tag{12.20}$$

By Theorem 12.9, we also have

$$\begin{aligned}\int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right) &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f \circ T(x_i, x^i) \\ &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f(x_1, \ldots, x_n).\end{aligned} \tag{12.21}$$

Then by the induction hypothesis,

$$\int_{X^i} d\mu^i\left(x_i\right) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \ldots, x_n) = \prod_{j \neq i} \bar{\int}_{X_j} d\mu_j\left(x_j\right) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \ldots, x_n) \tag{12.22}$$

where the ordering the integrals in the last product are inconsequential. Combining Eqs. (12.20) – (12.22) completes the proof.

■

Convention: We are now going to drop the bar above the integral sign with the understanding that $\int_X f d\mu = 0$ whenever $f : X \to \mathbb{C}$ is a measurable function such that $\int_X |f|\, d\mu = \infty$. However if f is a non-negative function (i.e. $f : X \to [0, \infty]$) non-integrable function we will interpret $\int_X f d\mu$ to be infinite.

Example 12.11. In this example we will show

$$\lim_{M \to \infty} \int_0^M \frac{\sin x}{x} dx = \pi/2. \tag{12.23}$$

To see this write $\frac{1}{x} = \int_0^\infty e^{-tx} dt$ and use Fubini-Tonelli to conclude that

$$\begin{aligned}\int_0^M \frac{\sin x}{x} dx &= \int_0^M \left[\int_0^\infty e^{-tx} \sin x\ dt\right] dx \\ &= \int_0^\infty \left[\int_0^M e^{-tx} \sin x\ dx\right] dt \\ &= \int_0^\infty \frac{1}{1+t^2}\left(1 - te^{-Mt} \sin M - e^{-Mt} \cos M\right) dt \\ &\to \int_0^\infty \frac{1}{1+t^2} dt = \frac{\pi}{2} \text{ as } M \to \infty,\end{aligned}$$

wherein we have used the dominated convergence theorem (for instance, take $g(t) := \frac{1}{1+t^2}\left(1 + te^{-t} + e^{-t}\right))$ to pass to the limit.

The next example is a refinement of this result.

Example 12.12. We have

$$\int_0^\infty \frac{\sin x}{x} e^{-\Lambda x} dx = \frac{1}{2}\pi - \arctan \Lambda \text{ for all } \Lambda > 0 \tag{12.24}$$

and for$\Lambda, M \in [0, \infty)$,

$$\left|\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx - \frac{1}{2}\pi + \arctan \Lambda\right| \le C \frac{e^{-M\Lambda}}{M} \tag{12.25}$$

where $C = \max_{x \ge 0} \frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (12.23) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,

$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\, dy\right| \le \left|\int_0^x 1 dy\right| = |x|$$

so $\left|\frac{\sin x}{x}\right| \le 1$ for all $x \neq 0$. Making use of the identity

$$\int_0^\infty e^{-tx} dt = 1/x$$

and Fubini's theorem,

$$\begin{aligned}\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x \, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\ &= \int_0^\infty dt \int_0^M dx \sin x \, e^{-(\Lambda+t)x} \\ &= \int_0^\infty \frac{1 - (\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2 + 1} dt \\ &= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\ &= \frac{1}{2}\pi - \arctan \Lambda - \varepsilon(M,\Lambda) \qquad (12.26)\end{aligned}$$

where

$$\varepsilon(M,\Lambda) = \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt.$$

Since

$$\left| \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} \right| \le \frac{1+(\Lambda+t)}{(\Lambda+t)^2+1} \le C,$$

$$|\varepsilon(M,\Lambda)| \le \int_0^\infty e^{-M(\Lambda+t)} dt = C \frac{e^{-M\Lambda}}{M}.$$

This estimate along with Eq. (12.26) proves Eq. (12.25) from which Eq. (12.23) follows by taking $\Lambda \to \infty$ and Eq. (12.24) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

Lemma 12.13. *Suppose that X is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ is a C^1 – functions such that $\lim_{x\to-\infty} \varphi(x) = 0$ and either $\varphi'(x) \ge 0$ for all x or $\int_{\mathbb{R}} |\varphi'(x)| dx < \infty$. Then*

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^\infty \varphi'(y) P(X > y) \, dy.$$

Similarly if $X \ge 0$ and $\varphi : [0,\infty) \to \mathbb{R}$ is a C^1 – function such that $\varphi(0) = 0$ and either $\varphi' \ge 0$ or $\int_0^\infty |\varphi'(x)| dx < \infty$, then

$$\mathbb{E}[\varphi(X)] = \int_0^\infty \varphi'(y) P(X > y) \, dy.$$

Proof. By the fundamental theorem of calculus for all $M < \infty$ and $x \in \mathbb{R}$,

$$\varphi(x) = \varphi(-M) + \int_{-M}^x \varphi'(y) \, dy. \qquad (12.27)$$

Under the stated assumptions on φ, we may use either the monotone or the dominated convergence theorem to let $M \to \infty$ in Eq. (12.27) to find,

$$\varphi(x) = \int_{-\infty}^x \varphi'(y) \, dy = \int_{\mathbb{R}} 1_{y<x} \varphi'(y) \, dy \text{ for all } x \in \mathbb{R}.$$

Therefore,

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left[\int_{\mathbb{R}} 1_{y<X} \varphi'(y) \, dy\right] = \int_{\mathbb{R}} \mathbb{E}[1_{y<X}] \varphi'(y) \, dy = \int_{-\infty}^\infty \varphi'(y) P(X > y) \, dy,$$

where we applied Fubini's theorem for the second equality. The proof of the second assertion is similar and will be left to the reader. ■

Example 12.14. Here are a couple of examples involving Lemma 12.13.

1. Suppose X is a random variable, then
$$\mathbb{E}\left[e^X\right] = \int_{-\infty}^\infty P(X > y) e^y dy = \int_0^\infty P(X > \ln u) \, du, \qquad (12.28)$$
where we made the change of variables, $u = e^y$, to get the second equality.
2. If $X \ge 0$ and $p \ge 1$, then
$$\mathbb{E}X^p = p \int_0^\infty y^{p-1} P(X > y) \, dy. \qquad (12.29)$$

12.4 Fubini's Theorem and Completions*

Notation 12.15 *Given $E \subset X \times Y$ and $x \in X$, let*

$${}_xE := \{y \in Y : (x,y) \in E\}.$$

Similarly if $y \in Y$ is given let

$$E_y := \{x \in X : (x,y) \in E\}.$$

If $f : X \times Y \to \mathbb{C}$ is a function let $f_x = f(x,\cdot)$ and $f^y := f(\cdot,y)$ so that $f_x : Y \to \mathbb{C}$ and $f^y : X \to \mathbb{C}$.

Theorem 12.16. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are **complete** σ – finite measure spaces. Let $(X\times Y,\mathcal{L},\lambda)$ be the completion of $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$. If f is $\mathcal{L}$ – measurable and (a) $f \ge 0$ or (b) $f \in L^1(\lambda)$ then f_x is $\mathcal{N}$ – measurable for μ-a.e. x and f^y is $\mathcal{M}$ – measurable for ν-a.e. y and in case (b) $f_x \in L^1(\nu)$ and $f^y \in L^1(\mu)$ for μ-a.e. x and ν-a.e. y respectively. Moreover,*

$$\left(x \to \int_Y f_x d\nu\right) \in L^1(\mu) \text{ and } \left(y \to \int_X f^y d\mu\right) \in L^1(\nu)$$

and

$$\int_{X\times Y} f d\lambda = \int_Y d\nu \int_X d\mu\, f = \int_X d\mu \int_Y d\nu\, f.$$

Proof. If $E \in \mathcal{M} \otimes \mathcal{N}$ is a $\mu \otimes \nu$ null set (i.e. $(\mu \otimes \nu)(E) = 0$), then

$$0 = (\mu \otimes \nu)(E) = \int\limits_X \nu({}_xE) d\mu(x) = \int\limits_X \mu(E_y) d\nu(y).$$

This shows that

$$\mu(\{x : \nu({}_xE) \neq 0\}) = 0 \text{ and } \nu(\{y : \mu(E_y) \neq 0\}) = 0,$$

i.e. $\nu({}_xE) = 0$ for μ-a.e. x and $\mu(E_y) = 0$ for ν-a.e. y. If h is $\mathcal{L}$ measurable and $h = 0$ for λ − a.e., then there exists $E \in \mathcal{M} \otimes \mathcal{N}$ such that $\{(x, y) : h(x, y) \neq 0\} \subset E$ and $(\mu \otimes \nu)(E) = 0$. Therefore $|h(x, y)| \le 1_E(x, y)$ and $(\mu \otimes \nu)(E) = 0$. Since

$$\{h_x \neq 0\} = \{y \in Y : h(x, y) \neq 0\} \subset {}_xE \text{ and}$$
$$\{h_y \neq 0\} = \{x \in X : h(x, y) \neq 0\} \subset E_y$$

we learn that for μ-a.e. x and ν-a.e. y that $\{h_x \neq 0\} \in \mathcal{M}$, $\{h_y \neq 0\} \in \mathcal{N}$, $\nu(\{h_x \neq 0\}) = 0$ and a.e. and $\mu(\{h_y \neq 0\}) = 0$. This implies $\int_Y h(x, y) d\nu(y)$ exists and equals 0 for μ-a.e. x and similarly that $\int_X h(x, y) d\mu(x)$ exists and equals 0 for ν-a.e. y. Therefore

$$0 = \int_{X\times Y} h d\lambda = \int_Y \left(\int_X h d\mu\right) d\nu = \int_X \left(\int_Y h d\nu\right) d\mu.$$

For general $f \in L^1(\lambda)$, we may choose $g \in L^1(\mathcal{M}\otimes\mathcal{N}, \mu\otimes\nu)$ such that $f(x, y) = g(x, y)$ for λ− a.e. (x, y). Define $h := f - g$. Then $h = 0$, λ− a.e. Hence by what we have just proved and Theorem 12.6 $f = g + h$ has the following properties:

1. For μ-a.e. x, $y \to f(x, y) = g(x, y) + h(x, y)$ is in $L^1(\nu)$ and
$$\int_Y f(x, y) d\nu(y) = \int_Y g(x, y) d\nu(y).$$
2. For ν-a.e. y, $x \to f(x, y) = g(x, y) + h(x, y)$ is in $L^1(\mu)$ and
$$\int_X f(x, y) d\mu(x) = \int_X g(x, y) d\mu(x).$$

From these assertions and Theorem 12.6, it follows that

$$\begin{aligned}\int_X d\mu(x) \int_Y d\nu(y) f(x, y) &= \int_X d\mu(x) \int_Y d\nu(y) g(x, y)\\ &= \int_Y d\nu(y) \int_Y d\nu(x) g(x, y)\\ &= \int_{X\times Y} g(x, y) d(\mu \otimes \nu)(x, y)\\ &= \int_{X\times Y} f(x, y) d\lambda(x, y).\end{aligned}$$

Similarly it is shown that

$$\int_Y d\nu(y) \int_X d\mu(x) f(x, y) = \int_{X\times Y} f(x, y) d\lambda(x, y).$$

■

12.5 Exercises

Exercise 12.1. Prove Theorem **??**. Suggestion, to get started define

$$\pi(A) := \int_{X_1} d\mu(x_1) \ldots \int_{X_n} d\mu(x_n)\, 1_A(x_1, \ldots, x_n)$$

and then show Eq. (**??**) holds. Use the case of two factors as the model of your proof.

Exercise 12.2. Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1, 2, 3$ be σ − finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1, x_2), x_3) = (x_1, x_2, x_3).$$

1. Show F is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ − measurable and F^{-1} is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ − measurable. That is
$$F : ((X_1 \times X_2)\times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2)\otimes\mathcal{M}_3) \to (X_1\times X_2\times X_3, \mathcal{M}_1\otimes\mathcal{M}_2\otimes\mathcal{M}_3)$$
is a “measure theoretic isomorphism.”
2. Let $\pi := F_* [(\mu_1 \otimes \mu_2) \otimes \mu_3]$, i.e. $\pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3] (F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then π is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.

3. Let $f : X_1 \times X_2 \times X_3 \to [0, \infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function. Verify the identity,
$$\int_{X_1 \times X_2 \times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1, x_2, x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

Exercise 12.3. Prove the second assertion of Theorem **??**. That is show m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$. **Hint:** Look at the proof of Theorem **??**.

Exercise 12.4. (Part of Folland Problem 2.46 on p. 69.) Let $X = [0,1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel σ – field on X, m be Lebesgue measure on $[0,1]$ and ν be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x,x) \in X^2 : x \in X\}$ be the diagonal in X^2. Show
$$\int_X \left[\int_X 1_D(x,y) d\nu(y) \right] dm(x) \neq \int_X \left[\int_X 1_D(x,y) dm(x) \right] d\nu(y)$$
by explicitly computing both sides of this equation.

Exercise 12.5. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 12.6. Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

Exercise 12.7. Folland Problem 2.55 on p. 77. (Explicit integrations.)

Exercise 12.8. Folland Problem 2.56 on p. 77. Let $f \in L^1((0,a), dm)$, $g(x) = \int_x^a \frac{f(t)}{t} dt$ for $x \in (0,a)$, show $g \in L^1((0,a), dm)$ and
$$\int_0^a g(x) dx = \int_0^a f(t) dt.$$

Exercise 12.9. Show $\int_0^\infty \left| \frac{\sin x}{x} \right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0,\infty), m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

Exercise 12.10. Folland Problem 2.57 on p. 77.

Exercise 12.11. Folland Problem 2.58 on p. 77.

Exercise 12.12. Folland Problem 2.60 on p. 77. Properties of the Γ – function.

Exercise 12.13. Folland Problem 2.61 on p. 77. Fractional integration.

Exercise 12.14. Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on S^{n-1}.

Exercise 12.15. Folland Problem 2.64 on p. 80. On the integrability of $|x|^a \left|\log|x|\right|^b$ for x near 0 and x near ∞ in $\mathbb{R}^n$.

Exercise 12.16. Show, using Problem 12.14 that
$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d} \delta_{ij} \sigma\left(S^{d-1}\right).$$

Hint: show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of i and therefore
$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d} \sum_{j=1}^d \int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

Part III

Metric, Banach, and Hilbert Space Basics

13

Metric Spaces

Definition 13.1. *A function $d : X \times X \to [0, \infty)$ is called a metric if*

1. *(Symmetry) $d(x, y) = d(y, x)$ for all $x, y \in X$,*
2. *(Non-degenerate) $d(x, y) = 0$ if and only if $x = y \in X$, and*
3. *(Triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.*

As primary examples, any normed space $(X, \|\cdot\|)$ (see Definition 4.24) is a metric space with $d(x, y) := \|x - y\|$. Thus the space $\ell^p(\mu)$ (as in Theorem 4.25) is a metric space for all $p \in [1, \infty]$. Also any subset of a metric space is a metric space. For example a surface Σ in $\mathbb{R}^3$ is a metric space with the distance between two points on Σ being the usual distance in $\mathbb{R}^3$.

Definition 13.2. *Let (X, d) be a metric space. The* ***open ball*** *$B(x, \delta) \subset X$ centered at $x \in X$ with radius $\delta > 0$ is the set*

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\}.$$

We will often also write $B(x, \delta)$ as $B_x(\delta)$. We also define the ***closed ball*** *centered at $x \in X$ with radius $\delta > 0$ as the set $C_x(\delta) := \{y \in X : d(x, y) \leq \delta\}$.*

Definition 13.3. *A sequence $\{x_n\}_{n=1}^\infty$ in a metric space (X, d) is said to be* ***convergent*** *if there exists a point $x \in X$ such that $\lim_{n\to\infty} d(x, x_n) = 0$. In this case we write $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ as $n \to \infty$.*

Exercise 13.1. Show that x in Definition 13.3 is necessarily unique.

Definition 13.4. *A set $E \subset X$ is* ***bounded*** *if $E \subset B(x, R)$ for some $x \in X$ and $R < \infty$. A set $F \subset X$ is* ***closed*** *iff every convergent sequence $\{x_n\}_{n=1}^\infty$ which is contained in F has its limit back in F. A set $V \subset X$ is* ***open*** *iff V^c is closed. We will write $F \sqsubset X$ to indicate F is a closed subset of X and $V \subset_o X$ to indicate the V is an open subset of X. We also let τ_d denote the collection of open subsets of X relative to the metric d.*

Exercise 13.2. Let $\mathcal{F}$ be a collection of closed subsets of X, show $\cap\mathcal{F} := \cap_{F\in\mathcal{F}} F$ is closed. Also show that finite unions of closed sets are closed, i.e. if $\{F_k\}_{k=1}^K$ are closed sets then $\cup_{k=1}^K F_k$ is closed. (By taking complements, this shows that the collection of open sets, τ_d, is closed under finite intersections and arbitrary unions.) Show by example that a countable union of closed sets need not be closed.

Exercise 13.3. Show that $V \subset X$ is open iff for every $x \in V$ there is a $\delta > 0$ such that $B_x(\delta) \subset V$. In particular show $B_x(\delta)$ is open for all $x \in X$ and $\delta > 0$. **Hint:** by definition V is not open iff V^c is not closed.

Definition 13.5. *A subset $A \subset X$ is a* ***neighborhood*** *of x if there exists an open set $V \subset_o X$ such that $x \in V \subset A$. We will say that $A \subset X$ is an* ***open neighborhood*** *of x if A is open and $x \in A$.*

The following "continuity" facts of the metric d will be used frequently in the remainder of this book.

Lemma 13.6. *For any non empty subset $A \subset X$, let $d_A(x) := \inf\{d(x, a) | a \in A\}$, then*

$$|d_A(x) - d_A(y)| \leq d(x, y) \; \forall x, y \in X \tag{13.1}$$

and in particular if $x_n \to x$ in X then $d_A(x_n) \to d_A(x)$ as $n \to \infty$. Moreover the set $F_\varepsilon := \{x \in X | d_A(x) \geq \varepsilon\}$ is closed in X.

Proof. Let $a \in A$ and $x, y \in X$, then

$$d_A(x) \leq d(x, a) \leq d(x, y) + d(y, a).$$

Take the infimum over a in the above equation shows that

$$d_A(x) \leq d(x, y) + d_A(y) \quad \forall x, y \in X.$$

Therefore, $d_A(x) - d_A(y) \leq d(x, y)$ and by interchanging x and y we also have that $d_A(y) - d_A(x) \leq d(x, y)$ which implies Eq. (13.1). If $x_n \to x \in X$, then by Eq. (13.1),

$$|d_A(x) - d_A(x_n)| \leq d(x, x_n) \to 0 \text{ as } n \to \infty$$

so that $\lim_{n\to\infty} d_A(x_n) = d_A(x)$. Now suppose that $\{x_n\}_{n=1}^\infty \subset F_\varepsilon$ and $x_n \to x$ in X, then

$$d_A(x) = \lim_{n\to\infty} d_A(x_n) \geq \varepsilon$$

since $d_A(x_n) \geq \varepsilon$ for all n. This shows that $x \in F_\varepsilon$ and hence F_ε is closed. ■

Corollary 13.7. *The function d satisfies,*

$$|d(x,y) - d(x',y')| \leq d(y,y') + d(x,x').$$

In particular $d : X \times X \to [0,\infty)$ *is "continuous" in the sense that* $d(x,y)$ *is close to* $d(x',y')$ *if* x *is close to* x' *and* y *is close to* y'. *(The notion of continuity will be developed shortly.)*

Proof. By Lemma 13.6 for single point sets and the triangle inequality for the absolute value of real numbers,

$$\begin{aligned}|d(x,y) - d(x',y')| &\leq |d(x,y) - d(x,y')| + |d(x,y') - d(x',y')| \\ &\leq d(y,y') + d(x,x').\end{aligned}$$

■

Example 13.8. Let $x \in X$ and $\delta > 0$, then $C_x(\delta)$ and $B_x(\delta)^c$ are closed subsets of X. For example if $\{y_n\}_{n=1}^\infty \subset C_x(\delta)$ and $y_n \to y \in X$, then $d(y_n, x) \leq \delta$ for all n and using Corollary 13.7 it follows $d(y,x) \leq \delta$, i.e. $y \in C_x(\delta)$. A similar proof shows $B_x(\delta)^c$ is closed, see Exercise 13.3.

Lemma 13.9 (Approximating open sets from the inside by closed sets). *Let* A *be a closed subset of* X *and* $F_\varepsilon := \{x \in X | d_A(x) \geq \varepsilon\} \sqsubset X$ *be as in Lemma 13.6. Then* $F_\varepsilon \uparrow A^c$ *as* $\varepsilon \downarrow 0$.

Proof. It is clear that $d_A(x) = 0$ for $x \in A$ so that $F_\varepsilon \subset A^c$ for each $\varepsilon > 0$ and hence $\cup_{\varepsilon>0} F_\varepsilon \subset A^c$. Now suppose that $x \in A^c \subset_o X$. By Exercise 13.3 there exists an $\varepsilon > 0$ such that $B_x(\varepsilon) \subset A^c$, i.e. $d(x,y) \geq \varepsilon$ for all $y \in A$. Hence $x \in F_\varepsilon$ and we have shown that $A^c \subset \cup_{\varepsilon>0} F_\varepsilon$. Finally it is clear that $F_\varepsilon \subset F_{\varepsilon'}$ whenever $\varepsilon' \leq \varepsilon$. ■

Definition 13.10. *Given a set* A *contained in a metric space* X, *let* $\bar{A} \subset X$ *be the* ***closure of*** A *defined by*

$$\bar{A} := \{x \in X : \exists \{x_n\} \subset A \ni x = \lim_{n\to\infty} x_n\}.$$

That is to say $\bar{A}$ *contains all* ***limit points*** *of* A. *We say* A *is* ***dense in*** X *if* $\bar{A} = X$, *i.e. every element* $x \in X$ *is a limit of a sequence of elements from* A. *A metric space is said to be* ***separable*** *if it contains a countable dense subset,* D.

Exercise 13.4. Given $A \subset X$, show $\bar{A}$ is a closed set and in fact

$$\bar{A} = \cap\{F : A \subset F \subset X \text{ with } F \text{ closed}\}. \tag{13.2}$$

That is to say $\bar{A}$ is the smallest closed set containing A.

Exercise 13.5. If D is a dense subset of a metric space (X,d) and $E \subset X$ is a subset such that to every point $x \in D$ there exists $\{x_n\}_{n=1}^\infty \subset E$ with $x = \lim_{n\to\infty} x_n$, then E is also a dense subset of X. If points in E well approximate every point in D and the points in D well approximate the points in X, then the points in E also well approximate all points in X.

Exercise 13.6. Suppose (X,d) is a metric space which contains an uncountable subset $\Lambda \subset X$ with the property that there exists $\varepsilon > 0$ such that $d(a,b) \geq \varepsilon$ for all $a, b \in \Lambda$ with $a \neq b$. Show that (X,d) is **not** separable.

13.1 Metric spaces as topological spaces

Let (X,d) be a metric space and let $\tau = \tau_d$ denote the collection of open subsets of X. (Recall $V \subset X$ is open iff V^c is closed iff for all $x \in V$ there exists an $\varepsilon = \varepsilon_x > 0$ such that $B(x, \varepsilon_x) \subset V$ iff V can be written as a (possibly uncountable) union of open balls.) Although we will stick with metric spaces in this chapter, it will be useful to introduce the definitions needed here in the more general context of a general "topological space," i.e. a space equipped with a collection of "open sets."

Definition 13.11 (Topological Space). *Let* X *be a set. A* ***topology*** *on* X *is a collection of subsets* (τ) *of* X *with the following properties;*

1. τ *contains both the empty set* $(\emptyset)$ *and* X.
2. τ *is closed under arbitrary unions.*
3. τ *is closed under finite intersections.*

The elements $V \in \tau$ *are called* ***open*** *subsets of* X. *A subset* $F \subset X$ *is said to be* ***closed*** *if* F^c *is open. I will write* $V \subset_o X$ *to indicate that* $V \subset X$ *and* $V \in \tau$ *and similarly* $F \sqsubset X$ *will denote* $F \subset X$ *and* F *is closed. Given* $x \in X$ *we say that* $V \subset X$ *is an* ***open neighborhood*** *of* x *if* $V \in \tau$ *and* $x \in V$. *Let* $\tau_x = \{V \in \tau : x \in V\}$ *denote the collection of open neighborhoods of* x.

Of course every metric space (X,d) is also a topological space where we take $\tau = \tau_d$.

Definition 13.12. *Let* (X,τ) *be a topological space and* A *be a subset of* X.

1. *The* ***closure*** *of* A *is the smallest closed set* $\bar{A}$ *containing* A, *i.e.*

$$\bar{A} := \cap\{F : A \subset F \sqsubset X\}.$$

(Because of Exercise 13.4 this is consistent with Definition 13.10 for the closure of a set in a metric space.)

2. *The **interior** of A is the largest open set A^o contained in A, i.e.*

$$A^o = \cup\{V \in \tau : V \subset A\}.$$

3. *$A \subset X$ is a **neighborhood of a point** $x \in X$ if $x \in A^o$.*
4. *The **accumulation points** of A is the set*

$$\mathrm{acc}(A) = \{x \in X : V \cap [A \setminus \{x\}] \neq \emptyset \text{ for all } V \in \tau_x\}.$$

5. *The **boundary** of A is the set $\mathrm{bd}(A) := \bar{A} \setminus A^o$.*
6. *A is **dense** in X if $\bar{A} = X$ and X is said to be **separable** if there exists a countable dense subset of X.*

Remark 13.13. The relationships between the interior and the closure of a set are:

$$(A^o)^c = \bigcap\{V^c : V \in \tau \text{ and } V \subset A\} = \bigcap\{C : C \text{ is closed } C \supset A^c\} = \overline{A^c}$$

and similarly, $(\bar{A})^c = (A^c)^o$. Hence the boundary of A may be written as

$$\mathrm{bd}(A) := \bar{A} \setminus A^o = \bar{A} \cap (A^o)^c = \bar{A} \cap \overline{A^c}, \tag{13.3}$$

which is to say $\mathrm{bd}(A)$ consists of the points in both the closures of A and A^c.

13.1.1 Continuity

Suppose now that (X, ρ) and (Y, d) are two metric spaces and $f : X \to Y$ is a function.

Definition 13.14. *A function $f : X \to Y$ is **continuous at** $x \in X$ if for all $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$d(f(x), f(x')) < \varepsilon \text{ provided that } \rho(x, x') < \delta. \tag{13.4}$$

*The function f is said to be **continuous** if f is continuous at all points $x \in X$.*

The following lemma gives two other characterizations of continuity of a function at a point.

Lemma 13.15 (Local Continuity Lemma). *Suppose that (X, ρ) and (Y, d) are two metric spaces and $f : X \to Y$ is a function defined in a neighborhood of a point $x \in X$. Then the following are equivalent:*

1. *f is continuous at $x \in X$.*
2. *For all neighborhoods $A \subset Y$ of $f(x)$, $f^{-1}(A)$ is a neighborhood of $x \in X$.*
3. *For all sequences $\{x_n\}_{n=1}^\infty \subset X$ such that $x = \lim_{n\to\infty} x_n$, $\{f(x_n)\}$ is convergent in Y and*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Proof. 1 $\implies$ 2. If $A \subset Y$ is a neighborhood of $f(x)$, there exists $\varepsilon > 0$ such that $B_{f(x)}(\varepsilon) \subset A$ and because f is continuous there exists a $\delta > 0$ such that Eq. (13.4) holds. Therefore

$$B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right) \subset f^{-1}(A)$$

showing $f^{-1}(A)$ is a neighborhood of x.

2 $\implies$ 3. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ and $x = \lim_{n\to\infty} x_n$. Then for any $\varepsilon > 0$, $B_{f(x)}(\varepsilon)$ is a neighborhood of $f(x)$ and so $f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ is a neighborhood of x which must contain $B_x(\delta)$ for some $\delta > 0$. Because $x_n \to x$, it follows that $x_n \in B_x(\delta) \subset f^{-1}\left(B_{f(x)}(\varepsilon)\right)$ for a.a. n and this implies $f(x_n) \in B_{f(x)}(\varepsilon)$ for a.a. n, i.e. $d(f(x), f(x_n)) < \varepsilon$ for a.a. n. Since $\varepsilon > 0$ is arbitrary it follows that $\lim_{n\to\infty} f(x_n) = f(x)$.

3. $\implies$ 1. We will show not 1. $\implies$ not 3. If f is not continuous at x, there exists an $\varepsilon > 0$ such that for all $n \in \mathbb{N}$ there exists a point $x_n \in X$ with $\rho(x_n, x) < \frac{1}{n}$ yet $d(f(x_n), f(x)) \geq \varepsilon$. Hence $x_n \to x$ as $n \to \infty$ yet $f(x_n)$ does not converge to $f(x)$. ■

Here is a global version of the previous lemma.

Lemma 13.16 (Global Continuity Lemma). *Suppose that (X, ρ) and (Y, d) are two metric spaces and $f : X \to Y$ is a function defined on all of X. Then the following are equivalent:*

1. *f is continuous.*
2. *$f^{-1}(V) \in \tau_\rho$ for all $V \in \tau_d$, i.e. $f^{-1}(V)$ is open in X if V is open in Y.*
3. *$f^{-1}(C)$ is closed in X if C is closed in Y.*
4. *For all convergent sequences $\{x_n\} \subset X$, $\{f(x_n)\}$ is convergent in Y and*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Proof. Since $f^{-1}(A^c) = \left[f^{-1}(A)\right]^c$, it is easily seen that 2. and 3. are equivalent. So because of Lemma 13.15 it only remains to show 1. and 2. are equivalent. If f is continuous and $V \subset Y$ is open, then for every $x \in f^{-1}(V)$, V is a neighborhood of $f(x)$ and so $f^{-1}(V)$ is a neighborhood of x. Hence $f^{-1}(V)$ is a neighborhood of all of its points and from this and Exercise 13.3 it follows that $f^{-1}(V)$ is open. Conversely, if $x \in X$ and $A \subset Y$ is a neighborhood of $f(x)$ then there exists $V \subset_o X$ such that $f(x) \in V \subset A$. Hence $x \in f^{-1}(V) \subset f^{-1}(A)$ and by assumption $f^{-1}(V)$ is open showing $f^{-1}(A)$ is a neighborhood of x. Therefore f is continuous at x and since $x \in X$ was arbitrary, f is continuous. ■

Definition 13.17 (Continuity at a point in topological terms). *Let (X, τ_X) and (Y, τ_Y) be topological spaces. A function $f : X \to Y$ is* ***continuous at a point*** *$x \in X$ if for every open neighborhood V of $f(x)$ there is an open neighborhood U of x such that $U \subset f^{-1}(V)$. See Figure 13.1.*

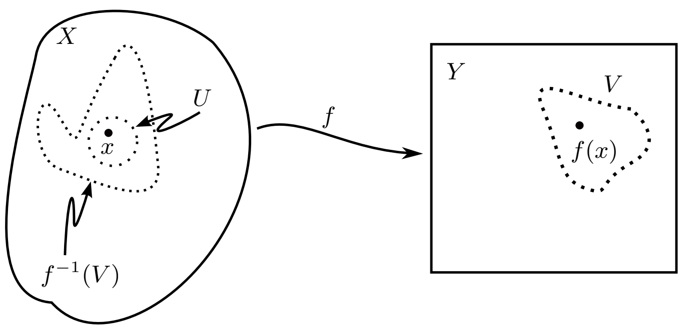


Fig. 13.1. Checking that a function is continuous at $x \in X$.

Definition 13.18 (Global continuity in topological terms). *Let (X, τ_X) and (Y, τ_Y) be topological spaces. A function $f : X \to Y$ is* ***continuous*** *if*

$$f^{-1}(\tau_Y) := \left\{ f^{-1}(V) : V \in \tau_Y \right\} \subset \tau_X.$$

We will also say that f is τ_X/τ_Y –continuous or (τ_X, τ_Y) – continuous. Let $C(X, Y)$ denote the set of continuous functions from X to Y.

Exercise 13.7. Show $f : X \to Y$ is continuous (Definition 35.16) iff f is continuous at all points $x \in X$.

Exercise 13.8. Show $f : X \to Y$ is continuous iff $f^{-1}(C)$ is closed in X for all closed subsets C of Y.

Definition 13.19. *A map $f : X \to Y$ between topological spaces is called a* ***homeomorphism*** *provided that f is bijective, f is continuous and $f^{-1} : Y \to X$ is continuous. If there exists $f : X \to Y$ which is a homeomorphism, we say that X and Y are homeomorphic. (As topological spaces X and Y are essentially the same.)*

Example 13.20. The function d_A defined in Lemma 13.6 is continuous for each $A \subset X$. In particular, if $A = \{x\}$, it follows that $y \in X \to d(y, x)$ is continuous for each $x \in X$.

Exercise 13.9. Use Example 13.20 and Lemma 13.16 to recover the results of Example 13.8.

The next result shows that there are lots of continuous functions on a metric space (X, d).

Lemma 13.21 (Urysohn's Lemma for Metric Spaces). *Let (X, d) be a metric space and suppose that A and B are two disjoint closed subsets of X. Then*

$$f(x) = \frac{d_B(x)}{d_A(x) + d_B(x)} \text{ for } x \in X \tag{13.5}$$

defines a continuous function, $f : X \to [0, 1]$, such that $f(x) = 1$ for $x \in A$ and $f(x) = 0$ if $x \in B$.

Proof. By Lemma 13.6, d_A and d_B are continuous functions on X. Since A and B are closed, $d_A(x) > 0$ if $x \notin A$ and $d_B(x) > 0$ if $x \notin B$. Since $A \cap B = \emptyset$, $d_A(x) + d_B(x) > 0$ for all x and $(d_A + d_B)^{-1}$ is continuous as well. The remaining assertions about f are all easy to verify. ∎

Sometimes Urysohn's lemma will be use in the following form. Suppose $F \subset V \subset X$ with F being closed and V being open, then there exists $f \in C(X, [0, 1]))$ such that $f = 1$ on F while $f = 0$ on V^c. This of course follows from Lemma 13.21 by taking $A = F$ and $B = V^c$.

13.2 Completeness in Metric Spaces

Definition 13.22 (Cauchy sequences). *A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space (X, d) is* ***Cauchy*** *provided that*

$$\lim_{m,n\to\infty} d(x_n, x_m) = 0.$$

Exercise 13.10. Show that convergent sequences are always Cauchy sequences. The converse is not always true. For example, let $X = \mathbb{Q}$ be the set of rational numbers and $d(x, y) = |x - y|$. Choose a sequence $\{x_n\}_{n=1}^{\infty} \subset \mathbb{Q}$ which converges to $\sqrt{2} \in \mathbb{R}$, then $\{x_n\}_{n=1}^{\infty}$ is $(\mathbb{Q}, d)$ – Cauchy but not $(\mathbb{Q}, d)$ – convergent. The sequence does converge in $\mathbb{R}$ however.

Definition 13.23. *A metric space (X, d) is* ***complete*** *if all Cauchy sequences are convergent sequences.*

Exercise 13.11. Let (X, d) be a complete metric space. Let $A \subset X$ be a subset of X viewed as a metric space using $d|_{A\times A}$. Show that $(A, d|_{A\times A})$ is complete iff A is a closed subset of X.

Example 13.24. Examples 2. – 4. of complete metric spaces will be verified in Chapter 14 below.

1. $X = \mathbb{R}$ and $d(x, y) = |x - y|$, see Theorem 3.8 above.
2. $X = \mathbb{R}^n$ and $d(x, y) = \|x - y\|_2 = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$.
3. $X = \ell^p(\mu)$ for $p \in [1, \infty]$ and any weight function $\mu : X \to (0, \infty)$.
4. $X = C([0, 1], \mathbb{R})$ – the space of continuous functions from $[0, 1]$ to $\mathbb{R}$ and
$$d(f, g) := \max_{t \in [0,1]} |f(t) - g(t)|.$$
This is a special case of Lemma 14.4 below.
5. Let $X = C([0, 1], \mathbb{R})$ and
$$d(f, g) := \int_0^1 |f(t) - g(t)|\, dt.$$
You are asked in Exercise 14.13 to verify that (X, d) is a metric space which is **not** complete.

Exercise 13.12 (Completions of Metric Spaces). Suppose that (X, d) is a (not necessarily complete) metric space. Using the following outline show there exists a complete metric space $(\bar{X}, \bar{d})$ and an isometric map $i : X \to \bar{X}$ such that $i(X)$ is dense in $\bar{X}$, see Definition 13.10.

1. Let $\mathcal{C}$ denote the collection of Cauchy sequences $a = \{a_n\}_{n=1}^\infty \subset X$. Given two element $a, b \in \mathcal{C}$ show $d_{\mathcal{C}}(a, b) := \lim_{n\to\infty} d(a_n, b_n)$ exists, $d_{\mathcal{C}}(a, b) \geq 0$ for all $a, b \in \mathcal{C}$ and $d_{\mathcal{C}}$ satisfies the triangle inequality,
$$d_{\mathcal{C}}(a, c) \leq d_{\mathcal{C}}(a, b) + d_{\mathcal{C}}(b, c) \text{ for all } a, b, c \in \mathcal{C}.$$
Thus $(\mathcal{C}, d_{\mathcal{C}})$ would be a metric space if it were true that $d_{\mathcal{C}}(a, b) = 0$ iff $a = b$. This however is false, for example if $a_n = b_n$ for all $n \geq 100$, then $d_{\mathcal{C}}(a, b) = 0$ while a need not equal b.
2. Define two elements $a, b \in \mathcal{C}$ to be equivalent (write $a \sim b$) whenever $d_{\mathcal{C}}(a, b) = 0$. Show " $\sim$ " is an equivalence relation on $\mathcal{C}$ and that $d_{\mathcal{C}}(a', b') = d_{\mathcal{C}}(a, b)$ if $a \sim a'$ and $b \sim b'$. (**Hint:** see Corollary 13.7.)
3. Given $a \in \mathcal{C}$ let $\bar{a} := \{b \in \mathcal{C} : b \sim a\}$ denote the equivalence class containing a and let $\bar{X} := \{\bar{a} : a \in \mathcal{C}\}$ denote the collection of such equivalence classes. Show that $\bar{d}(\bar{a}, \bar{b}) := d_{\mathcal{C}}(a, b)$ is well defined on $\bar{X} \times \bar{X}$ and verify $(\bar{X}, \bar{d})$ is a metric space.
4. For $x \in X$ let $i(x) = \bar{a}$ where a is the constant sequence, $a_n = x$ for all n. Verify that $i : X \to \bar{X}$ is an isometric map and that $i(X)$ is dense in $\bar{X}$.
5. Verify $(\bar{X}, \bar{d})$ is complete. **Hint:** if $\{\bar{a}(m)\}_{m=1}^\infty$ is a Cauchy sequence in $\bar{X}$ choose $b_m \in X$ such that $\bar{d}(i(b_m), \bar{a}(m)) \leq 1/m$. Then show $\bar{a}(m) \to \bar{b}$ where $b = \{b_m\}_{m=1}^\infty$.

13.3 Supplementary Remarks

13.3.1 Word of Caution

Example 13.25. Let (X, d) be a metric space. It is always true that $\overline{B_x(\varepsilon)} \subset C_x(\varepsilon)$ since $C_x(\varepsilon)$ is a closed set containing $B_x(\varepsilon)$. However, it is not always true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$. For example let $X = \{1, 2\}$ and $d(1, 2) = 1$, then $B_1(1) = \{1\}$, $\overline{B_1(1)} = \{1\}$ while $C_1(1) = X$. For another counterexample, take
$$X = \left\{(x, y) \in \mathbb{R}^2 : x = 0 \text{ or } x = 1\right\}$$
with the usually Euclidean metric coming from the plane. Then
$$\begin{aligned} B_{(0,0)}(1) &= \left\{(0, y) \in \mathbb{R}^2 : |y| < 1\right\}, \\ \overline{B_{(0,0)}(1)} &= \left\{(0, y) \in \mathbb{R}^2 : |y| \leq 1\right\}, \text{ while} \\ C_{(0,0)}(1) &= \overline{B_{(0,0)}(1)} \cup \{(1, 0)\}. \end{aligned}$$

In spite of the above examples, Lemmas 13.26 and 13.27 below shows that for certain metric spaces of interest it is true that $\overline{B_x(\varepsilon)} = C_x(\varepsilon)$.

Lemma 13.26. *Suppose that $(X, |\cdot|)$ is a normed vector space and d is the metric on X defined by $d(x, y) = |x - y|$. Then*
$$\overline{B_x(\varepsilon)} = C_x(\varepsilon) \text{ and}$$
$$\operatorname{bd}(B_x(\varepsilon)) = \{y \in X : d(x, y) = \varepsilon\}.$$
where the boundary operation, $\operatorname{bd}(\cdot)$ *is defined in Definition 35.29 (BRUCE: Forward Reference.) below.*

Proof. We must show that $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. For $y \in C$, let $v = y - x$, then
$$|v| = |y - x| = d(x, y) \leq \varepsilon.$$
Let $\alpha_n = 1 - 1/n$ so that $\alpha_n \uparrow 1$ as $n \to \infty$. Let $y_n = x + \alpha_n v$, then $d(x, y_n) = \alpha_n d(x, y) < \varepsilon$, so that $y_n \in B_x(\varepsilon)$ and $d(y, y_n) = (1 - \alpha_n)|v| \to 0$ as $n \to \infty$. This shows that $y_n \to y$ as $n \to \infty$ and hence that $y \in \bar{B}$. ■

13.3.2 Riemannian Metrics

This subsection is not completely self contained and may safely be skipped.

Lemma 13.27. *Suppose that X is a Riemannian (or sub-Riemannian) manifold and d is the metric on X defined by*
$$d(x, y) = \inf\{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}$$

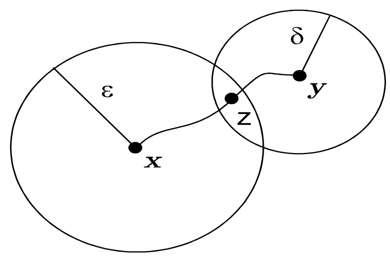


Fig. 13.2. An almost length minimizing curve joining x to y.

where $\ell(\sigma)$ is the length of the curve σ. We define $\ell(\sigma) = \infty$ if σ is not piecewise smooth.

Then

$$\overline{B_x(\varepsilon)} = C_x(\varepsilon) \text{ and}$$
$$\mathrm{bd}(B_x(\varepsilon)) = \{y \in X : d(x,y) = \varepsilon\}$$

where the boundary operation, bd$(\cdot)$ *is defined in Definition 35.29 below.*

Proof. Let $C := C_x(\varepsilon) \subset \overline{B_x(\varepsilon)} =: \bar{B}$. We will show that $C \subset \bar{B}$ by showing $\bar{B}^c \subset C^c$. Suppose that $y \in \bar{B}^c$ and choose $\delta > 0$ such that $B_y(\delta) \cap \bar{B} = \emptyset$. In particular this implies that

$$B_y(\delta) \cap B_x(\varepsilon) = \emptyset.$$

We will finish the proof by showing that $d(x,y) \geq \varepsilon + \delta > \varepsilon$ and hence that $y \in C^c$. This will be accomplished by showing: if $d(x,y) < \varepsilon + \delta$ then $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. If $d(x,y) < \max(\varepsilon,\delta)$ then either $x \in B_y(\delta)$ or $y \in B_x(\varepsilon)$. In either case $B_y(\delta) \cap B_x(\varepsilon) \neq \emptyset$. Hence we may assume that $\max(\varepsilon,\delta) \leq d(x,y) < \varepsilon + \delta$. Let $\alpha > 0$ be a number such that

$$\max(\varepsilon,\delta) \leq d(x,y) < \alpha < \varepsilon + \delta$$

and choose a curve σ from x to y such that $\ell(\sigma) < \alpha$. Also choose $0 < \delta' < \delta$ such that $0 < \alpha - \delta' < \varepsilon$ which can be done since $\alpha - \delta < \varepsilon$. Let $k(t) = d(y, \sigma(t))$ a continuous function on $[0,1]$ and therefore $k([0,1]) \subset \mathbb{R}$ is a connected set which contains 0 and $d(x,y)$. Therefore there exists $t_0 \in [0,1]$ such that $d(y, \sigma(t_0)) = k(t_0) = \delta'$. Let $z = \sigma(t_0) \in B_y(\delta)$ then

$$d(x,z) \leq \ell(\sigma|_{[0,t_0]}) = \ell(\sigma) - \ell(\sigma|_{[t_0,1]}) < \alpha - d(z,y) = \alpha - \delta' < \varepsilon$$

and therefore $z \in B_x(\varepsilon) \cap B_x(\delta) \neq \emptyset$. ■

Remark 13.28. Suppose again that X is a Riemannian (or sub-Riemannian) manifold and

$$d(x,y) = \inf\{\ell(\sigma) : \sigma(0) = x \text{ and } \sigma(1) = y\}.$$

Let σ be a curve from x to y and let $\varepsilon = \ell(\sigma) - d(x,y)$. Then for all $0 \leq u < v \leq 1$,

$$\begin{aligned} d(x,y) + \varepsilon = \ell(\sigma) &= \ell(\sigma|_{[0,u]}) + \ell(\sigma|_{[u,v]}) + \ell(\sigma|_{[v,1]}) \\ &\geq d(x,\sigma(u)) + \ell(\sigma|_{[u,v]}) + d(\sigma(v), y) \end{aligned}$$

and therefore, using the triangle inequality,

$$\begin{aligned} \ell(\sigma|_{[u,v]}) &\leq d(x,y) + \varepsilon - d(x,\sigma(u)) - d(\sigma(v), y) \\ &\leq d(\sigma(u), \sigma(v)) + \varepsilon. \end{aligned}$$

This leads to the following conclusions. If σ is within ε of a length minimizing curve from x to y then $\sigma|_{[u,v]}$ is within ε of a length minimizing curve from $\sigma(u)$ to $\sigma(v)$. In particular if σ is a length minimizing curve from x to y then $\sigma|_{[u,v]}$ is a length minimizing curve from $\sigma(u)$ to $\sigma(v)$.

13.4 Exercises

Exercise 13.13. Let (X,d) be a metric space. Suppose that $\{x_n\}_{n=1}^\infty \subset X$ is a sequence and set $\varepsilon_n := d(x_n, x_{n+1})$. Show that for $m > n$ that

$$d(x_n, x_m) \leq \sum_{k=n}^{m-1} \varepsilon_k \leq \sum_{k=n}^{\infty} \varepsilon_k.$$

Conclude from this that if

$$\sum_{k=1}^{\infty} \varepsilon_k = \sum_{n=1}^{\infty} d(x_n, x_{n+1}) < \infty$$

then $\{x_n\}_{n=1}^\infty$ is Cauchy. Moreover, show that if $\{x_n\}_{n=1}^\infty$ is a convergent sequence and $x = \lim_{n\to\infty} x_n$ then

$$d(x, x_n) \leq \sum_{k=n}^{\infty} \varepsilon_k.$$

Exercise 13.14. Show that (X,d) is a complete metric space iff every sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $\sum_{n=1}^\infty d(x_n, x_{n+1}) < \infty$ is a convergent sequence in X. You may find it useful to prove the following statements in the course of the proof.

1. If $\{x_n\}$ is Cauchy sequence, then there is a subsequence $y_j := x_{n_j}$ such that $\sum_{j=1}^{\infty} d(y_{j+1}, y_j) < \infty$.
2. If $\{x_n\}_{n=1}^{\infty}$ is Cauchy and there exists a subsequence $y_j := x_{n_j}$ of $\{x_n\}$ such that $x = \lim_{j\to\infty} y_j$ exists, then $\lim_{n\to\infty} x_n$ also exists and is equal to x.

Exercise 13.15. Suppose that $f : [0, \infty) \to [0, \infty)$ is a C^2 – function such that $f(0) = 0$, $f' > 0$ and $f'' \leq 0$ and (X, ρ) is a metric space. Show that $d(x, y) = f(\rho(x, y))$ is a metric on X. In particular show that

$$d(x,y) := \frac{\rho(x,y)}{1+\rho(x,y)}$$

is a metric on X. (Hint: use calculus to verify that $f(a + b) \leq f(a) + f(b)$ for all $a, b \in [0, \infty)$.)

Exercise 13.16. Let $\{(X_n, d_n)\}_{n=1}^{\infty}$ be a sequence of metric spaces, $X := \prod_{n=1}^{\infty} X_n$, and for $x = (x(n))_{n=1}^{\infty}$ and $y = (y(n))_{n=1}^{\infty}$ in X let

$$d(x,y) = \sum_{n=1}^{\infty} 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))}. \tag{13.6}$$

Show:

1. (X, d) is a metric space,
2. a sequence $\{x_k\}_{k=1}^{\infty} \subset X$ converges to $x \in X$ iff $x_k(n) \to x(n) \in X_n$ as $k \to \infty$ for each $n \in \mathbb{N}$ and
3. X is complete if X_n is complete for all n.

Exercise 13.17. Suppose (X, ρ) and (Y, d) are metric spaces and A is a dense subset of X.

1. Show that if $F : X \to Y$ and $G : X \to Y$ are two continuous functions such that $F = G$ on A then $F = G$ on X. **Hint:** consider the set $C := \{x \in X : F(x) = G(x)\}$.
2. Suppose $f : A \to Y$ is a function which is uniformly continuous, i.e. for every $\varepsilon > 0$ there exists a $\delta > 0$ such that
$$d(f(a), f(b)) < \varepsilon \text{ for all } a, b \in A \text{ with } \rho(a, b) < \delta.$$
Show there is a unique continuous function $F : X \to Y$ such that $F = f$ on A. **Hint:** each point $x \in X$ is a limit of a sequence consisting of elements from A.
3. Let $X = \mathbb{R} = Y$ and $A = \mathbb{Q} \subset X$, find a function $f : \mathbb{Q} \to \mathbb{R}$ which is continuous on $\mathbb{Q}$ but does **not** extend to a continuous function on $\mathbb{R}$.

14

Banach Spaces

Definition 14.1. *A **norm** on a vector space X is a function $\|\cdot\| : X \to [0, \infty)$ such that*

1. *(Homogeneity) $\|\lambda f\| = |\lambda| \, \|f\|$ for all $\lambda \in \mathbb{F}$ and $f \in X$.*
2. *(Triangle inequality) $\|f + g\| \leq \|f\| + \|g\|$ for all $f, g \in X$.*
3. *(Positive definite) $\|f\| = 0$ implies $f = 0$.*

*A function $p : X \to [0, \infty)$ satisfying properties 1. and 2. but not necessarily 3. above will be called a **semi-norm** on X.*

*A pair $(X, \|\cdot\|)$ where X is a vector space and $\|\cdot\|$ is a norm on X is called a **normed vector space.***

Let $(X, \|\cdot\|)$ be a normed vector space, then $d(x, y) := \|x - y\|$ is easily seen to be a metric on X. We say $\{x_n\}_{n=1}^\infty \subset X$ **converges to** $x \in X$ (and write $\lim_{n\to\infty} x_n = x$ or $x_n \to x$) if

$$0 = \lim_{n\to\infty} d(x, x_n) = \lim_{n\to\infty} \|x - x_n\| .$$

Similarly $\{x_n\}_{n=1}^\infty \subset X$ is said to be a **Cauchy sequence** if

$$0 = \lim_{m,n\to\infty} d(x_m, x_n) = \lim_{m,n\to\infty} \|x_m - x_n\| .$$

Definition 14.2 (Banach space). *A normed vector space $(X, \|\cdot\|)$ is a **Banach space** if the associated metric space (X, d) is complete, i.e. all Cauchy sequences are convergent.*

Remark 14.3. Since $\|x\| = d(x, 0)$, it follows from Lemma 13.6 that $\|\cdot\|$ is a continuous function on X and that

$$|\|x\| - \|y\|| \leq \|x - y\| \text{ for all } x, y \in X.$$

It is also easily seen that the vector addition and scalar multiplication are continuous on any normed space as the reader is asked to verify in Exercise 14.7. These facts will often be used in the sequel without further mention.

The next lemma contains a few simple examples of Banach spaces. We will see many more examples throughout the book.

Lemma 14.4. *Suppose that X is a set then the bounded functions, $\ell^\infty(X)$, on X is a Banach space with the norm*

$$\|f\| = \|f\|_\infty = \sup_{x\in X} |f(x)| .$$

Moreover if X is a metric space (more generally a topological space, see Chapter 35) the set $BC(X) \subset \ell^\infty(X) = B(X)$ is closed subspace of $\ell^\infty(X)$ and hence is also a Banach space.

Proof. Let $\{f_n\}_{n=1}^\infty \subset \ell^\infty(X)$ be a Cauchy sequence. Since for any $x \in X$, we have

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty \tag{14.1}$$

which shows that $\{f_n(x)\}_{n=1}^\infty \subset \mathbb{F}$ is a Cauchy sequence of numbers. Because $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) is complete, $f(x) := \lim_{n\to\infty} f_n(x)$ exists for all $x \in X$. Passing to the limit $n \to \infty$ in Eq. (14.1) implies

$$|f(x) - f_m(x)| \leq \lim\inf_{n\to\infty} \|f_n - f_m\|_\infty$$

and taking the supremum over $x \in X$ of this inequality implies

$$\|f - f_m\|_\infty \leq \lim\inf_{n\to\infty} \|f_n - f_m\|_\infty \to 0 \text{ as } m \to \infty$$

showing $f_m \to f$ in $\ell^\infty(X)$. For the second assertion, suppose that $\{f_n\}_{n=1}^\infty \subset BC(X) \subset \ell^\infty(X)$ and $f_n \to f \in \ell^\infty(X)$. We must show that $f \in BC(X)$, i.e. that f is continuous. To this end let $x, y \in X$, then

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\leq 2 \|f - f_n\|_\infty + |f_n(x) - f_n(y)| . \end{aligned} \tag{14.2}$$

Thus if $\varepsilon > 0$, we may choose n large so that $2\|f - f_n\|_\infty < \varepsilon/2$ and then for this n there exists an open neighborhood V_x of $x \in X$ such that $|f_n(x) - f_n(y)| < \varepsilon/2$ for $y \in V_x$. Thus $|f(x) - f(y)| < \varepsilon$ for $y \in V_x$ showing the limiting function f is continuous.

Alternative ending. From Eq. (14.2) we learn

$$\limsup_{y\to x} |f(x) - f(y)| \leq 2 \|f - f_n\|_\infty \to 0 \text{ as } n \to \infty,$$

where

$$\limsup_{y\to x}|f(x)-f(y)| := \lim_{\varepsilon\downarrow 0}\sup_{d(y,x)<\varepsilon}|f(x)-f(y)|$$

and we have used the continuity of f_n to assert $\limsup_{y\to x}|f_n(x)-f_n(y)|=0$. ■

Exercise 14.1. Let $Y = BC(\mathbb{R},\mathbb{C})$ be the Banach space of continuous bounded complex valued functions on $\mathbb{R}$ equipped with the uniform norm, $\|f\|_u := \sup_{x\in\mathbb{R}}|f(x)|$. Further let $C_0(\mathbb{R},\mathbb{C})$ denote those $f\in C(\mathbb{R},\mathbb{C})$ such that vanish at infinity, i.e. $\lim_{x\to\pm\infty}f(x)=0$. Also let $C_c(\mathbb{R},\mathbb{C})$ denote those $f\in C(\mathbb{R},\mathbb{C})$ with compact support, i.e. there exists $N<\infty$ such that $f(x)=0$ if $|x|\geq N$. Show $C_0(\mathbb{R},\mathbb{C})$ is a closed subspace of Y and that

$$\overline{C_c(\mathbb{R},\mathbb{C})} = C_0(\mathbb{R},\mathbb{C}).$$

Here is an application of Urysonhn's Lemma 13.21 and Lemma 14.4

Theorem 14.5 (Metric Space Tietze Extension Theorem). *Let (X,d) be a metric space, D be a closed subset of X, $-\infty<a<b<\infty$ and $f\in C(D,[a,b])$. (Here we are viewing D as a metric space with metric $d_D := d|_{D\times D}$.) Then there exists $F\in C(X,[a,b])$ such that $F|_D = f$.*

Proof.

1. By scaling and translation (i.e. by replacing f by $(b-a)^{-1}(f-a)$), it suffices to prove Theorem 14.5 with $a=0$ and $b=1$.
2. Suppose $\alpha\in(0,1]$ and $f:D\to[0,\alpha]$ is continuous function. Let $A := f^{-1}([0,\frac{1}{3}\alpha])$ and $B := f^{-1}([\frac{2}{3}\alpha,\alpha])$. By Lemma 13.21 there exists a function $\tilde{g}\in C(X,[0,\frac{\alpha}{3}])$ such that $\tilde{g}=0$ on A and $\tilde{g}=1$ on B. Letting $g := \frac{\alpha}{3}\tilde{g}$, we have $g\in C(X,[0,\frac{\alpha}{3}])$ such that $g=0$ on A and $g=\frac{\alpha}{3}$ on B. Further notice that
$$0\leq f(x)-g(x)\leq\frac{2}{3}\alpha \text{ for all } x\in D.$$
3. Now suppose $f:D\to[0,1]$ is a continuous function as in step 1. Let $g_1\in C(X,[0,1/3])$ be as in step 2 with $\alpha=1$ (see Figure 14.1) and let $f_1 := f-g_1|_D\in C(D,[0,2/3])$. Now apply step 2. with $f=f_1$ and $\alpha=2/3$ to find $g_2\in C(X,[0,\frac{1}{3}\frac{2}{3}])$ such that $f_2 := f-(g_1+g_2)|_D\in C(D,[0,\left(\frac{2}{3}\right)^2])$. Continue this way inductively to find $g_n\in C(X,[0,\frac{1}{3}\left(\frac{2}{3}\right)^{n-1}])$ such that
$$f-\sum_{n=1}^N g_n|_D =: f_N\in C\left(D,\left[0,\left(\frac{2}{3}\right)^N\right]\right). \tag{14.3}$$
4. Define $F := \sum_{n=1}^\infty g_n$. Since
$$\sum_{n=1}^\infty\|g_n\|_\infty\leq\sum_{n=1}^\infty\frac{1}{3}\left(\frac{2}{3}\right)^{n-1} = \frac{1}{3}\frac{1}{1-\frac{2}{3}} = 1,$$
the series defining F is uniformly convergent so $F\in C(X,[0,1])$ via Lemma 14.4. Passing to the limit in Eq. (14.3) shows $f=F|_D$.

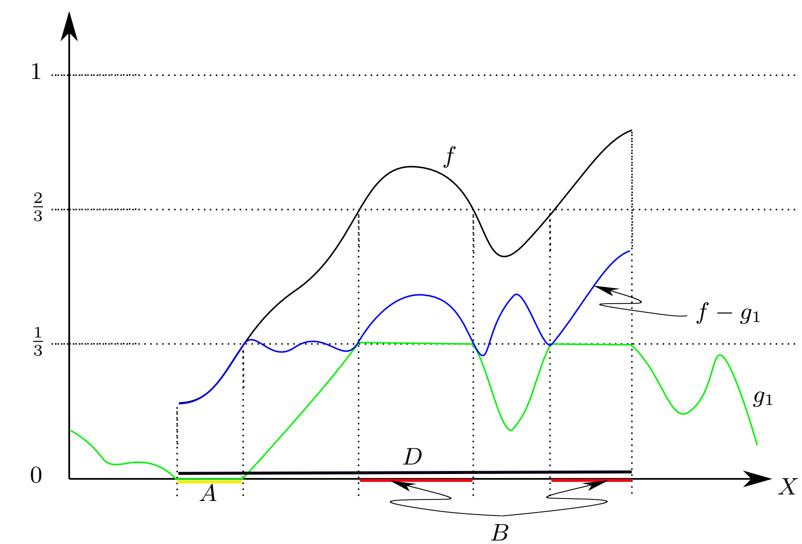


Fig. 14.1. Reducing f by subtracting off a globally defined function $g_1\in C\left(X,[0,\frac{1}{3}]\right)$. In this picture, D is depicted by the bold black line segment.

■

(The next theorem is a special case the scale of Banach spaces associated to a general measure space.)

Theorem 14.6 (Completeness of $\ell^p(\mu)$). *Let X be a set and $\mu: X\to(0,\infty)$ be a given function. Then for any $p\in[1,\infty]$, $(\ell^p(\mu),\|\cdot\|_p)$ is a Banach space.*

Proof. We have already proved this for $p=\infty$ in Lemma 14.4 so we now assume that $p\in[1,\infty)$. Let $\{f_n\}_{n=1}^\infty\subset\ell^p(\mu)$ be a Cauchy sequence. Since for any $x\in X$,

$$|f_n(x)-f_m(x)|\leq\frac{1}{\mu(x)}\|f_n-f_m\|_p\to 0 \text{ as } m,n\to\infty$$

it follows that $\{f_n(x)\}_{n=1}^{\infty}$ is a Cauchy sequence of numbers and $f(x) := \lim_{n\to\infty} f_n(x)$ exists for all $x \in X$. By Fatou's Lemma,

$$\begin{aligned}\|f_n - f\|_p^p &= \sum_X \mu \cdot \lim_{m\to\infty} \inf |f_n - f_m|^p \le \lim_{m\to\infty} \inf \sum_X \mu \cdot |f_n - f_m|^p \\ &= \lim_{m\to\infty} \inf \|f_n - f_m\|_p^p \to 0 \text{ as } n \to \infty.\end{aligned}$$

This then shows that $f = (f - f_n) + f_n \in \ell^p(\mu)$ (being the sum of two ℓ^p – functions) and that $f_n \xrightarrow{\ell^p} f$. ■

Remark 14.7. Let X be a set, Y be a Banach space and $\ell^\infty(X, Y)$ denote the bounded functions $f : X \to Y$ equipped with the norm

$$\|f\| = \|f\|_\infty = \sup_{x\in X} \|f(x)\|_Y .$$

If X is a metric space (or a general topological space, see Chapter 35), let $BC(X, Y)$ denote those $f \in \ell^\infty(X, Y)$ which are continuous. The same proof used in Lemma 14.4 shows that $\ell^\infty(X, Y)$ is a Banach space and that $BC(X, Y)$ is a closed subspace of $\ell^\infty(X, Y)$. Similarly, if $1 \le p < \infty$ we may define

$$\ell^p(X, Y) = \left\{ f : X \to Y : \|f\|_p = \left(\sum_{x\in X} \|f(x)\|_Y^p \right)^{1/p} < \infty \right\}.$$

The same proof as in Theorem 14.6 would then show that $\left(\ell^p(X, Y), \|\cdot\|_p\right)$ is a Banach space.

14.1 Bounded Linear Operators Basics

Definition 14.8. *Let X and Y be normed spaces and $T : X \to Y$ be a linear map. Then T is said to be bounded provided there exists $C < \infty$ such that $\|T(x)\|_Y \le C\|x\|_X$ for all $x \in X$. We denote the best constant by $\|T\|_{op} = \|T\|_{L(X,Y)}$, i.e.*

$$\|T\|_{L(X,Y)} = \sup_{x\neq 0} \frac{\|T(x)\|_Y}{\|x\|_X} = \sup_{x\neq 0} \{\|T(x)\|_Y : \|x\|_X = 1\}.$$

The number $\|T\|_{L(X,Y)}$ is called the operator norm of T.

In the future, we will usually drop the garnishing on the norms and simply write $\|x\|_X$ as $\|x\|$, $\|T\|_{L(X,Y)}$ as $\|T\|$, etc. The reader should be able to determine the norm that is to be used by context.

Proposition 14.9. *Suppose that X and Y are normed spaces and $T : X \to Y$ is a linear map. The the following are equivalent:*

1. *T is continuous.*
2. *T is continuous at 0.*
3. *T is bounded.*

Proof. 1. ⇒ 2. trivial. 2. ⇒ 3. If T continuous at 0 then there exist $\delta > 0$ such that $\|T(x)\| \le 1$ if $\|x\| \le \delta$. Therefore for any nonzero $x \in X$, $\|T(\delta x/\|x\|)\| \le 1$ which implies that $\|T(x)\| \le \frac{1}{\delta}\|x\|$ and hence $\|T\| \le \frac{1}{\delta} < \infty$. 3. ⇒ 1. Let $x \in X$ and $\varepsilon > 0$ be given. Then

$$\|Ty - Tx\| = \|T(y - x)\| \le \|T\| \; \|y - x\| < \varepsilon$$

provided $\|y - x\| < \varepsilon/\|T\| := \delta$. ■

Example 14.10 (An unbounded operator). Let $X = Y = \mathcal{P}$ – be the polynomial functions on $[0, 1]$ equipped with the uniform norm, $\|p\|_u := \sup_{x\in[0,1]} |p(x)|$ and let $D : X \to X$ be differentiation operator, $Dp = p'$ for all $p \in X$. Then $\|D\|_{op} = \infty$, i.e. D is unbounded. To see this is the case, let $p_n(x) = x^n$ so that $(Dp_n)(x) = nx^{n-1}$. Notice that

$$n = \|Dp_n\|_u \le \|D_{op}\| \cdot \|p_n\|_u = \|D_{op}\| \cdot 1 \text{ for all } n \in \mathbb{N}$$

from which if follows that $\|D\|_{op} = \infty$. [Compare this example with Example 14.12 below.]

Example 14.11 (Integral Operators). Suppose that $k : [0, 1] \times [0, 1] \to \mathbb{C}$ is a continuous function. For $f \in C([0, 1])$, let

$$Kf(x) = \int_0^1 k(x, y)f(y)dy.$$

Using the dominated convergence theorem one easily shows that $Kf \in C([0, 1])$ along with the linearity of the integral shows $K : C([0, 1]) \to C([0, 1])$ is a linear map.

1. Let us let $\|\cdot\|_\infty = \|\cdot\|_u$ denote the uniform norm on both copies of $C([0, 1])$ and compute $\|K\|_{op}$. Since

$$|Kf(x)| \le \int_0^1 |k(x, y)| \, |f(y)| \, dy \le \int_0^1 |k(x, y)| \, dy \cdot \|f\|_\infty \le A \|f\|_\infty$$

where

$$A := \sup_{x\in[0,1]} \int_0^1 |k(x, y)| \, dy < \infty, \tag{14.4}$$

we may conclude that $\|K\|_{op} \le A < \infty$. We may in fact show $\|K\| = A$. Since, by DCT, $x \to \int_0^1 |k(x,y)|\, dy$ is continuous, there exists $x_0 \in [0,1]$ such that

$$A = \sup_{x\in[0,1]} \int_0^1 |k(x,y)|\, dy = \int_0^1 |k(x_0,y)|\, dy.$$

Given $\varepsilon > 0$, let

$$f_\varepsilon(y) := \frac{\overline{k(x_0,y)}}{\varepsilon + |k(x_0,y)|} \in C([0,1])$$

so that

$$\|K\|_{op} \|f_\varepsilon\|_\infty \ge \|Kf_\varepsilon\|_\infty \ge |Kf_\varepsilon(x_0)| = Kf_\varepsilon(x_0) = \int_0^1 \frac{|k(x_0,y)|^2}{\varepsilon + |k(x_0,y)|} dy. \tag{14.5}$$

Since $\lim_{\varepsilon\downarrow 0} \|f_\varepsilon\|_\infty = 1$ and (by DCT)

$$\lim_{\varepsilon\downarrow 0} \int_0^1 \frac{|k(x_0,y)|^2}{\varepsilon + |k(x_0,y)|} dy = \int_0^1 |k(x_0,y)|\, dy = A,$$

we may let $\varepsilon \downarrow 0$ in Eq. (14.5) in order to conclude $\|K\|_{op} \ge A$.

2. We may also consider other norms on $C([0,1])$. Let (for now) $L^1([0,1])$ denote $C([0,1])$ with the norm

$$\|f\|_1 = \int_0^1 |f(x)|\, dx$$

and now consider $K : L^1([0,1], dm) \to C([0,1])$ is bounded as well. Indeed, let $M = \sup\{|k(x,y)| : x, y \in [0,1]\}$, then

$$|(Kf)(x)| \le \int_0^1 |k(x,y)f(y)|\, dy \le M \|f\|_1$$

which shows $\|Kf\|_\infty \le M \|f\|_1$ and hence,

$$\|K\|_{L^1\to C} \le \max\{|k(x,y)| : x, y \in [0,1]\} < \infty.$$

We can in fact show that $\|K\| = M$ as follows. Let $(x_0, y_0) \in [0,1]^2$ satisfying $|k(x_0,y_0)| = M$. Then given $\varepsilon > 0$, there exists a neighborhood $U = I \times J$ of (x_0,y_0) such that $|k(x,y) - k(x_0,y_0)| < \varepsilon$ for all $(x,y) \in U$. Let $f \in C_c(I, [0,\infty))$ such that $\int_0^1 f(x)dx = 1$. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha k(x_0,y_0) = M$, then

$$\begin{aligned} |(K\alpha f)(x_0)| &= \left|\int_0^1 k(x_0,y)\alpha f(y)dy\right| = \left|\int_I k(x_0,y)\alpha f(y)dy\right| \\ &\ge \operatorname{Re} \int_I \alpha k(x_0,y) f(y)dy \\ &\ge \int_I (M - \varepsilon) f(y)dy = (M-\varepsilon)\|\alpha f\|_{L^1} \end{aligned}$$

and hence

$$\|K\alpha f\|_C \ge (M-\varepsilon)\|\alpha f\|_{L^1}$$

showing that $\|K\| \ge M - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we learn that $\|K\| \ge M$ and hence $\|K\| = M$.

3. One may also view K as a map from $K : C([0,1]) \to L^1([0,1])$ in which case

$$\begin{aligned} \|Kf\|_1 &= \int_0^1 dx \left|\int_0^1 k(x,y)f(y)dy\right| \le \int_0^1 dx \int_0^1 dy\, |k(x,y)f(y)| \\ &\le \left[\int_0^1 dx \int_0^1 dy\, |k(x,y)|\right] \cdot \|f\|_\infty \end{aligned}$$

so that

$$\|K\|_{C\to L^1} \le \int_0^1 dx \int_0^1 dy\, |k(x,y)|\, .$$

Example 14.12 (Inverting differential operators). [This example used facts about L^p – spaces not yet covered in the notes!] Let Y denote those $f \in L^2([0,1], m)$ which have a C^1 – representative such that f' is absolutely continuous, $f'' \in L^2([0,1], m)$, $f(0) = 0 = f'(1)$ and for $f \in Y$ let $Lf = f''$. Let us try to invert L, i.e. given $g \in L^2([0,1], m)$ we wish to find $f \in Y$ such that $Lf = g$. Integrating $f'' = g$ shows,

$$\int_\sigma^1 g(y)\, dy = \int_\sigma^1 f''(y)\, dy = f'(1) - f'(\sigma) = -f'(\sigma)$$

and then integrating this equation implies

$$\begin{aligned} f(x) = f(x) - f(0) &= \int_0^x f'(\sigma)\, d\sigma \\ &= -\int_0^x \left[\int_\sigma^1 g(y)\, dy\right] d\sigma \\ &= -\int_{[0,1]^2} dy d\sigma 1_{0\le\sigma\le x, y\le 1} g(y) = -\int_0^1 \min(x,y)\, g(y)\, dy. \end{aligned}$$

This shows that

$$\left(L^{-1}g\right)(x) = -\int_0^1 \min(x,y)\, g(y)\, dy.$$

The function $\min(x,y)$ is the **Green's function** associated to L. If we let $K : L^2([0,1],m) \to L^2([0,1],m)$ be L^{-1} thought of as an operator from on $L^2([0,1],m)$, it is fairly easy to verify that K is bounded, see for example Exercise 18.5 below.

The next elementary theorem (referred to as the bounded linear transformation theorem, or B.L.T. theorem for short) is often useful when constructing bounded linear transformations.

Theorem 14.13 (B. L. T. Theorem). *Suppose that Z is a normed space, X is a Banach space, and $\mathcal{S} \subset Z$ is a dense linear subspace of Z. If $T : \mathcal{S} \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \le C\|z\|$ for all $z \in \mathcal{S}$), then T has a unique extension to an element $\bar{T} \in L(Z,X)$ and this extension still satisfies*

$$\left\|\bar{T}z\right\| \le C\|z\| \text{ for all } z \in \bar{\mathcal{S}}.$$

Exercise 14.2. Prove Theorem 14.13.

For the next three exercises, let $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ and $T : X \to Y$ be a linear transformation so that T is given by matrix multiplication by an $m \times n$ matrix. Let us identify the linear transformation T with this matrix.

Exercise 14.3. Assume the norms on X and Y are the ℓ^1 – norms, i.e. for $x \in \mathbb{R}^n$, $\|x\| = \sum_{j=1}^n |x_j|$. Then the operator norm of T is given by

$$\|T\| = \max_{1\le j\le n} \sum_{i=1}^m |T_{ij}|.$$

Exercise 14.4. Assume the norms on X and Y are the ℓ^∞ – norms, i.e. for $x \in \mathbb{R}^n$, $\|x\| = \max_{1\le j\le n}|x_j|$. Then the operator norm of T is given by

$$\|T\| = \max_{1\le i\le m} \sum_{j=1}^n |T_{ij}|.$$

Exercise 14.5. Assume the norms on X and Y are the ℓ^2 – norms, i.e. for $x \in \mathbb{R}^n$, $\|x\|^2 = \sum_{j=1}^n x_j^2$. Show $\|T\|^2$ is the largest eigenvalue of the matrix $T^{tr}T : \mathbb{R}^n \to \mathbb{R}^n$. **Hint:** Use the spectral theorem for symmetric real matrices.

Notation 14.14 *Let $L(X,Y)$ denote the bounded linear operators from X to Y and $L(X) = L(X,X)$. If $Y = \mathbb{F}$ we write X^* for $L(X,\mathbb{F})$ and call X^* the (continuous)* ***dual space*** *to X.*

Lemma 14.15. *Let X, Y be normed spaces, then the operator norm $\|\cdot\|$ on $L(X,Y)$ is a norm. Moreover if Z is another normed space and $T : X \to Y$ and $S : Y \to Z$ are linear maps, then $\|ST\| \le \|S\|\|T\|$, where $ST := S \circ T$.*

Proof. As usual, the main point in checking the operator norm is a norm is to verify the triangle inequality, the other axioms being easy to check. If $A, B \in L(X,Y)$ then the triangle inequality is verified as follows:

$$\begin{aligned}\|A+B\| &= \sup_{x\ne 0}\frac{\|Ax+Bx\|}{\|x\|} \le \sup_{x\ne 0}\frac{\|Ax\|+\|Bx\|}{\|x\|} \\ &\le \sup_{x\ne 0}\frac{\|Ax\|}{\|x\|} + \sup_{x\ne 0}\frac{\|Bx\|}{\|x\|} = \|A\| + \|B\|.\end{aligned}$$

For the second assertion, we have for $x \in X$, that

$$\|STx\| \le \|S\|\|Tx\| \le \|S\|\|T\|\|x\|.$$

From this inequality and the definition of $\|ST\|$, it follows that $\|ST\| \le \|S\|\|T\|$. ■

The reader is asked to prove the following continuity lemma in Exercise 14.11.

Lemma 14.16. *Let X, Y and Z be normed spaces. Then the maps*

$$(S,x) \in L(X,Y) \times X \longrightarrow Sx \in Y$$

and

$$(S,T) \in L(X,Y) \times L(Y,Z) \longrightarrow TS \in L(X,Z)$$

are continuous relative to the norms

$$\begin{aligned}\|(S,x)\|_{L(X,Y)\times X} &:= \|S\|_{L(X,Y)} + \|x\|_X \text{ and} \\ \|(S,T)\|_{L(X,Y)\times L(Y,Z)} &:= \|S\|_{L(X,Y)} + \|T\|_{L(Y,Z)}\end{aligned}$$

on $L(X,Y) \times X$ and $L(X,Y) \times L(Y,Z)$ respectively.

Proposition 14.17. *Suppose that X is a normed vector space and Y is a Banach space. Then $(L(X,Y), \|\cdot\|_{op})$ is a Banach space. In particular the dual space X^* is always a Banach space.*

Proof. Let $\{T_n\}_{n=1}^\infty$ be a Cauchy sequence in $L(X,Y)$. Then for each $x \in X$,

$$\|T_n x - T_m x\| \le \|T_n - T_m\|\|x\| \to 0 \text{ as } m,n \to \infty$$

showing $\{T_n x\}_{n=1}^\infty$ is Cauchy in Y. Using the completeness of Y, there exists an element $Tx \in Y$ such that

$$\lim_{n\to\infty} \|T_n x - Tx\| = 0.$$

The map $T : X \to Y$ is linear map, since for $x, x' \in X$ and $\lambda \in \mathbb{F}$ we have

$$T(x + \lambda x') = \lim_{n\to\infty} T_n (x + \lambda x') = \lim_{n\to\infty} [T_n x + \lambda T_n x'] = Tx + \lambda T x',$$

wherein we have used the continuity of the vector space operations in the last equality. Moreover,

$$\|Tx - T_n x\| \le \|Tx - T_m x\| + \|T_m x - T_n x\| \le \|Tx - T_m x\| + \|T_m - T_n\| \|x\|$$

and therefore

$$\begin{aligned}\|Tx - T_n x\| &\le \lim\inf_{m\to\infty} (\|Tx - T_m x\| + \|T_m - T_n\| \|x\|) \\ &= \|x\| \cdot \lim\inf_{m\to\infty} \|T_m - T_n\|.\end{aligned}$$

Hence

$$\|T - T_n\| \le \lim\inf_{m\to\infty} \|T_m - T_n\| \to 0 \text{ as } n \to \infty.$$

Thus we have shown that $T_n \to T$ in $L(X,Y)$ as desired. ∎

The following characterization of a Banach space will sometimes be useful in the sequel.

Theorem 14.18. *A normed space $(X, \|\cdot\|)$ is a Banach space iff every sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $\sum_{n=1}^\infty \|x_n\| < \infty$ implies $\lim_{N\to\infty} \sum_{n=1}^N x_n = s$ exists in X (that is to say every absolutely convergent series is a convergent series in X.) As usual we will denote s by $\sum_{n=1}^\infty x_n$.*

Proof. ($\Rightarrow$) If X is complete and $\sum_{n=1}^\infty \|x_n\| < \infty$ then sequence $s_N := \sum_{n=1}^N x_n$ for $N \in \mathbb{N}$ is Cauchy because (for $N > M$)

$$\|s_N - s_M\| \le \sum_{n=M+1}^N \|x_n\| \to 0 \text{ as } M, N \to \infty.$$

Therefore $s = \sum_{n=1}^\infty x_n := \lim_{N\to\infty} \sum_{n=1}^N x_n$ exists in X.

($\Longleftarrow$) Suppose that $\{s_n\}_{n=0}^\infty \subset X$ is a sequence such that

$$\infty > \sum_{n=0}^\infty d(s_n, s_{n-1}) = \sum_{n=0}^\infty \|s_n - s_{n-1}\|.$$

Then by assumption this implies $\sum_{n=0}^\infty (s_n - s_{n-1})$ exists in X, i.e.

$$\sum_{n=0}^\infty (s_n - s_{n-1}) = \lim_{N\to\infty} \sum_{n=0}^N (s_n - s_{n-1}) = \lim_{N\to\infty} (s_N - s_0).$$

This shows that $\{S_n\}_{n=0}^\infty$ is convergent in X and therefore we may appeal to Exercise 13.14 to see that (X,d) is complete, i.e. $(X, \|\cdot\|)$ is a Banach space. ∎

Example 14.19. Here is another proof of Proposition 14.17 which makes use of Theorem 14.18. Suppose that $T_n \in L(X,Y)$ is a sequence of operators such that $\sum_{n=1}^\infty \|T_n\| < \infty$. Then

$$\sum_{n=1}^\infty \|T_n x\| \le \sum_{n=1}^\infty \|T_n\| \|x\| < \infty$$

and therefore by the completeness of Y, $Sx := \sum_{n=1}^\infty T_n x = \lim_{N\to\infty} S_N x$ exists in Y, where $S_N := \sum_{n=1}^N T_n$. The reader should check that $S : X \to Y$ so defined is linear. Since,

$$\|Sx\| = \lim_{N\to\infty} \|S_N x\| \le \lim_{N\to\infty} \sum_{n=1}^N \|T_n x\| \le \sum_{n=1}^\infty \|T_n\| \|x\|,$$

S is bounded and

$$\|S\| \le \sum_{n=1}^\infty \|T_n\|. \tag{14.6}$$

Similarly,

$$\begin{aligned}\|Sx - S_M x\| &= \lim_{N\to\infty} \|S_N x - S_M x\| \\ &\le \lim_{N\to\infty} \sum_{n=M+1}^N \|T_n\| \|x\| = \sum_{n=M+1}^\infty \|T_n\| \|x\|\end{aligned}$$

and therefore,

$$\|S - S_M\| \le \sum_{n=M}^\infty \|T_n\| \to 0 \text{ as } M \to \infty.$$

For the remainder of this section let X be an infinite set, $\mu : X \to (0,\infty)$ be a given function and $p, q \in [1,\infty]$ such that $q = p/(p-1)$. It will also be convenient to define $\delta_x : X \to \mathbb{R}$ for $x \in X$ by

$$\delta_x(y) = \begin{cases} 1 \text{ if } y = x \\ 0 \text{ if } y \neq x. \end{cases}$$

Notation 14.20 *Let $c_0(X)$ denote those functions $f \in \ell^\infty(X)$ which "vanish at ∞," i.e. for every $\varepsilon > 0$ there exists a finite subset $\Lambda_\varepsilon \subset X$ such that $|f(x)| < \varepsilon$ whenever $x \notin \Lambda_\varepsilon$. Also let $c_f(X)$ denote those functions $f : X \to \mathbb{F}$ with finite support, i.e.*

$$c_f(X) := \{f \in \ell^\infty(X) : \#(\{x \in X : f(x) \neq 0\}) < \infty\}.$$

Exercise 14.6. Show $c_f(X)$ is a subspace of the Banach spaces $\left(\ell^p(\mu), \|\cdot\|_p\right)$ for $1 \leq p < \infty$, while the closure of $c_f(X)$ inside the Banach space, $(\ell^\infty(X), \|\cdot\|_\infty)$ is $c_0(X)$. Note from this it follows that $c_0(X)$ is a closed subspace of $\ell^\infty(X)$. (See Proposition 37.23 below where this last assertion is proved in a more general context.)

Theorem 14.21. *Let X be any set, $\mu : X \to (0, \infty)$ be a function, $p \in [1, \infty]$, $q := p/(p-1)$ be the conjugate exponent and for $f \in \ell^q(\mu)$ define $\phi_f : \ell^p(\mu) \to \mathbb{F}$ by*

$$\phi_f(g) := \sum_{x \in X} f(x) g(x) \mu(x).$$

Then

1. *$\phi_f(g)$ is well defined and $\phi_f \in \ell^p(\mu)^*$.*
2. *The map*
$$f \in \ell^q(\mu) \xrightarrow{\phi} \phi_f \in \ell^p(\mu)^* \tag{14.7}$$
is an isometric linear map of Banach spaces.
3. *If $p \in [1, \infty)$, then the map in Eq. (14.7) is also surjective and hence, $\ell^p(\mu)^*$ is isometrically isomorphic to $\ell^q(\mu)$.*
4. *When $p = \infty$, the map*
$$f \in \ell^1(\mu) \to \phi_f \in c_0(X)^*$$
is an isometric and surjective, i.e. $\ell^1(\mu)$ is isometrically isomorphic to $c_0(X)^*$.

(See Theorem 21.20 below for a continuation of this theorem.)

Proof.

1. By Holder's inequality,
$$\sum_{x \in X} |f(x)| |g(x)| \mu(x) \leq \|f\|_q \|g\|_p$$
which shows that ϕ_f is well defined. The $\phi_f : \ell^p(\mu) \to \mathbb{F}$ is linear by the linearity of sums and since
$$|\phi_f(g)| = \left|\sum_{x \in X} f(x) g(x) \mu(x)\right| \leq \sum_{x \in X} |f(x)| |g(x)| \mu(x) \leq \|f\|_q \|g\|_p,$$
we learn that
$$\|\phi_f\|_{\ell^p(\mu)^*} \leq \|f\|_q. \tag{14.8}$$
Therefore $\phi_f \in \ell^p(\mu)^*$.
2. The map ϕ in Eq. (14.7) is linear in f by the linearity properties of infinite sums. For $p \in (1, \infty)$, define $g(x) = \overline{\text{sgn}(f(x))} |f(x)|^{q-1}$ where
$$\text{sgn}(z) := \begin{cases} \frac{z}{|z|} & \text{if } z \neq 0 \\ 0 & \text{if } z = 0. \end{cases}$$
Then
$$\|g\|_p^p = \sum_{x \in X} |f(x)|^{(q-1)p} \mu(x) = \sum_{x \in X} |f(x)|^{\left(\frac{p}{p-1} - 1\right)p} \mu(x)$$
$$= \sum_{x \in X} |f(x)|^q \mu(x) = \|f\|_q^q$$
and
$$\phi_f(g) = \sum_{x \in X} f(x) \overline{\text{sgn}(f(x))} |f(x)|^{q-1} \mu(x) = \sum_{x \in X} |f(x)| |f(x)|^{q-1} \mu(x)$$
$$= \|f\|_q^{q\left(\frac{1}{q} + \frac{1}{p}\right)} = \|f\|_q \|f\|_q^{\frac{q}{p}} = \|f\|_q \|g\|_p.$$
Hence $\|\phi_f\|_{\ell^p(\mu)^*} \geq \|f\|_q$ which combined with Eq. (14.8) shows $\|\phi_f\|_{\ell^p(\mu)^*} = \|f\|_q$. For $p = \infty$, let $g(x) = \overline{\text{sgn}(f(x))}$, then $\|g\|_\infty = 1$ and
$$|\phi_f(g)| = \sum_{x \in X} f(x) \overline{\text{sgn}(f(x))} \mu(x)$$
$$= \sum_{x \in X} |f(x)| \mu(x) = \|f\|_1 \|g\|_\infty$$
which shows $\|\phi_f\|_{\ell^\infty(\mu)^*} \geq \|f\|_{\ell^1(\mu)}$. Combining this with Eq. (14.8) shows $\|\phi_f\|_{\ell^\infty(\mu)^*} = \|f\|_{\ell^1(\mu)}$. For $p = 1$,
$$|\phi_f(\delta_x)| = \mu(x) |f(x)| = |f(x)| \|\delta_x\|_1$$
and therefore $\|\phi_f\|_{\ell^1(\mu)^*} \geq |f(x)|$ for all $x \in X$. Hence $\|\phi_f\|_{\ell^1(\mu)^*} \geq \|f\|_\infty$ which combined with Eq. (14.8) shows $\|\phi_f\|_{\ell^1(\mu)^*} = \|f\|_\infty$.

. and 4. Suppose that $p\in[1,\infty)$ and $\lambda\in\ell^p(\mu)^*$ or $p=\infty$ and $\lambda\in c_0(X)^*$. We wish to find $f\in\ell^q(\mu)$ such that $\lambda=\phi_f$. If such an f exists, then $\lambda(\delta_x)=f(x)\mu(x)$ and so we must define $f(x):=\lambda(\delta_x)/\mu(x)$. As a preliminary estimate,

$$\begin{aligned}|f(x)|=\frac{|\lambda(\delta_x)|}{\mu(x)}&\le\frac{\|\lambda\|_{\ell^p(\mu)^*}\|\delta_x\|_{\ell^p(\mu)}}{\mu(x)}\\&=\frac{\|\lambda\|_{\ell^p(\mu)^*}[\mu(x)]^{\frac1p}}{\mu(x)}=\|\lambda\|_{\ell^p(\mu)^*}[\mu(x)]^{-\frac1q}.\end{aligned}$$

When $p=1$ and $q=\infty$, this implies $\|f\|_\infty\le\|\lambda\|_{\ell^1(\mu)^*}<\infty$. If $p\in(1,\infty]$ and $\Lambda\subset\subset X$, then

$$\begin{aligned}\|f\|^q_{\ell^q(\Lambda,\mu)}&:=\sum_{x\in\Lambda}|f(x)|^q\mu(x)=\sum_{x\in\Lambda}f(x)\overline{\operatorname{sgn}(f(x))}|f(x)|^{q-1}\mu(x)\\&=\sum_{x\in\Lambda}\frac{\lambda(\delta_x)}{\mu(x)}\overline{\operatorname{sgn}(f(x))}|f(x)|^{q-1}\mu(x)\\&=\sum_{x\in\Lambda}\lambda(\delta_x)\overline{\operatorname{sgn}(f(x))}|f(x)|^{q-1}\\&=\lambda\left(\sum_{x\in\Lambda}\overline{\operatorname{sgn}(f(x))}|f(x)|^{q-1}\delta_x\right)\\&\le\|\lambda\|_{\ell^p(\mu)^*}\left\|\sum_{x\in\Lambda}\overline{\operatorname{sgn}(f(x))}|f(x)|^{q-1}\delta_x\right\|_p.\end{aligned}$$

Since

$$\begin{aligned}\left\|\sum_{x\in\Lambda}\overline{\operatorname{sgn}(f(x))}|f(x)|^{q-1}\delta_x\right\|_p&=\left(\sum_{x\in\Lambda}|f(x)|^{(q-1)p}\mu(x)\right)^{1/p}\\&=\left(\sum_{x\in\Lambda}|f(x)|^q\mu(x)\right)^{1/p}=\|f\|^{q/p}_{\ell^q(\Lambda,\mu)}\end{aligned}$$

which is also valid for $p=\infty$ provided $\|f\|^{1/\infty}_{\ell^1(\Lambda,\mu)}:=1$. Combining the last two displayed equations shows

$$\|f\|^q_{\ell^q(\Lambda,\mu)}\le\|\lambda\|_{\ell^p(\mu)^*}\|f\|^{q/p}_{\ell^q(\Lambda,\mu)}$$

and solving this inequality for $\|f\|^q_{\ell^q(\Lambda,\mu)}$ (using $q-q/p=1$) implies $\|f\|_{\ell^q(\Lambda,\mu)}\le\|\lambda\|_{\ell^p(\mu)^*}$ Taking the supremum of this inequality on $\Lambda\subset\subset X$ shows $\|f\|_{\ell^q(\mu)}\le\|\lambda\|_{\ell^p(\mu)^*}$, i.e. $f\in\ell^q(\mu)$. Since $\lambda=\phi_f$ agree on $c_f(X)$ and $c_f(X)$ is a dense subspace of $\ell^p(\mu)$ for $p<\infty$ and $c_f(X)$ is dense subspace of $c_0(X)$ when $p=\infty$, it follows that $\lambda=\phi_f$.

■

14.2 General Sums in Banach Spaces

Definition 14.22. *Suppose X is a normed space.*

1. *Suppose that $\{x_n\}_{n=1}^\infty$ is a sequence in X, then we say $\sum_{n=1}^\infty x_n$ **converges** in X and $\sum_{n=1}^\infty x_n=s$ if*
$$\lim_{N\to\infty}\sum_{n=1}^N x_n=s\text{ in }X.$$
2. *Suppose that $\{x_\alpha:\alpha\in A\}$ is a given collection of vectors in X. We say the sum $\sum_{\alpha\in A}x_\alpha$ **converges** in X and write $s=\sum_{\alpha\in A}x_\alpha\in X$ if for all $\varepsilon>0$ there exists a finite set $\Gamma_\varepsilon\subset A$ such that $\left\|s-\sum_{\alpha\in\Lambda}x_\alpha\right\|<\varepsilon$ for any $\Lambda\subset\subset A$ such that $\Gamma_\varepsilon\subset\Lambda$.*

Warning: As usual if X is a Banach space and $\sum_{\alpha\in A}\|x_\alpha\|<\infty$ then $\sum_{\alpha\in A}x_\alpha$ exists in X, see Exercise 14.15. However, unlike the case of real valued sums the existence of $\sum_{\alpha\in A}x_\alpha$ does **not** imply $\sum_{\alpha\in\Lambda}\|x_\alpha\|<\infty$. See Proposition 18.22 below, from which one may manufacture counter-examples to this false premise.

Lemma 14.23. *Suppose that $\{x_\alpha\in X:\alpha\in A\}$ is a given collection of vectors in a normed space, X.*

1. *If $s=\sum_{\alpha\in A}x_\alpha\in X$ exists and $T:X\to Y$ is a bounded linear map between normed spaces, then $\sum_{\alpha\in A}Tx_\alpha$ exists in Y and*
$$Ts=T\sum_{\alpha\in A}x_\alpha=\sum_{\alpha\in A}Tx_\alpha.$$
2. *If $s=\sum_{\alpha\in A}x_\alpha$ exists in X then for every $\varepsilon>0$ there exists $\Gamma_\varepsilon\subset\subset A$ such that $\left\|\sum_{\alpha\in\Lambda}x_\alpha\right\|<\varepsilon$ for all $\Lambda\subset\subset A\setminus\Gamma_\varepsilon$.*
3. *If $s=\sum_{\alpha\in A}x_\alpha$ exists in X, the set $\Gamma:=\{\alpha\in A:x_a\neq0\}$ is at most countable. Moreover if Γ is infinite and $\{\alpha_n\}_{n=1}^\infty$ is an enumeration of Γ, then*
$$s=\sum_{n=1}^\infty x_{\alpha_n}:=\lim_{N\to\infty}\sum_{n=1}^N x_{\alpha_n}.\qquad(14.9)$$

4. *If we further assume that X is a Banach space and suppose for all $\varepsilon > 0$ there exists $\Gamma_\varepsilon \subset\subset A$ such that $\left\|\sum_{\alpha\in\Lambda} x_\alpha\right\| < \varepsilon$ whenever $\Lambda \subset\subset A \setminus \Gamma_\varepsilon$, then $\sum_{\alpha\in A} x_\alpha$ exists in X.*

Proof.

1. Let Γ_ε be as in Definition 14.22 and $\Lambda \subset\subset A$ such that $\Gamma_\varepsilon \subset \Lambda$. Then
$$\left\|Ts - \sum_{\alpha\in\Lambda} Tx_\alpha\right\| \le \|T\| \left\|s - \sum_{\alpha\in\Lambda} x_\alpha\right\| < \|T\|\,\varepsilon$$
which shows that $\sum_{\alpha\in\Lambda} Tx_\alpha$ exists and is equal to Ts.
2. Suppose that $s = \sum_{\alpha\in A} x_\alpha$ exists and $\varepsilon > 0$. Let $\Gamma_\varepsilon \subset\subset A$ be as in Definition 14.22. Then for $\Lambda \subset\subset A \setminus \Gamma_\varepsilon$,
$$\begin{aligned}\left\|\sum_{\alpha\in\Lambda} x_\alpha\right\| &= \left\|\sum_{\alpha\in\Gamma_\varepsilon\cup\Lambda} x_\alpha - \sum_{\alpha\in\Gamma_\varepsilon} x_\alpha\right\| \\ &\le \left\|\sum_{\alpha\in\Gamma_\varepsilon\cup\Lambda} x_\alpha - s\right\| + \left\|\sum_{\alpha\in\Gamma_\varepsilon} x_\alpha - s\right\| < 2\varepsilon.\end{aligned}$$
3. If $s = \sum_{\alpha\in A} x_\alpha$ exists in X, for each $n \in \mathbb{N}$ there exists a finite subset $\Gamma_n \subset A$ such that $\left\|\sum_{\alpha\in\Lambda} x_\alpha\right\| < \frac{1}{n}$ for all $\Lambda \subset\subset A \setminus \Gamma_n$. Without loss of generality we may assume $x_\alpha \neq 0$ for all $\alpha \in \Gamma_n$. Let $\Gamma_\infty := \cup_{n=1}^\infty \Gamma_n$ – a countable subset of A. Then for any $\beta \notin \Gamma_\infty$, we have $\{\beta\} \cap \Gamma_n = \emptyset$ and therefore
$$\|x_\beta\| = \left\|\sum_{\alpha\in\{\beta\}} x_\alpha\right\| \le \frac{1}{n} \to 0 \text{ as } n \to \infty.$$
Let $\{\alpha_n\}_{n=1}^\infty$ be an enumeration of Γ and define $\gamma_N := \{\alpha_n : 1 \le n \le N\}$. Since for any $M \in \mathbb{N}$, γ_N will eventually contain Γ_M for N sufficiently large, we have
$$\limsup_{N\to\infty} \left\|s - \sum_{n=1}^N x_{\alpha_n}\right\| \le \frac{1}{M} \to 0 \text{ as } M \to \infty.$$
Therefore Eq. (14.9) holds.
4. For $n \in \mathbb{N}$, let $\Gamma_n \subset\subset A$ such that $\left\|\sum_{\alpha\in\Lambda} x_\alpha\right\| < \frac{1}{n}$ for all $\Lambda \subset\subset A \setminus \Gamma_n$. Define $\gamma_n := \cup_{k=1}^n \Gamma_k \subset A$ and $s_n := \sum_{\alpha\in\gamma_n} x_\alpha$. Then for $m > n$,
$$\|s_m - s_n\| = \left\|\sum_{\alpha\in\gamma_m\setminus\gamma_n} x_\alpha\right\| \le 1/n \to 0 \text{ as } m, n \to \infty.$$
Therefore $\{s_n\}_{n=1}^\infty$ is Cauchy and hence convergent in X, because X is a Banach space. Let $s := \lim_{n\to\infty} s_n$. Then for $\Lambda \subset\subset A$ such that $\gamma_n \subset \Lambda$, we have
$$\left\|s - \sum_{\alpha\in\Lambda} x_\alpha\right\| \le \|s - s_n\| + \left\|\sum_{\alpha\in\Lambda\setminus\gamma_n} x_\alpha\right\| \le \|s - s_n\| + \frac{1}{n}.$$
Since the right side of this equation goes to zero as $n \to \infty$, it follows that $\sum_{\alpha\in A} x_\alpha$ exists and is equal to s.

■

14.3 Inverting Elements in $L(X)$

Definition 14.24. *A linear map $T : X \to Y$ is an* ***isometry*** *if $\|Tx\|_Y = \|x\|_X$ for all $x \in X$. T is said to be* ***invertible*** *if T is a bijection and T^{-1} is bounded.*

Notation 14.25 *We will write $GL(X,Y)$ for those $T \in L(X,Y)$ which are invertible. If $X = Y$ we simply write $L(X)$ and $GL(X)$ for $L(X,X)$ and $GL(X,X)$ respectively.*

Proposition 14.26. *Suppose X is a Banach space and $\Lambda \in L(X) := L(X,X)$ satisfies $\sum_{n=0}^\infty \|\Lambda^n\| < \infty$. Then $I - \Lambda$ is invertible and*
$$(I-\Lambda)^{-1} = \text{“}\frac{1}{I-\Lambda}\text{”} = \sum_{n=0}^\infty \Lambda^n \text{ and } \left\|(I-\Lambda)^{-1}\right\| \le \sum_{n=0}^\infty \|\Lambda^n\|.$$
In particular if $\|\Lambda\| < 1$ then the above formula holds and
$$\left\|(I-\Lambda)^{-1}\right\| \le \frac{1}{1-\|\Lambda\|}.$$

Proof. Since $L(X)$ is a Banach space and $\sum_{n=0}^\infty \|\Lambda^n\| < \infty$, it follows from Theorem 14.18 that
$$S := \lim_{N\to\infty} S_N := \lim_{N\to\infty} \sum_{n=0}^N \Lambda^n$$
exists in $L(X)$. Moreover, by Lemma 14.16,

$$(I-\Lambda)S = (I-\Lambda)\lim_{N\to\infty} S_N = \lim_{N\to\infty}(I-\Lambda)S_N$$
$$= \lim_{N\to\infty}(I-\Lambda)\sum_{n=0}^{N}\Lambda^n = \lim_{N\to\infty}(I-\Lambda^{N+1}) = I$$

and similarly $S(I-\Lambda)=I$. This shows that $(I-\Lambda)^{-1}$ exists and is equal to S. Moreover, $(I-\Lambda)^{-1}$ is bounded because

$$\left\|(I-\Lambda)^{-1}\right\| = \|S\| \le \sum_{n=0}^{\infty}\|\Lambda^n\|.$$

If we further assume $\|\Lambda\|<1$, then $\|\Lambda^n\|\le\|\Lambda\|^n$ and

$$\sum_{n=0}^{\infty}\|\Lambda^n\| \le \sum_{n=0}^{\infty}\|\Lambda\|^n = \frac{1}{1-\|\Lambda\|} < \infty.$$

■

Corollary 14.27. *Let X and Y be Banach spaces. Then $GL(X,Y)$ is an open (possibly empty) subset of $L(X,Y)$. More specifically, if $A\in GL(X,Y)$ and $B\in L(X,Y)$ satisfies*

$$\|B-A\| < \|A^{-1}\|^{-1} \tag{14.10}$$

then $B\in GL(X,Y)$

$$B^{-1} = \sum_{n=0}^{\infty}\left[I_X - A^{-1}B\right]^n A^{-1} \in L(Y,X), \tag{14.11}$$

$$\left\|B^{-1}\right\| \le \|A^{-1}\|\frac{1}{1-\|A^{-1}\|\,\|A-B\|} \tag{14.12}$$

and

$$\left\|B^{-1}-A^{-1}\right\| \le \frac{\|A^{-1}\|^2\,\|A-B\|}{1-\|A^{-1}\|\,\|A-B\|}. \tag{14.13}$$

In particular the map

$$A\in GL(X,Y)\to A^{-1}\in GL(Y,X) \tag{14.14}$$

is continuous.

Proof. Let A and B be as above, then

$$B = A-(A-B) = A\left[I_X - A^{-1}(A-B)\right] = A(I_X-\Lambda)$$

where $\Lambda : X\to X$ is given by

$$\Lambda := A^{-1}(A-B) = I_X - A^{-1}B.$$

Now

$$\|\Lambda\| = \left\|A^{-1}(A-B)\right\| \le \|A^{-1}\|\,\|A-B\| < \|A^{-1}\|\|A^{-1}\|^{-1} = 1.$$

Therefore $I-\Lambda$ is invertible and hence so is B (being the product of invertible elements) with

$$B^{-1} = (I_X-\Lambda)^{-1}A^{-1} = \left[I_X - A^{-1}(A-B)\right]^{-1}A^{-1}.$$

Taking norms of the previous equation gives

$$\left\|B^{-1}\right\| \le \left\|(I_X-\Lambda)^{-1}\right\|\,\|A^{-1}\| \le \|A^{-1}\|\frac{1}{1-\|\Lambda\|}$$
$$\le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\,\|A-B\|}$$

which is the bound in Eq. (14.12). The bound in Eq. (14.13) holds because

$$\left\|B^{-1}-A^{-1}\right\| = \left\|B^{-1}(A-B)A^{-1}\right\| \le \left\|B^{-1}\right\|\left\|A^{-1}\right\|\|A-B\|$$
$$\le \frac{\|A^{-1}\|^2\,\|A-B\|}{1-\|A^{-1}\|\,\|A-B\|}.$$

■

For an application of these results to linear ordinary differential equations, see Section 32.3.

14.4 Exercises

Exercise 14.7. Let $(X,\|\cdot\|)$ be a normed space over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$). Show the map

$$(\lambda,x,y)\in\mathbb{F}\times X\times X\to x+\lambda y\in X$$

is continuous relative to the norm on $\mathbb{F}\times X\times X$ defined by

$$\|(\lambda,x,y)\|_{\mathbb{F}\times X\times X} := |\lambda| + \|x\| + \|y\|.$$

(See Exercise 35.33 for more on the metric associated to this norm.) Also show that $\|\cdot\| : X\to[0,\infty)$ is continuous.

Exercise 14.8. Let $X=\mathbb{N}$ and for $p,q\in[1,\infty)$ let $\|\cdot\|_p$ denote the $\ell^p(\mathbb{N})$ – norm. Show $\|\cdot\|_p$ and $\|\cdot\|_q$ are inequivalent norms for $p\neq q$ by showing

$$\sup_{f\neq 0}\frac{\|f\|_p}{\|f\|_q} = \infty \text{ if } p<q.$$

Exercise 14.9. Suppose that $(X, \|\cdot\|)$ is a normed space and $S \subset X$ is a linear subspace.

1. Show the closure $\bar{S}$ of S is also a linear subspace.
2. Now suppose that X is a Banach space. Show that S with the inherited norm from X is a Banach space iff S is closed.

Exercise 14.10 (See Chapter 15). Folland Problem 5.9. Showing $C^k([0,1])$ is a Banach space.

Exercise 14.11. Suppose that X, Y and Z are Banach spaces and $Q : X \times Y \to Z$ is a bilinear form, i.e. we are assuming $x \in X \to Q(x, y) \in Z$ is linear for each $y \in Y$ and $y \in Y \to Q(x, y) \in Z$ is linear for each $x \in X$. Show Q is continuous relative to the product norm, $\|(x, y)\|_{X \times Y} := \|x\|_X + \|y\|_Y$, on $X \times Y$ iff there is a constant $M < \infty$ such that

$$\|Q(x, y)\|_Z \le M \|x\|_X \cdot \|y\|_Y \text{ for all } (x, y) \in X \times Y. \tag{14.15}$$

Then apply this result to prove Lemma 14.16.

The following exercise is very similar to Exercise 13.16.

Exercise 14.12. Let $d : C(\mathbb{R}) \times C(\mathbb{R}) \to [0, \infty)$ be defined by

$$d(f, g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\|f - g\|_n}{1 + \|f - g\|_n},$$

where $\|f\|_n := \sup\{|f(x)| : |x| \le n\} = \max\{|f(x)| : |x| \le n\}$.

1. Show that d is a metric on $C(\mathbb{R})$.
2. Show that a sequence $\{f_n\}_{n=1}^{\infty} \subset C(\mathbb{R})$ converges to $f \in C(\mathbb{R})$ as $n \to \infty$ iff f_n converges to f uniformly on bounded subsets of $\mathbb{R}$.
3. Show that $(C(\mathbb{R}), d)$ is a complete metric space.

Exercise 14.13. Let $X = C([0,1], \mathbb{R})$ and for $f \in X$, let

$$\|f\|_1 := \int_0^1 |f(t)|\, dt.$$

Show that $(X, \|\cdot\|_1)$ is normed space and show by example that this space is **not** complete. Hint: For the last assertion find a sequence of $\{f_n\}_{n=1}^{\infty} \subset X$ which is "trying" to converge to the function $f = 1_{[\frac{1}{2}, 1]} \notin X$.

Exercise 14.14. Let $(X, \|\cdot\|_1)$ be the normed space in Exercise 14.13. Compute the closure of A when

1. $A = \{f \in X : f(1/2) = 0\}$.
2. $A = \left\{f \in X : \sup_{t \in [0,1]} f(t) \le 5\right\}$. [Hint: you may use without proof that if $f_n \in A$ converges to $f \in X$ in $\|\cdot\|_1$ then there is a subsequence which converges for a.e. t.]
3. $A = \left\{f \in X : \int_0^{1/2} f(t)\, dt = 0\right\}$.

Exercise 14.15. Suppose $\{x_\alpha \in X : \alpha \in A\}$ is a given collection of vectors in a Banach space X. Show $\sum_{\alpha \in A} x_\alpha$ exists in X and

$$\left\| \sum_{\alpha \in A} x_\alpha \right\| \le \sum_{\alpha \in A} \|x_\alpha\|$$

if $\sum_{\alpha \in A} \|x_\alpha\| < \infty$. That is to say **"absolute convergence"** implies convergence in a Banach space.

Exercise 14.16. Suppose X is a Banach space and $\{f_n : n \in \mathbb{N}\}$ is a sequence in X such that $\lim_{n \to \infty} f_n = f \in X$. Show $s_N := \frac{1}{N} \sum_{n=1}^{N} f_n$ for $N \in \mathbb{N}$ is still a convergent sequence and

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f_n = \lim_{N \to \infty} s_N = f.$$

Exercise 14.17 (Dominated Convergence Theorem Again). Let X be a Banach space, A be a set and suppose $f_n : A \to X$ is a sequence of functions such that $f(\alpha) := \lim_{n \to \infty} f_n(\alpha)$ exists for all $\alpha \in A$. Further assume there exists a summable function $g : A \to [0, \infty)$ such that $\|f_n(\alpha)\| \le g(\alpha)$ for all $\alpha \in A$. Show $\sum_{\alpha \in A} f(\alpha)$ exists in X and

$$\lim_{n \to \infty} \sum_{\alpha \in A} f_n(\alpha) = \sum_{\alpha \in A} f(\alpha).$$

15

Hölder Spaces as Banach Spaces

In this section, we will assume that the reader has basic knowledge of the Riemann integral and differentiability properties of functions. The results used here may be found in Part VII below. (BRUCE: there are forward references in this section.)

Notation 15.1 *Let Ω be an open subset of $\mathbb{R}^d$, $BC(\Omega)$ and $BC(\bar{\Omega})$ be the bounded continuous functions on Ω and $\bar{\Omega}$ respectively. By identifying $f \in BC(\bar{\Omega})$ with $f|_\Omega \in BC(\Omega)$, we will consider $BC(\bar{\Omega})$ as a subset of $BC(\Omega)$. For $u \in BC(\Omega)$ and $0 < \beta \le 1$ let*

$$\|u\|_u := \sup_{x\in\Omega} |u(x)| \text{ and } [u]_\beta := \sup_{\substack{x,y\in\Omega \\ x\neq y}} \left\{ \frac{|u(x)-u(y)|}{|x-y|^\beta} \right\}.$$

If $[u]_\beta < \infty$, then u is ***Hölder continuous*** *with Hölder exponent*[1] *β. The collection of β – Hölder continuous function on Ω will be denoted by*

$$C^{0,\beta}(\Omega) := \{u \in BC(\Omega) : [u]_\beta < \infty\}$$

and for $u \in C^{0,\beta}(\Omega)$ let

$$\|u\|_{C^{0,\beta}(\Omega)} := \|u\|_u + [u]_\beta. \tag{15.1}$$

Remark 15.2. If $u : \Omega \to \mathbb{C}$ and $[u]_\beta < \infty$ for some $\beta > 1$, then u is constant on each connected component of Ω. Indeed, if $x \in \Omega$ and $h \in \mathbb{R}^d$ then

$$\left| \frac{u(x+th)-u(x)}{t} \right| \le [u]_\beta \, |th|^\beta / t \to 0 \text{ as } t \to 0$$

which shows $\partial_h u(x) = 0$ for all $x \in \Omega$. If $y \in \Omega$ is in the same connected component as x, then by Exercise 31.8 below there exists a smooth curve $\sigma : [0,1] \to \Omega$ such that $\sigma(0) = x$ and $\sigma(1) = y$. So by the fundamental theorem of calculus and the chain rule,

$$u(y) - u(x) = \int_0^1 \frac{d}{dt} u(\sigma(t))dt = \int_0^1 0 \; dt = 0.$$

This is why we do not talk about Hölder spaces with Hölder exponents larger than 1.

[1] If $\beta = 1$, u is is said to be Lipschitz continuous.

Lemma 15.3. *Suppose $u \in C^1(\Omega) \cap BC(\Omega)$ and $\partial_i u \in BC(\Omega)$ for $i = 1, 2, \ldots, d$, then $u \in C^{0,1}(\Omega)$, i.e. $[u]_1 < \infty$.*

The proof of this lemma is left to the reader as Exercise 15.1.

Theorem 15.4. *Let Ω be an open subset of $\mathbb{R}^d$. Then*

1. *Under the identification of $u \in BC\left(\bar{\Omega}\right)$ with $u|_\Omega \in BC(\Omega)$, $BC(\bar{\Omega})$ is a closed subspace of $BC(\Omega)$.*
2. *Every element $u \in C^{0,\beta}(\Omega)$ has a unique extension to a continuous function (still denoted by u) on $\bar{\Omega}$. Therefore we may identify $C^{0,\beta}(\Omega)$ with $C^{0,\beta}(\bar{\Omega}) \subset BC(\bar{\Omega})$. (In particular we may consider $C^{0,\beta}(\Omega)$ and $C^{0,\beta}(\bar{\Omega})$ to be the same when $\beta > 0$.)*
3. *The function $u \in C^{0,\beta}(\Omega) \to \|u\|_{C^{0,\beta}(\Omega)} \in [0,\infty)$ is a norm on $C^{0,\beta}(\Omega)$ which make $C^{0,\beta}(\Omega)$ into a Banach space.*

Proof. 1. The first item is trivial since for $u \in BC(\bar{\Omega})$, the sup-norm of u on $\bar{\Omega}$ agrees with the sup-norm on Ω and $BC(\bar{\Omega})$ is complete in this norm.

2. Suppose that $[u]_\beta < \infty$ and $x_0 \in \mathrm{bd}(\Omega)$. Let $\{x_n\}_{n=1}^\infty \subset \Omega$ be a sequence such that $x_0 = \lim_{n\to\infty} x_n$. Then

$$|u(x_n) - u(x_m)| \le [u]_\beta \, |x_n - x_m|^\beta \to 0 \text{ as } m, n \to \infty$$

showing $\{u(x_n)\}_{n=1}^\infty$ is Cauchy so that $\bar{u}(x_0) := \lim_{n\to\infty} u(x_n)$ exists. If $\{y_n\}_{n=1}^\infty \subset \Omega$ is another sequence converging to x_0, then

$$|u(x_n) - u(y_n)| \le [u]_\beta \, |x_n - y_n|^\beta \to 0 \text{ as } n \to \infty,$$

showing $\bar{u}(x_0)$ is well defined. In this way we define $\bar{u}(x)$ for all $x \in \mathrm{bd}(\Omega)$ and let $\bar{u}(x) = u(x)$ for $x \in \Omega$. Since a similar limiting argument shows

$$|\bar{u}(x) - \bar{u}(y)| \le [u]_\beta \, |x-y|^\beta \text{ for all } x, y \in \bar{\Omega}$$

it follows that $\bar{u}$ is still continuous and $[\bar{u}]_\beta = [u]_\beta$. In the sequel we will abuse notation and simply denote $\bar{u}$ by u.

3. For $u, v \in C^{0,\beta}(\Omega)$,

$$[v+u]_\beta = \sup_{\substack{x,y\in\Omega \\ x\neq y}} \left\{ \frac{|v(y)+u(y)-v(x)-u(x)|}{|x-y|^\beta} \right\}$$
$$\leq \sup_{\substack{x,y\in\Omega \\ x\neq y}} \left\{ \frac{|v(y)-v(x)|+|u(y)-u(x)|}{|x-y|^\beta} \right\} \leq [v]_\beta + [u]_\beta$$

and for $\lambda \in \mathbb{C}$ it is easily seen that $[\lambda u]_\beta = |\lambda|\,[u]_\beta$. This shows $[\cdot]_\beta$ is a semi-norm (see Definition 4.24) on $C^{0,\beta}(\Omega)$ and therefore $\|\cdot\|_{C^{0,\beta}(\Omega)}$ defined in Eq. (15.1) is a norm. To see that $C^{0,\beta}(\Omega)$ is complete, let $\{u_n\}_{n=1}^\infty$ be a $C^{0,\beta}(\Omega)$–Cauchy sequence. Since $BC(\bar{\Omega})$ is complete, there exists $u \in BC(\bar{\Omega})$ such that $\|u-u_n\|_\infty \to 0$ as $n\to\infty$. For $x, y \in \Omega$ with $x \neq y$,

$$\frac{|u(x)-u(y)|}{|x-y|^\beta} = \lim_{n\to\infty} \frac{|u_n(x)-u_n(y)|}{|x-y|^\beta}$$
$$\leq \limsup_{n\to\infty} [u_n]_\beta \leq \lim_{n\to\infty} \|u_n\|_{C^{0,\beta}(\Omega)} < \infty,$$

and so we see that $u \in C^{0,\beta}(\Omega)$. Similarly,

$$\frac{|u(x)-u_n(x)-(u(y)-u_n(y))|}{|x-y|^\beta} = \lim_{m\to\infty} \frac{|(u_m-u_n)(x)-(u_m-u_n)(y)|}{|x-y|^\beta}$$
$$\leq \limsup_{m\to\infty} [u_m-u_n]_\beta \to 0 \text{ as } n\to\infty,$$

showing $[u-u_n]_\beta \to 0$ as $n\to\infty$ and therefore $\lim_{n\to\infty}\|u-u_n\|_{C^{0,\beta}(\Omega)} = 0$. ■

Notation 15.5 *(BRUCE: $C_0(\Omega)$ is not defined until Chapter 37.) Since Ω and $\bar{\Omega}$ are locally compact Hausdorff spaces, we may define $C_0(\Omega)$ and $C_0(\bar{\Omega})$ as in Definition 37.22. We will also let*

$$C_0^{0,\beta}(\Omega) := C^{0,\beta}(\Omega) \cap C_0(\Omega) \text{ and } C_0^{0,\beta}(\bar{\Omega}) := C^{0,\beta}(\Omega) \cap C_0(\bar{\Omega}).$$

It has already been shown in Proposition 37.23 that $C_0(\Omega)$ and $C_0(\bar{\Omega})$ are closed subspaces of $BC(\Omega)$ and $BC(\bar{\Omega})$ respectively. The next proposition describes the relation between $C_0(\Omega)$ and $C_0(\bar{\Omega})$.

Proposition 15.6. *Each $u \in C_0(\Omega)$ has a unique extension to a continuous function on $\bar{\Omega}$ given by $\bar{u} = u$ on Ω and $\bar{u} = 0$ on $\mathrm{bd}(\Omega)$ and the extension $\bar{u}$ is in $C_0(\bar{\Omega})$. Conversely if $u \in C_0(\bar{\Omega})$ and $u|_{\mathrm{bd}(\Omega)} = 0$, then $u|_\Omega \in C_0(\Omega)$. In this way we may identify $C_0(\Omega)$ with those $u \in C_0(\bar{\Omega})$ such that $u|_{\mathrm{bd}(\Omega)} = 0$.*

Proof. Any extension $u \in C_0(\Omega)$ to an element $\bar{u} \in C(\bar{\Omega})$ is necessarily unique, since Ω is dense inside $\bar{\Omega}$. So define $\bar{u} = u$ on Ω and $\bar{u} = 0$ on $\mathrm{bd}(\Omega)$. We must show $\bar{u}$ is continuous on $\bar{\Omega}$ and $\bar{u} \in C_0(\bar{\Omega})$. For the continuity assertion it is enough to show $\bar{u}$ is continuous at all points in $\mathrm{bd}(\Omega)$. For any $\varepsilon > 0$, by assumption, the set $K_\varepsilon := \{x \in \Omega : |u(x)| \geq \varepsilon\}$ is a compact subset of Ω. Since $\mathrm{bd}(\Omega) = \bar{\Omega} \setminus \Omega$, $\mathrm{bd}(\Omega) \cap K_\varepsilon = \emptyset$ and therefore the distance, $\delta := d(K_\varepsilon, \mathrm{bd}(\Omega))$, between K_ε and $\mathrm{bd}(\Omega)$ is positive. So if $x \in \mathrm{bd}(\Omega)$ and $y \in \bar{\Omega}$ and $|y-x| < \delta$, then $|\bar{u}(x)-\bar{u}(y)| = |u(y)| < \varepsilon$ which shows $\bar{u} : \bar{\Omega} \to \mathbb{C}$ is continuous. This also shows $\{|\bar{u}| \geq \varepsilon\} = \{|u| \geq \varepsilon\} = K_\varepsilon$ is compact in Ω and hence also in $\bar{\Omega}$. Since $\varepsilon > 0$ was arbitrary, this shows $\bar{u} \in C_0(\bar{\Omega})$. Conversely if $u \in C_0(\bar{\Omega})$ such that $u|_{\mathrm{bd}(\Omega)} = 0$ and $\varepsilon > 0$, then $K_\varepsilon := \left\{x \in \bar{\Omega} : |u(x)| \geq \varepsilon\right\}$ is a compact subset of $\bar{\Omega}$ which is contained in Ω since $\mathrm{bd}(\Omega) \cap K_\varepsilon = \emptyset$. Therefore K_ε is a compact subset of Ω showing $u|_\Omega \in C_0(\bar{\Omega})$. ■

Definition 15.7. *Let Ω be an open subset of $\mathbb{R}^d$, $k \in \mathbb{N} \cup \{0\}$ and $\beta \in (0,1]$. Let $BC^k(\Omega)$ ($BC^k(\bar{\Omega})$) denote the set of k – times continuously differentiable functions u on Ω such that $\partial^\alpha u \in BC(\Omega)$ ($\partial^\alpha u \in BC(\bar{\Omega})$)[2] for all $|\alpha| \leq k$. Similarly, let $BC^{k,\beta}(\Omega)$ denote those $u \in BC^k(\Omega)$ such that $[\partial^\alpha u]_\beta < \infty$ for all $|\alpha| = k$. For $u \in BC^k(\Omega)$ let*

$$\|u\|_{C^k(\Omega)} = \sum_{|\alpha|\leq k} \|\partial^\alpha u\|_u \text{ and}$$
$$\|u\|_{C^{k,\beta}(\overline{\Omega})} = \sum_{|\alpha|\leq k} \|\partial^\alpha u\|_u + \sum_{|\alpha|=k} [\partial^\alpha u]_\beta.$$

Theorem 15.8. *The spaces $BC^k(\Omega)$ and $BC^{k,\beta}(\Omega)$ equipped with $\|\cdot\|_{C^k(\Omega)}$ and $\|\cdot\|_{C^{k,\beta}(\overline{\Omega})}$ respectively are Banach spaces and $BC^k(\bar{\Omega})$ is a closed subspace of $BC^k(\Omega)$ and $BC^{k,\beta}(\Omega) \subset BC^k(\bar{\Omega})$. Also*

$$C_0^{k,\beta}(\Omega) = C_0^{k,\beta}(\bar{\Omega}) = \{u \in BC^{k,\beta}(\Omega) : \partial^\alpha u \in C_0(\Omega)\ \forall\ |\alpha| \leq k\}$$

is a closed subspace of $BC^{k,\beta}(\Omega)$.

Proof. Suppose that $\{u_n\}_{n=1}^\infty \subset BC^k(\Omega)$ is a Cauchy sequence, then $\{\partial^\alpha u_n\}_{n=1}^\infty$ is a Cauchy sequence in $BC(\Omega)$ for $|\alpha| \leq k$. Since $BC(\Omega)$ is complete, there exists $g_\alpha \in BC(\Omega)$ such that $\lim_{n\to\infty}\|\partial^\alpha u_n - g_\alpha\|_\infty = 0$ for all $|\alpha| \leq k$. Letting $u := g_0$, we must show $u \in C^k(\Omega)$ and $\partial^\alpha u = g_\alpha$ for all $|\alpha| \leq k$. This will be done by induction on $|\alpha|$. If $|\alpha| = 0$ there is nothing to prove. Suppose that we have verified $u \in C^l(\Omega)$ and $\partial^\alpha u = g_\alpha$ for all $|\alpha| \leq l$ for some $l < k$. Then for $x \in \Omega$, $i \in \{1, 2, \ldots, d\}$ and $t \in \mathbb{R}$ sufficiently small,

$$\partial^a u_n(x+te_i) = \partial^a u_n(x) + \int_0^t \partial_i \partial^a u_n(x+\tau e_i)\,d\tau.$$

[2] To say $\partial^\alpha u \in BC(\bar{\Omega})$ means that $\partial^\alpha u \in BC(\Omega)$ and $\partial^\alpha u$ extends to a continuous function on $\bar{\Omega}$.

Letting $n \to \infty$ in this equation gives

$$\partial^a u(x + te_i) = \partial^a u(x) + \int_0^t g_{\alpha+e_i}(x + \tau e_i) d\tau$$

from which it follows that $\partial_i \partial^\alpha u(x)$ exists for all $x \in \Omega$ and $\partial_i \partial^\alpha u = g_{\alpha+e_i}$. This completes the induction argument and also the proof that $BC^k(\Omega)$ is complete. It is easy to check that $BC^k(\bar{\Omega})$ is a closed subspace of $BC^k(\Omega)$ and by using Exercise 15.1 and Theorem 15.4 that $BC^{k,\beta}(\Omega)$ is a subspace of $BC^k(\bar{\Omega})$. The fact that $C_0^{k,\beta}(\Omega)$ is a closed subspace of $BC^{k,\beta}(\Omega)$ is a consequence of (BRUCE: forward reference.) Proposition 37.23. To prove $BC^{k,\beta}(\Omega)$ is complete, let $\{u_n\}_{n=1}^\infty \subset BC^{k,\beta}(\Omega)$ be a $\|\cdot\|_{C^{k,\beta}(\overline{\Omega})}$ – Cauchy sequence. By the completeness of $BC^k(\Omega)$ just proved, there exists $u \in BC^k(\Omega)$ such that $\lim_{n\to\infty} \|u - u_n\|_{C^k(\Omega)} = 0$. An application of Theorem 15.4 then shows $\lim_{n\to\infty} \|\partial^\alpha u_n - \partial^\alpha u\|_{C^{0,\beta}(\Omega)} = 0$ for $|\alpha| = k$ and therefore $\lim_{n\to\infty} \|u - u_n\|_{C^{k,\beta}(\overline{\Omega})} = 0$. ■

The reader is asked to supply the proof of the following lemma.

Lemma 15.9. *The following inclusions hold. For any* $\beta \in [0, 1]$

$$BC^{k+1,0}(\Omega) \subset BC^{k,1}(\Omega) \subset BC^{k,\beta}(\Omega)$$
$$BC^{k+1,0}(\bar{\Omega}) \subset BC^{k,1}(\bar{\Omega}) \subset BC^{k,\beta}(\Omega).$$

15.1 Exercises

Exercise 15.1. Prove Lemma 15.3.

L^p-Space Basics

Let $(X, \mathcal{M}, \mu)$ be a measure space and for $0 < p < \infty$ and a measurable function $f : X \to \mathbb{C}$ let

$$\|f\|_p := \left(\int_X |f|^p \, d\mu\right)^{1/p}. \tag{16.1}$$

When $p = \infty$, let

$$\begin{aligned} \|f\|_\infty &= \inf\{a \geq 0 : \mu(|f| > a) = 0\} &(16.2)\\ &= \inf\{a \geq 0 : |f(x)| \leq a \text{ for } \mu - \text{a.e. } x\}, &(16.3)\end{aligned}$$

wherein we have used

$$\{a \geq 0 : \mu(|f| > a) = 0\} = \{a \geq 0 : |f(x)| \leq a \text{ for } \mu - \text{a.e. } x\}.$$

For $0 < p \leq \infty$, let

$$L^p(X, \mathcal{M}, \mu) = \{f : X \to \mathbb{C} : f \text{ is measurable and } \|f\|_p < \infty\}/\sim$$

where $f \sim g$ iff $f = g$ a.e. Notice that $\|f - g\|_p = 0$ iff $f \sim g$ and if $f \sim g$ then $\|f\|_p = \|g\|_p$. In general we will (by abuse of notation) use f to denote both the function f and the equivalence class containing f.

16.1 $L^p(\mu)$ is a Banach Space

The next theorem is a generalization Theorem 4.29 to general integrals and the proof is essentially identical to the proof of Theorem 4.29.

Theorem 16.1 (Hölder's inequality). *Suppose that $1 \leq p \leq \infty$ and $q := \frac{p}{p-1}$, or equivalently $p^{-1} + q^{-1} = 1$. If f and g are measurable functions then*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \tag{16.4}$$

Assuming $p \in (1, \infty)$ and $\|f\|_p \cdot \|g\|_q < \infty$, equality holds in Eq. (16.4) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of L^1 which happens iff

$$|g|^q \|f\|_p^p = \|g\|_q^q |f|^p \quad a.e. \tag{16.5}$$

Proof. The cases where $\|f\|_q = 0$ or ∞ or $\|g\|_p = 0$ or ∞ are easy to deal with and are left to the reader. So we will now assume that $0 < \|f\|_q, \|g\|_p < \infty$. Let $s = |f|/\|f\|_p$ and $t = |g|/\|g\|_q$ then Lemma 4.28 implies

$$\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p} + \frac{1}{q} \frac{|g|^q}{\|g\|^q} \tag{16.6}$$

with equality iff $|g/\|g\|_q| = |f|^{p-1}/\|f\|_p^{(p-1)} = |f|^{p/q}/\|f\|_p^{p/q}$, i.e. $|g|^q \|f\|_p^p = \|g\|_q^q |f|^p$. Integrating Eq. (16.6) implies

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (16.5) holds. The proof is finished since it is easily checked that equality holds in Eq. (16.4) when $|f|^p = c|g|^q$ of $|g|^q = c|f|^p$ for some constant c. ∎

Remark 16.2. If we put $s = |f|$ and $t = |g|$ in the above proof we would have arrived at the inequality,

$$\|fg\|_1 \leq \frac{1}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q$$

which is not longer scale invariant. For example if $\lambda > 0$ and replace f by λf in the above inequality we would learn,

$$\|fg\|_1 \leq \frac{\lambda^{p-1}}{p}\|f\|_p^p + \lambda^{-1}\frac{1}{q}\|g\|_q^q. \tag{16.7}$$

The idea of the proof of Theorem 16.1 is to normalize f and g in by there norms to manifestly arrive at a scale invariant inequality.

Alternatively, one may simply minimize the right side of Eq. (16.7) over $\lambda > 0$ to arrive at the same result. Indeed, by the first derivative test we look for λ such that

$$0 = \frac{p-1}{p}\lambda^{p-2}\|f\|_p^p - \lambda^{-2}\frac{1}{q}\|g\|_q^q \iff \lambda^p \|f\|_p^p = \|g\|_q^q \iff \lambda = \frac{\|g\|_q^{q/p}}{\|f\|_p}.$$

Plugging this λ back into Eq. (16.7) shows

$$\|fg\|_1 \le \frac{\left(\frac{\|g\|_q^{q/p}}{\|f\|_p}\right)^{p-1}}{p}\|f\|_p^p + \frac{\|f\|_p}{\|g\|_q^{q/p}}\frac{1}{q}\|g\|_q^q$$
$$= \frac{1}{p}\|f\|_p\|g\|_q + \frac{1}{q}\|f\|_p\|g\|_q = \|f\|_p\|g\|_q$$

by which we recover the estimate in Eq. (16.4).

The following corollary is an easy extension of Hölder's inequality.

Corollary 16.3. *Suppose that $f_i : X \to \mathbb{C}$ are measurable functions for $i = 1, \ldots, n$ and $p_1, \ldots, p_n$ and r are positive numbers such that $\sum_{i=1}^n p_i^{-1} = r^{-1}$, then*

$$\left\|\prod_{i=1}^n f_i\right\|_r \le \prod_{i=1}^n \|f_i\|_{p_i} \text{ where } \sum_{i=1}^n p_i^{-1} = r^{-1}.$$

Proof. To prove this inequality, start with $n = 2$, then for any $p \in [1, \infty]$,

$$\|fg\|_r^r = \int_X |f|^r |g|^r d\mu \le \|f^r\|_p \|g^r\|_{p^*}$$

where $p^* = \frac{p}{p-1}$ is the conjugate exponent. Let $p_1 = pr$ and $p_2 = p^* r$ so that $p_1^{-1} + p_2^{-1} = r^{-1}$ as desired. Then the previous equation states that

$$\|fg\|_r \le \|f\|_{p_1} \|g\|_{p_2}$$

as desired. The general case is now proved by induction. Indeed,

$$\left\|\prod_{i=1}^{n+1} f_i\right\|_r = \left\|\prod_{i=1}^n f_i \cdot f_{n+1}\right\|_r \le \left\|\prod_{i=1}^n f_i\right\|_q \|f_{n+1}\|_{p_{n+1}}$$

where $q^{-1} + p_{n+1}^{-1} = r^{-1}$. Since $\sum_{i=1}^n p_i^{-1} = q^{-1}$, we may now use the induction hypothesis to conclude

$$\left\|\prod_{i=1}^n f_i\right\|_q \le \prod_{i=1}^n \|f_i\|_{p_i},$$

which combined with the previous displayed equation proves the generalized form of Holder's inequality. ■

Theorem 16.4 (Minkowski's Inequality). *If $1 \le p \le \infty$ and $f, g \in L^p$ then*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p. \tag{16.8}$$

Moreover, assuming f and g are not identically zero, equality holds in Eq. (16.8) iff $\operatorname{sgn}(f) \doteq \operatorname{sgn}(g)$ a.e. (see the notation in Definition 4.30) when $p = 1$ and $f = cg$ a.e. for some $c > 0$ for $p \in (1, \infty)$.

Proof. When $p = \infty$, $|f| \le \|f\|_\infty$ a.e. and $|g| \le \|g\|_\infty$ a.e. so that $|f+g| \le |f| + |g| \le \|f\|_\infty + \|g\|_\infty$ a.e. and therefore

$$\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty.$$

When $p < \infty$,

$$|f+g|^p \le (2\max(|f|, |g|))^p = 2^p \max(|f|^p, |g|^p) \le 2^p (|f|^p + |g|^p),$$
$$\|f+g\|_p^p \le 2^p \left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

In case $p = 1$,

$$\|f+g\|_1 = \int_X |f+g| d\mu \le \int_X |f| \, d\mu + \int_X |g| d\mu$$

with equality iff $|f| + |g| = |f+g|$ a.e. which happens iff $\operatorname{sgn}(f) \doteq \operatorname{sgn}(g)$ a.e. In case $p \in (1, \infty)$, we may assume $\|f+g\|_p$, $\|f\|_p$ and $\|g\|_p$ are all positive since otherwise the theorem is easily verified. Now

$$|f+g|^p = |f+g||f+g|^{p-1} \le (|f| + |g|)|f+g|^{p-1}$$

with equality iff $\operatorname{sgn}(f) \doteq \operatorname{sgn}(g)$. Integrating this equation and applying Holder's inequality with $q = p/(p-1)$ gives

$$\int_X |f+g|^p d\mu \le \int_X |f| \; |f+g|^{p-1} d\mu + \int_X |g| \; |f+g|^{p-1} d\mu$$
$$\le (\|f\|_p + \|g\|_p) \, \| \, |f+g|^{p-1} \, \|_q \tag{16.9}$$

with equality iff

$$\operatorname{sgn}(f) \doteq \operatorname{sgn}(g) \text{ and}$$
$$\left(\frac{|f|}{\|f\|_p}\right)^p = \frac{|f+g|^p}{\|f+g\|_p^p} = \left(\frac{|g|}{\|g\|_p}\right)^p \text{ a.e.} \tag{16.10}$$

Therefore

$$\||f+g|^{p-1}\|_q^q = \int_X (|f+g|^{p-1})^q d\mu = \int_X |f+g|^p d\mu. \tag{16.11}$$

Combining Eqs. (16.9) and (16.11) implies

$$\|f+g\|_p^p \le \|f\|_p \|f+g\|_p^{p/q} + \|g\|_p \|f+g\|_p^{p/q} \tag{16.12}$$

with equality iff Eq. (16.10) holds which happens iff $f = cg$ a.e. with $c > 0$. Solving for $\|f+g\|_p$ in Eq. (16.12) gives Eq. (16.8). ■

Corollary 16.5. *Suppose that $\{f_n\}_{n=1}^\infty$ is any sequence of non-negative measurable functions on $(X, \mathcal{M}, \mu)$ and $1 \le p \le \infty$. Then*

$$\left\| \sum_{n=1}^\infty f_n \right\|_p \le \sum_{n=1}^\infty \|f_n\|_p .$$

Proof. First notice that Eq. (16.8) remains valid if f and g are any non-negative measurable functions independent of whether $f, g \in L^p$. Therefore it follows (by induction) that

$$\left\| \sum_{n=1}^N f_n \right\|_p \le \sum_{n=1}^N \|f_n\|_p \le \sum_{n=1}^\infty \|f_n\|_p \text{ for all } N \in \mathbb{N}.$$

For $p < \infty$, an application of the monotone convergence theorem now completes the proof;

$$\left\| \sum_{n=1}^\infty f_n \right\|_p = \left\| \lim_{N\to\infty} \sum_{n=1}^N f_n \right\|_p = \lim_{N\to\infty} \left\| \sum_{n=1}^N f_n \right\|_p \le \lim_{N\to\infty} \sum_{n=1}^N \|f_n\|_p = \sum_{n=1}^\infty \|f_n\|_p .$$

When $p = \infty$ we have $f_n \le \|f_n\|_\infty$ a.e. implies $\sum_{n=1}^\infty f_n \le \sum_{n=1}^\infty \|f_n\|_\infty$ a.e. and this implies $\|\sum_{n=1}^\infty f_n\|_\infty \le \sum_{n=1}^\infty \|f_n\|_\infty$. ■

Corollary 16.6. *Suppose that $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ and $f_n \to f$ in $L^p(\mu)$. Then there exists a subsequence $\{f_{n_k}\}$ which is a.e. convergent and $\lim_{k\to\infty} f_{n_k}(x) = f(x)$ for μ – a.e. x.*

Proof. By passing to a subsequence if necessary we may assume $\sum_{n=2}^\infty \|f_n - f_{n-1}\|_p < \infty$. From Corollary 16.5 we know $\|\sum_{n=2}^\infty |f_n - f_{n-1}|\|_p < \infty$ which implies

$$\sum_{n=2}^\infty |f_n - f_{n-1}| < \infty \text{ a.e.}$$

By completeness of the complex numbers it then follows that $F := \lim_{n\to\infty} f_n$ exists a.e. If $p < \infty$ we have by Fatou's lemma that

$$\|F - f\|_p = \left\| \liminf_{n\to\infty} |f_n - f| \right\|_p \le \liminf_{n\to\infty} \||f_n - f|\|_p = 0$$

which shows that $F = f$ a.e. When $p = \infty$ life is even easier since $|f_n - f| \le \|f_n - f\|_\infty$ a.e. which shows for μ – a.e. x that $\lim_{n\to\infty} f_n(x) = f(x)$. There is no need to pass to a subsequence in the $p = \infty$ case. ■

The next theorem gives another example of using Hölder's inequality

Theorem 16.7. *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces, $p \in [1, \infty]$, $q = p/(p-1)$ and $k : X \times Y \to \mathbb{C}$ be a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Assume there exist finite constants C_1 and C_2 such that*

$$\int_X |k(x,y)|\, d\mu(x) \le C_1 \text{ for } \nu\text{-a.e. } y \text{ and}$$
$$\int_Y |k(x,y)|\, d\nu(y) \le C_2 \text{ for } \mu\text{-a.e. } x.$$

If $f \in L^p(\nu)$, then

$$\int_Y |k(x,y) f(y)|\, d\nu(y) < \infty \text{ for } \mu\text{-a.e. } x,$$

$x \to Kf(x) := \int_Y k(x,y) f(y) d\nu(y) \in L^p(\mu)$ and

$$\|Kf\|_{L^p(\mu)} \le C_1^{1/p} C_2^{1/q} \|f\|_{L^p(\nu)} \tag{16.13}$$

Proof. Suppose $p \in (1, \infty)$ to begin with and let $q = p/(p-1)$, then by Hölder's inequality,

$$\begin{aligned}
\int_Y |k(x,y) f(y)|\, d\nu(y) &= \int_Y |k(x,y)|^{1/q} |k(x,y)|^{1/p} |f(y)|\, d\nu(y) \\
&\le \left[\int_Y |k(x,y)|\, d\nu(y) \right]^{1/q} \left[\int_Y |k(x,y)| \, |f(y)|^p \, d\nu(y) \right]^{1/p} \\
&\le C_2^{1/q} \left[\int_Y |k(x,y)| \, |f(y)|^p \, d\nu(y) \right]^{1/p} .
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left\| \int_Y |k(\cdot,y) f(y)|\, d\nu(y) \right\|_{L^p(\mu)}^p &= \int_X d\mu(x) \left[\int_Y |k(x,y) f(y)|\, d\nu(y) \right]^p \\
&\le C_2^{p/q} \int_X d\mu(x) \int_Y d\nu(y)\, |k(x,y)| \, |f(y)|^p \\
&= C_2^{p/q} \int_Y d\nu(y)\, |f(y)|^p \int_X d\mu(x)\, |k(x,y)| \\
&\le C_2^{p/q} C_1 \int_Y d\nu(y)\, |f(y)|^p = C_2^{p/q} C_1 \|f\|_{L^p(\nu)}^p ,
\end{aligned}$$

wherein we used Tonelli's theorem in third line. From this it follows that $\int_Y |k(x,y) f(y)|\, d\nu(y) < \infty$ for μ-a.e. x,

$$x \to Kf(x) := \int_Y k(x,y) f(y) d\nu(y) \in L^p(\mu)$$

and that Eq. (16.13) holds.

Similarly if $p = \infty$,

$$\int_Y |k(x,y)f(y)|\, d\nu(y) \le \|f\|_{L^\infty(\nu)} \cdot \int_Y |k(x,y)|\, d\nu(y) \le C_2 \|f\|_{L^\infty(\nu)} \text{ for } \mu - \text{a.e. } x.$$

so that $\|Kf\|_{L^\infty(\mu)} \le C_2 \|f\|_{L^\infty(\nu)}$. If $p = 1$, then

$$\begin{aligned}\int_X d\mu(x) \int_Y d\nu(y)\, |k(x,y)f(y)| &= \int_Y d\nu(y)\, |f(y)| \int_X d\mu(x)\, |k(x,y)| \\ &\le C_1 \int_Y d\nu(y)\, |f(y)|\end{aligned}$$

which shows $\|Kf\|_{L^1(\mu)} \le C_1 \|f\|_{L^1(\nu)}$. ■

Theorem 16.8 (Completeness of $L^p(\mu)$). *For $1 \le p \le \infty$, $L^p(\mu)$ equipped with the L^p – norm, $\|\cdot\|_p$ (see Eq. (16.1)), is a Banach space.*

Proof. By Minkowski's Theorem 16.4, $\|\cdot\|_p$ satisfies the triangle inequality. As above the reader may easily check the remaining conditions that ensure $\|\cdot\|_p$ is a norm. So we are left to prove the completeness of $L^p(\mu)$ for $1 \le p \le \infty$.

Suppose that $f_n \in L^p(\mu)$ such that $\sum_{n=1}^\infty \|f_n\|_p < \infty$. Then by Corollary 16.5 it follows that

$$\left\|\sum_{n=1}^\infty |f_n|\right\|_p \le \sum_{n=1}^\infty \||f_n|\|_p = \sum_{n=1}^\infty \|f_n\|_p < \infty$$

from which it follows that $\sum_{n=1}^\infty |f_n| < \infty$ a.e. So by completeness of the complex numbers we know $S := \sum_{n=1}^\infty f_n 1_{\sum_{n=1}^\infty |f_n| < \infty}$ exists. Moreover,

$$\left|S - \sum_{n=1}^N f_n\right| = \left|\sum_{n=N+1}^\infty f_n\right| \le \sum_{n=N+1}^\infty |f_n| \text{ a.e.}$$

and therefore by Corollary 16.5,

$$\left\|S - \sum_{n=1}^N f_n\right\|_p \le \sum_{n=N+1}^\infty \|f_n\|_p \to 0 \text{ as } N \to \infty.$$

This show that $S \in L^p(\mu)$ and that $\sum_{n=1}^\infty f_n$ is convergent in $L^p(\mu)$ to S. By Theorem 14.18, this suffices to show $L^p(\mu)$ is complete.

Alternate proof using convergence results in Section 16.3 below. We assume $1 \le p < \infty$ as the case $p = \infty$ is discussed in more detail in Theorem 16.9 below. Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality (Lemma 16.18), $\{f_n\}$ is L^0-Cauchy (i.e. Cauchy in measure) and by Theorem 16.20 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j \to f$ a.e. By Fatou's Lemma,

$$\begin{aligned}\|g_j - f\|_p^p &= \int \lim_{k\to\infty} \inf |g_j - g_k|^p d\mu \le \lim_{k\to\infty} \inf \int |g_j - g_k|^p d\mu \\ &= \lim_{k\to\infty} \inf \|g_j - g_k\|_p^p \to 0 \text{ as } j \to \infty.\end{aligned}$$

In particular, $\|f\|_p \le \|g_j - f\|_p + \|g_j\|_p < \infty$ so the $f \in L^p$ and $g_j \xrightarrow{L^p} f$. The proof is finished because,

$$\|f_n - f\|_p \le \|f_n - g_j\|_p + \|g_j - f\|_p \to 0 \text{ as } j, n \to \infty.$$

■

Theorem 16.9. *Let $\|\cdot\|_\infty$ be as defined in Eq. (16.2), then $(L^\infty(X, \mathcal{M}, \mu), \|\cdot\|_\infty)$ is a Banach space. A sequence $\{f_n\}_{n=1}^\infty \subset L^\infty$ converges to $f \in L^\infty$ iff there exists $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $f_n \to f$ uniformly on E^c. Moreover, bounded simple functions are dense in L^∞.*

Proof. By Minkowski's Theorem 16.4, $\|\cdot\|_\infty$ satisfies the triangle inequality. The reader may easily check the remaining conditions that ensure $\|\cdot\|_\infty$ is a norm. Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a sequence such $f_n \to f \in L^\infty$, i.e. $\|f - f_n\|_\infty \to 0$ as $n \to \infty$. Then for all $k \in \mathbb{N}$, there exists $N_k < \infty$ such that

$$\mu\left(|f - f_n| > k^{-1}\right) = 0 \text{ for all } n \ge N_k.$$

Let

$$E = \cup_{k=1}^\infty \cup_{n \ge N_k} \left\{|f - f_n| > k^{-1}\right\}.$$

Then $\mu(E) = 0$ and for $x \in E^c$, $|f(x) - f_n(x)| \le k^{-1}$ for all $n \ge N_k$. This shows that $f_n \to f$ uniformly on E^c. Conversely, if there exists $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $f_n \to f$ uniformly on E^c, then for any $\varepsilon > 0$,

$$\mu(|f - f_n| \ge \varepsilon) = \mu(\{|f - f_n| \ge \varepsilon\} \cap E^c) = 0$$

for all n sufficiently large. That is to say $\limsup_{n\to\infty} \|f - f_n\|_\infty \le \varepsilon$ for all $\varepsilon > 0$. The density of simple functions follows from the approximation Theorem **??**.

(**Completeness of** L^∞.) Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a Cauchy sequence so that $\varepsilon_{m,n} := \|f_m - f_n\|_\infty \to 0$ as $m, n \to \infty$. Let $E_{m,n} = \{|f_n - f_m| > \varepsilon_{m,n}\}$ and $E := \cup E_{m,n}$, then $\mu(E) = 0$ and

$$\sup_{x \in E^c} |f_m(x) - f_n(x)| \le \varepsilon_{m,n} \to 0 \text{ as } m, n \to \infty.$$

Therefore, $f := \lim_{n\to\infty} f_n$ exists on E^c and the limit is uniform on E^c. Letting $f = \lim_{n\to\infty} 1_{E^c} f_n$, it then follows that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$.

(**Alternative proof of completeness.**) Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a sequence such that $\sum_{n=1}^\infty \|f_n\|_\infty < \infty$. Let $M_n := \|f_n\|_\infty$, $E_n := \{|f_n| > M_n\}$, and $E := \cup_{n=1}^\infty E_n$ so that $\mu(E) = 0$. Then

$$\sum_{n=1}^\infty \sup_{x\in E^c} |f_n(x)| \le \sum_{n=1}^\infty M_n < \infty$$

which shows that $S_N(x) = \sum_{n=1}^N f_n(x)$ converges uniformly to $S(x) := \sum_{n=1}^\infty f_n(x)$ on E^c, i.e. $\lim_{n\to\infty} \|S - S_n\|_\infty = 0$. The completeness of $L^\infty(\mu)$ now follows from Theorem 14.18. ■

16.2 Density Results

Theorem 16.10 (Density Theorem). *Let $p \in [1,\infty)$, $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that*

1. *$\mathbb{M} \subset L^p(\mu, \mathbb{R})$ and $\sigma(\mathbb{M}) = \mathcal{B}$.*
2. *There exists $\psi_k \in \mathbb{M}$ such that $\psi_k \to 1$ boundedly.*

Then to every function $f \in L^p(\mu, \mathbb{R})$, there exist $\varphi_n \in \mathbb{M}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(\mu)} = 0$, i.e. $\mathbb{M}$ is dense in $L^p(\mu, \mathbb{R})$.

Proof. Fix $k \in \mathbb{N}$ for the moment and let $\mathbb{H}$ denote those bounded $\mathcal{B}$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{M}$ such that $\lim_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{M} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated convergence theorem, $\lim_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} = 0$.[1] (Take the dominating function to be $g = [2C|\psi_k|]^p$ where C is a constant bounding all of the $\{|f_n|\}_{n=1}^\infty$.) We may now choose $\varphi_n \in \mathbb{M}$ such that $\|\varphi_n - \psi_k f_n\|_{L^p(\mu)} \le \frac{1}{n}$ then

$$\begin{aligned}\limsup_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} \le & \limsup_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} \\ & + \limsup_{n\to\infty} \|\psi_k f_n - \varphi_n\|_{L^p(\mu)} = 0 \qquad (16.14)\end{aligned}$$

which implies $f \in \mathbb{H}$.

[1] It is at this point that the proof would break down if $p = \infty$.

An application of Dynkin's Multiplicative System Theorem 11.2 or Theorem 11.7 now shows $\mathbb{H}$ contains all bounded measurable functions on Ω. Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f|\le k\}} f - f\right\|_{L^p(\mu)} = 0$. (Take the dominating function to be $g = [2C|f|]^p$ where C is a bound on all of the $|\psi_k|$.) Using this and what we have just proved, there exists $\varphi_k \in \mathbb{M}$ such that

$$\left\|\psi_k 1_{\{|f|\le k\}} f - \varphi_k\right\|_{L^p(\mu)} \le \frac{1}{k}.$$

The same line of reasoning used in Eq. (16.14) now implies $\lim_{k\to\infty} \|f - \varphi_k\|_{L^p(\mu)} = 0$. ■

Example 16.11. Let μ be a measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu([-M, M]) < \infty$ for all $M < \infty$. Then, $C_c(\mathbb{R}, \mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ with compact support) is dense in $L^p(\mu)$ for all $1 \le p < \infty$. To see this, apply Theorem 16.10 with $\mathbb{M} = C_c(\mathbb{R}, \mathbb{R})$ and $\psi_k := 1_{[-k,k]}$.

Theorem 16.12. *Suppose $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{B} \subset 2^\Omega$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and μ is σ – finite on $\mathcal{A}$. Let $\mathbb{S}(\mathcal{A}, \mu)$ denote the measurable simple functions, $\varphi : \Omega \to \mathbb{R}$ such $\{\varphi = y\} \in \mathcal{A}$ for all $y \in \mathbb{R}$ and $\mu(\{\varphi \ne 0\}) < \infty$. Then $\mathbb{S}(\mathcal{A}, \mu)$ is dense subspace of $L^p(\mu)$.*

Proof. Let $\mathbb{M} := \mathbb{S}(\mathcal{A}, \mu)$. By assumption there exists $\Omega_k \in \mathcal{A}$ such that $\mu(\Omega_k) < \infty$ and $\Omega_k \uparrow \Omega$ as $k \to \infty$. If $A \in \mathcal{A}$, then $\Omega_k \cap A \in \mathcal{A}$ and $\mu(\Omega_k \cap A) < \infty$ so that $1_{\Omega_k \cap A} \in \mathbb{M}$. Therefore $1_A = \lim_{k\to\infty} 1_{\Omega_k \cap A}$ is $\sigma(\mathbb{M})$ – measurable for every $A \in \mathcal{A}$. So we have shown that $\mathcal{A} \subset \sigma(\mathbb{M}) \subset \mathcal{B}$ and therefore $\mathcal{B} = \sigma(\mathcal{A}) \subset \sigma(\mathbb{M}) \subset \mathcal{B}$, i.e. $\sigma(\mathbb{M}) = \mathcal{B}$. The theorem now follows from Theorem 16.10 after observing $\psi_k := 1_{\Omega_k} \in \mathbb{M}$ and $\psi_k \to 1$ boundedly. ■

Theorem 16.13 (Separability of L^p – Spaces). *Suppose, $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{B}$ is a countable algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and μ is σ – finite on $\mathcal{A}$. Then $L^p(\mu)$ is separable and*

$$\mathbb{D} = \{\sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q},\ A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty\}$$

is a countable dense subset.

Proof. It is left to reader to check $\mathbb{D}$ is dense in $\mathbb{S}(\mathcal{A}, \mu)$ relative to the $L^p(\mu)$ – norm. Once this is done, the proof is then complete since $\mathbb{S}(\mathcal{A}, \mu)$ is a dense subspace of $L^p(\mu)$ by Theorem 16.12. ■

Example 16.14. Let μ be a measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu([-M, M]) < \infty$ for all $M < \infty$. Then $L^p(\mu)$ is separable for all $1 \le p < \infty$. This follows from an application of Theorem 16.13 with $\mathcal{A}$ being the countable algebra generated by

$$\{(a, b] : -\infty < a < b < \infty \text{ with } a, b \in \mathbb{Q}\}.$$

Exercise 16.1. Let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$, then $\mathbb{D} := \text{span}\{e^{i\lambda \cdot x} : \lambda \in \mathbb{R}^d\}$ is a dense subspace of $L^p(\mu)$ for all $1 \le p < \infty$. **Hints:** By the d - dimensional generalization of Example 16.11 (see Theorem 31.8), $C_c(\mathbb{R}^d)$ is a dense subspace of $L^p(\mu)$. For $f \in C_c(\mathbb{R}^d)$ and $N \in \mathbb{N}$, let

$$f_N(x) := \sum_{n \in \mathbb{Z}^d} f(x + 2\pi N n).$$

Show $f_N \in BC(\mathbb{R}^d)$ and $x \to f_N(Nx)$ is 2π – periodic, so by Exercise 7.15 (see also Theorem 7.42 or Exercise 37.13), $x \to f_N(Nx)$ can be approximated uniformly by trigonometric polynomials. Use this fact to conclude that $f_N \in \bar{\mathbb{D}}^{L^p(\mu)}$. After this show $f_N \to f$ in $L^p(\mu)$.

Exercise 16.2. Suppose that μ and ν are two finite measures on $\mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} e^{i\lambda \cdot x} d\mu(x) = \int_{\mathbb{R}^d} e^{i\lambda \cdot x} d\nu(x) \tag{16.15}$$

for all $\lambda \in \mathbb{R}^d$. Show $\mu = \nu$.

Hint: Perhaps the easiest way to do this is to use Exercise 16.1 with the measure μ being replaced by $\mu + \nu$. Alternatively, use the method of proof of Exercise 16.1 to show Eq. (16.15) implies $\int_{\mathbb{R}^d} f d\mu(x) = \int_{\mathbb{R}^d} f d\nu(x)$ for all $f \in C_c(\mathbb{R}^d)$ and then apply Corollary 11.35.

Exercise 16.3. Again let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$. Further assume that $C_M := \int_{\mathbb{R}^d} e^{M|x|} d\mu(x) < \infty$ for all $M \in (0, \infty)$. Let $\mathcal{P}(\mathbb{R}^d)$ be the space of polynomials, $\rho(x) = \sum_{|\alpha| \le N} \rho_\alpha x^\alpha$ with $\rho_\alpha \in \mathbb{C}$, on $\mathbb{R}^d$. (Notice that $|\rho(x)|^p \le Ce^{M|x|}$ for some constant $C = C(\rho, p, M)$, so that $\mathcal{P}(\mathbb{R}^d) \subset L^p(\mu)$ for all $1 \le p < \infty$.) Show $\mathcal{P}(\mathbb{R}^d)$ is dense in $L^p(\mu)$ for all $1 \le p < \infty$. Here is a possible outline.

Outline: Fix a $\lambda \in \mathbb{R}^d$ and let $f_n(x) = (\lambda \cdot x)^n / n!$ for all $n \in \mathbb{N}$.

1. Use calculus to verify $\sup_{t \ge 0} t^\alpha e^{-Mt} = \left(\frac{\alpha}{Me}\right)^\alpha$ for all $\alpha \ge 0$ where $(0/M)^0 := 1$. Use this estimate along with the identity
$$|\lambda \cdot x|^{pn} \le |\lambda|^{pn} |x|^{pn} = \left(|x|^{pn} e^{-M|x|}\right) |\lambda|^{pn} e^{M|x|}$$
to find an estimate on $\|f_n\|_p$.
2. Use your estimate on $\|f_n\|_p$ to show $\sum_{n=0}^\infty \|f_n\|_p < \infty$ and conclude
$$\lim_{N \to \infty} \left\| e^{i\lambda \cdot (\cdot)} - \sum_{n=0}^N i^n f_n \right\|_p = 0.$$
3. Now finish by appealing to Exercise 16.1.

Exercise 16.4. Again let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$ but now assume there exists an $\varepsilon > 0$ such that $C := \int_{\mathbb{R}^d} e^{\varepsilon |x|} d\mu(x) < \infty$. Also let $q > 1$ and $h \in L^q(\mu)$ be a function such that $\int_{\mathbb{R}^d} h(x) x^\alpha d\mu(x) = 0$ for all $\alpha \in \mathbb{N}_0^d$. (As mentioned in Exercise 16.4, $\mathcal{P}(\mathbb{R}^d) \subset L^p(\mu)$ for all $1 \le p < \infty$, so $x \to h(x) x^\alpha$ is in $L^1(\mu)$.) Show $h(x) = 0$ for μ– a.e. x using the following outline.

Outline: Fix a $\lambda \in \mathbb{R}^d$, let $f_n(x) = (\lambda \cdot x)^n / n!$ for all $n \in \mathbb{N}$, and let $p = q/(q-1)$ be the conjugate exponent to q.

1. Use calculus to verify $\sup_{t \ge 0} t^\alpha e^{-\varepsilon t} = (\alpha/\varepsilon)^\alpha e^{-\alpha}$ for all $\alpha \ge 0$ where $(0/\varepsilon)^0 := 1$. Use this estimate along with the identity
$$|\lambda \cdot x|^{pn} \le |\lambda|^{pn} |x|^{pn} = \left(|x|^{pn} e^{-\varepsilon |x|}\right) |\lambda|^{pn} e^{\varepsilon |x|}$$
to find an estimate on $\|f_n\|_p$.
2. Use your estimate on $\|f_n\|_p$ to show there exists $\delta > 0$ such that $\sum_{n=0}^\infty \|f_n\|_p < \infty$ when $|\lambda| \le \delta$ and conclude for $|\lambda| \le \delta$ that $e^{i\lambda \cdot x} = L^p(\mu)$-$\sum_{n=0}^\infty i^n f_n(x)$. Conclude from this that
$$\int_{\mathbb{R}^d} h(x) e^{i\lambda \cdot x} d\mu(x) = 0 \text{ when } |\lambda| \le \delta.$$
3. Let $\lambda \in \mathbb{R}^d$ ($|\lambda|$ not necessarily small) and set $g(t) := \int_{\mathbb{R}^d} e^{it\lambda \cdot x} h(x) d\mu(x)$ for $t \in \mathbb{R}$. Show $g \in C^\infty(\mathbb{R})$ and
$$g^{(n)}(t) = \int_{\mathbb{R}^d} (i\lambda \cdot x)^n e^{it\lambda \cdot x} h(x) d\mu(x) \text{ for all } n \in \mathbb{N}.$$
4. Let $T = \sup\{\tau \ge 0 : g|_{[0,\tau]} \equiv 0\}$. By Step 2., $T \ge \delta$. If $T < \infty$, then
$$0 = g^{(n)}(T) = \int_{\mathbb{R}^d} (i\lambda \cdot x)^n e^{iT\lambda \cdot x} h(x) d\mu(x) \text{ for all } n \in \mathbb{N}.$$
Use Step 3. with h replaced by $e^{iT\lambda \cdot x} h(x)$ to conclude
$$g(T + t) = \int_{\mathbb{R}^d} e^{i(T+t)\lambda \cdot x} h(x) d\mu(x) = 0 \text{ for all } t \le \delta / |\lambda|.$$
This violates the definition of T and therefore $T = \infty$ and in particular we may take $T = 1$ to learn
$$\int_{\mathbb{R}^d} h(x) e^{i\lambda \cdot x} d\mu(x) = 0 \text{ for all } \lambda \in \mathbb{R}^d.$$
5. Use Exercise 16.1 to conclude that
$$\int_{\mathbb{R}^d} h(x) g(x) d\mu(x) = 0$$
for all $g \in L^p(\mu)$. Now choose g judiciously to finish the proof.

16.3 Convergence in Measure

As usual let $(X, \mathcal{M}, \mu)$ be a fixed measure space, assume $1 \le p \le \infty$ and let $\{f_n\}_{n=1}^\infty \cup \{f\}$ be a collection of complex valued measurable functions on X. We have the following notions of convergence and Cauchy sequences.

Definition 16.15. *1. $f_n \to f$ a.e. if there is a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ and $\lim_{n\to\infty} 1_{E^c} f_n = 1_{E^c} f$.*

2. $f_n \to f$ in μ – measure if $\lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0$ for all $\varepsilon > 0$. We will abbreviate this by saying $f_n \to f$ in L^0 or by $f_n \overset{\mu}{\to} f$.

3. $f_n \to f$ in L^p iff $f \in L^p$ and $f_n \in L^p$ for all n, and $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Definition 16.16. *1. $\{f_n\}$ is a.e. Cauchy if there is a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ and$\{1_{E^c} f_n\}$ is a pointwise Cauchy sequences.*

2. $\{f_n\}$ is Cauchy in μ – measure (or L^0 – Cauchy) if $\lim_{m,n\to\infty} \mu(|f_n - f_m| > \varepsilon) = 0$ for all $\varepsilon > 0$.

3. $\{f_n\}$ is Cauchy in L^p if $\lim_{m,n\to\infty} \|f_n - f_m\|_p = 0$.

Lemma 16.17 (Limits in measure are unique). *The following hold;*

1. If $f_n \overset{\mu}{\to} f$ and $f_n \overset{\mu}{\to} g$, then $f = g$ a.e.
2. If $f_n \overset{\mu}{\to} f$ and $g_n \overset{\mu}{\to} g$ then $f_n + g_n \overset{\mu}{\to} f + g$.
3. If $f_n \overset{\mu}{\to} f$ then $\{f_n\}_{n=1}^\infty$ is Cauchy in measure.

Proof. 1. and 2. One of the basic tricks here is to observe that if $\varepsilon > 0$ and $a, b \ge 0$ such that $a + b \ge \varepsilon$, then either $a \ge \varepsilon/2$ or $b \ge \varepsilon/2$. For for example if $\varepsilon < |f \pm g| \le |f| + |g|$, then $|f| > \varepsilon/2$ or $|g| > \varepsilon/2$ and it follows that

$$\{|f \pm g| > \varepsilon\} \subset \{|f| > \varepsilon/2\} \cup \{|g| > \varepsilon/2\}.$$

Therefore,

$$\begin{aligned} \mu(|f - g| > \varepsilon) &= \mu(|f - f_n + f_n - g| > \varepsilon) \\ &\le \mu\left(|f - f_n| > \frac{\varepsilon}{2}\right) + \mu\left(|f_n - g| > \frac{\varepsilon}{2}\right) \to 0 \text{ as } n \to \infty \end{aligned}$$

and we have shown $\mu(|f - g| > \varepsilon) = 0$ for all $\varepsilon > 0$. Hence

$$\mu(|f - g| > 0) = \mu\left(\cup_{n=1}^\infty \left\{|f - g| > \frac{1}{n}\right\}\right) \le \sum_{n=1}^\infty \mu\left(|f - g| > \frac{1}{n}\right) = 0,$$

i.e. $f = g$ a.e. The second assertion is proved similarly.

3. Suppose $f_n \overset{\mu}{\to} f$, $\varepsilon > 0$ and $m, n \in \mathbb{N}$, then $|f_n - f_m| \le |f - f_n| + |f_m - f|$. So by the basic trick,

$$\mu(|f_n - f_m| > \varepsilon) \le \mu(|f_n - f| > \varepsilon/2) + \mu(|f_m - f| > \varepsilon/2) \to 0 \text{ as } m, n \to \infty.$$

■

Lemma 16.18 (Chebyshev's inequality again). *Let $p \in [1, \infty)$ and $f \in L^p$, then for all $\varepsilon > 0$,*

$$\mu(|f| \ge \varepsilon) \le \frac{1}{\varepsilon^p} \int_{\{|f| \ge \varepsilon\}} |f|^p \, d\mu \le \frac{1}{\varepsilon^p} \|f\|_p^p.$$

In particular if $\{f_n\} \subset L^p$ is L^p – convergent (Cauchy) then $\{f_n\}$ is also convergent (Cauchy) in measure.

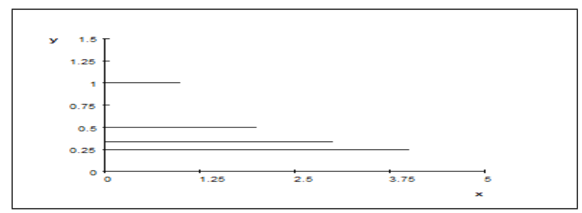
Proof. This is simply a repear of Chebyshev's inequality (see Remark 7.22 or Eq. **??**) using $\mu(|f| \ge \varepsilon) = \mu(|f|^p \ge \varepsilon^p)$. Here is the argument again for completeness;

$$\begin{aligned} \mu(|f| \ge \varepsilon) = \mu(|f|^p \ge \varepsilon^p) &= \int_X 1_{\frac{|f|^p}{\varepsilon^p} \ge 1} d\mu \\ &\le \int_X 1_{\frac{|f|^p}{\varepsilon^p} \ge 1} \cdot \frac{|f|^p}{\varepsilon^p} d\mu = \frac{1}{\varepsilon^p} \int_{\{|f| \ge \varepsilon\}} |f|^p \, d\mu = \frac{1}{\varepsilon^p} \|f\|_p^p. \end{aligned}$$

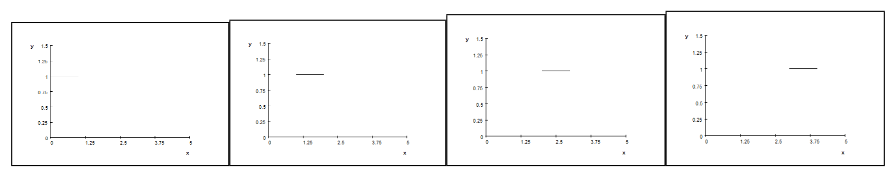
Given this ineqaulity if $\{f_n\}$ is L^p – Cauchy, then

$$\mu(|f_n - f_m| \ge \varepsilon) \le \frac{1}{\varepsilon^p} \|f_n - f_m\|_p^p \to 0 \text{ as } m, n \to \infty$$

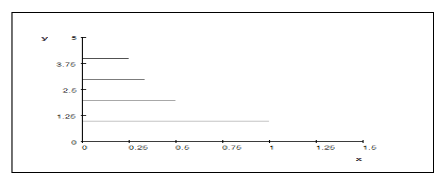
showing $\{f_n\}$ is L^0 – Cauchy. A similar argument holds for the L^p – convergent case. ■



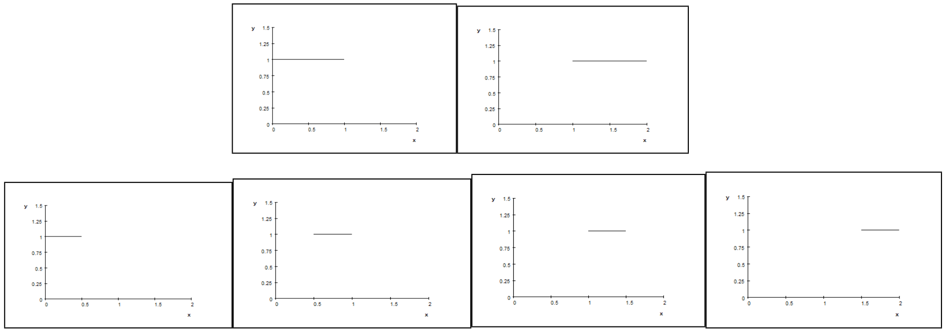
Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \nrightarrow 0$ in L^1, $f_n \overset{m}{\to} 0$.



Above is a sequence of functions where $f_n \to 0$ a.e., yet $f_n \nrightarrow 0$ in L^1. or in measure.



Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \overset{m}{\to} 0$ but $f_n \nrightarrow 0$ in L^1.



Above is a sequence of functions where $f_n \to 0$ in L^1, $f_n \nrightarrow 0$ a.e., and $f_n \stackrel{m}{\to} 0$.

Lemma 16.19. *Suppose $a_n \in \mathbb{C}$ and $|a_{n+1} - a_n| \le \varepsilon_n$ and $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Then $\lim_{n\to\infty} a_n = a \in \mathbb{C}$ exists and $|a - a_n| \le \delta_n := \sum_{k=n}^{\infty} \varepsilon_k$.*

Proof. (This is a special case of Exercise 13.13.) Let $m > n$ then

$$|a_m - a_n| = \left|\sum_{k=n}^{m-1} (a_{k+1} - a_k)\right| \le \sum_{k=n}^{m-1} |a_{k+1} - a_k| \le \sum_{k=n}^{\infty} \varepsilon_k := \delta_n . \tag{16.16}$$

So $|a_m - a_n| \le \delta_{\min(m,n)} \to 0$ as $m, n \to \infty$, i.e. $\{a_n\}$ is Cauchy. Let $m \to \infty$ in (16.16) to find $|a - a_n| \le \delta_n$. ∎

Theorem 16.20 (L^0 - Completeness). *Suppose $\{f_n\}$ is L^0-Cauchy. Then there exists a subsequence $g_j = f_{n_j}$ of $\{f_n\}$ such that $\lim g_j := f$ exists a.e. and $f_n \stackrel{\mu}{\to} f$ as $n \to \infty$. Moreover if g is a measurable function such that $f_n \stackrel{\mu}{\to} g$ as $n \to \infty$, then $f = g$ a.e.*

Proof. Let $\varepsilon_n > 0$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum_{k=n}^{\infty} \varepsilon_k$. Choose $g_j = f_{n_j}$ such that $\{n_j\}$ is a subsequence of $\mathbb{N}$ and

$$\mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \le \varepsilon_j.$$

Let $E_j = \{|g_{j+1} - g_j| > \varepsilon_j\}$,

$$F_N = \bigcup_{j=N}^{\infty} E_j = \bigcup_{j=N}^{\infty} \{|g_{j+1} - g_j| > \varepsilon_j\}$$

and

$$E := \bigcap_{N=1}^{\infty} F_N = \bigcap_{N=1}^{\infty} \bigcup_{j=N}^{\infty} E_j = \{|g_{j+1} - g_j| > \varepsilon_j \text{ i.o.}\}.$$

Then $\mu(E) = 0$ by Lemma **??** or the computation

$$\mu(E) \le \sum_{j=N}^{\infty} \mu(E_j) \le \sum_{j=N}^{\infty} \varepsilon_j = \delta_N \to 0 \text{ as } N \to \infty.$$

If $x \notin F_N$, i.e. $|g_{j+1}(x) - g_j(x)| \le \varepsilon_j$ for all $j \ge N$, then by Lemma 16.19, $f(x) = \lim_{j\to\infty} g_j(x)$ exists and $|f(x) - g_j(x)| \le \delta_j$ for all $j \ge N$. Therefore, since $E^c = \bigcup_{N=1}^{\infty} F_N^c$, $\lim_{j\to\infty} g_j(x) = f(x)$ exists for all $x \notin E$. Moreover, $\{x : |f(x) - g_j(x)| > \delta_j\} \subset F_j$ for all $j \ge N$ and hence

$$\mu(|f - g_j| > \delta_j) \le \mu(F_j) \le \delta_j \to 0 \text{ as } j \to \infty.$$

Therefore $g_j \stackrel{\mu}{\to} f$ as $j \to \infty$. Since

$$\begin{aligned}\{|f_n - f| > \varepsilon\} &= \{|f - g_j + g_j - f_n| > \varepsilon\} \\ &\subset \{|f - g_j| > \varepsilon/2\} \cup \{|g_j - f_n| > \varepsilon/2\},\end{aligned}$$

$$\mu(\{|f_n - f| > \varepsilon\}) \le \mu(\{|f - g_j| > \varepsilon/2\}) + \mu(|g_j - f_n| > \varepsilon/2)$$

and

$$\mu(\{|f_n - f| > \varepsilon\}) \le \lim_{j\to\infty} \sup \mu(|g_j - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$

If there is another function g such that $f_n \stackrel{\mu}{\to} g$ as $n \to \infty$, then arguing as above

$$\mu(|f - g| > \varepsilon) \le \mu(\{|f - f_n| > \varepsilon/2\}) + \mu(|g - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$

Hence

$$\mu(|f - g| > 0) = \mu(\cup_{n=1}^{\infty} \{|f - g| > \frac{1}{n}\}) \le \sum_{n=1}^{\infty} \mu(|f - g| > \frac{1}{n}) = 0,$$

i.e. $f = g$ a.e. ∎

Corollary 16.21 (Dominated Convergence Theorem). *Suppose $\{f_n\}$, $\{g_n\}$, and g are in L^1 and $f \in L^0$ are functions such that*

$$|f_n| \le g_n \text{ a.e.}, \; f_n \xrightarrow{\mu} f, \; g_n \xrightarrow{\mu} g, \text{ and } \int g_n \to \int g \text{ as } n \to \infty.$$

Then $f \in L^1$ and $\lim_{n\to\infty} \|f - f_n\|_1 = 0$, i.e. $f_n \to f$ in L^1. In particular $\lim_{n\to\infty} \int f_n = \int f$.

Proof. First notice that $|f| \le g$ a.e. and hence $f \in L^1$ since $g \in L^1$. To see that $|f| \le g$, use Theorem 16.20 to find subsequences $\{f_{n_k}\}$ and $\{g_{n_k}\}$ of $\{f_n\}$ and $\{g_n\}$ respectively which are almost everywhere convergent. Then

$$|f| = \lim_{k\to\infty} |f_{n_k}| \le \lim_{k\to\infty} g_{n_k} = g \text{ a.e.}$$

If (for sake of contradiction) $\lim_{n\to\infty} \|f - f_n\|_1 \ne 0$ there exists $\varepsilon > 0$ and a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that

$$\int |f - f_{n_k}| \ge \varepsilon \text{ for all } k. \tag{16.17}$$

Using Theorem 16.20 again, we may assume (by passing to a further subsequences if necessary) that $f_{n_k} \to f$ and $g_{n_k} \to g$ almost everywhere. Noting, $|f - f_{n_k}| \le g + g_{n_k} \to 2g$ and $\int (g + g_{n_k}) \to \int 2g$, an application of the dominated convergence Theorem 10.27 implies $\lim_{k\to\infty} \int |f - f_{n_k}| = 0$ which contradicts Eq. (16.17). ∎

Exercise 16.5 (Fatou's Lemma). If $f_n \ge 0$ and $f_n \to f$ in measure, then $\int f \le \liminf_{n\to\infty} \int f_n$.

Theorem 16.22 (Egoroff's Theorem). *Suppose $\mu(X) < \infty$ and $f_n \to f$ a.e. Then for all $\varepsilon > 0$ there exists $E \in \mathcal{M}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. In particular $f_n \xrightarrow{\mu} f$ as $n \to \infty$.*

Proof. Let $f_n \to f$ a.e. Then $\mu(\{|f_n - f| > \frac{1}{k} \text{ i.o. } n\}) = 0$ for all $k > 0$, i.e.

$$\lim_{N\to\infty} \mu\left(\bigcup_{n\ge N} \{|f_n - f| > \frac{1}{k}\}\right) = \mu\left(\bigcap_{N=1}^{\infty} \bigcup_{n\ge N} \{|f_n - f| > \frac{1}{k}\}\right) = 0.$$

Let $E_k := \bigcup_{n\ge N_k} \{|f_n - f| > \frac{1}{k}\}$ and choose an increasing sequence $\{N_k\}_{k=1}^{\infty}$ such that $\mu(E_k) < \varepsilon 2^{-k}$ for all k. Setting $E := \cup E_k$, $\mu(E) < \sum_k \varepsilon 2^{-k} = \varepsilon$ and if $x \notin E$, then $|f_n - f| \le \frac{1}{k}$ for all $n \ge N_k$ and all k. That is $f_n \to f$ uniformly on E^c. ∎

Exercise 16.6. Show that Egoroff's Theorem remains valid when the assumption $\mu(X) < \infty$ is replaced by the assumption that $|f_n| \le g \in L^1$ for all n. **Hint:** make use of Theorem 16.22 applied to $f_n|_{X_k}$ where $X_k := \left\{|g| \ge k^{-1}\right\}$.

17

Bochner Integral

Throughout this chapter we will assume that $(\Omega, \mathcal{F}, \mu)$ be a fixed measure space, X is a separable real or complex Banach space, $\mathcal{B} = \mathcal{B}(X) = \mathcal{B}_X$ is the Borel σ–algebra, and X^* is the continuous dual of X. In this chapter we will define the *Bochner integral*, $\int_\Omega f d\mu \in X$, of a measurable function $f : \Omega \to X$. (Shortly we will further assume that μ is a σ – finite on $\mathcal{F}$.)

17.1 Banach Valued L^p – Spaces

Notice that, since $\|\cdot\|_X : X \to [0, \infty)$ is continuous, it is $\mathcal{B}_X/\mathcal{B}_{[0,\infty)}$ – measurable. Before getting down to business we need to address a couple more measure theoretic properties of the Borel σ – algebra $(\mathcal{B}_X)$ on X.

Proposition 17.1. *Let X be a separable Banach space, and $D \subset X$ be a countable dense set, and*

$$\mathcal{V} := \{B(x, \varepsilon) : x \in D \text{ and } \varepsilon \in \mathbb{Q} \cap (0, \infty)\}.$$

Then $\mathcal{B}_X = \sigma(\mathcal{V})$.

Proof. As $\mathcal{V}$ consists of open sets it follows that $\sigma(\mathcal{V}) \subset \sigma(\text{open sets}) =: \mathcal{B}_X$. Conversely if $V \subset X$ is any open set we have $V = \cup_{W \in \mathcal{V}: W \subset V} W$ which shows V is a countable union of elements from $\mathcal{V}$. Therefore $V \in \sigma(\mathcal{V})$ and hence $\sigma(\mathcal{V})$ contains all open sets and therefore contains $\mathcal{B}_X$. ∎

Proposition 17.2. *Suppose that X and Y are two separable Banach spaces and equip $X \times Y$ with the norm*[1] $\|(x, y)\|_{X \times Y} := \max(\|x\|_X, \|y\|_Y)$ *and let $\mathcal{B}_{X \times Y}$ be the corresponding Borel σ – algebra. Then $\mathcal{B}_{X \times Y} = \mathcal{B}_X \otimes \mathcal{B}_Y$.*

Proof. Let $\pi : X \times Y \to X$ and $\alpha : X \times Y \to Y$ be the projection maps onto the first and second factor respectively. Each of these maps are continuous and hence Borel measurable and therefore $\pi^{-1}(\mathcal{B}_X) \subset \mathcal{B}_{X \times Y}$ and $\alpha^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_{X \times Y}$. (Indeed if τ_X denotes the open subset of X we find,

$$\begin{aligned}\{A \times Y : A \in \mathcal{B}_X\} &= \left\{\pi^{-1}(A) : A \in \mathcal{B}_X\right\} \\ &= \pi^{-1}(\mathcal{B}_X) = \pi^{-1}(\sigma(\tau_X)) = \sigma\left(\pi^{-1}(\tau_X)\right)\end{aligned}$$

[1] Any other equivalent norm would work just as well.

wherein we have used Lemma 9.3 for the last equality. Since $\pi^{-1}(V) = V \times Y$ is open in $X \times Y$ for all $V \in \tau_X$ it follows that $\pi^{-1}(\tau_X) \subset \mathcal{B}_{X \times Y}$.) Therefore

$$\{A \times Y : A \in \mathcal{B}_X\} = \sigma\left(\pi^{-1}(\tau_X)\right) \subset \mathcal{B}_{X \times Y}.$$

Similarly one shows $\{X \times B : B \in \mathcal{B}_Y\} \subset \mathcal{B}_{X \times Y}$ and therefore

$$A \times B = (A \times Y) \cap (X \times B) \in \mathcal{B}_{X \times Y} \text{ for all } A \in \mathcal{B}_X \text{ and } B \in \mathcal{B}_Y.$$

This shows that $\mathcal{B}_X \otimes \mathcal{B}_Y \subset \mathcal{B}_{X \times Y}$ even if X and Y are not separable.

For the converse inclusion it suffices to show if W is an open subset of $X \times Y$ then $W \in \mathcal{B}_X \otimes \mathcal{B}_Y$. To see this is the case let D_X and D_Y be countable dense subsets of X and Y respectively. One easily shows that $D_X \times D_Y$ is a countable dense subset of $X \times Y$ and therefore for any open subset $W \subset X \times Y$ we have,

$$W = \cup\{B_{X \times Y}((x, y), \varepsilon) \subset W : (x, y) \in D_X \times D_Y, \ \varepsilon \in \mathbb{Q} \cap (0, \infty)\}.$$

As $B_{X \times Y}((x, y), \varepsilon) = B_X(x, \varepsilon) \times B_Y(y, \varepsilon) \in \mathcal{B}_X \otimes \mathcal{B}_Y$ it follows that $W \in \mathcal{B}_X \otimes \mathcal{B}_Y$ being it is the countable union of sets in $\mathcal{B}_X \otimes \mathcal{B}_Y$. ∎

Corollary 17.3. *Suppose that X is a separable Banach space, then the vector addition map $X \times X \ni (x, y) \to x + y \in X$ is $\mathcal{B}_X \otimes \mathcal{B}_X/\mathcal{B}_X$ – measurable and the scalar multiplication map, $\mathbb{F} \times X \ni (\lambda, x) \to \lambda x \in X$ is $\mathcal{B}_\mathbb{F} \otimes \mathcal{B}_X/\mathcal{B}_X$ – measurable.*

Proof. Since each map is continuous and hence Borel – measurable the result is now a direct consequence of Proposition 17.2 which asserts the Borel and the product σ – algebras are one in the same. ∎

Corollary 17.4. *Suppose that X is a separable Banach space. If $f, g : \Omega \to X$ and $\lambda : \Omega \to \mathbb{F}$ are measurable maps, then $f + \lambda g : \Omega \to X$ is also measurable.*

Proof. Let $H : \Omega \to X \times X$ be the map defined by $H(\omega) := (f(\omega), g(\omega))$. As $\pi \circ H = f$ and $\alpha \circ H = g$ are both $\mathcal{F}/\mathcal{B}_X$ – measurable maps it follows (see Corollary 9.19) that H is This map is $\mathcal{F}/\mathcal{B}_X \otimes \mathcal{B}_X$ – measurable. Since $f + g$ is the composition of H with with vector addition which is also a measurable map by Corollary 17.3 we learn that $f + g$ is $\mathcal{F}/\mathcal{B}_X$ – measurable. A very similar argument shows that λg is $\mathcal{F}/\mathcal{B}_X$ – measurable as well. ∎

Proposition 17.5. *Suppose that X is a separable Banach space and $f_n : \Omega \to X$ is a sequence of $\mathcal{F}/\mathcal{B}_X$ – measurable functions such that $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ exists for all $\omega \in \Omega$. Then $f : \Omega \to X$ is a $\mathcal{F}/\mathcal{B}_X$ – measurable function.*

Proof. According to Proposition 17.1 it suffices to show $f^{-1}(B(x,\varepsilon)) \in \mathcal{F}$ for all $x \in D_X$ and $\varepsilon \in (0,\infty) \cap \mathbb{Q}$. To see this is the case choose $0 < \varepsilon_k < \varepsilon$ such that $\varepsilon_k \uparrow \varepsilon$ as $k \uparrow \infty$ and observe that

$$\begin{aligned} f^{-1}(B(x,\varepsilon)) &= \cup_{k=1}^{\infty} \{f_n \in B(x,\varepsilon_k) \text{ a.a}\} \\ &:= \cup_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n\geq N} \{f_n \in B(x,\varepsilon_k)\} \\ &= \cup_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n\geq N} f_n^{-1}(B(x,\varepsilon_k)) \in \mathcal{F}. \end{aligned}$$

■

Remark 17.6. Proposition 17.5 holds even if X is not separable. To prove this in full generality suppose that $V \subset X$ is an arbitrary open set. For $\varepsilon > 0$ let $V_\varepsilon := \{x \in V : d_{V^c}(x) > \varepsilon\}$ which is still open in X. We then have

$$\begin{aligned} f^{-1}(V) &= \cup_{k=1}^{\infty} \left\{f_n \in V_{1/k} \text{ a.a}\right\} \\ &:= \cup_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n\geq N} \left\{f_n \in V_{1/k}\right\} \\ &= \cup_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n\geq N} f_n^{-1}\left(V_{1/k}\right) \in \mathcal{F}. \end{aligned}$$

Remark 17.7. Later we will show that the Borel σ – field $\mathcal{B}$ on X is the same as the σ – field $(\sigma(X^*))$ which is generated by X^* – the continuous linear functionals on X. (This is done in Proposition **??**.) As a consequence $F : \Omega \to X$ is $\mathcal{F}/\mathcal{B}$ measurable iff $\varphi \circ F : \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$ – measurable for all $\varphi \in X^*$. This result gives an each proof that the sum of measurable functions is measurable and that scalar multiplication by a measurable functions preserves measurability. We may also conclude that the set $\{F \neq G\} = \{F - G \neq 0\}$ is measurable if F and G are measurable funcitons. Also note that $\|\cdot\| : X \to [0,\infty)$ is continuous and hence measurable and hence $\omega \to \|F(\omega)\|_X$ is the composition of two measurable functions and therefore measurable.

Definition 17.8. *For $1 \leq p < \infty$ let $L^p(\mu; X)$ denote the space of measurable functions $F : \Omega \to X$ such that $\int_\Omega \|F\|^p d\mu < \infty$. For $F \in L^p(\mu; X)$, define*

$$\|F\|_{L^p(\mu;X)} = \left(\int_\Omega \|F\|_X^p d\mu\right)^{\frac{1}{p}} = \|\|F(\cdot)\|_X\|_{L^p(\mu)}.$$

As usual in L^p – spaces we will identify two measurable functions, $F, G : \Omega \to X$, if $F = G$ a.e.

Theorem 17.9. *For each $p \in [0,\infty)$, the space $(L^p(\mu; X), \|\cdot\|_{L^p})$ is a Banach space.*

Proof. It is straightforward to check that $\|\cdot\|_{L^p}$ is a norm. For example,

$$\begin{aligned} \|F+G\|_{L^p(\mu;X)} &= \|\|F(\cdot)+G(\cdot)\|_X\|_{L^p(\mu)} \leq \|\|F(\cdot)\|_X + \|G(\cdot)\|_X\|_{L^p(\mu)} \\ &\leq \|\|F(\cdot)\|_X\|_{L^p(\mu)} + \|\|G(\cdot)\|_X\|_{L^p(\mu)} \quad \text{(by Minikowski's inequality)} \\ &= \|F\|_{L^p} + \|G\|_{L^p}. \end{aligned}$$

So the main point is to prove completeness of the norm.

If $\{F_n\}_{n=1}^{\infty} \subset L^p(\mu; X)$ is a sequence such that $\sum_{n=1}^{\infty} \|F_n\|_{L^p(\mu;X)} < \infty$, then by Corollary 16.5

$$\left\|\sum_{n=1}^{\infty} \|F_n(\cdot)\|_X\right\|_{L^p(\mu)} \leq \sum_{n=1}^{\infty} \|\|F_n(\cdot)\|_X\|_{L^p(\mu)} = \sum_{n=1}^{\infty} \|F_n\|_{L^p(\mu;X)} < \infty.$$

This inequality implies $\sum_{n=1}^{\infty} \|F_n(\omega)\|_X < \infty$ for μ – a.e. ω. If we let E be the exceptional null set;

$$E := \left\{\omega \in \Omega : \sum_{n=1}^{\infty} \|F_n(\omega)\|_X = \infty\right\},.$$

then $S(\omega) := \sum_{n=1}^{\infty} 1_{E^c}(\omega) \cdot F_n(\omega)$ is convergent in X for all $\omega \in \Omega$. By Proposition 17.5 we know that S is measurable. Moreover, if we let $S_N := \sum_{n=1}^{N} F_n$, then (using Corollary 16.5 again)

$$\begin{aligned} \|S - S_N\|_{L^p(\mu;X)} &= \|\|S - S_N\|_X\|_{L^p(\mu)} = \left\|\left\|\sum_{n=N+1}^{\infty} 1_{E^c} \cdot F_n\right\|_X\right\|_{L^p(\mu)} \\ &\leq \left\|\sum_{n=N+1}^{\infty} \|1_{E^c} \cdot F_n\|_X\right\|_{L^p(\mu)} \leq \sum_{n=N+1}^{\infty} \|\|1_{E^c} \cdot F_n\|_X\|_{L^p(\mu)} \\ &= \sum_{n=N+1}^{\infty} \|F_n\|_{L^p(\mu;X)} \to 0 \text{ as } N \to \infty. \end{aligned}$$

This shows that $S_N \to S$ in $L^p(\mu; X)$ and hence we have shown $L^p(\mu; X)$ is complete because of Theorem 14.18. ■

Remark 17.10. The same proof as Corollary 16.6 shows that every $L^p(\mu; X)$ – convergent sequence has a subsequence which is convergent almost everywhere.

We say a function $F : \Omega \to X$ is a simple function if F is measurable and has finite range. If F also satisfies, $\mu(F \neq 0) < \infty$ we say that F is a μ – simple function and let $\mathcal{S}(\mu; X)$ denote the vector space of μ – simple functions.

Proposition 17.11. *For each $1 \le p < \infty$ the μ – simple functions, $\mathcal{S}(\mu; X)$, are dense inside of $L^p(\mu; X)$.*

Proof. Let $\mathbb{D} := \{x_n\}_{n=1}^\infty$ be a countable dense subset of $X \setminus \{0\}$. For each $\varepsilon > 0$ and $n \in \mathbb{N}$ let

$$B_n^\varepsilon := \left\{x \in X : \|x - x_n\| \le \min\left(\varepsilon, \frac{1}{2}\|x_n\|\right)\right\}$$

and then define $A_n^\varepsilon := B_n^\varepsilon \setminus \left(\cup_{k=1}^{n-1} B_k^\varepsilon\right)$. Thus $\{A_n^\varepsilon\}_{n=1}^\infty$ is a partition of $X \setminus \{0\}$ with the added property that $\|y - x_n\| \le \varepsilon$ and $\frac{1}{2}\|x_n\| \le \|y\| \le \frac{3}{2}\|x_n\|$ for all $y \in A_n^\varepsilon$.

Given $F \in L^p(\mu; X)$ let

$$F_\varepsilon := \sum_{n=1}^\infty x_n \cdot 1_{F \in A_n^\varepsilon} = \sum_{n=1}^\infty x_n \cdot 1_{F^{-1}(A_n^\varepsilon)}.$$

For $\omega \in F^{-1}(A_n^\varepsilon)$, i.e. $F(\omega) \in A_n^\varepsilon$, we have

$$\|F_\varepsilon(\omega)\| = \|x_n\| \le 2\|F(\omega)\| \text{ and}$$
$$\|F_\varepsilon(\omega) - F(\omega)\| = \|x_n - F(\omega)\| \le \varepsilon.$$

Putting these two estimates together shows,

$$\|F_\varepsilon - F\| \le \varepsilon \text{ and } \|F_\varepsilon - F\| \le \|F_\varepsilon\| + \|F\| \le 3\|F\|.$$

Hence we may now apply the dominated convergence theorem in order to show

$$\lim_{\varepsilon \downarrow 0} \|F - F_\varepsilon\|_{L^p(\mu;X)} = 0.$$

We are not quite done yet since F_ε typically has countable rather than finite range. To remedy this defect, to each $N \in \mathbb{N}$ let

$$F_\varepsilon^N := \sum_{n=1}^N x_n \cdot 1_{F^{-1}(A_n^\varepsilon)}.$$

Then it is clear that $\lim_{N\to\infty} F_\varepsilon^N = F_\varepsilon$ and that $\left\|F_\varepsilon^N\right\| \le \|F_\varepsilon\| \le 2\|F\|$ for all N. Therefore another application of the dominated convergence theorem implies, $\lim_{N\to\infty}\left\|F_\varepsilon^N - F_\varepsilon\right\|_{L^p(\mu;X)} = 0$. Thus any $F \in L^p(\mu; X)$ may be arbitrarily well approximated by one of the $F_\varepsilon^N \in \mathcal{S}(\mu; X)$ with ε sufficiently small and N sufficiently large. ■

For later purposes it will be useful to record a result based on the partitions $\{A_n^\varepsilon\}_{n=1}^\infty$ of $X \setminus \{0\}$ introduced in the above proof.

Lemma 17.12. *Suppose that $F : \Omega \to X$ is a measurable function such that $\mu(F \neq 0) > 0$. Then there exists $B \in \mathcal{F}$ and $\varphi \in X^*$ such that $\mu(B) > 0$ and $\inf_{\omega \in B} \varphi \circ F(\omega) > 0$.*

Proof. Let $\varepsilon > 0$ be chosen arbitrarily, for example you might take $\varepsilon = 1$ and let $\{A_n := A_n^\varepsilon\}_{n=1}^\infty$ be the partition of $X \setminus \{0\}$ introduced in the proof of Proposition 17.11 above. Since $\{F \neq 0\} = \sum_{n=1}^\infty \{F \in A_n\}$ and $\mu(F \neq 0) > 0$, it follows that that $\mu(F \in A_n) > 0$ for some $n \in \mathbb{N}$. We now let $B := \{F \in A_n\} = F^{-1}(A_n)$ and choose $\varphi \in X^*$ such that $\varphi(x_n) = \|x_n\|$ and $\|\varphi\|_{X^*} = 1$. For $\omega \in B$ we have $F(\omega) \in A_n$ and therefore $\|F(\omega) - x_n\| \le \frac{1}{2}\|x_n\|$ and hence,

$$|\varphi(F(\omega)) - \|x_n\|| = |\varphi(F(\omega)) - \varphi(x_n)| \le \|\varphi\|_{X^*}\|F(\omega) - x_n\| \le \frac{1}{2}\|x_n\|.$$

From this inequality we see that $\varphi(F(\omega)) \ge \frac{1}{2}\|x_n\| > 0$ for all $\omega \in B$. ■

17.2 Bochner Integral

Definition 17.13. *To each $F \in \mathcal{S}(\mu; X)$, let*

$$I(F) = \sum_{x \in X} x\mu(F^{-1}(\{x\})) = \sum_{x \in X} x\mu(\{F = x\})$$
$$= \sum_{x \in F(\Omega)} x\mu(F = x) \in X.$$

The following proposition is straightforward to prove.

Proposition 17.14. *The map $I : \mathcal{S}(\mu; X) \to X$ is linear and satisfies for all $F \in \mathcal{S}(\mu; X)$,*

$$\|I(F)\|_X \le \int_\Omega \|F\| d\mu \text{ and} \tag{17.1}$$

$$\varphi(I(F)) = \int_X \varphi \circ F \, d\mu \; \forall \varphi \in X^*. \tag{17.2}$$

Proof. If $0 \neq c \in \mathbb{R}$ and $F \in \mathcal{S}(\mu; X)$, then

$$I(cF) = \sum_{x \in X} x\mu(cF = x) = \sum_{x \in X} x\mu\left(F = \frac{x}{c}\right)$$
$$= \sum_{y \in X} cy \; \mu(F = y) = cI(F)$$

and if $c = 0$, $I(0F) = 0 = 0I(F)$. If $F, G \in \mathcal{S}(\mu; X)$,

$$\begin{aligned} I(F+G) &= \sum_x x\mu(F+G=x) \\ &= \sum_x x \sum_{y+z=x} \mu(F=y, G=z) \\ &= \sum_{y,z} (y+z)\mu(F=y,G=z) \\ &= \sum_y y\mu(F=y) + \sum_z z\mu(G=z) = I(F)+I(G). \end{aligned}$$

Equation (17.1) is a consequence of the following computation:

$$\|I(F)\|_X = \|\sum_{x\in X} x\mu(F=x)\| \le \sum_{x\in X} \|x\|\mu(F=x) = \int_\Omega \|F\| d\mu$$

and Eq. (17.2) follows from:

$$\begin{aligned} \varphi(I(F)) &= \varphi(\sum_{x\in X} x\mu(\{F=x\})) \\ &= \sum_{x\in X} \varphi(x)\mu(\{F=x\}) = \int_X \varphi\circ F\, d\mu. \end{aligned}$$

■

Theorem 17.15 (Bochner Integral). *There is a unique continuous linear map* $\bar{I} : L^1(\Omega,\mathcal{F},\mu;X) \to X$ *such that* $\bar{I}|_{\mathcal{S}(\mu;X)} = I$ *where* I *is defined in Definition 17.13. Moreover, for all* $F \in L^1(\Omega,\mathcal{F},\mu;X)$,

$$\|\bar{I}(F)\|_X \le \int_\Omega \|F\| d\mu \qquad (17.3)$$

and $\bar{I}(F)$ *is the unique element in* X *such that*

$$\varphi(\bar{I}(F)) = \int_X \varphi\circ F\, d\mu \;\forall \varphi\in X^*. \qquad (17.4)$$

The map $\bar{I}(F)$ *will be denoted suggestively by* $\int_X F d\mu$ *or* $\mu(F)$ *so that Eq. (17.4) may be written as*

$$\varphi\left(\int_X F d\mu\right) = \int_X \varphi\circ F\, d\mu \;\forall\varphi\in X^* \text{ or}$$
$$\varphi(\mu(F)) = \mu(\varphi\circ F)\;\;\forall\varphi\in X^*$$

Proof. The existence of a continuous linear map $\bar{I} : L^1(\Omega,\mathcal{F},\mu;X) \to X$ such that $\bar{I}|_{\mathcal{S}(\mu;X)} = I$ and Eq. (17.3) holds follows from Propositions 17.14 and 17.11 and the bounded linear transformation Theorem 32.4. If $\varphi\in X^*$ and $F\in L^1(\Omega,\mathcal{F},\mu;X)$, choose $F_n \in \mathcal{S}(\mu;X)$ such that $F_n \to F$ in $L^1(\Omega,\mathcal{F},\mu;X)$ as $n\to\infty$. Then $\bar{I}(F) = \lim_{n\to\infty} I(F_n)$ and hence by Eq. (17.2),

$$\varphi(\bar{I}(F)) = \varphi(\lim_{n\to\infty} I(F_n)) = \lim_{n\to\infty}\varphi(I(F_n)) = \lim_{n\to\infty}\int_X \varphi\circ F_n\, d\mu.$$

This proves Eq. (17.4) since

$$\begin{aligned} \left|\int_\Omega (\varphi\circ F - \varphi\circ F_n) d\mu\right| &\le \int_\Omega |\varphi\circ F - \varphi\circ F_n|\, d\mu \\ &\le \int_\Omega \|\varphi\|_{X^*} \|\varphi\circ F - \varphi\circ F_n\|_X\, d\mu \\ &= \|\varphi\|_{X^*}\|F-F_n\|_{L^1} \to 0 \text{ as } n\to\infty. \end{aligned}$$

The fact that $\bar{I}(F)$ is determined by Eq. (17.4) is a consequence of the Hahn-Banach Theorem 21.7 below. ■

Example 17.16. Suppose that $x\in X$ and $f\in L^1(\mu;\mathbb{R})$, then $F(\omega) := f(\omega)x$ defines an element of $L^1(\mu;X)$ and

$$\int_\Omega F d\mu = \left(\int_\Omega f d\mu\right) x. \qquad (17.5)$$

To prove this just observe that $\|F\| = |f|\,\|x\| \in L^1(\mu)$ and then choose $f_n \in \mathcal{S}(\mu;\mathbb{C})$ approximating f in $L^1(\mu)$. Then $f_n x \to F$ is $L^1(\mu : X)$ and using the easily proved fact that $I(f_n x) = \int_\Omega f_n d\mu \cdot x$ it follows that therefore,

$$\int_\Omega F d\mu = \lim_{n\to\infty} I(f_n x) = \lim_{n\to\infty}\int_\Omega f_n d\mu\cdot x = \left(\int_\Omega f d\mu\right)x.$$

Alternatively, for $\varphi\in X^*$ we have

$$\begin{aligned} \varphi\left(\left(\int_\Omega f d\mu\right)x\right) &= \left(\int_\Omega f d\mu\right)\cdot\varphi(x) \\ &= \left(\int_\Omega f\varphi(x)\, d\mu\right) = \int_\Omega \varphi\circ F\, d\mu. \end{aligned}$$

Since $\varphi\left(\int_\Omega F d\mu\right) = \int_\Omega \varphi\circ F\, d$ for all $\varphi\in X^*$ it follows that Eq. (17.5) is correct.

Remark 17.17. The separability assumption on X may be relaxed by assuming that $F : \Omega \to X$ has separable essential range. In this case we may still define $\int_X F d\mu$ by applying the above formalism with X replaced by the separable Banach space, $X_0 := \overline{\text{span}(\text{essran}_\mu(F))}$. For example if Ω is a compact topological space and $F : \Omega \to X$ is a continuous map, then $\int_\Omega F d\mu$ is always defined.

18

Hilbert Space Basics

Definition 18.1. *Let H be a complex vector space. An inner product on H is a function, $\langle\cdot|\cdot\rangle : H \times H \to \mathbb{C}$, such that*

1. $\langle ax+by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$ *i.e.* $x \to \langle x|z\rangle$ *is linear.*
2. $\overline{\langle x|y\rangle} = \langle y|x\rangle$.
3. $\|x\|^2 := \langle x|x\rangle \geq 0$ *with equality* $\|x\|^2 = 0$ *iff* $x = 0$.

Notice that combining properties (1) and (2) that $x \to \langle z|x\rangle$ is conjugate linear for fixed $z \in H$, i.e.

$$\langle z|ax+by\rangle = \bar{a}\langle z|x\rangle + \bar{b}\langle z|y\rangle.$$

The following identity will be used frequently in the sequel without further mention,

$$\begin{aligned}\|x+y\|^2 &= \langle x+y|x+y\rangle = \|x\|^2 + \|y\|^2 + \langle x|y\rangle + \langle y|x\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle. \end{aligned} \tag{18.1}$$

Theorem 18.2 (Schwarz Inequality). *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, then for all $x, y \in H$*

$$|\langle x|y\rangle| \leq \|x\|\|y\|$$

and equality holds iff x and y are linearly dependent.

Proof. If $y = 0$, the result holds trivially. So assume that $y \neq 0$ and observe; if $x = \alpha y$ for some $\alpha \in \mathbb{C}$, then $\langle x|y\rangle = \alpha \left\|y\right\|^2$ and hence

$$|\langle x|y\rangle| = |\alpha| \left\|y\right\|^2 = \|x\|\|y\|.$$

Now suppose that $x \in H$ is arbitrary, let $z := x - \|y\|^{-2}\langle x|y\rangle y$. (So z is the "orthogonal projection" of x onto y, see Figure 18.1.) Then

$$\begin{aligned} 0 \leq \|z\|^2 &= \left\|x - \frac{\langle x|y\rangle}{\|y\|^2}y\right\|^2 = \|x\|^2 + \frac{|\langle x|y\rangle|^2}{\|y\|^4}\|y\|^2 - 2\mathrm{Re}\langle x|\frac{\langle x|y\rangle}{\|y\|^2}y\rangle \\ &= \|x\|^2 - \frac{|\langle x|y\rangle|^2}{\|y\|^2} \end{aligned}$$

from which it follows that $0 \leq \|y\|^2\|x\|^2 - |\langle x|y\rangle|^2$ with equality iff $z = 0$ or equivalently iff $x = \|y\|^{-2}\langle x|y\rangle y$. ■

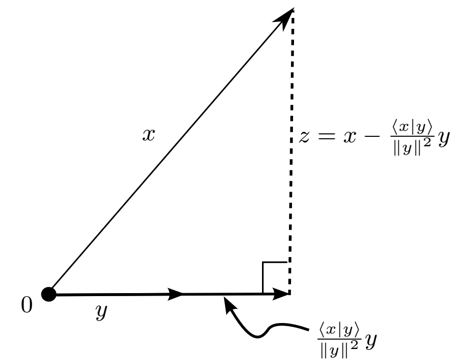


Fig. 18.1. The picture behind the proof of the Schwarz inequality.

Corollary 18.3. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space and $\|x\| := \sqrt{\langle x|x\rangle}$. Then the **Hilbertian norm**, $\|\cdot\|$, is a norm on H. Moreover $\langle\cdot|\cdot\rangle$ is continuous on $H \times H$, where H is viewed as the normed space $(H, \|\cdot\|)$.*

Proof. If $x, y \in H$, then, using Schwarz's inequality,

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.\end{aligned}$$

Taking the square root of this inequality shows $\|\cdot\|$ satisfies the triangle inequality.

Checking that $\|\cdot\|$ satisfies the remaining axioms of a norm is now routine and will be left to the reader. The continuity of the inner product follows from Theorem 18.2 as in Exercise 14.11. For completeness, if $x, y, \Delta x, \Delta y \in H$, then

$$\begin{aligned}|\langle x+\Delta x|y+\Delta y\rangle - \langle x|y\rangle| &= |\langle x|\Delta y\rangle + \langle \Delta x|y\rangle + \langle \Delta x|\Delta y\rangle| \\ &\leq \|x\|\,\|\Delta y\| + \|y\|\,\|\Delta x\| + \|\Delta y\|\,\|\Delta x\|\end{aligned}$$

with the latter expression clearly going to zero as Δx and Δy go to zero proving the inner product is continuous.

■

Remark 18.4 (Polarization identity). It is sometimes useful to know that the inner product $\langle\cdot|\cdot\rangle$ may be reconstructed from knowledge of its associated norm

$\|\cdot\|$. For example in the real case we have from Eq. (18.1) that

$$\langle x|y\rangle = \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right). \tag{18.2}$$

Similarly if we are working over $\mathbb{C}$, then from Eq. (18.1) or direct computation,

$$2\,\mathrm{Re}\langle x|y\rangle = \|x+y\|^2 - \|x\|^2 - \|y\|^2$$

and

$$-2\mathrm{Re}\langle x|y\rangle = \|x-y\|^2 - \|x\|^2 - \|y\|^2.$$

Subtracting these two equations gives the "polarization identity,"

$$4\mathrm{Re}\langle x|y\rangle = \|x+y\|^2 - \|x-y\|^2.$$

Replacing y by iy in this equation then implies that

$$4\mathrm{Im}\langle x|y\rangle = \|x+iy\|^2 - \|x-iy\|^2$$

from which we find

$$\langle x|y\rangle = \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\varepsilon y\|^2 \tag{18.3}$$

where $G = \{\pm 1, \pm i\}$ – a cyclic subgroup of $S^1 \subset \mathbb{C}$.

Definition 18.5. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, we say $x, y \in H$ are **orthogonal** and write $x \perp y$ iff $\langle x|y\rangle = 0$. More generally if $A \subset H$ is a set, $x \in H$ is **orthogonal to** A (write $x \perp A$) iff $\langle x|y\rangle = 0$ for all $y \in A$. Let $A^\perp = \{x \in H : x \perp A\}$ be the set of vectors orthogonal to A. A subset $S \subset H$ is an **orthogonal set** if $x \perp y$ for all distinct elements $x, y \in S$. If S further satisfies, $\|x\| = 1$ for all $x \in S$, then S is said to be an **orthonormal set.***

Proposition 18.6. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space then*

1. ***(Parallelogram Law)***
$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{18.4}$$
for all $x, y \in H$. [See Proposition 18.43 for a "converse" to the parallelogram law.]
2. ***(Pythagorean Theorem)** If $S \subset\subset H$ is a finite orthogonal set, then*
$$\left\|\sum_{x\in S} x\right\|^2 = \sum_{x\in S}\|x\|^2. \tag{18.5}$$
3. *If $A \subset H$ is a set, then $A^\perp$ is a **closed** linear subspace of H.*

Proof. I will assume that H is a complex Hilbert space, the real case being easier. Items 1. and 2. are proved by the following elementary computations;

$$\begin{aligned}\|x+y\|^2 &+ \|x-y\|^2 \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle + \|x\|^2 + \|y\|^2 - 2\mathrm{Re}\langle x|y\rangle \\ &= 2\|x\|^2 + 2\|y\|^2,\end{aligned}$$

and

$$\begin{aligned}\left\|\sum_{x\in S} x\right\|^2 &= \langle\sum_{x\in S} x|\sum_{y\in S} y\rangle = \sum_{x,y\in S}\langle x|y\rangle \\ &= \sum_{x\in S}\langle x|x\rangle = \sum_{x\in S}\|x\|^2.\end{aligned}$$

Item 3. is a consequence of the continuity of $\langle\cdot|\cdot\rangle$ and the fact that

$$A^\perp = \cap_{x\in A}\,\mathrm{Nul}(\langle\cdot|x\rangle)$$

where $\mathrm{Nul}(\langle\cdot|x\rangle) = \{y \in H : \langle y|x\rangle = 0\}$ – a closed subspace of H. ∎

Definition 18.7. *A **Hilbert space** is an inner product space $(H, \langle\cdot|\cdot\rangle)$ such that the induced Hilbertian norm is complete.*

Example 18.8. Suppose X is a set and $\mu : X \to (0, \infty)$, then $H := \ell^2(\mu)$ is a Hilbert space when equipped with the inner product,

$$\langle f|g\rangle := \sum_{x\in X} f(x)\,\bar{g}(x)\,\mu(x).$$

In Exercise 18.9 you will show every Hilbert space H is "equivalent" to a Hilbert space of this form with $\mu \equiv 1$. More generally, if $(X, \mathcal{M}, \mu)$ is a general measure space then $L^2(\mu)$ is a Hilbert space when equipped with the inner product,

$$\langle f|g\rangle := \int_X f(x)\,\bar{g}(x)\,d\mu(x).$$

Definition 18.9. *A subset C of a vector space X is said to be convex if for all $x, y \in C$ the line segment $[x, y] := \{tx + (1-t)y : 0 \le t \le 1\}$ joining x to y is contained in C as well. (Notice that any vector subspace of X is convex.)*

Theorem 18.10 (Best Approximation Theorem). *Suppose that H is a Hilbert space and $M \subset H$ is a closed convex subset of H. Then for any $x \in H$ there exists a unique $y \in M$ such that*

$$\|x-y\| = d(x, M) = \inf_{z\in M}\|x-z\|.$$

Moreover, if M is a vector subspace of H, then the point y may also be characterized as the unique point in M such that $(x-y) \perp M$.

Proof. Let $x \in H$, $\delta := d(x, M)$, $y, z \in M$, and, referring to Figure 18.2, let $w = z + (y - x)$ and $c = (z + y)/2 \in M$. It then follows by the parallelogram law (Eq. (18.4) with $a = (y - x)$ and $b = (z - x)$) and the fact that $c \in M$ that

$$\begin{aligned} 2\|y - x\|^2 + 2\|z - x\|^2 &= \|w - x\|^2 + \|y - z\|^2 \\ &= \|z + y - 2x\|^2 + \|y - z\|^2 \\ &= 4\|x - c\|^2 + \|y - z\|^2 \\ &\geq 4\delta^2 + \|y - z\|^2 . \end{aligned}$$

Thus we have shown for all $y, z \in M$ that,

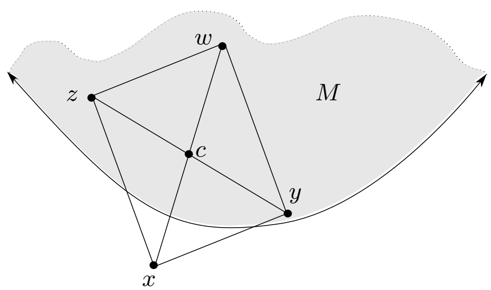


Fig. 18.2. In this figure $y, z \in M$ and by convexity, $c = (z + y)/2 \in M$.

$$\|y - z\|^2 \leq 2\|y - x\|^2 + 2\|z - x\|^2 - 4\delta^2. \tag{18.6}$$

Uniqueness. If $y, z \in M$ minimize the distance to x, then $\|y - x\| = \delta = \|z - x\|$ and it follows from Eq. (18.6) that $y = z$.

Existence. Let $y_n \in M$ be chosen such that $\|y_n - x\| = \delta_n \to \delta = d(x, M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (18.6) shows

$$\|y_n - y_m\|^2 \leq 2\delta_m^2 + 2\delta_n^2 - 4\delta^2 \to 0 \text{ as } m, n \to \infty.$$

Therefore, by completeness of H, $\{y_n\}_{n=1}^\infty$ is convergent. Because M is closed, $y := \lim_{n\to\infty} y_n \in M$ and because the norm is continuous,

$$\|y - x\| = \lim_{n\to\infty} \|y_n - x\| = \delta = d(x, M).$$

So y is the desired point in M which is closest to x.

Orthogonality property. Now suppose M is a closed subspace of H and $x \in H$. Let $y \in M$ be the closest point in M to x. Then for $w \in M$, the function

$$g(t) := \|x - (y + tw)\|^2 = \|x - y\|^2 - 2t\mathrm{Re}\langle x - y | w\rangle + t^2\|w\|^2$$

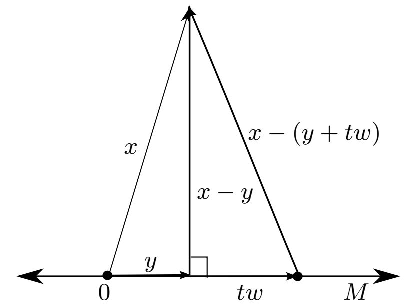


Fig. 18.3. The orthogonality relationships of closest points.

has a minimum at $t = 0$ and therefore $0 = g'(0) = -2\mathrm{Re}\langle x - y | w\rangle$. Since $w \in M$ is arbitrary, this implies that $(x - y) \perp M$, see Figure 18.3.Finally suppose $y \in M$ is any point such that $(x - y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2$$

which shows $d(x, M)^2 \geq \|x - y\|^2$. That is to say y is the point in M closest to x.

Alternate organization. Let $\delta = d(x, M)$ and $\{y_n\} \subset M$ be a minimizing sequence, i.e. $\|y_n - x\| = \delta_n \to \delta = d(x, M)$. We start with the identity,

$$\|(x - y) + (x - z)\|^2 + \|(x - y) - (x - z)\|^2 = 2 \cdot \left[\|x - y\|^2 + \|x - z\|^2\right]$$

which is equivalent to

$$4\left\|x - \frac{y + z}{2}\right\|^2 + \|z - y\|^2 = 2\left[\|x - y\|^2 + \|x - z\|^2\right].$$

As $\frac{y+z}{2} \in M$ we may conclude with that

$$4\delta^2 + \|z - y\|^2 \leq 2\left[\|x - y\|^2 + \|x - z\|^2\right]. \tag{18.7}$$

Taking $y = y_m$ and $z = y_n$ in Eq. (18.7) shows,

$$4\delta^2 + \|y_n - y_m\|^2 \leq 2\left[\delta_m^2 + \delta_n^2\right] \to 2\left[\delta^2 + \delta^2\right] = 4\delta^2 \text{ as } m, n \to \infty$$

from which it follows that $\{y_n\}_{n=1}^\infty$ is a Cauchy sequence. Therefore a desired minimizer is $y := \lim_{n\to\infty} y_n$ which is in M because M is closed. If z were another minimizer, it would follows from Eq. (18.7) that

$$4\delta^2 + \|z - y\|^2 \leq 2\left[\delta^2 + \delta^2\right] = 4\delta^2.$$

This inequality forces $\|z-y\|^2 = 0$ or equivalently $y = z$. ∎

Definition 18.11. *Suppose that $A : H \to H$ is a bounded operator. The* ***adjoint*** *of A, denoted A^*, is the unique operator $A^* : H \to H$ such that $\langle Ax|y\rangle = \langle x|A^*y\rangle$. (The proof that A^* exists and is unique will be given in Proposition 18.18 below.) A bounded operator $A : H \to H$ is* ***self - adjoint*** *or* ***Hermitian*** *if $A = A^*$.*

Example 18.12 (L^2 – spaces need to be introduced first). Let $(X, \mathcal{B}, \mu)$ be a σ – finite measure space and suppose $H = L^2(X, \mathcal{B}, \mu)$ and $K : H \to H$ is given by the integral operator of the form,

$$Kf(x) = \int_X k(x,y) f(y)\, d\mu(y) \text{ for all } f \in H$$

where $k \in L^2(\mu \otimes \mu)$. In this case if $f, g \in H$ then

$$\int_{X\times X} |k(x,y) f(y) \bar{g}(x)|\, d\mu(y)\, d\mu(x)$$
$$\le \left[\int_{X\times X} |k(x,y)|^2 d\mu(y)\, d\mu(x) \cdot \int_{X\times X} |f(y)\bar{g}(x)|^2 d\mu(y)\, d\mu(x)\right]^{1/2} < \infty$$

and hence by Fubini's theorem,

$$\begin{aligned}\langle Kf|g\rangle &= \int_X d\mu(x) \int_X d\mu(y)\, k(x,y) f(y) \bar{g}(x) \\ &= \int_X d\mu(y) f(y) \overline{\int_X d\mu(x)\, \bar{k}(x,y) g(x)} = \langle f|K^*g\rangle\end{aligned}$$

where

$$(K^*g)(y) := \int_X d\mu(x)\, \bar{k}(x,y) g(x).$$

Thus the integral kernel for K^* is gotten by interchanging the arguments of k and taking the complex conjugate of the result.

Definition 18.13. *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection of H onto M is the function $P_M : H \to H$ such that for $x \in H$, $P_M(x)$ is the unique element in M such that $(x - P_M(x)) \perp M$.*

Theorem 18.14 (Projection Theorem). *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection P_M satisfies:*

1. *P_M is linear and hence we will write $P_M x$ rather than $P_M(x)$.*
2. *$P_M^2 = P_M$ (P_M is a projection).*
3. *$P_M^* = P_M$ (P_M is self-adjoint).*
4. *$\operatorname{Ran}(P_M) = M$ and $\operatorname{Nul}(P_M) = M^\perp$.*
5. *If $N \subset M \subset H$ is another closed subspace, the $P_N P_M = P_M P_N = P_N$.*
6. *Provided $M \neq \{0\}$, $\|P_M\|_{op} = 1$.*

Proof.

1. Let $x_1, x_2 \in H$ and $\alpha \in \mathbb{F}$, then $P_M x_1 + \alpha P_M x_2 \in M$ and
$$P_M x_1 + \alpha P_M x_2 - (x_1 + \alpha x_2) = [P_M x_1 - x_1 + \alpha(P_M x_2 - x_2)] \in M^\perp$$
showing $P_M x_1 + \alpha P_M x_2 = P_M(x_1 + \alpha x_2)$, i.e. P_M is linear.
2. Obviously $\operatorname{Ran}(P_M) = M$ and $P_M x = x$ for all $x \in M$. Therefore $P_M^2 = P_M$.
3. Let $x, y \in H$, then since $(x - P_M x)$ and $(y - P_M y)$ are in $M^\perp$,
$$\begin{aligned}\langle P_M x|y\rangle &= \langle P_M x|P_M y + y - P_M y\rangle = \langle P_M x|P_M y\rangle \\ &= \langle P_M x + (x - P_M x)|P_M y\rangle = \langle x|P_M y\rangle.\end{aligned}$$
4. We have already seen, $\operatorname{Ran}(P_M) = M$ and $P_M x = 0$ iff $x = x - 0 \in M^\perp$, i.e. $\operatorname{Nul}(P_M) = M^\perp$.
5. If $N \subset M \subset H$ it is clear that $P_M P_N = P_N$ since $P_M = Id$ on $N = \operatorname{Ran}(P_N) \subset M$. Taking adjoints gives the other identity, namely that $P_N P_M = P_N$.
Alternative proof 1 of $P_N P_M = P_N$. If $x \in H$, then $(x - P_M x) \perp M$ and therefore $(x - P_M x) \perp N$. We also have $(P_M x - P_N P_M x) \perp N$ and therefore,
$$x - P_N P_M x = (x - P_M x) + (P_M x - P_N P_M x) \in N^\perp$$
which shows $P_N P_M x = P_N x$.
Alternative proof 2 of $P_N P_M = P_N$. If $x \in H$ and $n \in N$, we have
$$\langle P_N P_M x|n\rangle = \langle P_M x|P_N n\rangle = \langle P_M x|n\rangle = \langle x|P_M n\rangle = \langle x|n\rangle.$$
Since this holds for all n we may conclude that $P_N P_M x = P_N x$.
6. Given $z \in H$ we have,
$$\|z\|^2 = \|P_M z + (z - P_M z)\|^2 = \|P_M z\|^2 + \|z - P_M z\|^2 \ge \|P_M z\|^2$$
with equality for any $z \in M$. This shows that $\|P_M\|_{op} = 1$.

∎

Lemma 18.15. *Suppose that $\Gamma \subset_f H \setminus \{0\}$ is a finite orthogonal collection of elements of H. If $M := \operatorname{span}(\Gamma)$, then M is a closed subspace and*

$$P_M z = \sum_{h\in\Gamma} \frac{\langle z|h\rangle}{\|h\|^2} h \text{ for all } z \in H.$$

Proof. If $z \in M$ then $z = \sum a_h(z) h$ for some $a_h(z) \in \mathbb{C}$. Taking the inner product of this equation with $k \in \Gamma$ (using the orthogonality assumptions) shows $\langle z|k\rangle = a_k(z) \|k\|^2$, i.e.

$$a_k(z) = \frac{\langle z|k\rangle}{\|k\|^2} \text{ for all } k \in \Gamma.$$

Let $Pz := \sum_{h\in\Gamma} \frac{\langle z|h\rangle}{\|h\|^2} h$ for all $z \in H$ and observe that $\langle Pz|k\rangle = \langle z|k\rangle$ for all $k \in \Gamma$, i.e. $\langle z - Pz|k\rangle = 0$ for all $k \in \Gamma$. From this it follows that $(z - Pz) \perp M$. So to finish the proof it only remains to show M is closed. [We will later see that M is closed simply by the virtue that $\dim M < \infty$.]

Let us first observe that

$$\|z\|^2 = \|Pz + (z - Pz)\|^2 = \|Pz\|^2 + \|z - Pz\|^2 \geq \|Pz\|^2$$

for all $z \in H$ and therefore P is bounded. Hence if $\{z_n\} \subset M$ and $z_n \to z \in H$, then

$$z = \lim_{n\to\infty} z_n = \lim_{n\to\infty} Pz_n = P \lim_{n\to\infty} z_n = Pz \in M.$$

■

Corollary 18.16. *If $M \subset H$ is a proper closed subspace of a Hilbert space H, then $H = M \oplus M^\perp$.*

Proof. Given $x \in H$, let $y = P_M x$ so that $x - y \in M^\perp$. Then $x = y + (x - y) \in M + M^\perp$. If $x \in M \cap M^\perp$, then $x \perp x$, i.e. $\|x\|^2 = \langle x|x\rangle = 0$. So $M \cap M^\perp = \{0\}$. ■

Exercise 18.1. Suppose M is a subset of H, then $M^{\perp\perp} = \overline{\operatorname{span}(M)}$ where (as usual), $\operatorname{span}(M)$ denotes all finite linear combinations of elements from M.

Theorem 18.17 (Riesz Theorem). *Let H^* be the dual space of H (Notation 14.14), i.e. $f \in H^*$ iff $f : H \to \mathbb{F}$ is linear and continuous. The map*

$$z \in H \xrightarrow{j} \langle \cdot|z\rangle \in H^* \tag{18.8}$$

is a conjugate linear[1] isometric isomorphism, where for $f \in H^*$ we let,

$$\|f\|_{H^*} := \sup_{x\in H\setminus\{0\}} \frac{|f(x)|}{\|x\|} = \sup_{\|x\|=1} |f(x)|.$$

[1] Recall that j is conjugate linear if

$$j(z_1 + \alpha z_2) = jz_1 + \bar{\alpha} jz_2$$

for all $z_1, z_2 \in H$ and $\alpha \in \mathbb{C}$.

Proof. Let $f \in H^*$ and $M = \mathrm{Nul}(f)$ – a closed proper subspace of H since f is continuous. If $f = 0$, then clearly $f(\cdot) = \langle \cdot|0\rangle$. If $f \neq 0$ there exists $y \in H \setminus M$. Then for any $\alpha \in \mathbb{C}$ we have $e := \alpha(y - P_M y) \in M^\perp$. We now choose α so that $f(e) = 1$. Hence if $x \in H$,

$$f(x - f(x) e) = f(x) - f(x) f(e) = f(x) - f(x) = 0,$$

which shows $x - f(x) e \in M$. As $e \in M^\perp$ it follows that

$$0 = \langle x - f(x) e|e\rangle = \langle x|e\rangle - f(x) \|e\|^2$$

which shows $f(\cdot) = \langle \cdot|z\rangle = jz$ where $z := e/\|e\|^2$ and thus j is surjective.

The map j is conjugate linear by the axioms of the inner products. Moreover, for $x, z \in H$,

$$|\langle x|z\rangle| \leq \|x\| \|z\| \text{ for all } x \in H$$

with equality when $x = z$. This implies that $\|jz\|_{H^*} = \|\langle \cdot|z\rangle\|_{H^*} = \|z\|$. Therefore j is isometric and this implies j is injective.

■

Proposition 18.18 (Adjoints). *Let H and K be Hilbert spaces and $A : H \to K$ be a bounded operator. Then there exists a unique bounded operator $A^* : K \to H$ such that*

$$\langle Ax|y\rangle_K = \langle x|A^* y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{18.9}$$

Moreover, for all $A, B \in L(H, K)$ and $\lambda \in \mathbb{C}$,

1. *$(A + \lambda B)^* = A^* + \bar{\lambda} B^*$,*
2. *$A^{**} := (A^*)^* = A$,*
3. *$\|A^*\| = \|A\|$ and*
4. *$\|A^* A\| = \|A\|^2$.*
5. *If $K = H$, then $(AB)^* = B^* A^*$. In particular $A \in L(H)$ has a bounded inverse iff A^* has a bounded inverse and $(A^*)^{-1} = (A^{-1})^*$.*

Proof. For each $y \in K$, the map $x \to \langle Ax|y\rangle_K$ is in H^* and therefore there exists, by Theorem 18.17, a unique vector $z \in H$ (we will denote this z by $A^*(y)$) such that

$$\langle Ax|y\rangle_K = \langle x|z\rangle_H \text{ for all } x \in H.$$

This shows there is a unique map $A^* : K \to H$ such that $\langle Ax|y\rangle_K = \langle x|A^*(y)\rangle_H$ for all $x \in H$ and $y \in K$.

To see A^* is linear, let $y_1, y_2 \in K$ and $\lambda \in \mathbb{C}$, then for any $x \in H$,

$$\begin{aligned}\langle Ax|y_1 + \lambda y_2\rangle_K &= \langle Ax|y_1\rangle_K + \bar{\lambda}\langle Ax|y_2\rangle_K \\ &= \langle x|A^*(y_1)\rangle_K + \bar{\lambda}\langle x|A^*(y_2)\rangle_H \\ &= \langle x|A^*(y_1) + \lambda A^*(y_2)\rangle_H\end{aligned}$$

and by the uniqueness of $A^*(y_1+\lambda y_2)$ we find

$$A^*(y_1+\lambda y_2)=A^*(y_1)+\lambda A^*(y_2).$$

This shows A^* is linear and so we will now write A^*y instead of $A^*(y)$.

Since

$$\langle A^*y|x\rangle_H=\overline{\langle x|A^*y\rangle_H}=\overline{\langle Ax|y\rangle_K}=\langle y|Ax\rangle_K$$

it follows that $A^{**}=A$. The assertion that $(A+\lambda B)^*=A^*+\bar\lambda B^*$ is Exercise 18.2.

Items 3. and 4. Making use of Schwarz's inequality (Theorem 18.2), we have

$$\begin{aligned}\|A^*\| &= \sup_{k\in K:\|k\|=1}\|A^*k\| \\ &= \sup_{k\in K:\|k\|=1}\sup_{h\in H:\|h\|=1}|\langle A^*k|h\rangle| \\ &= \sup_{h\in H:\|h\|=1}\sup_{k\in K:\|k\|=1}|\langle k|Ah\rangle| = \sup_{h\in H:\|h\|=1}\|Ah\|=\|A\|\end{aligned}$$

so that $\|A^*\|=\|A\|$. Since

$$\|A^*A\|\le\|A^*\|\,\|A\|=\|A\|^2$$

and

$$\begin{aligned}\|A\|^2 &= \sup_{h\in H:\|h\|=1}\|Ah\|^2=\sup_{h\in H:\|h\|=1}|\langle Ah|Ah\rangle| \\ &= \sup_{h\in H:\|h\|=1}|\langle h|A^*Ah\rangle|\le\sup_{h\in H:\|h\|=1}\|A^*Ah\|=\|A^*A\| \qquad (18.10)\end{aligned}$$

we also have $\|A^*A\|\le\|A\|^2\le\|A^*A\|$ which shows $\|A\|^2=\|A^*A\|$.

Alternatively, from Eq. (18.10),

$$\|A\|^2\le\|A^*A\|\le\|A\|\,\|A^*\| \qquad (18.11)$$

which then implies $\|A\|\le\|A^*\|$. Replacing A by A^* in this last inequality shows $\|A^*\|\le\|A\|$ and hence that $\|A^*\|=\|A\|$. Using this identity back in Eq. (18.11) proves $\|A\|^2=\|A^*A\|$.

Now suppose that $K=H$. Then

$$\langle ABh|k\rangle=\langle Bh|A^*k\rangle=\langle h|B^*A^*k\rangle$$

which shows $(AB)^*=B^*A^*$. If A^{-1} exists then

$$\begin{aligned}\left(A^{-1}\right)^*A^* &= \left(AA^{-1}\right)^*=I^*=I \text{ and} \\ A^*\left(A^{-1}\right)^* &= \left(A^{-1}A\right)^*=I^*=I.\end{aligned}$$

This shows that A^* is invertible and $(A^*)^{-1}=\left(A^{-1}\right)^*$. Similarly if A^* is invertible then so is $A=A^{**}$. ■

Exercise 18.2. Let H,K,M be Hilbert spaces, $A,B\in L(H,K)$, $C\in L(K,M)$ and $\lambda\in\mathbb{C}$. Show $(A+\lambda B)^*=A^*+\bar\lambda B^*$ and $(CA)^*=A^*C^*\in L(M,H)$.

Exercise 18.3. Let $H=\mathbb{C}^n$ and $K=\mathbb{C}^m$ equipped with the usual inner products, i.e. $\langle z|w\rangle_H=z\cdot\bar w$ for $z,w\in H$. Let A be an $m\times n$ matrix thought of as a linear operator from H to K. Show the matrix associated to $A^*:K\to H$ is the conjugate transpose of A.

Lemma 18.19. *Suppose $A:H\to K$ is a bounded operator, then:*

1. $\mathrm{Nul}(A^*)=\mathrm{Ran}(A)^\perp$.
2. $\overline{\mathrm{Ran}(A)}=\mathrm{Nul}(A^*)^\perp$.
3. *if $K=H$ and $V\subset H$ is an A – invariant subspace (i.e. $A(V)\subset V$), then $V^\perp$ is A^* – invariant.*

Proof. An element $y\in K$ is in $\mathrm{Nul}(A^*)$ iff $0=\langle A^*y|x\rangle=\langle y|Ax\rangle$ for all $x\in H$ which happens iff $y\in\mathrm{Ran}(A)^\perp$. Because, by Exercise 18.1, $\overline{\mathrm{Ran}(A)}=\mathrm{Ran}(A)^{\perp\perp}$, and so by the first item, $\overline{\mathrm{Ran}(A)}=\mathrm{Nul}(A^*)^\perp$. Now suppose $A(V)\subset V$ and $y\in V^\perp$, then

$$\langle A^*y|x\rangle=\langle y|Ax\rangle=0 \text{ for all } x\in V$$

which shows $A^*y\in V^\perp$. ■

Definition 18.20 (Strong Convergence). *Let X be a Banach space. We say a sequence of operators $\{A_n\}_{n=1}^\infty\subset L(X)$ **converges strongly** to $A\in L(X)$ if $\lim_{n\to\infty}A_nx=Ax$ for all $x\in X$. We abbreviate this by writing $A_n\overset{s}{\to}A$ as $n\to\infty$. [**Note well** that strong convergence is weaker than norm convergence, i.e. if $\|A-A_n\|_{op}\to 0$ then $A_n\overset{s}{\to}A$ as $n\to\infty$ but no necessarily the other way around unless* $\dim X<\infty$.]

Exercise 18.4. Let $(H,\langle\cdot|\cdot\rangle)$ be a Hilbert space and suppose that $\{P_n\}_{n=1}^\infty$ is a sequence of orthogonal projection operators on H such that $P_n(H)\subset P_{n+1}(H)$ for all n. Let $M:=\cup_{n=1}^\infty P_n(H)$ (a subspace of H) and let P denote orthonormal projection onto $\bar M$. Show $\lim_{n\to\infty}P_nx=Px$ for all $x\in H$. **Hint:** first prove the result for $x\in M^\perp$, then for $x\in M$ and then for $x\in\bar M$.

Exercise 18.5. Let $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ be σ – finite measure spaces, $f\in L^2(\nu)$ and $k\in L^2(\mu\otimes\nu)$. Show

$$\int|k(x,y)f(y)|\,d\nu(y)<\infty \text{ for } \mu\text{ – a.e. } x \qquad (18.12)$$

and define

$$Kf(x):=\int_Y k(x,y)f(y)d\nu(y) \qquad (18.13)$$

when the integral is defined and set $Kf(x) = 0$ otherwise. Show $Kf \in L^2(\mu)$ and $K : L^2(\nu) \to L^2(\mu)$ is a bounded operator with $\|K\|_{op} \le \|k\|_{L^2(\mu\otimes\nu)}$. [Remember that $L^2(\nu)$ and $L^2(\mu)$ consists of equivalence classes of functions.]

Exercise 18.6. Let $K : L^2(\nu) \to L^2(\mu)$ be the operator defined in Exercise 18.5. Show $K^* : L^2(\mu) \to L^2(\nu)$ is the operator given by

$$K^* g(y) = \int_X \bar{k}(x,y) g(x) d\mu(x).$$

18.1 Hilbert Space Basis

Proposition 18.21 (Bessel's Inequality). *Let T be an orthonormal set, then for any $x \in H$,*

$$\sum_{v\in T} |\langle x|v\rangle|^2 \le \|x\|^2 \text{ for all } x \in H. \tag{18.14}$$

In particular the set $T_x := \{v \in T : \langle x|v\rangle \ne 0\}$ is at most countable for all $x \in H$.

Proof. Let $\Gamma \subset\subset T$ be any finite set and $M = \operatorname{span}\Gamma$. From Lemma 18.15 we know that $P_M x = \sum_{v\in\Gamma} \langle x|v\rangle v$. As $\|P_M\|_{op} = 1$ it now follows that

$$\sum_{v\in\Gamma} |\langle x|v\rangle|^2 = \|P_M x\|^2 \le \|x\|^2.$$

Taking the supremum of this inequality over $\Gamma \subset\subset T$ then proves Eq. (18.14). ■

Proposition 18.22. *Suppose $T \subset H$ is an orthogonal set. Then $s = \sum_{v\in T} v$ exists in H (see Definition 14.22) iff $\sum_{v\in T} \|v\|^2 < \infty$. (In particular T must be at most a countable set.) Moreover, if $\sum_{v\in T} \|v\|^2 < \infty$, then*

1. *$\|s\|^2 = \sum_{v\in T} \|v\|^2$ and*
2. *$\langle s|x\rangle = \sum_{v\in T} \langle v|x\rangle$ for all $x \in H$.*

Similarly if $\{v_n\}_{n=1}^\infty$ is an orthogonal set, then $s = \sum_{n=1}^\infty v_n$ exists in H iff $\sum_{n=1}^\infty \|v_n\|^2 < \infty$. In particular if $\sum_{n=1}^\infty v_n$ exists, then it is independent of rearrangements of $\{v_n\}_{n=1}^\infty$.

Proof. Suppose $s = \sum_{v\in T} v$ exists. Then there exists $\Gamma \subset\subset T$ such that

$$\sum_{v\in\Lambda} \|v\|^2 = \left\|\sum_{v\in\Lambda} v\right\|^2 \le 1$$

for all $\Lambda \subset\subset T \setminus \Gamma$, wherein the first inequality we have used Pythagorean's theorem. Taking the supremum over such Λ shows that $\sum_{v\in T\setminus\Gamma} \|v\|^2 \le 1$ and therefore

$$\sum_{v\in T} \|v\|^2 \le 1 + \sum_{v\in\Gamma} \|v\|^2 < \infty.$$

Conversely, suppose that $\sum_{v\in T} \|v\|^2 < \infty$. Then for all $\varepsilon > 0$ there exists $\Gamma_\varepsilon \subset\subset T$ such that if $\Lambda \subset\subset T \setminus \Gamma_\varepsilon$,

$$\left\|\sum_{v\in\Lambda} v\right\|^2 = \sum_{v\in\Lambda} \|v\|^2 < \varepsilon^2. \tag{18.15}$$

Hence by Lemma 14.23, $\sum_{v\in T} v$ exists.

For item 1, let Γ_ε be as above and set $s_\varepsilon := \sum_{v\in\Gamma_\varepsilon} v$. Then

$$|\|s\| - \|s_\varepsilon\|| \le \|s - s_\varepsilon\| < \varepsilon$$

and by Eq. (18.15),

$$0 \le \sum_{v\in T} \|v\|^2 - \|s_\varepsilon\|^2 = \sum_{v\notin\Gamma_\varepsilon} \|v\|^2 \le \varepsilon^2.$$

Letting $\varepsilon \downarrow 0$ we deduce from the previous two equations that $\|s_\varepsilon\| \to \|s\|$ and $\|s_\varepsilon\|^2 \to \sum_{v\in T} \|v\|^2$ as $\varepsilon \downarrow 0$ and therefore $\|s\|^2 = \sum_{v\in T} \|v\|^2$. Item 2. is a special case of Lemma 14.23.

Alternative proof of item 1. We could use the last result to prove Item 1. Indeed, if $\sum_{v\in T} \|v\|^2 < \infty$, then T is countable and so we may write $T = \{v_n\}_{n=1}^\infty$. Then $s = \lim_{N\to\infty} s_N$ with s_N as above. Since the norm, $\|\cdot\|$, is continuous on H,

$$\|s\|^2 = \lim_{N\to\infty} \|s_N\|^2 = \lim_{N\to\infty} \left\|\sum_{n=1}^N v_n\right\|^2 = \lim_{N\to\infty} \sum_{n=1}^N \|v_n\|^2$$
$$= \sum_{n=1}^\infty \|v_n\|^2 = \sum_{v\in T} \|v\|^2.$$

Countable Case. Let $s_N := \sum_{n=1}^N v_n$ and suppose that $\lim_{N\to\infty} s_N = s$ exists in H. Since $\{s_N\}_{N=1}^\infty$ is Cauchy, we have (say $N > M$) that

$$\sum_{n=M+1}^{N} \|v_n\|^2 = \|s_N - s_M\|^2 \to 0 \text{ as } M, N \to \infty$$

which shows that $\sum_{n=1}^{\infty} \|v_n\|^2$ is convergent, i.e. $\sum_{n=1}^{\infty} \|v_n\|^2 < \infty$. Conversely if $\sum_{n=1}^{\infty} \|v_n\|^2 < \infty$, then

$$\|s_N - s_M\|^2 = \sum_{n=M+1}^{N} \|v_n\|^2 \to 0 \text{ as } M, N \to \infty$$

which shows $\{s_N\}_{N=1}^{\infty}$ is Cauchy and hence convergent. Finally, suppose that $\varphi : \mathbb{N} \to \mathbb{N}$ is a bijective map and let $s_N^{\varphi} := \sum_{n=1}^{N} v_{\varphi(n)}$. If we let

$$\Gamma_N = \{1, 2, \dots, N\} \triangle \{\varphi(1), \varphi(2), \dots, \varphi(N)\},$$

then $\min \Gamma_N \to \infty$ as $N \to \infty$ as φ is a bijection. Therefore

$$\|s_N - s_N^{\varphi}\|^2 = \sum_{n \in \Gamma_N} \|v_n\|^2 \le \sum_{n=\min \Gamma_N}^{\infty} \|v_n\|^2 \to 0 \text{ as } N \to \infty.$$

which shows the limit is independent of rearrangements. ∎

Corollary 18.23. *Suppose H is a Hilbert space, $\beta \subset H$ is an orthonormal set and $M = \overline{\operatorname{span} \beta}$. Then*

$$P_M x = \sum_{u \in \beta} \langle x|u \rangle u, \tag{18.16}$$

$$\sum_{u \in \beta} |\langle x|u \rangle|^2 = \|P_M x\|^2 \text{ and} \tag{18.17}$$

$$\sum_{u \in \beta} \langle x|u \rangle \langle u|y \rangle = \langle P_M x|y \rangle \tag{18.18}$$

for all $x, y \in H$.

Proof. By Bessel's inequality, $\sum_{u \in \beta} |\langle x|u \rangle|^2 \le \|x\|^2$ for all $x \in H$ and hence by Proposition 18.21, $Px := \sum_{u \in \beta} \langle x|u \rangle u$ exists in H and for all $x, y \in H$,

$$\langle Px|y \rangle = \sum_{u \in \beta} \langle \langle x|u \rangle u|y \rangle = \sum_{u \in \beta} \langle x|u \rangle \langle u|y \rangle. \tag{18.19}$$

Taking $y \in \beta$ in Eq. (18.19) gives $\langle Px|y \rangle = \langle x|y \rangle$, i.e. that $\langle x - Px|y \rangle = 0$ for all $y \in \beta$. So $(x - Px) \perp \operatorname{span} \beta$ and by continuity we also have $(x - Px) \perp M = \overline{\operatorname{span} \beta}$. Since Px is also in M, it follows from the definition of P_M that $Px = P_M x$ proving Eq. (18.16). Equations (18.17) and (18.18) now follow from (18.19), Proposition 18.22 and the fact that $\langle P_M x|y \rangle = \langle P_M^2 x|y \rangle = \langle P_M x|P_M y \rangle$ for all $x, y \in H$. ∎

Definition 18.24 (Basis). *Let H be a Hilbert space. A **basis** β of H is a maximal orthonormal subset $\beta \subset H$.*

Proposition 18.25. *Every Hilbert space has an orthonormal basis.*

Proof. Let $\mathcal{F}$ be the collection of all orthonormal subsets of H ordered by inclusion. If $\Phi \subset \mathcal{F}$ is linearly ordered then $\cup \Phi$ is an upper bound. By Zorn's Lemma (see Theorem 2.14) there exists a maximal element $\beta \in \mathcal{F}$. [See Proposition 18.30 below for a more down to earth proof in the case that H is separable.] ∎

An orthonormal set $\beta \subset H$ is said to be **complete** if $\beta^{\perp} = \{0\}$. That is to say if $\langle x|u \rangle = 0$ for all $u \in \beta$ then $x = 0$.

Lemma 18.26. *Let β be an orthonormal subset of H then the following are equivalent:*

1. *β is a basis,*
2. *β is complete and*
3. *$\overline{\operatorname{span} \beta} = H$.*

Proof. (1. $\iff$ 2.) If β is not complete, then there exists a unit vector $x \in \beta^{\perp} \setminus \{0\}$. The set $\beta \cup \{x\}$ is an orthonormal set properly containing β, so β is not maximal. Conversely, if β is not maximal, there exists an orthonormal set $\beta_1 \subset H$ such that $\beta \subsetneq \beta_1$. Then if $x \in \beta_1 \setminus \beta$, we have $\langle x|u \rangle = 0$ for all $u \in \beta$ showing β is not complete.

(2. $\iff$ 3.) If β is not complete and $x \in \beta^{\perp} \setminus \{0\}$, then $\overline{\operatorname{span} \beta} \subset x^{\perp}$ which is a proper subspace of H. Conversely if $\overline{\operatorname{span} \beta}$ is a proper subspace of H, $\beta^{\perp} = \overline{\operatorname{span} \beta}^{\perp}$ is a non-trivial subspace by Corollary 18.16 and β is not complete. ∎

Theorem 18.27. *Let $\beta \subset H$ be an orthonormal set. Then the following are equivalent:*

1. *β is complete, i.e. β is an orthonormal basis for H.*
2. *$x = \sum_{u \in \beta} \langle x|u \rangle u$ for all $x \in H$.*
3. *$\langle x|y \rangle = \sum_{u \in \beta} \langle x|u \rangle \langle u|y \rangle$ for all $x, y \in H$.*
4. *$\|x\|^2 = \sum_{u \in \beta} |\langle x|u \rangle|^2$ for all $x \in H$.*

Proof. Let $M = \overline{\operatorname{span}\,\beta}$ and $P = P_M$.
(1) ⇒ (2) By assumption $M = H$ and by Corollary 18.23,

$$\sum_{u\in\beta} \langle x|u\rangle u = P_M x = P_H x = x.$$

(2) ⇒ (3) is a consequence of Proposition 18.22, i.e. by the continuity of the inner product.
(3) ⇒ (4) is obvious, just take $y = x$.
(4) ⇒ (1) If $x \in \beta^\perp$, then by 4), $\|x\| = 0$, i.e. $x = 0$. This shows that β is complete. ■

Suppose $\Gamma := \{u_n\}_{n=1}^\infty$ is a collection of vectors in an inner product space $(H, \langle\cdot|\cdot\rangle)$. The standard **Gram-Schmidt** process produces from Γ an orthonormal subset, $\beta = \{v_n\}_{n=1}^\infty$, such that every element $u_n \in \Gamma$ is a finite linear combination of elements from β. Recall the procedure is to define v_n inductively by setting

$$\tilde{v}_{n+1} := v_{n+1} - \sum_{j=1}^{n} \langle u_{n+1}|v_j\rangle v_j = v_{n+1} - P_n v_{n+1}$$

where P_n is orthogonal projection onto $M_n := \operatorname{span}(\{v_k\}_{k=1}^n)$. If $v_{n+1} := 0$, let $\tilde{v}_{n+1} = 0$, otherwise set $v_{n+1} := \|\tilde{v}_{n+1}\|^{-1}\tilde{v}_{n+1}$. Finally re-index the resulting sequence so as to throw out those v_n with $v_n = 0$. The result is an orthonormal subset, $\beta \subset H$, with the desired properties.

Definition 18.28. *A subset, Γ, of a normed space X is said to be **total** if $\operatorname{span}(\Gamma)$ is dense in X.*

Remark 18.29. Suppose that $\{u_n\}_{n=1}^\infty$ is a **total** subset of H. Let $\{v_n\}_{n=1}^\infty$ be the vectors found by performing Gram-Schmidt on the set $\{u_n\}_{n=1}^\infty$. Then $\beta := \{v_n\}_{n=1}^\infty$ is an orthonormal basis for H. Indeed, if $h \in H$ is orthogonal to β then h is orthogonal to $\{u_n\}_{n=1}^\infty$ and hence also $\overline{\operatorname{span}\{u_n\}_{n=1}^\infty} = H$. In particular h is orthogonal to itself and so $h = 0$. This generalizes the corresponding results for finite dimensional inner product spaces.

Proposition 18.30. *A Hilbert space H is separable (see Definition 13.10 or Definition 35.42) iff H has a countable orthonormal basis $\beta \subset H$. Moreover, if H is separable, all orthonormal bases of H are countable. (See Proposition 4.14 in Conway's, "A Course in Functional Analysis," for a more general version of this proposition.)*

Proof. Let $\mathbb{D} \subset H$ be a countable dense set $\mathbb{D} = \{u_n\}_{n=1}^\infty$. By Gram-Schmidt process there exists $\beta = \{v_n\}_{n=1}^\infty$ an orthonormal set such that $\operatorname{span}\{v_n : n = 1, 2 \dots, N\} \supseteq \operatorname{span}\{u_n : n = 1, 2 \dots, N\}$. So if $\langle x|v_n\rangle = 0$ for all n then $\langle x|u_n\rangle = 0$ for all n. Since $\mathbb{D} \subset H$ is dense we may choose $\{w_k\} \subset \mathbb{D}$ such that $x = \lim_{k\to\infty} w_k$ and therefore $\langle x|x\rangle = \lim_{k\to\infty}\langle x|w_k\rangle = 0$. That is to say $x = 0$ and β is complete. Conversely if $\beta \subset H$ is a countable orthonormal basis, then the countable set

$$\mathbb{D} = \left\{ \sum_{u\in\beta} a_u u : a_u \in \mathbb{Q} + i\mathbb{Q} : \#\{u : a_u \neq 0\} < \infty \right\}$$

is dense in H. Finally let $\beta = \{u_n\}_{n=1}^\infty$ be an orthonormal basis and $\beta_1 \subset H$ be another orthonormal basis. Then the sets

$$B_n = \{v \in \beta_1 : \langle v|u_n\rangle \neq 0\}$$

are countable for each $n \in \mathbb{N}$ and hence $B := \bigcup_{n=1}^\infty B_n$ is a countable subset of β_1. Suppose there exists $v \in \beta_1 \setminus B$, then $\langle v|u_n\rangle = 0$ for all n and since $\beta = \{u_n\}_{n=1}^\infty$ is an orthonormal basis, this implies $v = 0$ which is impossible since $\|v\| = 1$. Therefore $\beta_1 \setminus B = \emptyset$ and hence $\beta_1 = B$ is countable. ■

Notation 18.31 *If $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ are two functions, let $f \otimes g : X \times Y \to \mathbb{C}$ be defined by $f \otimes g\,(x, y) := f(x)\,g(y)$.*

Proposition 18.32. *Suppose X and Y are sets and $\mu : X \to (0, \infty)$ and $\nu : Y \to (0, \infty)$ are given weight functions. If $\beta \subset \ell^2(\mu)$ and $\gamma \subset \ell^2(\nu)$ are orthonormal bases, then*

$$\beta \otimes \gamma := \{f \otimes g : f \in \beta \text{ and } g \in \gamma\}$$

is an orthonormal basis for $\ell^2(\mu \otimes \nu)$.

Proof. Let $f, f' \in \ell^2(\mu)$ and $g, g' \in \ell^2(\nu)$, then by the Tonelli's Theorem 4.22 for sums and Hölder's inequality,

$$\begin{aligned}\sum_{X\times Y} \left| f \otimes g \cdot \overline{f' \otimes g'} \right| \mu \otimes \nu &= \sum_X \left| f\bar{f'} \right| \mu \cdot \sum_Y \left| g\bar{g'} \right| \nu \\ &\le \|f\|_{\ell^2(\mu)} \|f'\|_{\ell^2(\mu)} \|g\|_{\ell^2(\nu)} \|g'\|_{\ell^2(\nu)} = 1 < \infty.\end{aligned}$$

So by Fubini's Theorem 4.23 for sums,

$$\begin{aligned}\langle f \otimes g|f' \otimes g'\rangle_{\ell^2(\mu\otimes\nu)} &= \sum_X f\ \bar{f}'\mu \cdot \sum_Y g\bar{g}'\nu \\ &= \langle f|f'\rangle_{\ell^2(\mu)} \langle g|g'\rangle_{\ell^2(\nu)} = \delta_{f,f'}\delta_{g,g'}.\end{aligned}$$

Therefore, $\beta \otimes \gamma$ is an orthonormal subset of $\ell^2(\mu \otimes \nu)$. So it only remains to show $\beta \otimes \gamma$ is complete. We will give two proofs of this fact. Let $F \in \ell^2(\mu \otimes \nu)$.

In the first proof we will verify item 4. of Theorem 18.27 while in the second we will verify item 1 of Theorem 18.27.

First Proof. By Tonelli's Theorem,

$$\sum_{x\in X}\mu(x)\sum_{y\in Y}\nu(y)\left|F(x,y)\right|^2=\|F\|^2_{\ell^2(\mu\otimes\nu)}<\infty$$

and since $\mu>0$, it follows that

$$\sum_{y\in Y}\left|F(x,y)\right|^2\nu(y)<\infty \text{ for all } x\in X,$$

i.e. $F(x,\cdot)\in\ell^2(\nu)$ for all $x\in X$. By the completeness of γ,

$$\sum_{y\in Y}\left|F(x,y)\right|^2\nu(y)=\langle F(x,\cdot)|F(x,\cdot)\rangle_{\ell^2(\nu)}=\sum_{g\in\gamma}\left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2$$

and therefore,

$$\begin{aligned}\|F\|^2_{\ell^2(\mu\otimes\nu)}&=\sum_{x\in X}\mu(x)\sum_{y\in Y}\nu(y)\left|F(x,y)\right|^2\\&=\sum_{x\in X}\sum_{g\in\gamma}\left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2\mu(x).\end{aligned}\tag{18.20}$$

and in particular, $x\to\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}$ is in $\ell^2(\mu)$. So by the completeness of β and the Fubini and Tonelli theorems, we find

$$\begin{aligned}\sum_X\left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2\mu(x)&=\sum_{f\in\beta}\left|\sum_{x\in X}\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\bar f(x)\mu(x)\right|^2\\&=\sum_{f\in\beta}\left|\sum_{x\in X}\left(\sum_{y\in Y}F(x,y)\bar g(y)\nu(y)\right)\bar f(x)\mu(x)\right|^2\\&=\sum_{f\in\beta}\left|\sum_{(x,y)\in X\times Y}F(x,y)\overline{f\otimes g}(x,y)\,\mu\otimes\nu(x,y)\right|^2\\&=\sum_{f\in\beta}\left|\langle F|f\otimes g\rangle_{\ell^2(\mu\otimes\nu)}\right|^2.\end{aligned}$$

Combining this result with Eq. (18.20) shows

$$\|F\|^2_{\ell^2(\mu\otimes\nu)}=\sum_{f\in\beta,\ g\in\gamma}\left|\langle F|f\otimes g\rangle_{\ell^2(\mu\otimes\nu)}\right|^2$$

as desired.

Second Proof. Suppose, for all $f\in\beta$ and $g\in\gamma$ that $\langle F|f\otimes g\rangle=0$, i.e.

$$\begin{aligned}0=\langle F|f\otimes g\rangle_{\ell^2(\mu\otimes\nu)}&=\sum_{x\in X}\mu(x)\sum_{y\in Y}\nu(y)F(x,y)\bar f(x)\bar g(y)\\&=\sum_{x\in X}\mu(x)\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\bar f(x).\end{aligned}\tag{18.21}$$

Since

$$\sum_{x\in X}\left|\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}\right|^2\mu(x)\le\sum_{x\in X}\mu(x)\sum_{y\in Y}\left|F(x,y)\right|^2\nu(y)<\infty,\tag{18.22}$$

it follows from Eq. (18.21) and the completeness of β that $\langle F(x,\cdot)|g\rangle_{\ell^2(\nu)}=0$ for all $x\in X$. By the completeness of γ we conclude that $F(x,y)=0$ for all $(x,y)\in X\times Y$. ■

Definition 18.33. *A linear map $U:H\to K$ is an* ***isometry*** *if $\|Ux\|_K=\|x\|_H$ for all $x\in H$ and U is* ***unitary*** *if U is also surjective.*

Exercise 18.7. Let $U:H\to K$ be a linear map between complex Hilbert spaces. Show the following are equivalent:

1. $U:H\to K$ is an isometry,
2. $\langle Ux|Ux'\rangle_K=\langle x|x'\rangle_H$ for all $x,x'\in H$,
3. $U^*U=id_H$.

Hint: use the polarization identities in Remark 18.4.,

Exercise 18.8. Let $U:H\to K$ be a linear map, show the following are equivalent:

1. $U:H\to K$ is unitary
2. $U^*U=id_H$ and $UU^*=id_K$.
3. U is invertible and $U^{-1}=U^*$.

Exercise 18.9. Let H be a Hilbert space. Use Theorem 18.27 to show there exists a set X and a unitary map $U:H\to\ell^2(X)$. Moreover, if H is separable and $\dim(H)=\infty$, then X can be taken to be $\mathbb{N}$ so that H is unitarily equivalent to $\ell^2=\ell^2(\mathbb{N})$.

18.2 $\mathbf{L^2}$-Orthonormal Basis Examples

Example 18.34. 1. Let $H = L^2([-1,1], dm)$, $A := \{1, x, x^2, x^3 \dots\}$ and $\beta \subset H$ be the result of doing the Gram-Schmidt procedure on A. By the Stone-Weierstrass theorem or by Exercise 16.3 directly, A is total in H. Hence by Remark 18.29, β is an orthonormal basis for H. The basis, β, consists of polynomials which up to normalization are the so called "**Legendre polynomials.**"
2. Let $H = L^2(\mathbb{R}, e^{-\frac{1}{2}x^2}dx)$ and $A := \{1, x, x^2, x^3 \dots\}$. Again by Exercise 16.3, A is total in H and hence the Gram-Schmidt procedure applied to A produces an orthonormal basis, β, of polynomial functions for H. This basis consists, up to normalizations, of the so called "**Hermite polynomials**" on $\mathbb{R}$.

Remark 18.35 (An Interesting Phenomena). Let $H = L^2([-1,1], dm)$ and $B := \{1, x^3, x^6, x^9, \dots\}$. Then again B is total in H by the same argument as in item 2. Example 18.34. This is true even though B is a proper subset of A. Notice that A is an algebraic basis for the polynomials on $[-1,1]$ while B is not! The following computations may help relieve some of the reader's anxiety. Let $f \in L^2([-1,1], dm)$, then, making the change of variables $x = y^{1/3}$, shows that

$$\int_{-1}^{1} |f(x)|^2 dx = \int_{-1}^{1} \left|f(y^{1/3})\right|^2 \frac{1}{3} y^{-2/3} dy = \int_{-1}^{1} \left|f(y^{1/3})\right|^2 d\mu(y) \qquad (18.23)$$

where $d\mu(y) = \frac{1}{3}y^{-2/3}dy$. Since $\mu([-1,1]) = m([-1,1]) = 2$, μ is a finite measure on $[-1,1]$ and hence by Exercise 16.3 $A := \{1, x, x^2, x^3 \dots\}$ is total (see Definition 18.28) in $L^2([-1,1], d\mu)$. In particular for any $\varepsilon > 0$ there exists a polynomial $p(y)$ such that

$$\int_{-1}^{1} \left|f(y^{1/3}) - p(y)\right|^2 d\mu(y) < \varepsilon^2.$$

However, by Eq. (18.23) we have

$$\varepsilon^2 > \int_{-1}^{1} \left|f(y^{1/3}) - p(y)\right|^2 d\mu(y) = \int_{-1}^{1} \left|f(x) - p(x^3)\right|^2 dx.$$

Alternatively, if $f \in C([-1,1])$, then $g(y) = f(y^{1/3})$ is back in $C([-1,1])$. Therefore for any $\varepsilon > 0$, there exists a polynomial $p(y)$ such that

$$\begin{aligned} \varepsilon > \|g - p\|_\infty &= \sup\{|g(y) - p(y)| : y \in [-1,1]\} \\ &= \sup\left\{\left|g(x^3) - p(x^3)\right| : x \in [-1,1]\right\} \\ &= \sup\left\{\left|f(x) - p(x^3)\right| : x \in [-1,1]\right\}. \end{aligned}$$

This gives another proof the polynomials in x^3 are dense in $C([-1,1])$ and hence in $L^2([-1,1])$.

Exercise 18.10. Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces such that $L^2(\mu)$ and $L^2(\nu)$ are separable. If $\{f_n\}_{n=1}^\infty$ and $\{g_m\}_{m=1}^\infty$ are orthonormal bases for $L^2(\mu)$ and $L^2(\nu)$ respectively, then $\beta := \{f_n \otimes g_m : m, n \in \mathbb{N}\}$ is an orthonormal basis for $L^2(\mu \otimes \nu)$. (Recall that $f \otimes g(x,y) := f(x) g(y)$, see Notation **??**.) **Hint:** model your proof on the proof of Proposition 18.32.

Definition 18.36 (External direct sum of Hilbert spaces). *Suppose that $\{H_n\}_{n=1}^\infty$ is a sequence of Hilbert spaces. Let $\oplus_{n=1}^\infty H_n$ denote the space of sequences, $f \in \prod_{n=1}^\infty H_n$ such that*

$$\|f\| = \sqrt{\sum_{n=1}^\infty \|f(n)\|_{H_n}^2} < \infty.$$

It is easily seen that $(\oplus_{n=1}^\infty H_n, \|\cdot\|)$ is a Hilbert space with inner product defined, for all $f, g \in \oplus_{n=1}^\infty H_n$, by

$$\langle f|g\rangle_{\oplus_{n=1}^\infty H_n} = \sum_{n=1}^\infty \langle f(n)|g(n)\rangle_{H_n}.$$

Exercise 18.11. Suppose H is a Hilbert space and $\{H_n : n \in \mathbb{N}\}$ are closed subspaces of H such that $H_n \perp H_m$ for all $m \neq n$ and if $f \in H$ with $f \perp H_n$ for all $n \in \mathbb{N}$, then $f = 0$. For $f \in \oplus_{n=1}^\infty H_n$, show the sum $\sum_{n=1}^\infty f(n)$ is convergent in H and the map $U : \oplus_{n=1}^\infty H_n \to H$ defined by $Uf := \sum_{n=1}^\infty f(n)$ is unitary.

Exercise 18.12. Suppose $(X, \mathcal{M}, \mu)$ is a measure space and $X = \coprod_{n=1}^\infty X_n$ with $X_n \in \mathcal{M}$ and $\mu(X_n) > 0$ for all n. Then $U : L^2(X, \mu) \to \oplus_{n=1}^\infty L^2(X_n, \mu)$ defined by $(Uf)(n) := f 1_{X_n}$ is unitary.

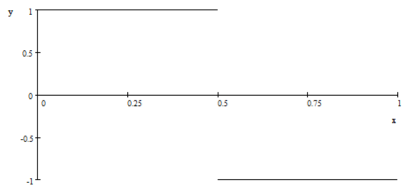
Exercise 18.13 (Haar Basis). In this problem, let L^2 denote $L^2([0,1], m)$ with the standard inner product,

$$\psi(x) = 1_{[0,1/2)}(x) - 1_{[1/2,1)}(x)$$

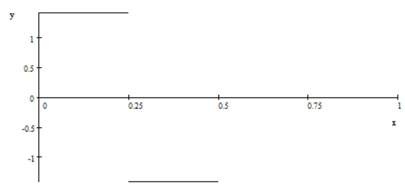
and for $k, j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ with $0 \le j < 2^k$ let

$$\begin{aligned} \psi_{kj}(x) &= 2^{k/2}\psi(2^k x - j) \\ &= 2^{k/2}\left(1_{2^{-k}[j,j+1/2)}(x) - 1_{2^{-k}[j+1/2,j+1)}(x)\right). \end{aligned}$$

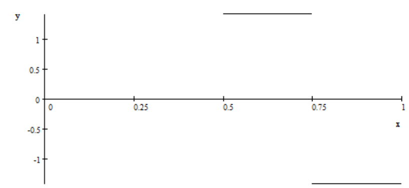
The following pictures shows the graphs of $\psi_{0,0}$, $\psi_{1,0}, \psi_{1,1}, \psi_{2,1}, \psi_{2,2}$ and $\psi_{2,3}$ respectively.



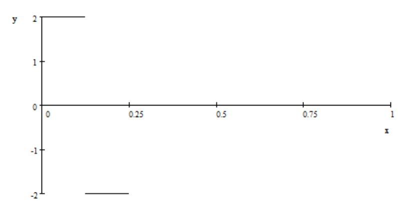
Plot of $\psi_0, 0$.



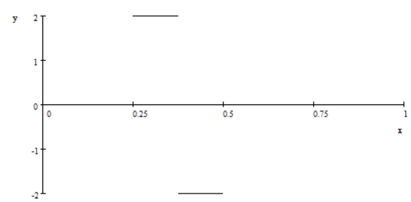
Plot of $\psi_1 0$.



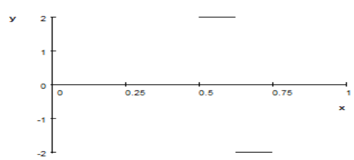
Plot of $\psi_1 1$.



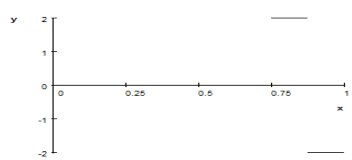
Plot of $\psi_2 0$.



Plot of $\psi_2 1$.



Plot of $\psi_2 2$.



Plot of $\psi_2 3$.

1. Let $M_0 = \text{span}(\{\mathbf{1}\})$ and for $n \in \mathbb{N}$ let
$$M_n := \text{span}\left(\{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k < n \text{ and } 0 \le j < 2^k\right\}\right),$$
where $\mathbf{1}$ denotes the constant function 1. Show
$$M_n = \text{span}\left(\{1_{[j2^{-n},(j+1)2^{-n})} : \text{and } 0 \le j < 2^n\right).$$
2. Show $\beta := \{\mathbf{1}\} \cup \left\{\psi_{kj} : 0 \le k \text{ and } 0 \le j < 2^k\right\}$ is an orthonormal set. **Hint:** show $\psi_{k+1,j} \in M_k^\perp$ for all $0 \le j < 2^{k+1}$ and show $\left\{\psi_{kj} : 0 \le j < 2^k\right\}$ is an orthonormal set for fixed k.
3. Show $\cup_{n=1}^\infty M_n$ is a dense subspace of L^2 and therefore β is an orthonormal basis for L^2. **Hint:** see Theorem 31.15.
4. For $f \in L^2$, let
$$H_n f := \langle f|\mathbf{1}\rangle \mathbf{1} + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \langle f|\psi_{kj}\rangle \psi_{kj}.$$
Show (compare with Exercise 40.8)
$$H_n f = \sum_{j=0}^{2^n-1} \left(2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(x)dx\right) 1_{[j2^{-n},(j+1)2^{-n})}$$
and use this to show $\|f - H_n f\|_\infty \to 0$ as $n \to \infty$ for all $f \in C([0,1])$. **Hint**: Compute orthogonal projection onto M_n using a judiciously chosen basis for M_n.

18.3 Fourier Series

In this subsection we will let $d\theta$, dx, $d\alpha$, etc. denote Lebesgue measure on $\mathbb{R}^d$ normalized so that the cube, $Q := (-\pi, \pi]^d$, has measure one, i.e. $d\theta = (2\pi)^{-d} dm(\theta)$ where m is standard Lebesgue measure on $\mathbb{R}^d$. Further let $\langle \cdot|\cdot\rangle$ denote the inner product on the Hilbert space, $H := L^2([-\pi,\pi]^d)$, given by
$$\langle f|g\rangle := \int_Q f(\theta)\bar{g}(\theta)d\theta = \left(\frac{1}{2\pi}\right)^d \int_Q f(\theta)\bar{g}(\theta)dm(\theta)$$
and define $\varphi_k(\theta) := e^{ik\cdot\theta}$ for all $k \in \mathbb{Z}^d$. For $f \in L^1(Q)$, we will write $\tilde{f}(k)$ for the **Fourier coefficient**,
$$\tilde{f}(k) := \langle f|\varphi_k\rangle = \int_Q f(\theta)e^{-ik\cdot\theta}d\theta. \tag{18.24}$$

Notation 18.37 *Let $C_{per}(\mathbb{R}^d)$ denote the 2π – periodic functions in $C(\mathbb{R}^d)$, that is $f \in C_{per}(\mathbb{R}^d)$ iff $f \in C(\mathbb{R}^d)$ and $f(\theta + 2\pi e_i) = f(\theta)$ for all $\theta \in \mathbb{R}^d$ and $i = 1, 2, \ldots, d$.*

Since any 2π – periodic functions on $\mathbb{R}^d$ may be identified with function on the d - dimensional torus, $\mathbb{T}^d \cong \mathbb{R}^d/(2\pi\mathbb{Z})^d \cong (S^1)^d$, I may also write $C(\mathbb{T}^d)$ for $C_{per}(\mathbb{R}^d)$ and $L^p(\mathbb{T}^d)$ for $L^p(Q)$ where elements in $f \in L^p(Q)$ are to be thought of as there extensions to 2π – periodic functions on $\mathbb{R}^d$.

Theorem 18.38 (Fourier Series). *The functions $\beta := \{\varphi_k : k \in \mathbb{Z}^d\}$ form an orthonormal basis for H, i.e. if $f \in H$ then*

$$f = \sum_{k\in\mathbb{Z}^d} \langle f|\varphi_k\rangle \varphi_k = \sum_{k\in\mathbb{Z}^d} \tilde{f}(k)\varphi_k \tag{18.25}$$

where the convergence takes place in $L^2([-\pi,\pi]^d)$.

Proof. Simple computations show $\beta := \{\varphi_k : k \in \mathbb{Z}^d\}$ is an orthonormal set. We now claim that β is an orthonormal basis. To see this recall that $C_c((-\pi,\pi)^d)$ is dense in $L^2((-\pi,\pi)^d, dm)$. Any $f \in C_c((-\pi,\pi)^d)$ may be extended to be a continuous 2π – periodic function on $\mathbb{R}$ and hence by Exercise 7.15 (see also Theorem 7.42, Exercise 37.13 and Remark 37.46), f may uniformly (and hence in L^2) be approximated by a trigonometric polynomial. Therefore β is a total orthonormal set, i.e. β is an orthonormal basis.

This may also be proved by first proving the case $d = 1$ as above and then using Exercise 18.10 inductively to get the result for any d. ■

We will discuss Fourier series and the related Fourier transform in more detail later, see Section 40.1.

18.3.1 Fourier Series Exercises

Exercise 18.14. Show that if $f \in L^1\left([-\pi,\pi]^d\right)$ and $\tilde{f}(k) = 0$ for all k then $f = 0$ a.e.

Exercise 18.15. Show $\sum_{k=1}^\infty k^{-2} = \pi^2/6$, by taking $f(x) = x$ on $[-\pi,\pi]$ and computing $\|f\|_2^2$ directly and then in terms of the Fourier Coefficients $\tilde{f}$ of f.

Exercise 18.16 (Riemann Lebesgue Lemma for Fourier Series). Show for $f \in L^1\left([-\pi,\pi]^d\right)$ that $\tilde{f} \in c_0(\mathbb{Z}^d)$, i.e. $\tilde{f} : \mathbb{Z}^d \to \mathbb{C}$ and $\lim_{k\to\infty} \tilde{f}(k) = 0$. **Hint:** If $f \in L^2\left([-\pi,\pi]^d\right)$, this follows from Bessel's inequality. Now use a density argument.

Exercise 18.17. Suppose $f \in L^1([-\pi,\pi]^d)$ is a function such that $\tilde{f} \in \ell^1(\mathbb{Z}^d)$ and set

$$g(x) := \sum_{k\in\mathbb{Z}^d} \tilde{f}(k)e^{ik\cdot x} \text{ (pointwise).}$$

1. Show $g \in C_{per}(\mathbb{R}^d)$.
2. Show $g(x) = f(x)$ for m – a.e. x in $[-\pi,\pi]^d$. **Hint**: Show $\tilde{g}(k) = \tilde{f}(k)$ and apply Exercise 18.14.
3. Conclude that $f \in L^1([-\pi,\pi]^d) \cap L^\infty([-\pi,\pi]^d)$ and in particular $f \in L^p([-\pi,\pi]^d)$ for all $p \in [1,\infty]$.

Notation 18.39 *Given a multi-index $\alpha \in \mathbb{Z}_+^d$, let $|\alpha| = \alpha_1 + \cdots + \alpha_d$,*

$$x^\alpha := \prod_{j=1}^d x_j^{\alpha_j}, \text{ and } \partial_x^\alpha = \left(\frac{\partial}{\partial x}\right)^\alpha := \prod_{j=1}^d \left(\frac{\partial}{\partial x_j}\right)^{\alpha_j}.$$

Further for $k \in \mathbb{N}_0$, let $f \in C_{per}^k(\mathbb{R}^d)$ iff $f \in C^k(\mathbb{R}^d) \cap C_{per}(\mathbb{R}^d)$, $\partial_x^\alpha f(x)$ exists and is continuous for $|\alpha| \le k$.

Exercise 18.18 (Smoothness implies decay). Suppose $m \in \mathbb{N}_0$, α is a multi-index such that $|\alpha| \le 2m$ and $f \in C_{per}^{2m}(\mathbb{R}^d)$[2].

1. Using integration by parts, show (using Notation 18.39) that
$$(ik)^\alpha \tilde{f}(k) = \langle \partial^\alpha f|\varphi_k\rangle \text{ for all } k \in \mathbb{Z}^d.$$
Note: This equality implies
$$\left|\tilde{f}(k)\right| \le \frac{1}{k^\alpha}\|\partial^\alpha f\|_H \le \frac{1}{k^\alpha}\|\partial^\alpha f\|_\infty.$$
2. Now let $\Delta f = \sum_{i=1}^d \partial^2 f/\partial x_i^2$, Working as in part 1) show
$$\langle (1-\Delta)^m f|\varphi_k\rangle = (1+|k|^2)^m \tilde{f}(k). \tag{18.26}$$

Remark 18.40. Suppose that m is an even integer, α is a multi-index and $f \in C_{per}^{m+|\alpha|}(\mathbb{R}^d)$, then

$$\begin{aligned}
\left(\sum_{k\in\mathbb{Z}^d} |k^\alpha| \left|\tilde{f}(k)\right|\right)^2 &= \left(\sum_{k\in\mathbb{Z}^d} |\langle \partial^\alpha f|e_k\rangle| (1+|k|^2)^{m/2}(1+|k|^2)^{-m/2}\right)^2 \\
&= \left(\sum_{k\in\mathbb{Z}^d} \left|\langle (1-\Delta)^{m/2}\partial^\alpha f|e_k\rangle\right| (1+|k|^2)^{-m/2}\right)^2 \\
&\le \sum_{k\in\mathbb{Z}^d} \left|\langle (1-\Delta)^{m/2}\partial^\alpha f|e_k\rangle\right|^2 \cdot \sum_{k\in\mathbb{Z}^d} (1+|k|^2)^{-m} \\
&= C_m \left\|(1-\Delta)^{m/2}\partial^\alpha f\right\|_H^2
\end{aligned}$$

[2] We view $C_{per}(\mathbb{R}^d)$ as a subspace of $H = L^2\left([-\pi,\pi]^d\right)$ by identifying $f \in C_{per}(\mathbb{R}^d)$ with $f|_{[-\pi,\pi]^d} \in H$.

where $C_m := \sum_{k\in\mathbb{Z}^d}(1+|k|^2)^{-m} < \infty$ iff $m > d/2$. So the smoother f is the faster $\tilde{f}$ decays at infinity. The next problem is the converse of this assertion and hence smoothness of f corresponds to decay of $\tilde{f}$ at infinity and visa-versa.

Exercise 18.19 (A Sobolev Imbedding Theorem). Suppose $s \in \mathbb{R}$ and $\{c_k \in \mathbb{C} : k \in \mathbb{Z}^d\}$ are coefficients such that

$$\sum_{k\in\mathbb{Z}^d} |c_k|^2 (1+|k|^2)^s < \infty.$$

Show if $s > \frac{d}{2} + m$, the function f defined by

$$f(x) = \sum_{k\in\mathbb{Z}^d} c_k e^{ik\cdot x}$$

is in $C^m_{per}(\mathbb{R}^d)$. **Hint**: Work as in the above remark to show

$$\sum_{k\in\mathbb{Z}^d} |c_k| \, |k^\alpha| < \infty \text{ for all } |\alpha| \le m.$$

Exercise 18.20 (Poisson Summation Formula). Let $F \in L^1(\mathbb{R}^d)$,

$$E := \left\{ x \in \mathbb{R}^d : \sum_{k\in\mathbb{Z}^d} |F(x+2\pi k)| = \infty \right\}$$

and set

$$\hat{F}(k) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} F(x) e^{-ik\cdot x} dx \text{ for } k \in \mathbb{Z}^d.$$

Further assume $\hat{F} \in \ell^1(\mathbb{Z}^d)$. [This can be achieved by assuming F is sufficiently differentiable with the derivatives being integrable like in Exercise 18.18.]

1. Show $m(E) = 0$ and $E + 2\pi k = E$ for all $k \in \mathbb{Z}^d$. **Hint**: Compute $\int_{[-\pi,\pi]^d} \sum_{k\in\mathbb{Z}^d} |F(x+2\pi k)| \, dx$.
2. Let
$$f(x) := \begin{cases} \sum_{k\in\mathbb{Z}^d} F(x+2\pi k) & \text{for } x \notin E \\ 0 & \text{if } x \in E. \end{cases}$$
Show $f \in L^1([-\pi,\pi]^d)$ and $\tilde{f}(k) = (2\pi)^{-d/2} \hat{F}(k)$.
3. Using item 2) and the assumptions on F, show
$$f(x) = \sum_{k\in\mathbb{Z}^d} \tilde{f}(k) e^{ik\cdot x} = \sum_{k\in\mathbb{Z}^d} (2\pi)^{-d/2} \hat{F}(k) e^{ik\cdot x} \text{ for } m-\text{a.e. } x,$$
i.e.
$$\sum_{k\in\mathbb{Z}^d} F(x+2\pi k) = (2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \hat{F}(k) e^{ik\cdot x} \text{ for } m-\text{a.e. } x \tag{18.27}$$
and form this conclude that $f \in L^1([-\pi,\pi]^d) \cap L^\infty([-\pi,\pi]^d)$.
Hint: see the hint for item 2. of Exercise 18.17.
4. Suppose we now assume that $F \in C(\mathbb{R}^d)$ and F satisfies $|F(x)| \le C(1+|x|)^{-s}$ for some $s > d$ and $C < \infty$. Under these added assumptions on F, show Eq. (18.27) holds for **all** $x \in \mathbb{R}^d$ and in particular
$$\sum_{k\in\mathbb{Z}^d} F(2\pi k) = (2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \hat{F}(k).$$

For notational simplicity, in the remaining problems we will assume that $d = 1$.

Exercise 18.21 (Heat Equation 1.). Let $(t,x) \in [0,\infty) \times \mathbb{R} \to u(t,x)$ be a continuous function such that $u(t,\cdot) \in C_{per}(\mathbb{R})$ for all $t \ge 0$, $\dot{u} := u_t$, u_x, and u_{xx} exists and are continuous when $t > 0$. Further assume that u satisfies the heat equation $\dot{u} = \frac{1}{2} u_{xx}$. Let $\tilde{u}(t,k) := \langle u(t,\cdot)|\varphi_k\rangle$ for $k \in \mathbb{Z}$. Show for $t > 0$ and $k \in \mathbb{Z}$ that $\tilde{u}(t,k)$ is differentiable in t and $\frac{d}{dt}\tilde{u}(t,k) = -k^2 \tilde{u}(t,k)/2$. Use this result to show

$$u(t,x) = \sum_{k\in\mathbb{Z}} e^{-\frac{t}{2}k^2} \tilde{f}(k) e^{ikx} \tag{18.28}$$

where $f(x) := u(0,x)$ and as above

$$\tilde{f}(k) = \langle f|\varphi_k\rangle = \int_{-\pi}^{\pi} f(y) e^{-iky} dy = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{-iky} dm(y).$$

Notice from Eq. (18.28) that $(t,x) \to u(t,x)$ is C^∞ for $t > 0$.

Exercise 18.22 (Heat Equation 2.). Let $q_t(x) := \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} e^{-\frac{t}{2}k^2} e^{ikx}$. Show that Eq. (18.28) may be rewritten as

$$u(t,x) = \int_{-\pi}^{\pi} q_t(x-y) f(y) dy$$

and

$$q_t(x) = \sum_{k\in\mathbb{Z}} p_t(x + k2\pi)$$

where $p_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}$. Also show $u(t,x)$ may be written as

$$u(t,x) = p_t * f(x) := \int_{\mathbb{R}^d} p_t(x-y) f(y) dy.$$

Hint: To show $q_t(x) = \sum_{k\in\mathbb{Z}} p_t(x+k2\pi)$, use the Poisson summation formula (Exercise 18.20) and the Gaussian integration identity,

$$\hat{p}_t(\omega) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} p_t(x)e^{i\omega x}dx = \frac{1}{\sqrt{2\pi}}e^{-\frac{t}{2}\omega^2}. \tag{18.29}$$

Equation (18.29) will be discussed in Example 43.4 below.

Exercise 18.23 (Wave Equation). Let $u \in C^2(\mathbb{R}\times\mathbb{R})$ be such that $u(t,\cdot) \in C_{per}(\mathbb{R})$ for all $t \in \mathbb{R}$. Further assume that u solves the wave equation, $u_{tt} = u_{xx}$. Let $f(x) := u(0,x)$ and $g(x) = \dot{u}(0,x)$. Show $\tilde{u}(t,k) := \langle u(t,\cdot), \varphi_k\rangle$ for $k \in \mathbb{Z}$ is twice continuously differentiable in t and $\frac{d^2}{dt^2}\tilde{u}(t,k) = -k^2\tilde{u}(t,k)$. Use this result to show

$$u(t,x) = \sum_{k\in\mathbb{Z}}\left(\tilde{f}(k)\cos(kt) + \tilde{g}(k)\frac{\sin kt}{k}\right)e^{ikx} \tag{18.30}$$

with the sum converging absolutely. Also show that $u(t,x)$ may be written as

$$u(t,x) = \frac{1}{2}\left[f(x+t) + f(x-t)\right] + \frac{1}{2}\int_{-t}^{t} g(x+\tau)d\tau. \tag{18.31}$$

Hint: To show Eq. (18.30) implies (18.31) use

$$\cos kt = \frac{e^{ikt}+e^{-ikt}}{2},$$
$$\sin kt = \frac{e^{ikt}-e^{-ikt}}{2i}, \text{ and}$$
$$\frac{e^{ik(x+t)}-e^{ik(x-t)}}{ik} = \int_{-t}^{t} e^{ik(x+\tau)}d\tau.$$

18.4 Exercises

Exercise 18.24 (The Mean Ergodic Theorem). Let $U : H \to H$ be a unitary operator on a Hilbert space H, $M = \mathrm{Nul}(U-I)$, $P = P_M$ be orthogonal projection onto M, and $S_n = \frac{1}{n}\sum_{k=0}^{n-1} U^k$. Show $S_n \to P_M$ **strongly** by which we mean $\lim_{n\to\infty} S_n x = P_M x$ for all $x \in H$. [See Exercise 18.27 for some explicit examples of U and the resulting P_M.]

Hints: 1. Show H is the orthogonal direct sum of M and $\overline{\mathrm{Ran}(U-I)}$ by first showing $\mathrm{Nul}(U^*-I) = \mathrm{Nul}(U-I)$ and then using Lemma 18.19. 2. Verify the result for $x \in \mathrm{Nul}(U-I)$ and $x \in \mathrm{Ran}(U-I)$. 3. Use a limiting argument to verify the result for $x \in \overline{\mathrm{Ran}(U-I)}$.

Exercise 18.25 (A "Martingale" Convergence Theorem). Suppose that $\{M_n\}_{n=1}^\infty$ is an increasing sequence of closed subspaces of a Hilbert space, H, $P_n := P_{M_n}$, and $\{x_n\}_{n=1}^\infty$ is a sequence of elements from H such that $x_n = P_n x_{n+1}$ for all $n \in \mathbb{N}$. Show;

1. $P_m x_n = x_m$ for all $1 \le m \le n < \infty$,
2. $(x_n - x_m) \perp M_m$ for all $n \ge m$,
3. $\|x_n\|$ is increasing as n increases,
4. if $\sup_n \|x_n\| = \lim_{n\to\infty}\|x_n\| < \infty$, then $x := \lim_{n\to\infty} x_n$ exists in M and that $x_n = P_n x$ for all $n \in \mathbb{N}$. (**Hint:** show $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence.)

Remark 18.41. Let $H = \ell^2 := L^2(\mathbb{N}, \text{counting measure})$,

$$M_n = \{(a(1), \dots, a(n), 0, 0, \dots) : a(i) \in \mathbb{C} \text{ for } 1 \le i \le n\},$$

and $x_n(i) = 1_{i\le n}$, then $x_m = P_m x_n$ for all $n \ge m$ while $\|x_n\|^2 = n \uparrow \infty$ as $n \to \infty$. Thus, we can not drop the assumption that $\sup_n \|x_n\| < \infty$ in Exercise 18.25.

Remark 18.42. See Definition 36.19 and the exercises in Section 23.4 for more on the notion of weak and strong convergence. See section **??** for more Hilbert space problems involving weak and strong convergence.

Exercise 18.26. Let $f \in L^1((-\pi,\pi])$ which we extend to a 2π – periodic function on $\mathbb{R}$ and continue to denote by f. If there exists $q \in \mathbb{N}$ such that $f\left(x + \frac{2\pi}{q}\right) = f(x)$ for m – a.e. x, then $\tilde{f}(k) = 0$ unless q divides k.

Exercise 18.27. In this problem we assume the notation from subsection 18.3 with $d = 1$. For simplicity of notation we identify $L^2((-\pi,\pi], d\theta)$ with 2π – periodic functions on $\mathbb{R}$ via,

$$L^2((-\pi,\pi],d\theta) \ni f \longleftrightarrow \sum_{n\in\mathbb{Z}} f(x+n2\pi)\, 1_{(-\pi,\pi]}(x+n2\pi) \in L^2_{per}(\mathbb{R}).$$

Given $\alpha \in \mathbb{R}$ let $(U_\alpha f)(\theta) = f(\theta + \alpha 2\pi)$ wherein we have used the above identification. If $\alpha \notin \mathbb{Q}$ show

$$M_\alpha = \mathrm{Nul}(U_\alpha - I) = \mathbb{C}\cdot 1.$$

If $\alpha \in \mathbb{Q}$ write $\alpha = \frac{p}{q}$ where $\gcd(q,p) = 1$, i.e. p and q are relatively prime. In this case show $M_\alpha = \mathrm{Nul}(U_\alpha - I)$ consists of those $f \in L^2_{per}(\mathbb{R})$ such that $f\left(x+\frac{2\pi}{q}\right) = f(x)$ for m – a.e. x. [Consequently, combining this exercise with Exercise 18.24 shows,

$$\frac{1}{n}\sum_{k=0}^{n-1} U_\alpha^k \xrightarrow{s} P_M$$

where M_α depends on α as described above.]

18.5 Supplement 1: Converse of the Parallelogram Law

Proposition 18.43 (Parallelogram Law Converse). *If $(X, \|\cdot\|)$ is a normed space such that Eq. (18.4) holds for all $x, y \in X$, then there exists a unique inner product on $\langle\cdot|\cdot\rangle$ such that $\|x\| := \sqrt{\langle x|x\rangle}$ for all $x \in X$. In this case we say that $\|\cdot\|$ is a Hilbertian norm.*

Proof. If $\|\cdot\|$ is going to come from an inner product $\langle\cdot|\cdot\rangle$, it follows from Eq. (18.1) that

$$2\mathrm{Re}\langle x|y\rangle = \|x+y\|^2 - \|x\|^2 - \|y\|^2$$

and

$$-2\mathrm{Re}\langle x|y\rangle = \|x-y\|^2 - \|x\|^2 - \|y\|^2.$$

Subtracting these two equations gives the "polarization identity,"

$$4\mathrm{Re}\langle x|y\rangle = \|x+y\|^2 - \|x-y\|^2.$$

Replacing y by iy in this equation then implies that

$$4\mathrm{Im}\langle x|y\rangle = \|x+iy\|^2 - \|x-iy\|^2$$

from which we find

$$\langle x|y\rangle = \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\varepsilon y\|^2 \tag{18.32}$$

where $G = \{\pm 1, \pm i\}$ – a cyclic subgroup of $S^1 \subset \mathbb{C}$. Hence, if $\langle\cdot|\cdot\rangle$ is going to exist we must define it by Eq. (18.32) and the uniqueness has been proved.

For existence, define $\langle x|y\rangle$ by Eq. (18.32) in which case,

$$\begin{aligned}\langle x|x\rangle &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\varepsilon x\|^2 = \frac{1}{4}\left[\|2x\|^2 + i\|x+ix\|^2 - i\|x-ix\|^2\right] \\ &= \|x\|^2 + \frac{i}{4}\left|1+i|^2\right|\|x\|^2 - \frac{i}{4}\left|1-i|^2\right|\|x\|^2 = \|x\|^2.\end{aligned}$$

So to finish the proof, it only remains to show that $\langle x|y\rangle$ defined by Eq. (18.32) is an inner product.

Since

$$\begin{aligned}4\langle y|x\rangle &= \sum_{\varepsilon\in G}\varepsilon\|y+\varepsilon x\|^2 = \sum_{\varepsilon\in G}\varepsilon\|\varepsilon\left(y+\varepsilon x\right)\|^2 \\ &= \sum_{\varepsilon\in G}\varepsilon\|\varepsilon y+\varepsilon^2 x\|^2 \\ &= \|y+x\|^2 - \|-y+x\|^2 + i\|iy-x\|^2 - i\|-iy-x\|^2 \\ &= \|x+y\|^2 - \|x-y\|^2 + i\|x-iy\|^2 - i\|x+iy\|^2 \\ &= 4\overline{\langle x|y\rangle}\end{aligned}$$

it suffices to show $x \to \langle x|y\rangle$ is linear for all $y \in H$. (The rest of this proof may safely be skipped by the reader.) For this we will need to derive an identity from Eq. (18.4). To do this we make use of Eq. (18.4) three times to find

$$\begin{aligned}\|x+y+z\|^2 &= -\|x+y-z\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|x-y-z\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|y+z-x\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= -\|y+z+x\|^2 + 2\|y+z\|^2 + 2\|x\|^2 \\ &\quad - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2.\end{aligned}$$

Solving this equation for $\|x+y+z\|^2$ gives

$$\|x+y+z\|^2 = \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2. \tag{18.33}$$

Using Eq. (18.33), for $x, y, z \in H$,

$$\begin{aligned}4\,\mathrm{Re}\langle x+z|y\rangle &= \|x+z+y\|^2 - \|x+z-y\|^2 \\ &= \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2 \\ &\quad - \left(\|z-y\|^2 + \|x-y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2\right) \\ &= \|z+y\|^2 - \|z-y\|^2 + \|x+y\|^2 - \|x-y\|^2 \\ &= 4\,\mathrm{Re}\langle x|y\rangle + 4\,\mathrm{Re}\langle z|y\rangle.\end{aligned} \tag{18.34}$$

Now suppose that $\delta \in G$, then since $|\delta| = 1$,

$$\begin{aligned}4\langle\delta x|y\rangle &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|\delta x+\varepsilon y\|^2 = \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\delta^{-1}\varepsilon y\|^2 \\ &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\delta\|x+\delta\varepsilon y\|^2 = 4\delta\langle x|y\rangle\end{aligned} \tag{18.35}$$

where in the third inequality, the substitution $\varepsilon \to \varepsilon\delta$ was made in the sum. So Eq. (18.35) says $\langle \pm ix|y\rangle = \pm i\langle x|y\rangle$ and $\langle -x|y\rangle = -\langle x|y\rangle$. Therefore

$$\mathrm{Im}\langle x|y\rangle = \mathrm{Re}\left(-i\langle x|y\rangle\right) = \mathrm{Re}\langle -ix|y\rangle$$

which combined with Eq. (18.34) shows

$$\begin{aligned}\mathrm{Im}\langle x+z|y\rangle &= \mathrm{Re}\langle -ix-iz|y\rangle = \mathrm{Re}\langle -ix|y\rangle + \mathrm{Re}\langle -iz|y\rangle \\ &= \mathrm{Im}\langle x|y\rangle + \mathrm{Im}\langle z|y\rangle\end{aligned}$$

and therefore (again in combination with Eq. (18.34)),

$$\langle x+z|y\rangle = \langle x|y\rangle + \langle z|y\rangle \text{ for all } x, y \in H.$$

Because of this equation and Eq. (18.35) to finish the proof that $x \to \langle x|y\rangle$ is linear, it suffices to show $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$. Now if $\lambda = m \in \mathbb{N}$, then

$$\langle mx|y\rangle = \langle x + (m-1)x|y\rangle = \langle x|y\rangle + \langle (m-1)x|y\rangle$$

so that by induction $\langle mx|y\rangle = m\langle x|y\rangle$. Replacing x by x/m then shows that $\langle x|y\rangle = m\langle m^{-1}x|y\rangle$ so that $\langle m^{-1}x|y\rangle = m^{-1}\langle x|y\rangle$ and so if $m, n \in \mathbb{N}$, we find

$$\langle \frac{n}{m}x|y\rangle = n\langle \frac{1}{m}x|y\rangle = \frac{n}{m}\langle x|y\rangle$$

so that $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$ and $\lambda \in \mathbb{Q}$. By continuity, it now follows that $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$. ∎

18.6 Supplement 2. Non-complete inner product spaces

Part of Theorem 18.27 goes through when H is a not necessarily complete inner product space. We have the following proposition.

Proposition 18.44. *Let $(H, \langle\cdot|\cdot\rangle)$ be a not necessarily complete inner product space and $\beta \subset H$ be an orthonormal set. Then the following two conditions are equivalent:*

1. $x = \sum_{u\in\beta} \langle x|u\rangle u$ *for all* $x \in H$.
2. $\|x\|^2 = \sum_{u\in\beta} |\langle x|u\rangle|^2$ *for all* $x \in H$.

Moreover, either of these two conditions implies that $\beta \subset H$ is a maximal orthonormal set. However $\beta \subset H$ being a maximal orthonormal set is not sufficient (without completeness of H) to show that items 1. and 2. hold!

Proof. As in the proof of Theorem 18.27, 1) implies 2). For 2) implies 1) let $\Lambda \subset\subset \beta$ and consider

$$\begin{aligned}\left\|x - \sum_{u\in\Lambda}\langle x|u\rangle u\right\|^2 &= \|x\|^2 - 2\sum_{u\in\Lambda}|\langle x|u\rangle|^2 + \sum_{u\in\Lambda}|\langle x|u\rangle|^2 \\ &= \|x\|^2 - \sum_{u\in\Lambda}|\langle x|u\rangle|^2 .\end{aligned}$$

Since $\|x\|^2 = \sum_{u\in\beta}|\langle x|u\rangle|^2$, it follows that for every $\varepsilon > 0$ there exists $\Lambda_\varepsilon \subset\subset \beta$ such that for all $\Lambda \subset\subset \beta$ such that $\Lambda_\varepsilon \subset \Lambda$,

$$\left\|x - \sum_{u\in\Lambda}\langle x|u\rangle u\right\|^2 = \|x\|^2 - \sum_{u\in\Lambda}|\langle x|u\rangle|^2 < \varepsilon$$

showing that $x = \sum_{u\in\beta}\langle x|u\rangle u$. Suppose $x = (x_1, x_2, \dots, x_n, \dots) \in \beta^\perp$. If 2) is valid then $\|x\|^2 = 0$, i.e. $x = 0$. So β is maximal. Let us now construct a counterexample to prove the last assertion. Take $H = \mathrm{Span}\{e_i\}_{i=1}^\infty \subset \ell^2$ and let $\tilde{u}_n = e_1 - (n+1)e_{n+1}$ for $n = 1, 2 \dots$. Applying Gram-Schmidt to $\{\tilde{u}_n\}_{n=1}^\infty$ we construct an orthonormal set $\beta = \{u_n\}_{n=1}^\infty \subset H$. I now claim that $\beta \subset H$ is maximal. Indeed if $x = (x_1, x_2, \dots, x_n, \dots) \in \beta^\perp$ then $x \perp u_n$ for all n, i.e.

$$0 = \langle x|\tilde{u}_n\rangle = x_1 - (n+1)x_{n+1}.$$

Therefore $x_{n+1} = (n+1)^{-1} x_1$ for all n. Since $x \in \mathrm{Span}\{e_i\}_{i=1}^\infty$, $x_N = 0$ for some N sufficiently large and therefore $x_1 = 0$ which in turn implies that $x_n = 0$ for all n. So $x = 0$ and hence β is maximal in H. On the other hand, β is not maximal in ℓ^2. In fact the above argument shows that $\beta^\perp$ in ℓ^2 is given by the span of $v = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots)$. Let P be the orthogonal projection of ℓ^2 onto the $\mathrm{Span}(\beta) = v^\perp$. Then

$$\sum_{n=1}^\infty \langle x|u_n\rangle u_n = Px = x - \frac{\langle x|v\rangle}{\|v\|^2}v,$$

so that $\sum_{n=1}^\infty \langle x|u_n\rangle u_n = x$ iff $x \in \mathrm{Span}(\beta) = v^\perp \subset \ell^2$. For example if $x = (1, 0, 0, \dots) \in H$ (or more generally for $x = e_i$ for any i), $x \notin v^\perp$ and hence $\sum_{i=1}^\infty \langle x|u_n\rangle u_n \neq x$. ∎

19

Compactness in Metric Space

Let (X, d) be a metric space and let $\tau = \tau_d$ denote the collection of open subsets of X. (Recall $V \subset X$ is open iff V^c is closed iff for all $x \in V$ there exists an $\varepsilon = \varepsilon_x > 0$ such that $B(x, \varepsilon_x) \subset V$ iff V can be written as a (possibly uncountable) union of open balls – see Section 13.1.) Although we will stick with metric spaces in this chapter, it will be useful to introduce the definitions needed here in the more general context of a general "topological space," i.e. a space equipped with a collection of "open sets."

19.1 General Topological Notions

Definition 19.1. *Let (X, τ) be a topological space and $A \subset X$. A collection of subsets $\mathcal{U} \subset \tau$ is an **open cover** of A if $A \subset \bigcup \mathcal{U} := \bigcup_{U \in \mathcal{U}} U$.*

Definition 19.2. *The subset A of a topological space, (X, τ), is said to be **compact** if every open cover (Definition 19.1) of A has finite a sub-cover, i.e. if $\mathcal{U}$ is an open cover of A there exists $\mathcal{U}_0 \subset\subset \mathcal{U}$ such that $\mathcal{U}_0$ is a cover of A. (We will write $A \sqsubset\sqsubset X$ to denote that $A \subset X$ and A is compact.) A subset $A \subset X$ is **precompact** if $\bar{A}$ is compact.*

Exercise 19.1. Suppose $f : X \to Y$ is continuous and $K \subset X$ is compact, then $f(K)$ is a compact subset of Y. Give an example of continuous map, $f : X \to Y$, and a compact subset K of Y such that $f^{-1}(K)$ is not compact.

Definition 19.3. *Let (X, d) be a metric space. We say a subset $A \subset X$ is bounded iff for all (any) $o \in X$, $\sup_{a \in A} d(o, a) < \infty$. In other words A should be contained in a finite radius ball in X.*

Exercise 19.2 (Dini's Theorem). Let (X, τ) be a compact topological space and $f_n : X \to [0, \infty)$ be a sequence of continuous functions such that $f_n(x) \downarrow 0$ as $n \to \infty$ for each $x \in X$. Show that in fact $f_n \downarrow 0$ uniformly in x, i.e. $\sup_{x \in X} f_n(x) \downarrow 0$ as $n \to \infty$. **Hint:** Given $\varepsilon > 0$, consider the open sets $V_n := \{x \in X : f_n(x) < \varepsilon\}$.

Proposition 19.4. *Suppose that $K \subset X$ is a compact set and $F \subset K$ is a closed subset. Then F is compact. If $\{K_i\}_{i=1}^n$ is a finite collections of compact subsets of X then $K = \cup_{i=1}^n K_i$ is also a compact subset of X.*

Proof. Let $\mathcal{U} \subset \tau$ be an open cover of F, then $\mathcal{U} \cup \{F^c\}$ is an open cover of K. The cover $\mathcal{U} \cup \{F^c\}$ of K has a finite subcover which we denote by $\mathcal{U}_0 \cup \{F^c\}$ where $\mathcal{U}_0 \subset\subset \mathcal{U}$. Since $F \cap F^c = \emptyset$, it follows that $\mathcal{U}_0$ is the desired subcover of F. For the second assertion suppose $\mathcal{U} \subset \tau$ is an open cover of K. Then $\mathcal{U}$ covers each compact set K_i and therefore there exists a finite subset $\mathcal{U}_i \subset\subset \mathcal{U}$ for each i such that $K_i \subset \cup \mathcal{U}_i$. Then $\mathcal{U}_0 := \cup_{i=1}^n \mathcal{U}_i$ is a finite cover of K. ■

Definition 19.5. *A collection $\mathcal{F}$ of closed subsets of (X, τ) has the **finite intersection property** if $\cap \mathcal{F}_0 \neq \emptyset$ for all $\mathcal{F}_0 \subset\subset \mathcal{F}$.*

The notion of compactness may be expressed in terms of closed sets as follows.

Proposition 19.6. *A topological space (X, τ) is compact iff every family of closed sets $\mathcal{F} \subset 2^X$ having the **finite intersection property** satisfies $\bigcap \mathcal{F} \neq \emptyset$.*

Proof. ($\Rightarrow$) Suppose that X is compact and $\mathcal{F} \subset 2^X$ is a collection of closed sets such that $\bigcap \mathcal{F} = \emptyset$. Let

$$\mathcal{U} = \mathcal{F}^c := \{C^c : C \in \mathcal{F}\} \subset \tau,$$

then $\mathcal{U}$ is a cover of X and hence has a finite subcover, $\mathcal{U}_0$. Let $\mathcal{F}_0 = \mathcal{U}_0^c \subset\subset \mathcal{F}$, then $\cap \mathcal{F}_0 = \emptyset$ so that $\mathcal{F}$ does not have the finite intersection property.

($\Leftarrow$) If X is not compact, there exists an open cover $\mathcal{U}$ of X with no finite subcover. Let

$$\mathcal{F} = \mathcal{U}^c := \{U^c : U \in \mathcal{U}\},$$

then $\mathcal{F}$ is a collection of closed sets with the finite intersection property while $\bigcap \mathcal{F} = \emptyset$. ■

Exercise 19.3 ($BC(X)$ is closed). Let (X, τ) be a topological space and $\ell^\infty(X)$ denote the Banach space of bounded complex functions on X equipped with the norm,

$$\|f\|_\infty = \sup_{x \in X} |f(x)|.$$

Suppose that $f \in \ell^\infty(X)$ and $\{f_n\} \subset \ell^\infty(X)$ with $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$. If each f_n is continuous at some point $x \in X$, then f is continuous at $x \in X$. In particular, this shows that $BC(X)$ – bounded continuous functions on X is a closed subspace of $\ell^\infty(X)$.

19.2 Compactness in Metric Spaces

For the rest of this chapter we will assume that (X,d) is a fixed metric space and $\tau = \tau_d$. For $x \in X$ and $\varepsilon > 0$ let

$$B'_x(\varepsilon) := B_x(\varepsilon) \setminus \{x\}$$

be the ball centered at x of radius $\varepsilon > 0$ with x deleted. Recall from Definition **??** that a point $x \in X$ is an accumulation point of a subset $E \subset X$ if $\emptyset \neq E \cap V \setminus \{x\}$ for all open neighborhoods, V, of x. The proof of the following elementary lemma is left to the reader.

Lemma 19.7. *Let $E \subset X$ be a subset of a metric space (X,d). Then the following are equivalent:*

1. *$x \in X$ is an accumulation point of E.*
2. *$B'_x(\varepsilon) \cap E \neq \emptyset$ for all $\varepsilon > 0$.*
3. *$B_x(\varepsilon) \cap E$ is an infinite set for all $\varepsilon > 0$.*
4. *There exists $\{x_n\}_{n=1}^\infty \subset E \setminus \{x\}$ with $\lim_{n\to\infty} x_n = x$.*

Definition 19.8. *A subset A of a metric space (X,d) is ε – **bounded** $(\varepsilon > 0)$ if there exists a finite cover of A by balls of radius ε and it is **totally bounded** if it is ε – bounded for all $\varepsilon > 0$.*

Exercise 19.4. Given $n \in \mathbb{N}$ let $\mathbb{R}^n$ be equipped with the usual Euclidean distance. Show that every bounded subset of $\mathbb{R}^n$ is totally bounded where $A \subset \mathbb{R}^n$ is bounded iff $A \subset B_0(R)$ for some $R < \infty$.

Exercise 19.5. Let $X = \ell^2$ with $d(x,y) := \|y - x\|_{\ell^2}$ where

$$\|x\|_{\ell^2}^2 = \sum_{i=1}^\infty |x_i|^2 .$$

Further let $e_n := (0,\dots,0,1,0,0,\dots)$ where the 1 appears in the n^{th} – slot be the standard basis for ℓ^2. Show that $A := \{e_n\}_{n=1}^\infty$ is not totally bounded.

Theorem 19.9. *Let (X,d) be a metric space. The following are equivalent.*

(a) X is compact.
(b) Every infinite subset of X has an accumulation point.
(c) Every sequence $\{x_n\}_{n=1}^\infty \subset X$ has a convergent subsequence.
(d) X is totally bounded and complete.

Proof. The proof will consist of showing that $a \Rightarrow b \Rightarrow c \Rightarrow d \Rightarrow a$.

$(a \Rightarrow b)$ We will show that **not** $b \Rightarrow$ **not** a. Suppose there exists an infinite subset $E \subset X$ which has no accumulation points. Then for all $x \in X$ there exists $\delta_x > 0$ such that $V_x := B_x(\delta_x)$ satisfies $(V_x \setminus \{x\}) \cap E = \emptyset$. Clearly $\mathcal{V} = \{V_x\}_{x\in X}$ is a cover of X, yet $\mathcal{V}$ has no finite sub cover. Indeed, for each $x \in X$, $V_x \cap E \subset \{x\}$ and hence if $\Lambda \subset\subset X$, $\cup_{x\in\Lambda} V_x$ can only contain a finite number of points from E (namely $\Lambda \cap E$). Thus for any $\Lambda \subset\subset X$, $E \nsubseteq \cup_{x\in\Lambda} V_x$ and in particular $X \neq \cup_{x\in\Lambda} V_x$. (See Figure 19.1.)

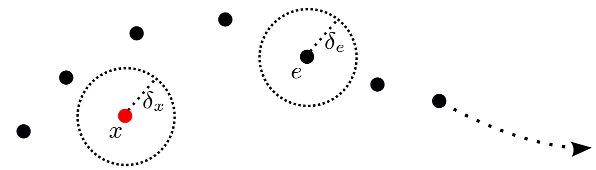


Fig. 19.1. The black dots represents an infinite set, E, with no accumulation points. For each $x \in X \setminus E$ we choose $\delta_x > 0$ so that $B_x(\delta_x) \cap E = \emptyset$ and for $x \in E$ so that $B_x(\delta_x) \cap E = \{x\}$.

$(b \Rightarrow c)$ Let $\{x_n\}_{n=1}^\infty \subset X$ be a sequence and $E := \{x_n : n \in \mathbb{N}\}$. If $\#(E) < \infty$, then $\{x_n\}_{n=1}^\infty$ has a subsequence $\{x_{n_k}\}_{k=1}^\infty$ which is constant and hence convergent. On the other hand if $\#(E) = \infty$ then by assumption E has an accumulation point and hence by Lemma 19.7, $\{x_n\}_{n=1}^\infty$ has a convergent subsequence.

$(c \Rightarrow d)$ Suppose $\{x_n\}_{n=1}^\infty \subset X$ is a Cauchy sequence. By assumption there exists a subsequence $\{x_{n_k}\}_{k=1}^\infty$ which is convergent to some point $x \in X$. Since $\{x_n\}_{n=1}^\infty$ is Cauchy it follows that $x_n \to x$ as $n \to \infty$ showing X is complete. We now show that X is totally bounded. Let $\varepsilon > 0$ be given and choose an arbitrary point $x_1 \in X$. If possible choose $x_2 \in X$ such that $d(x_2, x_1) \geq \varepsilon$, then if possible choose $x_3 \in X$ such that $d_{\{x_1,x_2\}}(x_3) \geq \varepsilon$ and continue inductively choosing points $\{x_j\}_{j=1}^n \subset X$ such that $d_{\{x_1,\dots,x_{n-1}\}}(x_n) \geq \varepsilon$. (See Figure 19.2.) This process must terminate, for otherwise we would produce a sequence $\{x_n\}_{n=1}^\infty \subset X$ which can have no convergent subsequences. Indeed, the x_n have been chosen so that $d(x_n, x_m) \geq \varepsilon > 0$ for every $m \neq n$ and hence no subsequence of $\{x_n\}_{n=1}^\infty$ can be Cauchy.

$(d \Rightarrow a)$ For sake of contradiction, assume there exists an open cover $\mathcal{V} = \{V_\alpha\}_{\alpha\in A}$ of X with no finite subcover. Since X is totally bounded for each $n \in \mathbb{N}$ there exists $\Lambda_n \subset\subset X$ such that

$$X = \bigcup_{x\in\Lambda_n} B_x(1/n) \subset \bigcup_{x\in\Lambda_n} C_x(1/n).$$

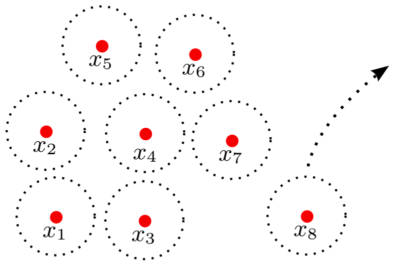


Fig. 19.2. Constructing a set with out an accumulation point.

Choose $x_1 \in \Lambda_1$ such that no finite subset of $\mathcal{V}$ covers $K_1 := C_{x_1}(1)$. Since $K_1 = \cup_{x\in\Lambda_2} K_1 \cap C_x(1/2)$, there exists $x_2 \in \Lambda_2$ such that $K_2 := K_1 \cap C_{x_2}(1/2)$ can not be covered by a finite subset of $\mathcal{V}$, see Figure 19.3. Continuing this way inductively, we construct sets $K_n = K_{n-1} \cap C_{x_n}(1/n)$ with $x_n \in \Lambda_n$ such that no K_n can be covered by a finite subset of $\mathcal{V}$. Now choose $y_n \in K_n$ for each n. Since $\{K_n\}_{n=1}^\infty$ is a decreasing sequence of closed sets such that $\operatorname{diam}(K_n) \le 2/n$, it follows that $\{y_n\}$ is a Cauchy and hence convergent with

$$y = \lim_{n\to\infty} y_n \in \cap_{m=1}^\infty K_m.$$

Since $\mathcal{V}$ is a cover of X, there exists $V \in \mathcal{V}$ such that $y \in V$. Since $K_n \downarrow \{y\}$ and $\operatorname{diam}(K_n) \to 0$, it now follows that $K_n \subset V$ for some n large. But this violates the assertion that K_n can not be covered by a finite subset of $\mathcal{V}$.

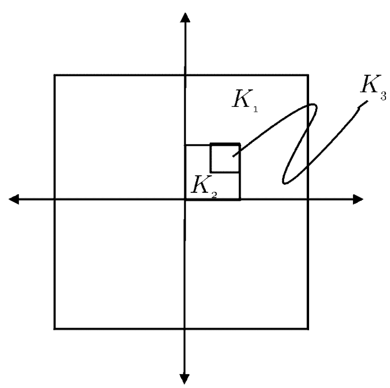


Fig. 19.3. Nested Sequence of cubes.

■

Proposition 19.10. *Suppose (X,d) is a metric space and $K \subset X$ is a compact subset. Then K is both closed and bounded.*

Proof. Let $o \in X$ and $f(x) := d(o,x)$. Then $f : X \to \mathbb{R}$ is a continuous function and therefore $f(K)$ is compact in $\mathbb{R}$. As the compact subsets of $\mathbb{R}$ are closed and bounded it follows that K is bounded as $\sup_{k\in K} d(o,k) = \sup_{k\in K} f(k) < \infty$. If K were not closed, we could find $\{x_n\}_{n=1}^\infty \subset K$ for which $x := \lim_{n\to\infty} x_n \notin K$. This sequence has no convergent subsequence to a point **in** K showing K is not compact. ■

Definition 19.11. *A topological space, (X,τ), is **second countable** if there exists a countable base $\mathcal{V}$ for τ, i.e. $\mathcal{V} \subset \tau$ is a countable set such that for every $W \in \tau$,*

$$W = \cup\{V : V \in \mathcal{V} \text{ such that } V \subset W\}.$$

Corollary 19.12. *Any compact metric space (X,d) is second countable and hence also separable by Exercise 35.12. (See Example 37.25 below for an example of a compact topological space which is not separable.)*

Proof. To each integer n, there exists $\Lambda_n \subset\subset X$ such that $X = \cup_{x\in\Lambda_n} B(x,1/n)$. The collection of open balls,

$$\mathcal{V} := \cup_{n\in\mathbb{N}} \cup_{x\in\Lambda_n} \{B(x,1/n)\}$$

forms a countable basis for the metric topology on X. To check this, suppose that $x_0 \in X$ and $\varepsilon > 0$ are given and choose $n \in \mathbb{N}$ such that $1/n < \varepsilon/2$ and $x \in \Lambda_n$ such that $d(x_0,x) < 1/n$. Then $B(x,1/n) \subset B(x_0,\varepsilon)$ because for $y \in B(x,1/n)$,

$$d(y,x_0) \le d(y,x) + d(x,x_0) < 2/n < \varepsilon.$$

■

Corollary 19.13. *The compact subsets of $\mathbb{R}^n$ are the closed and bounded sets.*

Proof. This is a consequence of Theorem 32.2 and Theorem 19.9. Here is another proof. If K is closed and bounded then K is complete (being the closed subset of a complete space) and K is contained in $[-M,M]^n$ for some positive integer M. For $\delta > 0$, let

$$\Lambda_\delta = \delta\mathbb{Z}^n \cap [-M,M]^n := \{\delta x : x \in \mathbb{Z}^n \text{ and } \delta|x_i| \le M \text{ for } i = 1,2,\dots,n\}.$$

We will show, by choosing $\delta > 0$ sufficiently small, that

$$K \subset [-M,M]^n \subset \cup_{x\in\Lambda_\delta} B(x,\varepsilon) \tag{19.1}$$

which shows that K is totally bounded. Hence by Theorem 19.9, K is compact. Suppose that $y \in [-M,M]^n$, then there exists $x \in \Lambda_\delta$ such that $|y_i - x_i| \le \delta$ for $i = 1,2,\dots,n$. Hence

$$d^2(x,y) = \sum_{i=1}^{n} (y_i - x_i)^2 \le n\delta^2$$

which shows that $d(x,y) \le \sqrt{n}\delta$. Hence if choose $\delta < \varepsilon/\sqrt{n}$ we have shows that $d(x,y) < \varepsilon$, i.e. Eq. (19.1) holds. ■

Example 19.14. Let $X = \ell^p(\mathbb{N})$ with $p \in [1,\infty)$ and $\mu \in \ell^p(\mathbb{N})$ such that $\mu(k) \ge 0$ for all $k \in \mathbb{N}$. The set

$$K := \{x \in X : |x(k)| \le \mu(k) \text{ for all } k \in \mathbb{N}\}$$

is compact. To prove this, let $\{x_n\}_{n=1}^\infty \subset K$ be a sequence. By compactness of closed bounded sets in $\mathbb{C}$, for each $k \in \mathbb{N}$ there is a subsequence of $\{x_n(k)\}_{n=1}^\infty \subset \mathbb{C}$ which is convergent. By Cantor's diagonalization trick, we may choose a subsequence $\{y_n\}_{n=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ such that $y(k) := \lim_{n\to\infty} y_n(k)$ exists for all $k \in \mathbb{N}$.[1] Since $|y_n(k)| \le \mu(k)$ for all n it follows that $|y(k)| \le \mu(k)$, i.e. $y \in K$. Finally

$$\lim_{n\to\infty} \|y - y_n\|_p^p = \lim_{n\to\infty} \sum_{k=1}^{\infty} |y(k) - y_n(k)|^p = \sum_{k=1}^{\infty} \lim_{n\to\infty} |y(k) - y_n(k)|^p = 0$$

wherein we have used the Dominated convergence theorem. (Note

$$|y(k) - y_n(k)|^p \le 2^p \mu^p(k)$$

and μ^p is summable.) Therefore $y_n \to y$ and we are done.

Alternatively, we can prove K is compact by showing that K is closed and totally bounded. It is simple to show K is closed, for if $\{x_n\}_{n=1}^\infty \subset K$ is a convergent sequence in X, $x := \lim_{n\to\infty} x_n$, then

$$|x(k)| \le \lim_{n\to\infty} |x_n(k)| \le \mu(k) \ \forall\ k \in \mathbb{N}.$$

This shows that $x \in K$ and hence K is closed. To see that K is totally bounded, let $\varepsilon > 0$ and choose N such that $\left(\sum_{k=N+1}^\infty |\mu(k)|^p\right)^{1/p} < \varepsilon$. Since $\prod_{k=1}^N C_{\mu(k)}(0) \subset \mathbb{C}^N$ is closed and bounded, it is compact. Therefore there exists a finite subset $\Lambda \subset \prod_{k=1}^N C_{\mu(k)}(0)$ such that

$$\prod_{k=1}^N C_{\mu(k)}(0) \subset \cup_{z\in\Lambda} B_z^N(\varepsilon)$$

where $B_z^N(\varepsilon)$ is the open ball centered at $z \in \mathbb{C}^N$ relative to the $\ell^p(\{1,2,3,\dots,N\})$ – norm. For each $z \in \Lambda$, let $\tilde{z} \in X$ be defined by $\tilde{z}(k) = z(k)$ if $k \le N$ and $\tilde{z}(k) = 0$ for $k \ge N+1$. I now claim that

$$K \subset \cup_{z\in\Lambda} B_{\tilde{z}}(2\varepsilon) \tag{19.2}$$

which, when verified, shows K is totally bounded. To verify Eq. (19.2), let $x \in K$ and write $x = u + v$ where $u(k) = x(k)$ for $k \le N$ and $u(k) = 0$ for $k < N$. Then by construction $u \in B_{\tilde{z}}(\varepsilon)$ for some $\tilde{z} \in \Lambda$ and

$$\|v\|_p \le \left(\sum_{k=N+1}^\infty |\mu(k)|^p\right)^{1/p} < \varepsilon.$$

So we have

$$\|x - \tilde{z}\|_p = \|u + v - \tilde{z}\|_p \le \|u - \tilde{z}\|_p + \|v\|_p < 2\varepsilon.$$

Exercise 19.6 (Extreme value theorem). Let (X,d) be a compact metric space and $f : X \to \mathbb{R}$ be a continuous function. Show $-\infty < \inf f \le \sup f < \infty$ and there exists $a, b \in X$ such that $f(a) = \inf f$ and $f(b) = \sup f$.

Exercise 19.7 (Uniform Continuity). Let (X,d) be a compact metric space, (Y,ρ) be a metric space and $f : X \to Y$ be a continuous function. Show that f is uniformly continuous, i.e. if $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(y), f(x)) < \varepsilon$ if $x, y \in X$ with $d(x,y) < \delta$.

19.3 Locally Compact Banach Spaces

Definition 19.15. *Let Y be a vector space. We say that two norms, $|\cdot|$ and $\|\cdot\|$, on Y are equivalent if there exists constants $\alpha, \beta \in (0,\infty)$ such that*

$$\|f\| \le \alpha |f| \text{ and } |f| \le \beta \|f\| \text{ for all } f \in Y.$$

Theorem 19.16. *Let $(Y, \|\cdot\|)$ be a finite dimensional normed vector space. Then;*

[1] The argument is as follows. Let $\{n_j^1\}_{j=1}^\infty$ be a subsequence of $\mathbb{N} = \{n\}_{n=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^1}(1)$ exists. Now choose a subsequence $\{n_j^2\}_{j=1}^\infty$ of $\{n_j^1\}_{j=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^2}(2)$ exists and similarly $\{n_j^3\}_{j=1}^\infty$ of $\{n_j^2\}_{j=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^3}(3)$ exists. Continue on this way inductively to get

$$\{n\}_{n=1}^\infty \supset \{n_j^1\}_{j=1}^\infty \supset \{n_j^2\}_{j=1}^\infty \supset \{n_j^3\}_{j=1}^\infty \supset \dots$$

such that $\lim_{j\to\infty} x_{n_j^k}(k)$ exists for all $k \in \mathbb{N}$. Let $m_j := n_j^j$ so that eventually $\{m_j\}_{j=1}^\infty$ is a subsequence of $\{n_j^k\}_{j=1}^\infty$ for all k. Therefore, we may take $y_j := x_{m_j}$.

1. *any other norm* $|\cdot|$ *on* Y *is equivalent to* $\|\cdot\|$, *i.e. all norms on* Y *are equivalent. (This is typically not true for norms on infinite dimensional spaces, see for example Exercise 14.8.)*
2. $(Y, \|\cdot\|)$ *is complete.*
3. *A subset* $B \subset Y$ *is compact iff* B *is closed and bounded relative to the given norm,* $\|\cdot\|$.

Proof. 1. Let $\{f_i\}_{i=1}^n$ be a real basis for Y and let $T : \mathbb{R}^n \to Y$ be the linear isomorphism of vector spaces defined by

$$Tz = \sum_{i=1}^n z_i f_i \text{ for all } z = (z_1, \ldots, z_n) \in \mathbb{R}^n.$$

Since,

$$\|Tz\| = \left\| \sum_{i=1}^n z_i f_i \right\| \le \sum_{i=1}^n \|z_i f_i\| \le \sum_{i=1}^n |z_i| \, \|f_i\| \le C \cdot \|z\|_2 \tag{19.3}$$

where

$$C := \sqrt{\sum_{i=1}^n \|f_i\|^2} \text{ and } \|z\|_2 := \sqrt{\sum_{i=1}^n |z_i|^2}$$

it follows that $T : (\mathbb{R}^n, \|\cdot\|_2) \to (Y, \|\cdot\|)$ bounded, i.e. continuous. Since the unit sphere, $S \subset \mathbb{R}^n$, is compact the continuous function $z \to \|Tz\|$ restricted to S attains a minimum on S, i.e.

$$\min_{\|z\|_2 = 1} \|Tz\| = \|Tz_0\| = \varepsilon > 0.$$

By homogeneity it now follows that $\|Tz\| \ge \varepsilon \|z\|_2$ for all $z \in \mathbb{R}^n$ and therefore

$$\varepsilon \|z\|_2 \le \|Tz\| \le C \cdot \|z\|_2 \text{ for all } z \in \mathbb{R}^n. \tag{19.4}$$

Evaluating Eq. (19.4) at $z = T^{-1}y$ implies,

$$\left\|T^{-1}y\right\|_2 \varepsilon \le \|y\| \le C \cdot \left\|T^{-1}y\right\|_2 \text{ for all } y \in Y. \tag{19.5}$$

Hence we have shown that every norm on Y is equivalent to the norm $y \to \left\|T^{-1}y\right\|_2$ and hence all norms on Y are equivalent.

2. The reader may now easily check that a sequence $\{y_n\}_{n=1}^\infty \subset Y$ is convergent (Cauchy) iff $\left\{T^{-1}y_n\right\}_{n=1}^\infty \subset \mathbb{R}^n$ is convergent (Cauchy). Thus if $\{y_n\}_{n=1}^\infty$ is Cauchy then $\left\{T^{-1}y_n\right\}_{n=1}^\infty \subset \mathbb{R}^n$ is Cauchy and hence convergent in $\mathbb{R}^n$ as $\mathbb{R}^n$ is complete. Therefore $\{y_n\}_{n=1}^\infty$ is convergent in Y.

3. Let B be a $\|\cdot\|$ – closed and bounded subset of Y and $\{y_n\}_{n=1}^\infty \subset B$. Then $\left\{z_n := T^{-1}y_n\right\}_{n=1}^\infty$ is a bounded sequence in $\mathbb{R}^n$ by Eq. (19.5). Since bounded subsets of $\mathbb{R}^n$ are pre-compact, we know that $\{z_n\}_{n=1}^\infty$ has a convergent subsequence in $\mathbb{R}^n$ and hence $\{y_n\}_{n=1}^\infty$ has a convergent subsequence in B. As B is closed this subsequence is convergent in B showing B is compact.

Conversely, from Proposition 19.10 we know that compact subsets of metric spaces are always closed and bounded For completeness I will give an argument here as well. 1) If B is not bounded we could find a sequence $\{y_n\}_{n=1}^\infty \subset B$ such that $\lim_{n\to\infty} \|y_n\| = \infty$. As $\{\|y_n\|\}_{n=1}^\infty$ has no convergent subsequence it follows that $\{y_n\}_{n=1}^\infty \subset B$ does not have a convergent subsequence either. 2) If B is not closed we could find $\{y_n\}_{n=1}^\infty \subset B$ such that $y = \lim_{n\to\infty} y_n \notin B$. Again there can be no convergent subsequence which converges to an element in B. So in either case B is not compact.

■

Corollary 19.17. *Any finite dimensional subspace,* Y, *of a normed vector space* $(X, \|\cdot\|)$ *is automatically closed.*

Proof. Suppose $\{y_n\}_{n=1}^\infty \subset Y$ and $y_n \to x \in X$. Since $\{y_n\}_{n=1}^\infty$ is convergent it is a Cauchy sequence in Y. Since $\dim Y < \infty$, we know Y is complete and therefore $\lim_{n\to\infty} y_n = y$ exists in Y. Since limits in a normed space are unique, it follows that $x = y \in Y$, i.e. Y is a closed subspace of X. ■

Lemma 19.18. *Let* $(X, \|\cdot\|)$ *be a Banach space,* $Y \subset X$ *be a proper closed subspace, and* $d_Y(x) := \inf_{y\in Y} \|x - y\|$ *be the distance from* x *to* Y. *Then* $d_Y(x) > 0$ *for all* $x \notin Y$ *and*

$$d_Y(ax + y) = |a| \, d_Y(x) \text{ for all } a \in \mathbb{C}, \ x \in X, \text{ and } y \in Y. \tag{19.6}$$

Proof. Since Y is proper and closed $X \setminus Y$ is a non-empty open subset of X. Therefore if $x \in X$, there exists $\varepsilon > 0$ such that $B_x(\varepsilon) \subset X \setminus Y$, i.e. $B_x(\varepsilon) \cap Y = \emptyset$. So for any $y \in Y$ we must have $\|x - y\| \ge \varepsilon$ and therefore that $d_Y(x) \ge \varepsilon > 0$. To prove Eq. (19.6) we may assume $a \ne 0$ otherwise both sides are zero. When $a \ne 0$ we have

$$\begin{aligned} d_Y(ax + y) &= \inf_{z\in Y} \|ax + y - z\| \\ &= |a| \inf_{z\in Y} \left\| x - \frac{1}{a}(z - y) \right\| = |a| \, d_Y(x) \end{aligned}$$

wherein we have used $\left\{\frac{1}{a}(z - y) : z \in Y\right\} = Y$. ■

Lemma 19.19. *Let* X *be a normed linear space and* Y *be a closed proper subspace of* X. *For any* $\alpha \in (0, 1)$ *there exists* $x \in X$ *such that* $\|x\| = 1$ *and* $\operatorname{dist}(x, Y) := d_Y(x) \ge \alpha$.

Proof. Let $z \in X \setminus Y$ and $\varepsilon := d_Y(z) > 0$ by Lemma 19.18. Choose $y \in Y$ so that $\varepsilon \geq \alpha \|z-y\|$ and set $x := \frac{z-y}{\|z-y\|}$. Then, again using Lemma 19.18,

$$d_Y(x) = d_Y\left(\frac{z-y}{\|z-y\|}\right) = \frac{1}{\|z-y\|} d_Y(z-y)$$
$$= \frac{1}{\|z-y\|} d_Y(z) = \frac{\varepsilon}{\|z-y\|} \geq \alpha.$$

∎

Proposition 19.20. *If the unit sphere in a Banach space, X, is compact then $\dim X < \infty$. [This implies again that a locally compact Banach space is finite dimensional.]*

Proof. Suppose that X is an infinite dimensional Banach space. We are going to construct a sequence $x_1, x_2, \ldots, x_n, \ldots$ such that $\|x_n\| = 1$, $\|x_i - x_j\| \geq 1/2$, $i \neq j$. Take x_1 to be any unit vector. Suppose vectors $x_1, \ldots, x_n$ are constructed. Let X_n be the linear span of $x_1, \ldots, x_n$. By Corollary 19.17, X_n is closed. By Lemma 19.19, there exists a unit vector x_{n+1} such that $d_{X_n}(x_{n+1}) \geq \frac{1}{2}$ and in particular $\|x_i - x_{n+1}\| \geq 1/2$ for $1 \leq i \leq n$. Now the sequence just constructed has no Cauchy subsequence. Hence the closed unit sphere is not compact. As spheres of different positive radii in X are homeomorphicly related by dilation, they are not compact either. In particular this implies that closed balls of positive radius are not compact as well. ∎

Theorem 19.21. *Suppose that $(X, \|\cdot\|)$ is a normed vector in which the unit ball, $V := B_0(1)$, is precompact. Then $\dim X < \infty$.*

Proof. First proof. This is a simple consequence of Proposition 19.20.
Second proof. Since $\bar{V}$ is compact, we may choose $\Lambda \subset\subset X$ such that

$$\bar{V} \subset \cup_{x \in \Lambda}\left(x + \frac{1}{2}V\right) \tag{19.7}$$

where, for any $\delta > 0$,

$$\delta V := \{\delta x : x \in V\} = B_0(\delta).$$

Let $Y := \mathrm{span}(\Lambda)$, then Eq. (19.7) implies,

$$V \subset \bar{V} \subset Y + \frac{1}{2}V.$$

Multiplying this equation by $\frac{1}{2}$ then shows

$$\frac{1}{2}V \subset \frac{1}{2}Y + \frac{1}{4}V = Y + \frac{1}{4}V$$

and hence

$$V \subset Y + \frac{1}{2}V \subset Y + Y + \frac{1}{4}V = Y + \frac{1}{4}V.$$

Continuing this way inductively then shows that

$$V \subset Y + \frac{1}{2^n}V \text{ for all } n \in \mathbb{N}. \tag{19.8}$$

Indeed, if Eq. (19.8) holds, then

$$V \subset Y + \frac{1}{2}V \subset Y + \frac{1}{2}\left(Y + \frac{1}{2^n}V\right) = Y + \frac{1}{2^{n+1}}V.$$

Hence if $x \in V$, there exists $y_n \in Y$ and $z_n \in B_0(2^{-n})$ such that $y_n + z_n \to x$. Since $\lim_{n\to\infty} z_n = 0$, it follows that $x = \lim_{n\to\infty} y_n \in \bar{Y}$. Since $\dim Y \leq \#(\Lambda) < \infty$, Corollary 19.17 implies $Y = \bar{Y}$ and so we have shown that $V \subset Y$. Since for any $x \in X$, $\frac{1}{2\|x\|}x \in V \subset Y$, we have $x \in Y$ for all $x \in X$, i.e. $X = Y$.
∎

Exercise 19.8. Suppose $(Y, \|\cdot\|_Y)$ is a normed space and $(X, \|\cdot\|_X)$ is a finite dimensional normed space. Show **every** linear transformation $T : X \to Y$ is necessarily bounded.

19.4 Compact Operators

Definition 19.22 (Compact Operator). *Let $A : X \to Y$ be a bounded operator between two Banach spaces. Then A is **compact** if $A[B_X(0,1)]$ is precompact in Y or equivalently for any $\{x_n\}_{n=1}^\infty \subset X$ such that $\|x_n\| \leq 1$ for all n the sequence $y_n := Ax_n \in Y$ has a convergent subsequence.*

Definition 19.23. *A bounded operator $A : X \to Y$ is said to have **finite rank** if $\mathrm{Ran}(A) \subset Y$ is finite dimensional.*

The following result is a simple consequence of Theorem 19.16 and Corollary 19.17.

Corollary 19.24. *If $A : X \to Y$ is a finite rank operator, then A is compact. In particular if either $\dim(X) < \infty$ or $\dim(Y) < \infty$ then any bounded operator $A : X \to Y$ is finite rank and hence compact.*

Theorem 19.25. *Let X and Y be Banach spaces and $\mathcal{K} := \mathcal{K}(X,Y)$ denote the compact operators from X to Y. Then $\mathcal{K}(X,Y)$ is a norm-closed subspace of $L(X,Y)$. In particular, operator norm limits of finite rank operators are compact.*

Proof. Using the sequential definition of compactness it is easily seen that $\mathcal{K}$ is a vector subspace of $L(X,Y)$. To finish the proof, we must show that $K \in L(X,Y)$ is compact if there exists $K_n \in \mathcal{K}(X,Y)$ such that $\lim_{n\to\infty} \|K_n - K\|_{op} = 0$.

First Proof. Let $U := B_0(1)$ be the unit ball in X. Given $\varepsilon > 0$, choose $N = N(\varepsilon)$ such that $\|K_N - K\| \le \varepsilon$. Using the fact that $K_N U$ is precompact, choose a finite subset $\Lambda \subset U$ such that $K_N U \subset \cup_{\sigma\in\Lambda} B_{K_N\sigma}(\varepsilon)$. Then given $y = Kx \in KU$ we have $K_N x \in B_{K_N\sigma}(\varepsilon)$ for some $\sigma \in \Lambda$ and for this σ;

$$\begin{aligned}\|y - K_N\sigma\| &= \|Kx - K_N\sigma\| \\ &\le \|Kx - K_N x\| + \|K_N x - K_N\sigma\| < \varepsilon\|x\| + \varepsilon < 2\varepsilon.\end{aligned}$$

This shows $KU \subset \cup_{\sigma\in\Lambda} B_{K_N\sigma}(2\varepsilon)$ and therefore is KU is 2ε – bounded for all $\varepsilon > 0$, i.e. KU is totally bounded and hence precompact.

Second Proof. Suppose $\{x_n\}_{n=1}^\infty$ is a bounded sequence in X. By compactness, there is a subsequence $\{x_n^1\}_{n=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ such that $\{K_1 x_n^1\}_{n=1}^\infty$ is convergent in Y. Working inductively, we may construct subsequences

$$\{x_n\}_{n=1}^\infty \supset \{x_n^1\}_{n=1}^\infty \supset \{x_n^2\}_{n=1}^\infty \cdots \supset \{x_n^m\}_{n=1}^\infty \supset \dots$$

such that $\{K_m x_n^m\}_{n=1}^\infty$ is convergent in Y for each m. By the usual Cantor's diagonalization procedure, let $\sigma_n := x_n^n$, then $\{\sigma_n\}_{n=1}^\infty$ is a subsequence of $\{x_n\}_{n=1}^\infty$ such that $\{K_m\sigma_n\}_{n=1}^\infty$ is convergent for all m. Since

$$\begin{aligned}\|K\sigma_n - K\sigma_l\| &\le \|(K - K_m)\sigma_n\| + \|K_m(\sigma_n - \sigma_l)\| + \|(K_m - K)\sigma_l)\| \\ &\le 2\|K - K_m\| + \|K_m(\sigma_n - \sigma_l)\|,\end{aligned}$$

$$\limsup_{n,l\to\infty} \|K\sigma_n - K\sigma_l\| \le 2\|K - K_m\| \to 0 \text{ as } m \to \infty,$$

which shows $\{K\sigma_n\}_{n=1}^\infty$ is Cauchy and hence convergent. ■

Example 19.26. Let $X = \ell^2 = Y$ and $\lambda_n \in \mathbb{C}$ such that $\lim_{n\to\infty} \lambda_n = 0$, then $A : X \to Y$ defined by $(Ax)(n) = \lambda_n x(n)$ is compact. To verify this claim, for each $m \in \mathbb{N}$ let $(A_m x)(n) = \lambda_n x(n) 1_{n\le m}$. In matrix language,

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \lambda_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix} \text{ and } A_m = \begin{pmatrix} \lambda_1 & 0 & \cdots & & & & \\ 0 & \lambda_2 & 0 & \cdots & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ & & & 0 & \lambda_m & 0 & \cdots \\ & & & \cdots & 0 & 0 & \ddots \\ & & & & & \ddots & \ddots \end{pmatrix}.$$

Then A_m is finite rank and $\|A - A_m\|_{op} = \max_{n>m} |\lambda_n| \to 0$ as $m \to \infty$. The claim now follows from Theorem 19.25.

We will see more examples of compact operators below in Section 19.5 and Exercise 19.19 below.

Lemma 19.27. *If $X \xrightarrow{A} Y \xrightarrow{B} Z$ are bounded operators between Banach spaces such the either A or B is compact then the composition $BA : X \to Z$ is also compact. In particular if $\dim X = \infty$ and $A \in L(X,Y)$ is an invertible operator such that*[2] *$A^{-1} \in L(Y,X)$, then A is **not** compact.*

Proof. Let $B_X(0,1)$ be the open unit ball in X. If A is compact and B is bounded, then $BA(B_X(0,1)) \subset B(\overline{AB_X(0,1)})$ which is compact since the image of compact sets under continuous maps are compact. Hence we conclude that $\overline{BA(B_X(0,1))}$ is compact, being the closed subset of the compact set $B(\overline{AB_X(0,1)})$. If A is continuous and B is compact, then $A(B_X(0,1))$ is a bounded set and so by the compactness of B, $BA(B_X(0,1))$ is a precompact subset of Z, i.e. BA is compact.

Alternatively: Suppose that $\{x_n\}_{n=1}^\infty \subset X$ is a bounded sequence. If A is compact, then $y_n := Ax_n$ has a convergent subsequence, $\{y_{n_k}\}_{k=1}^\infty$. Since B is continuous it follows that $z_{n_k} := By_{n_k} = BAx_{n_k}$ is a convergent subsequence of $\{BAx_n\}_{n=1}^\infty$. Similarly if A is bounded and B is compact then $y_n = Ax_n$ defines a bounded sequence inside of Y. By compactness of B, there is a subsequence $\{y_{n_k}\}_{k=1}^\infty$ for which $\{BAx_{n_k} = By_{n_k}\}_{k=1}^\infty$ is convergent in Z.

For the second statement, if A were compact then $I_X := A^{-1}A$ would be compact as well. As I_X takes the unit ball to the unit ball, the identity is compact iff $\dim X < \infty$. ■

Corollary 19.28. *Let X be a Banach space and $\mathcal{K}(X) := \mathcal{K}(X,X)$. Then $\mathcal{K}(X)$ is a norm-closed ideal of $L(X)$ which contains I_X iff $\dim X < \infty$.*

19.4.1 Compact Operators on a Hilbert Space

Lemma 19.29. *Suppose that $T, T_n \in L(X,Y)$ for $n \in \mathbb{N}$ where X and Y are normed spaces. If $T_n \xrightarrow{s} T$, $M = \sup_n \|T_n\| < \infty$,*[3] *and $x_n \to x$ in X as $n \to \infty$, then $T_n x_n \to Tx$ in Y as $n \to \infty$. Moreover if $K \subset X$ is a compact set then*

$$\lim_{n\to\infty} \sup_{x\in K} \|Tx - T_n x\| = 0. \tag{19.9}$$

Proof. 1. We have,

$$\begin{aligned}\|Tx - T_n x_n\| &\le \|Tx - T_n x\| + \|T_n x - T_n x_n\| \\ &\le \|Tx - T_n x\| + M\|x - x_n\| \to 0 \text{ as } n \to \infty.\end{aligned}$$

[2] Later we will see that A being one to one and onto automatically implies that A^{-1} is bounded by the open mapping Theorem 23.1.

[3] If X and Y are Banach spaces, the uniform boundedness principle covered below will show that $T_n \xrightarrow{s} T$ automatically implies $\sup_n \|T_n\| < \infty$.

2. For sake of contradiction, suppose that

$$\limsup_{n\to\infty}\sup_{x\in K}\|Tx-T_nx\|=\varepsilon>0.$$

In this case we can find $\{n_k\}_{k=1}^{\infty}\subset\mathbb{N}$ and $x_{n_k}\in K$ such that $\|Tx_{n_k}-T_{n_k}x_{n_k}\|\geq\varepsilon/2$. Since K is compact, by passing to a subsequence if necessary, we may assume $\lim_{k\to\infty}x_{n_k}=x$ exists in K. On the other hand by part **1.** we know that

$$\lim_{k\to\infty}\|Tx_{n_k}-T_{n_k}x_{n_k}\|=\left\|\lim_{k\to\infty}Tx_{n_k}-\lim_{k\to\infty}T_{n_k}x_{n_k}\right\|=\|Tx-Tx\|=0.$$

2 alternate proof. Given $\varepsilon>0$, there exists $\{x_1,\dots,x_N\}\subset K$ such that $K\subset\cup_{l=1}^{N}B_{x_l}(\varepsilon)$. If $x\in K$, choose l such that $x\in B_{x_l}(\varepsilon)$ in which case,

$$\begin{aligned}\|Tx-T_nx\|&\leq\|Tx-Tx_l\|+\|Tx_l-T_nx_l\|+\|T_nx_l-T_nx\|\\&\leq\left(\|T\|_{op}+M\right)\varepsilon+\|Tx_l-T_nx_l\|\end{aligned}$$

and therefore it follows that

$$\sup_{x\in K}\|Tx-T_nx\|\leq\left(\|T\|_{op}+M\right)\varepsilon+\max_{1\leq l\leq N}\|Tx_l-T_nx_l\|$$

and therefore,

$$\limsup_{n\to\infty}\sup_{x\in K}\|Tx-T_nx\|\leq\left(\|T\|_{op}+M\right)\varepsilon.$$

As $\varepsilon>0$ was arbitrary we conclude that Eq. (19.9) holds. ■

For the rest of this section, let H and B be Hilbert spaces and $U:=\{x\in H:\|x\|<1\}$ be the **open unit ball** in H.

Proposition 19.30. *A bounded operator $K:H\to B$ is compact iff there exists finite rank operators, $K_n:H\to B$, such that $\|K-K_n\|\to 0$ as $n\to\infty$.*

Proof. Suppose that $K:H\to B$. Since $\overline{K(U)}$ is compact it contains a countable dense subset and from this it follows that $\overline{K(H)}$ is a separable subspace of B. Let $\{\varphi_\ell\}$ be an orthonormal basis for $\overline{K(H)}\subset B$ and

$$P_ny=\sum_{\ell=1}^{n}\langle y|\varphi_\ell\rangle\varphi_\ell$$

be the orthogonal projection of y onto $\operatorname{span}\{\varphi_\ell\}_{\ell=1}^{n}$. Then $\lim_{n\to\infty}\|P_ny-y\|=0$ for all $y\in\overline{K(H)}$. Define $K_n:=P_nK$ – a finite rank operator on H. It then follows that

$$\begin{aligned}\limsup_{n\to\infty}\|K-K_n\|&=\limsup_{n\to\infty}\sup_{x\in U}\|Kx-K_nx\|\\&=\limsup_{n\to\infty}\sup_{x\in U}\|(I-P_n)Kx\|\\&\leq\limsup_{n\to\infty}\sup_{y\in\overline{K(U)}}\|(I-P_n)y\|=0\end{aligned}$$

by Lemma 19.29 along with the facts that $\overline{K(U)}$ is compact and $P_n\overset{s}{\to}I$. The converse direction follows from Corollary 19.24 and Theorem 19.25. ■

Corollary 19.31. *If K is compact then so is K^*.*

Proof. First Proof. Let $K_n=P_nK$ be as in the proof of Proposition 19.30, then $K_n^*=K^*P_n$ is still finite rank. Furthermore, using Proposition 18.18,

$$\|K^*-K_n^*\|=\|K-K_n\|\to 0\text{ as }n\to\infty$$

showing K^* is a limit of finite rank operators and hence compact.

Second Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a bounded sequence in B, then

$$\|K^*x_n-K^*x_m\|^2=\langle x_n-x_m|KK^*(x_n-x_m)\rangle\leq 2C\|KK^*(x_n-x_m)\|\tag{19.10}$$

where C is a bound on the norms of the x_n. Since $\{K^*x_n\}_{n=1}^{\infty}$ is also a bounded sequence, by the compactness of K there is a subsequence $\{x_n'\}$ of the $\{x_n\}$ such that KK^*x_n' is convergent and hence by Eq. (19.10), so is the sequence $\{K^*x_n'\}$.
■

19.5 Hilbert Schmidt Operators

In this section H and B will be Hilbert spaces.

Proposition 19.32. *Let H and B be a separable Hilbert spaces, $K:H\to B$ be a bounded linear operator, $\{e_n\}_{n=1}^{\infty}$ and $\{u_m\}_{m=1}^{\infty}$ be orthonormal basis for H and B respectively. Then:*

1. *$\sum_{n=1}^{\infty}\|Ke_n\|^2=\sum_{m=1}^{\infty}\|K^*u_m\|^2$ allowing for the possibility that the sums are infinite. In particular the **Hilbert Schmidt norm** of K,*

$$\|K\|_{HS}^2:=\sum_{n=1}^{\infty}\|Ke_n\|^2,$$

*is well defined independent of the choice of orthonormal basis $\{e_n\}_{n=1}^{\infty}$. We say $K:H\to B$ is a **Hilbert Schmidt operator** if $\|K\|_{HS}<\infty$ and let $HS(H,B)$ denote the space of Hilbert Schmidt operators from H to B.*

2. *For all* $K \in L(H,B)$, $\|K\|_{HS} = \|K^*\|_{HS}$ *and*

$$\|K\|_{HS} \geq \|K\|_{op} := \sup\{\|Kh\| : h \in H \text{ such that } \|h\| = 1\}.$$

3. *The set* $HS(H,B)$ *is a subspace of* $L(H,B)$ *(the bounded operators from* $H \to B$*),* $\|\cdot\|_{HS}$ *is a norm on* $HS(H,B)$ *for which* $(HS(H,B), \|\cdot\|_{HS})$ *is a Hilbert space, and the corresponding inner product is given by*

$$\langle K_1 | K_2 \rangle_{HS} = \sum_{n=1}^{\infty} \langle K_1 e_n | K_2 e_n \rangle. \tag{19.11}$$

4. *If* $K : H \to B$ *is a bounded finite rank operator, then* K *is Hilbert Schmidt.*
5. *Let* $P_N x := \sum_{n=1}^{N} \langle x | e_n \rangle e_n$ *be orthogonal projection onto* $\text{span}\{e_n : n \leq N\} \subset H$ *and for* $K \in HS(H,B)$*, let* $K_N := KP_N$*. Then*

$$\|K - K_N\|_{op}^2 \leq \|K - K_N\|_{HS}^2 \to 0 \text{ as } N \to \infty,$$

which shows that finite rank operators are dense in $(HS(H,B), \|\cdot\|_{HS})$*. In particular of* $HS(H,B) \subset \mathcal{K}(H,B)$ *– the space of compact operators from* $H \to B$*.*

6. *If* Y *is another Hilbert space and* $A : Y \to H$ *and* $C : B \to Y$ *are bounded operators, then*

$$\|KA\|_{HS} \leq \|K\|_{HS} \|A\|_{op} \text{ and } \|CK\|_{HS} \leq \|K\|_{HS} \|C\|_{op},$$

in particular $HS(H,H)$ *is an ideal in* $L(H)$*.*

Proof. Items 1. and 2. By Parseval's equality and Fubini's theorem for sums,

$$\sum_{n=1}^{\infty} \|Ke_n\|^2 = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |\langle Ke_n | u_m \rangle|^2$$
$$= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |\langle e_n | K^* u_m \rangle|^2 = \sum_{m=1}^{\infty} \|K^* u_m\|^2.$$

This proves $\|K\|_{HS}$ is well defined independent of basis and that $\|K\|_{HS} = \|K^*\|_{HS}$. For $x \in H \setminus \{0\}$, $x / \|x\|$ may be taken to be the first element in an orthonormal basis for H and hence

$$\left\| K \frac{x}{\|x\|} \right\| \leq \|K\|_{HS}.$$

Multiplying this inequality by $\|x\|$ shows $\|Kx\| \leq \|K\|_{HS} \|x\|$ and hence $\|K\|_{op} \leq \|K\|_{HS}$.

Item 3. For $K_1, K_2 \in L(H,B)$,

$$\begin{aligned}
\|K_1 + K_2\|_{HS} &= \sqrt{\sum_{n=1}^{\infty} \|K_1 e_n + K_2 e_n\|^2} \\
&\leq \sqrt{\sum_{n=1}^{\infty} [\|K_1 e_n\| + \|K_2 e_n\|]^2} \\
&= \|\{\|K_1 e_n\| + \|K_2 e_n\|\}_{n=1}^{\infty}\|_{\ell_2} \\
&\leq \|\{\|K_1 e_n\|\}_{n=1}^{\infty}\|_{\ell_2} + \|\{\|K_2 e_n\|\}_{n=1}^{\infty}\|_{\ell_2} \\
&= \|K_1\|_{HS} + \|K_2\|_{HS}.
\end{aligned}$$

From this triangle inequality and the homogeneity properties of $\|\cdot\|_{HS}$, we now easily see that $HS(H,B)$ is a subspace of $L(H,B)$ and $\|\cdot\|_{HS}$ is a norm on $HS(H,B)$. Since

$$\begin{aligned}
\sum_{n=1}^{\infty} |\langle K_1 e_n | K_2 e_n \rangle| &\leq \sum_{n=1}^{\infty} \|K_1 e_n\| \|K_2 e_n\| \\
&\leq \sqrt{\sum_{n=1}^{\infty} \|K_1 e_n\|^2} \sqrt{\sum_{n=1}^{\infty} \|K_2 e_n\|^2} = \|K_1\|_{HS} \|K_2\|_{HS},
\end{aligned}$$

the sum in Eq. (19.11) is well defined and is easily checked to define an inner product on $HS(H,B)$ such that $\|K\|_{HS}^2 = \langle K | K \rangle_{HS}$.

The proof that $\left(HS(H,B), \|\cdot\|_{HS}^2\right)$ is complete is very similar to the proof of Theorem 14.6. Indeed, suppose $\{K_m\}_{m=1}^{\infty}$ is a $\|\cdot\|_{HS}$ – Cauchy sequence in $HS(H,B)$. Because $L(H,B)$ is complete, there exists $K \in L(H,B)$ such that $\|K - K_m\|_{op} \to 0$ as $m \to \infty$. Thus, making use of Fatou's Lemma 4.12,

$$\begin{aligned}
\|K - K_m\|_{HS}^2 &= \sum_{n=1}^{\infty} \|(K - K_m) e_n\|^2 \\
&= \sum_{n=1}^{\infty} \liminf_{l \to \infty} \|(K_l - K_m) e_n\|^2 \\
&\leq \liminf_{l \to \infty} \sum_{n=1}^{\infty} \|(K_l - K_m) e_n\|^2 \\
&= \liminf_{l \to \infty} \|K_l - K_m\|_{HS}^2 \to 0 \text{ as } m \to \infty.
\end{aligned}$$

Hence $K \in HS(H,B)$ and $\lim_{m \to \infty} \|K - K_m\|_{HS}^2 = 0$.

Item 4. Since $\text{Nul}(K^*)^\perp = \overline{\text{Ran}(K)} = \text{Ran}(K)$,

$$\|K\|_{HS}^2 = \|K^*\|_{HS}^2 = \sum_{n=1}^{N} \|K^* v_n\|_H^2 < \infty$$

where $N := \dim \text{Ran}(K)$ and $\{v_n\}_{n=1}^N$ is an orthonormal basis for $\text{Ran}(K) = K(H)$.

Item 5. Simply observe,

$$\|K - K_N\|_{op}^2 \le \|K - K_N\|_{HS}^2 = \sum_{n>N} \|Ke_n\|^2 \to 0 \text{ as } N \to \infty.$$

Item 6. For $C \in L(B,Y)$ and $K \in L(H,B)$ then

$$\|CK\|_{HS}^2 = \sum_{n=1}^{\infty} \|CKe_n\|^2 \le \|C\|_{op}^2 \sum_{n=1}^{\infty} \|Ke_n\|^2 = \|C\|_{op}^2 \|K\|_{HS}^2$$

and for $A \in L(Y,H)$,

$$\|KA\|_{HS} = \|A^* K^*\|_{HS} \le \|A^*\|_{op} \|K^*\|_{HS} = \|A\|_{op} \|K\|_{HS}.$$

■

Remark 19.33. The separability assumptions made in Proposition 19.32 are unnecessary. In general, we define

$$\|K\|_{HS}^2 = \sum_{e \in \beta} \|Ke\|^2$$

where $\beta \subset H$ is an orthonormal basis. The same proof of Item 1. of Proposition 19.32 shows $\|K\|_{HS}$ is well defined and $\|K\|_{HS} = \|K^*\|_{HS}$. If $\|K\|_{HS}^2 < \infty$, then there exists a countable subset $\beta_0 \subset \beta$ such that $Ke = 0$ if $e \in \beta \setminus \beta_0$. Let $H_0 := \overline{\text{span}(\beta_0)}$ and $B_0 := \overline{K(H_0)}$. Then $K(H) \subset B_0$, $K|_{H_0^\perp} = 0$ and hence by applying the results of Proposition 19.32 to $K|_{H_0} : H_0 \to B_0$ one easily sees that the separability of H and B are unnecessary in Proposition 19.32.

Example 19.34. Let (X,μ) be a measure space, $H = L^2(X,\mu)$ and

$$k(x,y) := \sum_{i=1}^{n} f_i(x) g_i(y)$$

where

$$f_i, g_i \in L^2(X,\mu) \text{ for } i = 1, \dots, n.$$

Define

$$(Kf)(x) = \int_X k(x,y) f(y) d\mu(y),$$

then $K : L^2(X,\mu) \to L^2(X,\mu)$ is a finite rank operator and hence Hilbert Schmidt.

Exercise 19.9. Suppose that (X,μ) is a σ–finite measure space such that $H = L^2(X,\mu)$ is separable and $k : X \times X \to \mathbb{R}$ is a measurable function, such that

$$\|k\|_{L^2(X\times X,\mu\otimes\mu)}^2 := \int_{X\times X} |k(x,y)|^2 d\mu(x) d\mu(y) < \infty.$$

Define, for $f \in H$,

$$Kf(x) = \int_X k(x,y) f(y) d\mu(y),$$

when the integral makes sense. Show:

1. $Kf(x)$ is defined for μ–a.e. x in X.
2. The resulting function Kf is in H and $K : H \to H$ is linear.
3. $\|K\|_{HS} = \|k\|_{L^2(X\times X,\mu\otimes\mu)} < \infty$. (This implies $K \in HS(H,H)$.)

Exercise 19.10 (Converse to Exercise 19.9). Suppose that (X,μ) is a σ–finite measure space such that $H = L^2(X,\mu)$ is separable and $K : H \to H$ is a Hilbert Schmidt operator. Show there exists $k \in L^2(X \times X, \mu \otimes \mu)$ such that K is the integral operator associated to k, i.e.

$$Kf(x) = \int_X k(x,y) f(y) d\mu(y). \tag{19.12}$$

In fact you should show

$$k(x,y) := \sum_{n=1}^{\infty} \left(\overline{(K^* \varphi_n)}(y) \right) \varphi_n(x) \;\; (L^2(\mu\otimes\mu) \text{ – convergent sum}) \tag{19.13}$$

where $\{\varphi_n\}_{n=1}^\infty$ is any orthonormal basis for H.

19.6 Weak Convergence and Compactness in Hilbert Spaces

We have seen above that in infinite dimensions it is no longer true that norm bounded and closed sets are compact. In order to recover compactness we have to weaken our notion of convergence. [The practical price for doing this is that functions which are norm continuous need not be continuous relative to a weaker notion of convergence.] In this section we will always assume that H is a Hilbert space.

Definition 19.35 (Weak Convergence). *We say a sequence $\{x_n\}_{n=1}^\infty$ of a Hilbert space, H,* ***converges weakly*** *to $x \in H$ (and denote this by writing $x_n \xrightarrow{w} x \in H$ as $n \to \infty$) iff $\lim_{n\to\infty} \langle x_n | y \rangle = \langle x | y \rangle$ for all $y \in H$.*

Remark 19.36. Suppose that H is an infinite dimensional Hilbert space $\{x_n\}_{n=1}^\infty$ is an orthonormal subset of H. Then Bessel's inequality (Proposition 18.21) implies $x_n \overset{w}{\to} 0 \in H$ as $n \to \infty$. This points out the fact that if $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, it is no longer necessarily true that $\|x\| = \lim_{n\to\infty} \|x_n\|$, i.e. the norm is **not** weakly continuous. However we do always have $\|x\| \le \liminf_{n\to\infty} \|x_n\|$ because,

$$\|x\|^2 = \lim_{n\to\infty} \langle x_n|x\rangle \le \liminf_{n\to\infty} [\|x_n\| \|x\|] = \|x\| \liminf_{n\to\infty} \|x_n\|.$$

Exercise 19.11. Suppose that $\{x_n\}_{n=1}^\infty \subset H$ and $x_n \overset{w}{\to} x \in H$ as $n \to \infty$. Show $x_n \to x$ as $n \to \infty$ (i.e. $\lim_{n\to\infty} \|x - x_n\| = 0$) iff $\lim_{n\to\infty} \|x_n\| = \|x\|$.

Exercise 19.12. Suppose that $\{x_n\}_{n=1}^\infty$ and $\{y_n\}_{n=1}^\infty$ are sequences in H such that $y_n \overset{s}{\to} y \in H$ (i.e. $\lim_{n\to\infty} \|y - y_n\|_H = 0$), $x_n \overset{w}{\to} x \in H$, and[4] $M := \sup_n \|x_n\| < \infty$ as $n \to \infty$. Show $\lim_{n\to\infty} \langle x_n|y_n\rangle = \langle x|y\rangle$.

Exercise 19.13. Suppose that $K : H \to H$ is a compact operator, $x_n \overset{w}{\to} x \in H$, and[5] $M := \sup_n \|x_n\| < \infty$ as $n \to \infty$. Show $Kx_n \overset{s}{\to} Kx$ as $n \to \infty$. Combine this with Exercise 19.12 in order to show $f(x) := \langle Kx|x\rangle$ is weakly sequentially continuous on $C := \{x \in H : \|x\|_H \le 1\}$, i.e. if $\{x_n\} \subset C$ and $x_n \overset{w}{\to} x \in H$ then $\lim_{n\to\infty} f(x_n) = f(x)$.

Proposition 19.37. *Let H be a Hilbert space, $\beta \subset H$ be an orthonormal basis for H and $\{x_n\}_{n=1}^\infty \subset H$ be a bounded sequence, then the following are equivalent:*

1. $x_n \overset{w}{\to} x \in H$ *as* $n \to \infty$.
2. $\langle x|y\rangle = \lim_{n\to\infty} \langle x_n|y\rangle$ *for all* $y \in H$.
3. $\langle x|y\rangle = \lim_{n\to\infty} \langle x_n|y\rangle$ *for all* $y \in \beta$.

Moreover, if $c_y := \lim_{n\to\infty} \langle x_n|y\rangle$ exists for all $y \in \beta$, then $\sum_{y\in\beta} |c_y|^2 < \infty$ and $x_n \overset{w}{\to} x := \sum_{y\in\beta} c_y y \in H$ as $n \to \infty$.

Proof. 1. $\Longleftrightarrow$ 2. is simply the definition of weak convergence. 2. $\Longrightarrow$ 3. is trivial.

3. $\Longrightarrow$ 1. Let $M := \sup_n \|x_n\|$ and H_0 denote the **algebraic** span of β so that $\lim_{n\to\infty} \langle x_n|z\rangle = \langle x|z\rangle$ for all $x \in H_0$. For any $y \in H$ and $z \in H_0$

$$|\langle x - x_n|y\rangle| \le |\langle x - x_n|z\rangle| + |\langle x - x_n|y - z\rangle| \le |\langle x - x_n|z\rangle| + 2M \|y - z\|.$$

Passing to the limit in this equation implies $\limsup_{n\to\infty} |\langle x - x_n|y\rangle| \le 2M \|y - z\|$ which shows $\limsup_{n\to\infty} |\langle x - x_n|y\rangle| = 0$ since H_0 is dense in H.

To prove the last assertion, let $\Gamma \subset\subset \beta$. Then by Bessel's inequality (Proposition 18.21),

$$\sum_{y\in\Gamma} |c_y|^2 = \lim_{n\to\infty} \sum_{y\in\Gamma} |\langle x_n|y\rangle|^2 \le \liminf_{n\to\infty} \|x_n\|^2 \le M^2.$$

Since $\Gamma \subset\subset \beta$ was arbitrary, we conclude that $\sum_{y\in\beta} |c_y|^2 \le M < \infty$ and hence we may define $x := \sum_{y\in\beta} c_y y$. By construction we have

$$\langle x|y\rangle = c_y = \lim_{n\to\infty} \langle x_n|y\rangle \text{ for all } y \in \beta$$

and hence $x_n \overset{w}{\to} x \in H$ as $n \to \infty$ by what we have just proved. ∎

Theorem 19.38 (Weak Sequential Compactness). *Suppose $\{x_n\}_{n=1}^\infty$ is a bounded sequence in a Hilbert space, H. Then there exists a subsequence $y_k := x_{n_k}$ of $\{x_n\}_{n=1}^\infty$ and $x \in X$ such that $y_k \overset{w}{\to} x$ as $k \to \infty$.*

Proof. Let $H_0 := \overline{\operatorname{span}(x_k : k \in \mathbb{N})}$ and $M := \sup_n \|x_n\|_H < \infty$. Then H_0 is a closed separable Hilbert subspace of H and $\{x_k\}_{k=1}^\infty \subset H_0$. Let $\beta_0 := \{h_n\}_{n=1}^\infty$ be an orthonormal basis for H_0. (I am assuming $\dim H_0 = \infty$ as this is the more difficult case.) Since $|\langle x_k|h_n\rangle| \le \|x_k\| \|h_n\| \le M < \infty$, the sequence, $\{\langle x_k|h_n\rangle\}_{k=1}^\infty \subset \mathbb{C}$, is bounded and hence has a convergent sub-sequence for all $n \in \mathbb{N}$. By the Cantor's diagonalization argument we can find a a sub-sequence, $y_k := x_{n_k}$, of $\{x_n\}$ such that $\lim_{k\to\infty} \langle y_k|h_n\rangle$ exists for all $n \in \mathbb{N}$. Thus by Proposition 19.37, there exists a $x \in H_0$ such that $y_k \overset{w}{\to} x$ (weakly in H_0) as $k \to \infty$. For an arbitrary $z \in H$, decompose z as $z = z_0 + z_1$ where $z_0 \in H_0$ and $z_1 \in H_0^\perp$. Then

$$\lim_{k\to\infty} \langle y_k|z\rangle = \lim_{k\to\infty} \langle y_k|z_0\rangle = \langle x|z_0\rangle = \langle x|z\rangle$$

wherein we have use $\langle y_k|z_1\rangle = 0$ for all k and $\langle x|z_1\rangle = 0$. ∎

Exercise 19.14. Suppose that H is a Hilbert space and $K : H \to H$ is a non-zero compact operator. Show there exists

$$x_0 \in C := \{x \in H : \|x\|_H \le 1\}$$

such that

$$|\langle Kx_0|x_0\rangle| = \sup_{x\in C} |\langle Kx|x\rangle| =: M.$$

Hint: see Exercise 19.13. Also explain why we may assume that $\|x_0\| = 1$.

[4] The uniform boundedness principle covered below will show that the next hypothesis is redundant.

[5] The uniform boundedness principle covered below will show that the next hypothesis is redundant.

Exercise 19.15. Suppose that $A : H \to H$ is a bounded self-adjoint operator on H. Show;

1. $f(x) := \langle Ax|x\rangle \in \mathbb{R}$ for all $x \in H$.
2. If there exists $x_0 \in H$ with $\|x_0\| = 1$ such that
$$\lambda_0 := \sup_{\|x\|=1} \langle Ax|x\rangle = \langle Ax_0|x_0\rangle$$
then $Ax_0 = \lambda_0 x_0$. **Hint:** Given $y \in H$ let $c(t) := \frac{x_0+ty}{\|x_0+ty\|_H}$ for t near 0. Then apply the first derivative test to the function $g(t) = \langle Ac(t)|c(t)\rangle$.
3. If we further assume that A is compact, then A has at least one eigenvector.

Exercise 19.16 (Metrizing Weak Convergence). Suppose that H is a separable Hilbert space, $C := \{x \in H : \|x\| \le 1\}$ is the closed unit ball in H, and $\{e_\ell\}_{\ell=1}^\infty$ is an orthonormal basis for H. For $x, y \in H$ let

$$\rho(x,y) := \sum_{\ell=1}^\infty \frac{1}{2^\ell} |\langle x-y|e_\ell\rangle| . \tag{19.14}$$

Show;

1. (H,ρ) is a metric space.
2. Show if $\{x_n\}_{n=1}^\infty \subset C$ then $x_n \overset{w}{\to} x \in H$ iff$\rho(x_n,x) \to 0$.

Theorem 19.39 (Alaoglu's Theorem for Hilbert Spaces). *Suppose that H is a separable Hilbert space, $C := \{x \in H : \|x\| \le 1\}$ is the closed unit ball in H, and $\{e_n\}_{n=1}^\infty$ is an orthonormal basis for H, and ρ is the metric defined in Eq. (19.14). Then (C,ρ) is a compact metric space. (This theorem will be extended to Banach spaces, see Theorems 36.20 and 36.25 below.)*

Proof. This is a simple corollary of Theorem 19.38, Exercise 19.16, and Theorem 19.9. ■

Exercise 19.17 (Banach-Saks). Suppose that $\{x_n\}_{n=1}^\infty \subset H$, $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, and $c := \sup_n \|x_n\| < \infty$.[6] Show there exists a subsequence, $y_k = x_{n_k}$ such that
$$\lim_{N\to\infty} \left\| x - \frac{1}{N}\sum_{k=1}^N y_k \right\| = 0,$$
i.e. $\frac{1}{N}\sum_{k=1}^N y_k \to x$ as $N \to \infty$. **Hints:** 1. show it suffices to assume $x = 0$ and then choose $\{y_k\}_{k=1}^\infty$ so that $|\langle y_k|y_l\rangle| \le l^{-1}$ (or even smaller if you like) for all $k \le l$.

[6] The assumption that $c < \infty$ is superfluous because of the "uniform boundedness principle," see Theorem 23.9 below.

19.7 Arzela-Ascoli Theorem (Compactness in $C(X)$)

In this section, let (X,τ) be a topological space.

Definition 19.40. *Let $\mathcal{F} \subset C(X)$.*

1. $\mathcal{F}$ is **equicontinuous at** $x \in X$ iff for all $\varepsilon > 0$ there exists $U \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon$ for all $y \in U$ and $f \in \mathcal{F}$.
2. $\mathcal{F}$ is **equicontinuous** if $\mathcal{F}$ is equicontinuous at all points $x \in X$.
3. $\mathcal{F}$ is **pointwise bounded** if $\sup\{|f(x)| : f \in \mathcal{F}\} < \infty$ for all $x \in X$.

Theorem 19.41 (Ascoli-Arzela Theorem). *Let (X,τ) be a compact topological space and $\mathcal{F} \subset C(X)$. Then $\mathcal{F}$ is precompact in $C(X)$ iff $\mathcal{F}$ is equicontinuous and point-wise bounded.*

Proof. ($\Leftarrow$) Since $C(X) \subset \ell^\infty(X)$ is a complete metric space, we must show $\mathcal{F}$ is totally bounded. Let $\varepsilon > 0$ be given. By equicontinuity, for all $x \in X$, there exists $V_x \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon$ if $y \in V_x$ and $f \in \mathcal{F}$. Since X is compact we may choose $\Lambda \subset_f X$ such that $X = \cup_{x\in\Lambda} V_x$. We have now decomposed X into "blocks" $\{V_x\}_{x\in\Lambda}$ such that each $f \in \mathcal{F}$ is constant to within ε on V_x. Since $\sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} < \infty$, it is now evident that

$$\begin{aligned} M &:= \sup\{|f(x)| : x \in X \text{ and } f \in \mathcal{F}\} \\ &\le \sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} + \varepsilon < \infty. \end{aligned}$$

Let $D_M := \{z \in \mathbb{C} : |z| \le M\}$ be the closed M - disk in $\mathbb{C}$ centered at 0. Then D_M^Λ is a compact subset of $\ell^\infty(\Lambda)$. Therefore for all $\varepsilon > 0$ there exists a finite subset $\Gamma \subset_f D_M^\Lambda$ such that $D_M^\Lambda \subset \cup_{\varphi\in\Gamma} B_\varphi^{\ell^\infty(\Lambda)}(\varepsilon)$. By construction if $f \in \mathcal{F}$, then $f|_\Lambda \in D_M^\Lambda$ and therefore there exists $\varphi \in \Gamma$ such that $\|f|_\Lambda - \varphi\|_{\ell^\infty(\Lambda)} < \varepsilon$. This shows that $\mathcal{F} = \bigcup_{\varphi\in\Gamma} \mathcal{F}_\varphi$ where, for $\varphi \in \Gamma$,

$$\mathcal{F}_\varphi := \{f \in \mathcal{F} : \|f|_\Lambda - \varphi\|_{\ell^\infty(\Lambda)} < \varepsilon\}.$$

Let $\tilde{\Gamma} = \{\varphi \in \Gamma : \mathcal{F}_\varphi \ne \emptyset\}$ and for $\varphi \in \tilde{\Gamma}$ let $f_\varphi \in \mathcal{F}_\varphi$. If $f \in \mathcal{F}_\varphi$ and $y \in X$, choose $x \in \Lambda$ such that $y \in V_x$. Then we find

$$\begin{aligned} |f(y) - f_\varphi(y)| &\le |f(y) - f(x)| + |f(x) - \varphi(x)| \\ &\quad + |\varphi(x) - f_\varphi(x)| + |f_\varphi(x) - f_\varphi(y)| < 4\varepsilon. \end{aligned}$$

As $y \in X$ is arbitrary we have shown $\|f - f_\varphi\|_\infty \le 4\varepsilon$ for all $f \in \mathcal{F}_\varphi$, i.e.

$$\mathcal{F} = \bigcup_{\varphi\in\tilde{\Gamma}} \mathcal{F}_\varphi \subset \bigcup_{\varphi\in\tilde{\Gamma}} B_{f_\varphi}^{C(X)}(5\varepsilon).$$

This shows that $\mathcal{F}$ is 5ε – bounded for all $\varepsilon > 0$.

($\Rightarrow$) (*The rest of this proof may safely be skipped.) Let me first give the argument under the added restriction that $\tau = \tau_d$ for some metric, d, and X. Since $\|\cdot\|_\infty : C(X) \to [0,\infty)$ is a continuous function on $C(X)$ it is bounded on any compact subset $\mathcal{F} \subset C(X)$. This shows that $\sup\{\|f\|_\infty : f \in \mathcal{F}\} < \infty$ which clearly implies that $\mathcal{F}$ is pointwise bounded.[7] Suppose $\mathcal{F}$ were **not** equicontinuous at some point $x \in X$, i.e. there exists $\varepsilon > 0$ such that for all $V \in \tau_x$, $\sup_{y\in V} \sup_{f\in\mathcal{F}} |f(y) - f(x)| > \varepsilon$. Let $\{V_n = B_x(1/n)\}_{n=1}^\infty$. By the assumption that $\mathcal{F}$ is not equicontinuous at x, there exist $f_n \in \mathcal{F}$ and $x_n \in V_n$ such that $|f_n(x) - f_n(x_n)| \ge \varepsilon$ for all n. Since $\overline{\mathcal{F}}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some $f \in \mathcal{F}$. Because $x_n \to x$ as $n \to \infty$ we learn that

$$\begin{aligned}\varepsilon &\le |f_n(x) - f_n(x_n)| \le |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\ &\le 2\|f_n - f\|_\infty + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty\end{aligned}$$

which is a contradiction.

($\Rightarrow$) (Here is the general argument for arbitrary topological spaces.) As in the proof above, $\mathcal{F}$ is pointwise bounded. Suppose $\mathcal{F}$ were **not** equicontinuous at some point $x \in X$, i.e. there exists $\varepsilon > 0$ such that for all $V \in \tau_x$, $\sup_{y\in V} \sup_{f\in\mathcal{F}} |f(y) - f(x)| > \varepsilon$.[8] Equivalently said, to each $V \in \tau_x$ we may choose

$$f_V \in \mathcal{F} \text{ and } x_V \in V \ni |f_V(x) - f_V(x_V)| \ge \varepsilon.$$

Set $\mathcal{C}_V = \overline{\{f_W : W \in \tau_x \text{ and } W \subset V\}}^{\|\cdot\|_\infty} \subset \mathcal{F}$ and notice for any $\mathcal{V} \subset_f \tau_x$ that

$$\cap_{V\in\mathcal{V}} \mathcal{C}_V \supseteq \mathcal{C}_{\cap\mathcal{V}} \ne \emptyset,$$

so that $\{\mathcal{C}_V\}_V \in \tau_x \subset \mathcal{F}$ has the finite intersection property.[9] Since $\mathcal{F}$ is compact, it follows that there exists some

$$f \in \bigcap_{V\in\tau_x} \mathcal{C}_V \ne \emptyset.$$

Since f is continuous, there exists $V \in \tau_x$ such that $|f(x) - f(y)| < \varepsilon/3$ for all $y \in V$. Because $f \in \mathcal{C}_V$, there exists $W \subset V$ such that $\|f - f_W\| < \varepsilon/3$. We now arrive at a contradiction;

$$\begin{aligned}\varepsilon &\le |f_W(x) - f_W(x_W)| \\ &\le |f_W(x) - f(x)| + |f(x) - f(x_W)| + |f(x_W) - f_W(x_W)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

■

Exercise 19.18. Give an alternative proof of the implication, ($\Leftarrow$), in Theorem 19.41 by showing every subsequence $\{f_n : n \in \mathbb{N}\} \subset \mathcal{F}$ has a convergence subsequence.

Exercise 19.19. Suppose $k \in C\left([0,1]^2, \mathbb{R}\right)$ and for $f \in C([0,1], \mathbb{R})$, let

$$Kf(x) := \int_0^1 k(x,y) f(y)\, dy \text{ for all } x \in [0,1]. \tag{19.15}$$

Show K is a **compact operator** on $(C([0,1],\mathbb{R}), \|\cdot\|_\infty)$ which by definition means that $\overline{\{Kf : \|f\|_\infty \le 1\}}$ is compact in $C([0,1],\mathbb{R})$.

Exercise 19.20 (Continuation of Exercise 19.19). Suppose $k \in C\left([0,1]^2, \mathbb{R}\right)$ but now show that one can use the formula for K in Eq. (19.15) in order to define a linear operator from $L^2([0,1], m; \mathbb{R})$ to $C([0,1],\mathbb{R})$ and that this operator is still compact, i.e. $\overline{\{Kf : \|f\|_{L^2} \le 1\}}$ is compact in $C([0,1],\mathbb{R})$. Conclude that K is also a compact operator when viewed as an operator from $L^2([0,1], m; \mathbb{R})$ to $L^2([0,1], m; \mathbb{R})$.

The following result is a corollary of Theorem 19.41.

Corollary 19.42 (Locally Compact Ascoli-Arzela Theorem). *Let (X,d) be a metric space with the Heine Borel property, i.e. closed and bounded sets are compact. If $\{f_m\} \subset C(X)$ is a pointwise bounded sequence of functions such that $\{f_m|_K\}$ is equicontinuous for any compact subset $K \subset X$, then there exists a subsequence $\{m_n\} \subset \{m\}$ such that $\{g_n := f_{m_n}\}_{n=1}^\infty \subset C(X)$ is uniformly convergent on all compact subsets of X.*

---

[7] One could also prove that $\mathcal{F}$ is pointwise bounded by considering the continuous evaluation maps $e_x : C(X) \to \mathbb{R}$ given by $e_x(f) = f(x)$ for all $x \in X$.

[8] If X is first countable we could finish the proof with the following argument. Let $\{V_n\}_{n=1}^\infty$ be a neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By the assumption that $\mathcal{F}$ is not equicontinuous at x, there exist $f_n \in \mathcal{F}$ and $x_n \in V_n$ such that $|f_n(x) - f_n(x_n)| \ge \epsilon\ \forall\ n$. Since $\mathcal{F}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some $f \in \mathcal{F}$. Because $x_n \to x$ as $n \to \infty$ we learn that

$$\begin{aligned}\epsilon &\le |f_n(x) - f_n(x_n)| \le |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\ &\le 2\|f_n - f\| + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty\end{aligned}$$

which is a contradiction.

[9] If we are willing to use Net's described in Appendix ?? below we could finish the proof as follows. Since $\mathcal{F}$ is compact, the net $\{f_V\}_{V\in\tau_x} \subset \mathcal{F}$ has a cluster point $f \in \mathcal{F} \subset C(X)$. Choose a subnet $\{g_\alpha\}_{\alpha\in A}$ of $\{f_V\}_{V\in\tau_X}$ such that $g_\alpha \to f$ uniformly. Then, since $x_V \to x$ implies $x_{V_\alpha} \to x$, we may conclude from Eq. (36.1) that

$$\epsilon \le |g_\alpha(x) - g_\alpha(x_{V_\alpha})| \to |g(x) - g(x)| = 0$$

which is a contradiction.

Proof. Let $x \in X$ and $\{K_n = C_x(n)\}_{n=1}^\infty$, so that $\{K_n\}$ is a nested sequence of compact sets such that $K_n \uparrow X$ as $n \to \infty$. We may now apply Theorem 19.41 repeatedly to find a nested family of subsequences

$$\{f_m\} \supset \{g_m^1\} \supset \{g_m^2\} \supset \{g_m^3\} \supset \dots$$

such that the sequence $\{g_m^n\}_{m=1}^\infty \subset C(X)$ is uniformly convergent on K_n. Using Cantor's trick, define the subsequence $\{h_n\}$ of $\{f_m\}$ by $h_n := g_n^n$. Then $\{h_n\}$ is uniformly convergent on K_l for each $l \in \mathbb{N}$. Now if $K \subset X$ is an arbitrary compact set, there exists $l < \infty$ such that $K \subset K_l^o \subset K_l$ and therefore $\{h_n\}$ is uniformly convergent on K as well. ∎

Proposition 19.43. *Let $\Omega \subset_o \mathbb{R}^d$ such that $\bar{\Omega}$ is compact and $0 \le \alpha < \beta \le 1$. Then the inclusion map $i : C^\beta(\overline{\Omega}) \hookrightarrow C^\alpha(\overline{\Omega})$ is a compact operator. See Chapter 15 and Lemma 15.9 for the notation being used here.*

Let $\{u_n\}_{n=1}^\infty \subset C^\beta(\overline{\Omega})$ such that $\|u_n\|_{C^\beta} \le 1$, i.e. $\|u_n\|_\infty \le 1$ and

$$|u_n(x) - u_n(y)| \le |x-y|^\beta \text{ for all } x, y \in \overline{\Omega}.$$

By the Arzela-Ascoli Theorem 19.41, there exists a subsequence of $\{\tilde{u}_n\}_{n=1}^\infty$ of $\{u_n\}_{n=1}^\infty$ and $u \in C^o(\bar{\Omega})$ such that $\tilde{u}_n \to u$ in C^0. Since

$$|u(x) - u(y)| = \lim_{n\to\infty} |\tilde{u}_n(x) - \tilde{u}_n(y)| \le |x-y|^\beta,$$

$u \in C^\beta$ as well. Define $g_n := u - \tilde{u}_n \in C^\beta$, then

$$[g_n]_\beta + \|g_n\|_{C^0} = \|g_n\|_{C^\beta} \le 2$$

and $g_n \to 0$ in C^0. To finish the proof we must show that $g_n \to 0$ in C^α. Given $\delta > 0$,

$$[g_n]_\alpha = \sup_{x\ne y} \frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} \le A_n + B_n$$

where

$$\begin{aligned} A_n &= \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} : x \ne y \text{ and } |x-y| \le \delta\right\} \\ &= \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\beta} \cdot |x-y|^{\beta-\alpha} : x \ne y \text{ and } |x-y| \le \delta\right\} \\ &\le \delta^{\beta-\alpha} \cdot [g_n]_\beta \le 2\delta^{\beta-\alpha} \end{aligned}$$

and

$$B_n = \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} : |x-y| > \delta\right\} \le 2\delta^{-\alpha} \|g_n\|_{C^0} \to 0 \text{ as } n \to \infty.$$

Therefore,

$$\limsup_{n\to\infty} [g_n]_\alpha \le \limsup_{n\to\infty} A_n + \limsup_{n\to\infty} B_n \le 2\delta^{\beta-\alpha} + 0 \to 0 \text{ as } \delta \downarrow 0.$$

This proposition generalizes to the following theorem which the reader is asked to prove in Exercise 19.28 below.

Theorem 19.44. *Let Ω be a precompact open subset of $\mathbb{R}^d$, $\alpha, \beta \in [0,1]$ and $k, j \in \mathbb{N}_0$. If $j + \beta > k + \alpha$, then $C^{j,\beta}(\bar{\Omega})$ is compactly contained in $C^{k,\alpha}(\bar{\Omega})$.*

19.8 Exercises

19.8.1 Metric Spaces as Topological Spaces

Definition 19.45. *Two metrics d and ρ on a set X are said to be **equivalent** if there exists a constant $c \in (0,\infty)$ such that $c^{-1}\rho \le d \le c\rho$.*

Exercise 19.21. Suppose that d and ρ are two metrics on X.

1. Show $\tau_d = \tau_\rho$ if d and ρ are equivalent.
2. Show by example that it is possible for $\tau_d = \tau_\rho$ even though d and ρ are inequivalent.

Exercise 19.22. Let (X_i, d_i) for $i = 1, \dots, n$ be a finite collection of metric spaces and for $1 \le p \le \infty$ and $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $X := \prod_{i=1}^n X_i$, let

$$\rho_p(x,y) = \begin{cases} \left(\sum_{i=1}^n [d_i(x_i,y_i)]^p\right)^{1/p} & \text{if } p \ne \infty \\ \max_i d_i(x_i,y_i) & \text{if } p = \infty \end{cases}.$$

1. Show (X, ρ_p) is a metric space for $p \in [1,\infty]$. **Hint**: Minkowski's inequality.
2. Show for any $p, q \in [1,\infty]$, the metrics ρ_p and ρ_q are equivalent. **Hint:** This can be done with explicit estimates or you could use Theorem 19.16 below.

Exercise 19.23 (Tychonoff's Theorem for Compact Metric Spaces). Let $\{(X_n, d_n)\}_{n=1}^\infty$ be a sequence of compact metric spaces, $X := \prod_{n=1}^\infty X_n$, and for $x = (x(n))_{n=1}^\infty$ and $y = (y(n))_{n=1}^\infty$ in X let

$$d(x,y) = \sum_{n=1}^\infty 2^{-n} \frac{d_n(x(n), y(n))}{1 + d_n(x(n), y(n))},$$

as in Exercise 13.16. Further assume that the spaces X_n are compact for all n. Show (without using the general form of Tychonoff's Theorem 36.16 below) that (X, d) is compact. **Hint:** Either use Cantor's method to show every sequence $\{x_m\}_{m=1}^\infty \subset X$ has a convergent subsequence or alternatively show (X, d) is complete and totally bounded. (Compare with Example 19.14.)

Exercise 19.24. Let $X = \mathbb{R}$ and set $d(x,y) := |y-x|$ and $\rho(x,y) := \left|\tan^{-1}(y) - \tan^{-1}(x)\right|$ for all $x, y \in \mathbb{R}$. Show;

1. $\tau_\rho = \tau_d$, i.e. both metrics have the same open sets. The point here is to show $\tan^{-1} : \mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is a homeomorphism, i.e. is continuous with continuous inverse.
2. Show the sequence $\{n\}_{n=1}^\infty \subset \mathbb{R}$ is a ρ – Cauchy sequence which is not d – Cauchy.
3. $(\mathbb{R}, \rho)$ is **not** complete.

Moral: the notions of being Cauchy and complete are not purely topological notions but depend on the choice of metric inducing a given topology.

19.8.2 Arzela-Ascoli Theorem Problems

Exercise 19.25. Let (X, τ) be a compact topological space and $\mathcal{F} := \{f_n\}_{n=1}^\infty \subset C(X)$ is a sequence of functions which are equicontinuous and pointwise convergent. Show $f(x) := \lim_{n\to\infty} f_n(x)$ is continuous and that $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$, i.e. $f_n \to f$ uniformly as $n \to \infty$.

Exercise 19.26. Let $T \in (0,\infty)$ and $\mathcal{F} \subset C([0,T], \mathbb{R})$ be a family of functions such that:

1. $\dot{f}(t)$ exists for all $t \in (0,T)$ and $f \in \mathcal{F}$,
2. $\sup_{f\in\mathcal{F}} |f(0)| < \infty$, and
3. $M := \sup_{f\in\mathcal{F}} \sup_{t\in(0,T)} \left|\dot{f}(t)\right| < \infty$.

Show $\mathcal{F}$ is precompact in the Banach space $C([0,T])$ equipped with the norm $\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|$.

Exercise 19.27 (Peano's Existence Theorem). Suppose $Z : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is a bounded continuous function. Then for each $T < \infty$[10] there exists a solution to the differential equation

$$\dot{x}(t) = Z(t, x(t)) \text{ for } -T < t < T \text{ with } x(0) = x_0. \tag{19.16}$$

Do this by filling in the following outline for the proof.

1. Given $\varepsilon > 0$, show there exists a unique function $x_\varepsilon \in C([-\varepsilon,\infty) \to \mathbb{R}^d)$ such that $x_\varepsilon(t) := x_0$ for $-\varepsilon \le t \le 0$ and
$$x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_\varepsilon(\tau - \varepsilon))d\tau \text{ for all } t \ge 0. \tag{19.17}$$
Here
$$\int_0^t Z(\tau, x_\varepsilon(\tau-\varepsilon))d\tau = \left(\int_0^t Z_1(\tau, x_\varepsilon(\tau-\varepsilon))d\tau, \dots, \int_0^t Z_d(\tau, x_\varepsilon(\tau-\varepsilon))d\tau\right)$$
where $Z = (Z_1, \dots, Z_d)$ and the integrals are either the Lebesgue or the Riemann integral since they are equal on continuous functions. **Hint:** For $t \in [0,\varepsilon]$, it follows from Eq. (19.17) that
$$x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_0)d\tau.$$
Now that $x_\varepsilon(t)$ is known for $t \in [-\varepsilon, \varepsilon]$ it can be found by integration for $t \in [-\varepsilon, 2\varepsilon]$. The process can be repeated.
2. Then use Exercise 19.26 to show there exists $\{\varepsilon_k\}_{k=1}^\infty \subset (0,\infty)$ such that $\lim_{k\to\infty} \varepsilon_k = 0$ and x_{ε_k} converges to some $x \in C([0,T])$ with respect to the sup-norm: $\|x\|_\infty = \sup_{t\in[0,T]} |x(t)|$). Also show for this sequence that
$$\lim_{k\to\infty} \sup_{\varepsilon_k \le \tau \le T} |x_{\varepsilon_k}(\tau - \varepsilon_k) - x(\tau)| = 0.$$
3. Pass to the limit (**with justification**) in Eq. (19.17) with ε replaced by ε_k to show x satisfies
$$x(t) = x_0 + \int_0^t Z(\tau, x(\tau))d\tau \ \forall \ t \in [0,T].$$
4. Conclude from this that $\dot{x}(t)$ exists for $t \in (0,T)$ and that x solves Eq. (19.16).
5. Apply what you have just proved to the ODE,
$$\dot{y}(t) = -Z(-t, y(t)) \text{ for } 0 \le t < T \text{ with } y(0) = x_0.$$
Then extend $x(t)$ above to $(-T,T)$ by setting $x(t) = y(-t)$ if $t \in (-T, 0]$. Show x so defined solves Eq. (19.16) for $t \in (-T,T)$.

Exercise 19.28. Prove Theorem 19.44. **Hint:** First prove $C^{j,\beta}(\bar{\Omega}) \sqsubset\sqsubset C^{j,\alpha}(\bar{\Omega})$ is compact if $0 \le \alpha < \beta \le 1$. Then use Lemma 19.27 repeatedly to handle all of the other cases.

[10] Using Corollary 19.42, we may in fact allow $T = \infty$.

20

Some Spectral Theory

For this section let H and K be two Hilbert spaces over $\mathbb{C}$.

Exercise 20.1. Suppose $A : H \to H$ is a bounded self-adjoint operator. Show:

1. If λ is an eigenvalue of A, i.e. $Ax = \lambda x$ for some $x \in H \setminus \{0\}$, then $\lambda \in \mathbb{R}$.
2. If λ and μ are two distinct eigenvalues of A with eigenvectors x and y respectively, then $x \perp y$.

Unlike in finite dimensions, it is possible that an operator on a complex Hilbert space may have no eigenvalues, see Exercise 20.2, Example 20.6, and Lemma 20.7 below for a couple of examples. For this reason it is useful to generalize the notion of an eigenvalue as follows.

Definition 20.1. *Suppose X is a Banach space over $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) and $A \in L(X)$. We say $\lambda \in \mathbb{F}$ is in the **spectrum** of A if $A - \lambda I$ does **not** have a bounded[1] inverse. The **spectrum** will be denoted by $\sigma(A) \subset \mathbb{F}$. The **resolvent set** for A is $\rho(A) := \mathbb{F} \backslash \sigma(A)$.*

Exercise 20.2 (Multiplication Operators). Let $(X, \mathcal{M}, \mu)$ be a $\sigma -$ finite measure space, $a \in L^\infty(\mu)$ and let A be the bounded operator on $H := L^2(\mu)$ defined by $Af(x) = a(x) f(x)$ for all $f \in H$. (We will denote A by M_a in the future.) Show:

1. $\|A\|_{op} = \|a\|_{L^\infty(\mu)}$.
2. $A^* = M_{\bar{a}}$.
3. $\sigma(A) = \text{essran}(a)$ where $\sigma(A)$ is the spectrum of A and $\text{essran}(a)$ is the essential range of a, see Definitions 20.1 and 30.13 respectively.
4. Show λ is an eigenvalue for $A = M_a$ iff $\mu(\{a = \lambda\}) > 0$, i.e. iff a has a "flat spot of height λ." Thus if $\mu(\{a = \lambda\}) = 0$ for all $\lambda \in \mathbb{C}$, A has no eigenvalues.

Remark 20.2. If λ is an eigenvalue of A, then $A - \lambda I$ is not injective and hence not invertible. Therefore any eigenvalue of A is in the spectrum of A. If H is a Hilbert space and $A \in L(H)$, it follows from item 5. of Proposition 18.18 that $\lambda \in \sigma(A)$ iff $\bar{\lambda} \in \sigma(A^*)$, i.e.

$$\sigma(A^*) = \left\{\bar{\lambda} : \lambda \in \sigma(A)\right\}.$$

[1] It will follow by the open mapping Theorem 23.1 or the closed graph Theorem 23.4 that the word bounded may be omitted from this definition.

Exercise 20.3. Suppose X is a complex Banach space and $A \in GL(X)$. Show

$$\sigma\left(A^{-1}\right) = \sigma(A)^{-1} := \left\{\lambda^{-1} : \lambda \in \sigma(A)\right\}.$$

If we further assume A is both invertible and isometric, i.e. $\|Ax\| = \|x\|$ for all $x \in X$, then show

$$\sigma(A) \subset S^1 := \{z \in \mathbb{C} : |z| = 1\}.$$

Hint: working formally,

$$\left(A^{-1} - \lambda^{-1}\right)^{-1} = \frac{1}{\frac{1}{A} - \frac{1}{\lambda}} = \frac{1}{\frac{\lambda - A}{A\lambda}} = \frac{A\lambda}{\lambda - A}$$

from which you might expect that $\left(A^{-1} - \lambda^{-1}\right)^{-1} = -\lambda A (A - \lambda)^{-1}$ if $\lambda \in \rho(A)$.

Exercise 20.4. Suppose X is a Banach space and $A \in L(X)$. Use Corollary 14.27 to show $\sigma(A)$ is a closed subset of $\left\{\lambda \in \mathbb{F} : |\lambda| \leq \|A\| := \|A\|_{L(X)}\right\}$.

Lemma 20.3. *Suppose that $A \in L(H)$ is a normal operator, i.e. $0 = [A, A^*] := AA^* - A^*A$. Then $\lambda \in \sigma(A)$ iff*

$$\inf_{\|\psi\|=1} \|(A - \lambda 1)\psi\| = 0. \tag{20.1}$$

In other words, $\lambda \in \sigma(A)$ iff there is an "approximate sequence of eigenvectors" for (A, λ), i.e. there exists $\psi_n \in H$ such that $\|\psi_n\| = 1$ and $A\psi_n - \lambda\psi_n \to 0$ as $n \to \infty$.

Proof. By replacing A by $A - \lambda I$ we may assume that $\lambda = 0$. If $0 \notin \sigma(A)$, then

$$\inf_{\|\psi\|=1} \|A\psi\| = \inf \frac{\|A\psi\|}{\|\psi\|} = \inf \frac{\|\psi\|}{\|A^{-1}\psi\|} = 1/\left\|A^{-1}\right\| > 0.$$

Now suppose that $\inf_{\|\psi\|=1} \|A\psi\| = \varepsilon > 0$ or equivalently we have

$$\|A\psi\| \geq \varepsilon \|\psi\|$$

for all $\psi \in H$. Because A is normal,

$$\|A\psi\|^2 = \langle A\psi|A\psi\rangle = \langle A^*A\psi|\psi\rangle = \langle AA^*\psi|\psi\rangle = \langle A^*\psi|A^*\psi\rangle = \|A^*\psi\|^2.$$

Therefore we also have

$$\|A^*\psi\| = \|A\psi\| \geq \varepsilon\|\psi\| \ \forall \ \psi \in H. \tag{20.2}$$

This shows in particular that A and A^* are injective, $\mathrm{Ran}(A)$ is closed and hence by Lemma 18.19

$$\mathrm{Ran}(A) = \overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp = \{0\}^\perp = H.$$

Therefore A is algebraically invertible and the inverse is bounded by Eq. (20.2). ∎

Lemma 20.4. *Suppose that $A \in L(H)$ is self-adjoint (i.e. $A = A^*$) then*

$$\sigma(A) \subset \left[-\|A\|_{op}, \|A\|_{op}\right] \subset \mathbb{R}.$$

Proof. Writing $\lambda = \alpha + i\beta$ with $\alpha, \beta \in \mathbb{R}$, then

$$\begin{aligned}\|(A+\alpha+i\beta)\psi\|^2 &= \|(A+\alpha)\psi\|^2 + |\beta|^2\|\psi\|^2 + 2\,\mathrm{Re}((A+\alpha)\psi, i\beta\psi) \\ &= \|(A+\alpha)\psi\|^2 + |\beta|^2\|\psi\|^2 \end{aligned} \tag{20.3}$$

wherein we have used

$$\mathrm{Re}\left[i\beta((A+\alpha)\psi,\psi)\right] = \beta\,\mathrm{Im}((A+\alpha)\psi,\psi) = 0$$

since

$$((A+\alpha)\psi,\psi) = (\psi,(A+\alpha)\psi) = \overline{((A+\alpha)\psi,\psi)}.$$

Eq. (20.3) along with Lemma 20.3 shows that $\lambda \notin \sigma(A)$ if $\beta \neq 0$, i.e. $\sigma(A) \subset \mathbb{R}$. The fact that $\sigma(A)$ is now contained in $\left[-\|A\|_{op}, \|A\|_{op}\right]$ is a consequence of Exercise 20.3. ∎

Remark 20.5. It is not true that $\sigma(A) \subset \mathbb{R}$ implies $A = A^*$. For example let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ on $H = \mathbb{C}^2$, then $\sigma(A) = \{0\}$ yet $A \neq A^*$.

Example 20.6. Let $S \in L(H)$ be a (not necessarily) normal operator. The proof of Lemma 20.3 gives $\lambda \in \sigma(S)$ if Eq. (20.1) holds. However the converse is not always valid unless S is normal. For example, let $S : \ell^2 \to \ell^2$ be the shift, $S(\omega_1, \omega_2, \dots) = (0, \omega_1, \omega_2, \dots)$. Then for any $\lambda \in D := \{z \in \mathbb{C} : |z| < 1\}$,

$$\|(S-\lambda)\psi\| = \|S\psi - \lambda\psi\| \geq |\|S\psi\| - |\lambda|\,\|\psi\|| = (1-|\lambda|)\|\psi\|$$

and so there does not exists an approximate sequence of eigenvectors for (S, λ). However, as we will now show, $\sigma(S) = \bar{D}$.

To prove this it suffices to show by Remark 20.2 and Exercise 20.3 that $D \subset \sigma(S^*)$. For if this is the case then $\bar{D} \subset \sigma(S^*) \subset \bar{D}$ and hence $\sigma(S) = \bar{D}$ since $\bar{D}$ is invariant under complex conjugation.

A simple computation shows,

$$S^*(\omega_1, \omega_2, \dots) = (\omega_2, \omega_3, \dots)$$

and $\omega = (\omega_1, \omega_2, \dots)$ is an eigenvector for S^* with eigenvalue $\lambda \in \mathbb{C}$ iff

$$0 = (S^* - \lambda I)(\omega_1, \omega_2, \dots) = (\omega_2 - \lambda\omega_1, \omega_3 - \lambda\omega_2, \dots).$$

Solving these equation shows

$$\omega_2 = \lambda\omega_1, \ \omega_3 = \lambda\omega_2 = \lambda^2\omega_1, \dots, \ \omega_n = \lambda^{n-1}\omega_1, \dots.$$

Hence if $\lambda \in D$, we may let $\omega_1 = 1$ above to find

$$S^*(1, \lambda, \lambda^2, \dots) = \lambda(1, \lambda, \lambda^2, \dots)$$

where $(1, \lambda, \lambda^2, \dots) \in \ell^2$. Thus we have shown λ is an eigenvalue for S^* for all $\lambda \in D$ and hence $D \subset \sigma(S^*)$.

Lemma 20.7. *Let $H = \ell^2(\mathbb{Z})$ and let $A : H \to H$ be defined by*

$$Af(k) = i\left(f(k+1) - f(k-1)\right) \text{ for all } k \in \mathbb{Z}.$$

Then:

1. *A is a bounded self-adjoint operator.*
2. *A has no eigenvalues.*
3. *$\sigma(A) = [-2, 2]$.*

Proof. For another (simpler) proof of this lemma, see Exercise 40.1 below.
1. Since

$$\|Af\|_2 \leq \|f(\cdot+1)\|_2 + \|f(\cdot-1)\|_2 = 2\|f\|_2,$$

$\|A\|_{op} \leq 2 < \infty$. Moreover, for $f, g \in \ell^2(\mathbb{Z})$,

$$\begin{aligned}\langle Af|g\rangle &= \sum_k i\left(f(k+1) - f(k-1)\right)\bar{g}(k) \\ &= \sum_k if(k)\,\bar{g}(k-1) - \sum_k if(k)\,\bar{g}(k+1) \\ &= \sum_k f(k)\,\overline{Ag(k)} = \langle f|Ag\rangle,\end{aligned}$$

which shows $A = A^*$.

2. From Lemma 20.4, we know that $\sigma(A) \subset [-2,2]$. If $\lambda \in [-2,2]$ and $f \in H$ satisfies $Af = \lambda f$, then

$$f(k+1) = -i\lambda f(k) + f(k-1) \text{ for all } k \in \mathbb{Z}. \tag{20.4}$$

This is a second order difference equations which can be solved analogously to second order ordinary differential equations. The idea is to start by looking for a solution of the form $f(k) = \alpha^k$. Then Eq. (20.4) becomes, $\alpha^{k+1} = -i\lambda\alpha^k + \alpha^{k-1}$ or equivalently that

$$\alpha^2 + i\lambda\alpha - 1 = 0.$$

So we will have a solution if $\alpha \in \{\alpha_\pm\}$ where

$$\alpha_\pm = \frac{-i\lambda \pm \sqrt{4-\lambda^2}}{2}.$$

For $|\lambda| \neq 2$, there are two distinct roots and the general solution to Eq. (20.4) is of the form

$$f(k) = c_+\alpha_+^k + c_-\alpha_-^k \tag{20.5}$$

for some constants $c_\pm \in \mathbb{C}$ and $|\lambda| = 2$, the general solution has the form

$$f(k) = c\alpha_+^k + dk\alpha_+^k \tag{20.6}$$

Since in all cases, $|\alpha_\pm| = \frac{1}{4}\left(\lambda^2 + 4 - \lambda^2\right) = 1$, it follows that neither of these functions, f, will be in $\ell^2(\mathbb{Z})$ unless they are identically zero. This shows that A has no eigenvalues.

3. The above argument suggests a method for constructing approximate eigenfunctions. Namely, let $\lambda \in [-2,2]$ and define $f_n(k) := 1_{|k|\le n}\alpha^k$ where $\alpha = \alpha_+$. Then a simple computation shows

$$\lim_{n\to\infty} \frac{\|(A-\lambda I)f_n\|_2}{\|f_n\|_2} = 0 \tag{20.7}$$

and therefore $\lambda \in \sigma(A)$. ■

Exercise 20.5. Verify Eq. (20.7). Also show by explicit computations that

$$\lim_{n\to\infty} \frac{\|(A-\lambda I)f_n\|_2}{\|f_n\|_2} \neq 0$$

if $\lambda \notin [-2,2]$.

The next couple of results will be needed for the next section.

Theorem 20.8 (Rayleigh quotient). *Suppose $T \in L(H) := L(H,H)$ is a bounded self-adjoint operator, then*

$$\|T\| = \sup_{f\neq 0} \frac{|\langle f|Tf\rangle|}{\|f\|^2}.$$

Moreover if there exists a non-zero element $f \in H$ such that

$$\frac{|\langle Tf|f\rangle|}{\|f\|^2} = \|T\|,$$

then f is an eigenvector of T with $Tf = \lambda f$ and $\lambda \in \{\pm\|T\|\}$.

Proof. Let

$$M := \sup_{f\neq 0} \frac{|\langle f|Tf\rangle|}{\|f\|^2}.$$

We wish to show $M = \|T\|$. Since

$$|\langle f|Tf\rangle| \le \|f\|\|Tf\| \le \|T\|\|f\|^2,$$

we see $M \le \|T\|$. Conversely let $f, g \in H$ and compute

$$\begin{aligned}
&\langle f+g|T(f+g)\rangle - \langle f-g|T(f-g)\rangle \\
&\quad = \langle f|Tg\rangle + \langle g|Tf\rangle + \langle f|Tg\rangle + \langle g|Tf\rangle \\
&\quad = 2[\langle f|Tg\rangle + \langle Tg|f\rangle] = 2[\langle f|Tg\rangle + \overline{\langle f|Tg\rangle}] \\
&\quad = 4\mathrm{Re}\langle f|Tg\rangle.
\end{aligned}$$

Therefore, if $\|f\| = \|g\| = 1$, it follows that

$$|\mathrm{Re}\langle f|Tg\rangle| \le \frac{M}{4}\left\{\|f+g\|^2 + \|f-g\|^2\right\} = \frac{M}{4}\left\{2\|f\|^2 + 2\|g\|^2\right\} = M.$$

By replacing f be $e^{i\theta}f$ where θ is chosen so that $e^{i\theta}\langle f|Tg\rangle$ is real, we find

$$|\langle f|Tg\rangle| \le M \text{ for all } \|f\| = \|g\| = 1.$$

Hence

$$\|T\| = \sup_{\|f\|=\|g\|=1} |\langle f|Tg\rangle| \le M.$$

If $f \in H \setminus \{0\}$ and $\|T\| = |\langle Tf|f\rangle|/\|f\|^2$ then, using Schwarz's inequality,

$$\|T\| = \frac{|\langle Tf|f\rangle|}{\|f\|^2} \le \frac{\|Tf\|}{\|f\|} \le \|T\|. \tag{20.8}$$

This implies $|\langle Tf|f\rangle| = \|Tf\|\|f\|$ and forces equality in Schwarz's inequality. So by Theorem 18.2, Tf and f are linearly dependent, i.e. $Tf = \lambda f$ for some $\lambda \in \mathbb{C}$. Substituting this into (20.8) shows that $|\lambda| = \|T\|$. Since T is self-adjoint,

$$\lambda\|f\|^2 = \langle\lambda f|f\rangle = \langle Tf|f\rangle = \langle f|Tf\rangle = \langle f|\lambda f\rangle = \bar\lambda\langle f|f\rangle = \bar\lambda\,\|f\|^2,$$

which implies that $\lambda \in \mathbb{R}$ and therefore, $\lambda \in \{\pm\|T\|\}$. ■

20.1 The Spectral Theorem for Self Adjoint Compact Operators

For the rest of this section, $K \in \mathcal{K}(H) := \mathcal{K}(H,H)$ will be a self-adjoint compact operator or **S.A.C.O.** for short. Because of Proposition 19.30, we might expect compact operators to behave very much like finite dimensional matrices. This is typically the case as we will see below.

Example 20.9 (Model S.A.C.O.). Let $H = \ell_2$ and K be the diagonal matrix

$$K = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \lambda_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix},$$

where $\lim_{n\to\infty} |\lambda_n| = 0$ and $\lambda_n \in \mathbb{R}$. Then K is a self-adjoint compact operator. This assertion was proved in Example 19.26.

The main theorem (Theorem 20.11) of this subsection states that up to unitary equivalence, Example 20.9 is essentially the most general example of an S.A.C.O.

Proposition 20.10. *Let K be a S.A.C.O., then either $\lambda = \|K\|$ or $\lambda = -\|K\|$ is an eigenvalue of K.*

For those who have done Exercise 19.15, that exercise along with Theorem 20.8 constitutes a proof.. Without loss of generality we may assume that K is non-zero since otherwise the result is trivial. By Theorem 20.8, there exists $u_n \in H$ such that $\|u_n\| = 1$ and

$$\frac{|\langle u_n|Ku_n\rangle|}{\|u_n\|^2} = |\langle u_n|Ku_n\rangle| \longrightarrow \|K\| \text{ as } n\to\infty. \tag{20.9}$$

By passing to a subsequence if necessary, we may assume that $\lambda := \lim_{n\to\infty}\langle u_n|Ku_n\rangle$ exists and $\lambda \in \{\pm\|K\|\}$. By passing to a further subsequence if necessary, we may assume, using the compactness of K, that Ku_n is convergent as well. We now compute:

$$\begin{aligned} 0 \le \|Ku_n - \lambda u_n\|^2 &= \|Ku_n\|^2 - 2\lambda\langle Ku_n|u_n\rangle + \lambda^2 \\ &\le \lambda^2 - 2\lambda\langle Ku_n|u_n\rangle + \lambda^2 \\ &\to \lambda^2 - 2\lambda^2 + \lambda^2 = 0 \text{ as } n\to\infty. \end{aligned}$$

Hence

$$Ku_n - \lambda u_n \to 0 \text{ as } n\to\infty \tag{20.10}$$

and therefore

$$u := \lim_{n\to\infty} u_n = \frac{1}{\lambda}\lim_{n\to\infty} Ku_n$$

exists. By the continuity of the inner product, $\|u\| = 1 \neq 0$. By passing to the limit in Eq. (20.10) we find that $Ku = \lambda u$. ∎

Theorem 20.11 (Compact Operator Spectral Theorem). *Suppose that $K : H \to H$ is a non-zero S.A.C.O., then*

1. *there exists at least one eigenvalue $\lambda \in \{\pm\|K\|\}$.*
2. *There are at most countably many **non-zero** eigenvalues, $\{\lambda_n\}_{n=1}^N$, where $N = \infty$ is allowed. (Unless K is finite rank (i.e. $\dim \operatorname{Ran}(K) < \infty$), N will be infinite.)*
3. *The λ_n's (including multiplicities) may be arranged so that $|\lambda_n| \ge |\lambda_{n+1}|$ for all n. If $N = \infty$ then $\lim_{n\to\infty}|\lambda_n| = 0$. (In particular any eigenspace for K with **non-zero** eigenvalue is finite dimensional.)*
4. *The eigenvectors $\{\varphi_n\}_{n=1}^N$ can be chosen to be an O.N. set such that $H = \overline{\mathit{span}\{\varphi_n\}} \oplus \operatorname{Nul}(K)$.*
5. *Using the $\{\varphi_n\}_{n=1}^N$ above,*
$$Kf = \sum_{n=1}^N \lambda_n\langle f|\varphi_n\rangle\varphi_n \text{ for all } f \in H. \tag{20.11}$$
6. *The spectrum of K is $\sigma(K) = \{0\} \cup \{\lambda_n : n < N+1\}$ if $\dim H = \infty$, otherwise $\sigma(K) = \{\lambda_n : n \le N\}$ with $N \le \dim H$.*

Proof. We will find λ_n's and φ_n's recursively. Let $\lambda_1 \in \{\pm\|K\|\}$ and $\varphi_1 \in H$ such that $K\varphi_1 = \lambda_1\varphi_1$ as in Proposition 20.10.

Take $M_1 = \operatorname{span}(\varphi_1)$ so $K(M_1) \subset M_1$. By Lemma 18.19, $KM_1^\perp \subset M_1^\perp$. Define $K_1 : M_1^\perp \to M_1^\perp$ via $K_1 = K|_{M_1^\perp}$. Then K_1 is again a compact operator. If $K_1 = 0$, we are done. If $K_1 \neq 0$, by Proposition 20.10 there exists $\lambda_2 \in \{\pm\|K_1\|\}$ and $\varphi_2 \in M_1^\perp$ such that $\|\varphi_2\| = 1$ and $K_1\varphi_2 = K\varphi_2 = \lambda_2\varphi_2$. Let $M_2 := \operatorname{span}(\varphi_1,\varphi_2)$.

Again $K(M_2) \subset M_2$ and hence $K_2 := K|_{M_2^\perp} : M_2^\perp \to M_2^\perp$ is compact and if $K_2 = 0$ we are done. When $K_2 \neq 0$, we apply Proposition 20.10 again to find $\lambda_3 \in \{\pm\|K\|_2\}$ and $\varphi_3 \in M_2^\perp$ such that $\|\varphi_3\| = 1$ and $K_2\varphi_3 = K\varphi_3 = \lambda_3\varphi_3$.

Continuing this way indefinitely or until we reach a point where $K_n = 0$, we construct a sequence $\{\lambda_n\}_{n=1}^N$ of eigenvalues and orthonormal eigenvectors $\{\varphi_n\}_{n=1}^N$ such that $|\lambda_n| \ge |\lambda_{n+1}|$ with the further property that

$$|\lambda_n| = \sup_{\varphi\perp\{\varphi_1,\varphi_2,\dots\varphi_{n-1}\}} \frac{\|K\varphi\|}{\|\varphi\|}. \tag{20.12}$$

When $N < \infty$, the remaining results in the theorem are easily verified. So from now on let us assume that $N = \infty$.

If $\varepsilon := \lim_{n\to\infty} |\lambda_n| > 0$, then $\left\{\lambda_n^{-1}\varphi_n\right\}_{n=1}^{\infty}$ is a bounded sequence in H. Hence, by the compactness of K, there exists a subsequence $\{n_k : k \in \mathbb{N}\}$ of $\mathbb{N}$ such that $\left\{\varphi_{n_k} = \lambda_{n_k}^{-1} K\varphi_{n_k}\right\}_{k=1}^{\infty}$ is a convergent. However, since $\{\varphi_{n_k}\}_{k=1}^{\infty}$ is an orthonormal set, this is impossible and hence we must conclude that $\varepsilon := \lim_{n\to\infty} |\lambda_n| = 0$.

Let $M := \operatorname{span}\{\varphi_n\}_{n=1}^{\infty}$. Then $K(M) \subset M$ and hence, by Lemma 18.19, $K(M^\perp) \subset M^\perp$. Using Eq. (20.12),

$$\|K|_{M^\perp}\| \le \left\|K|_{M_n^\perp}\right\| = |\lambda_n| \longrightarrow 0 \text{ as } n \to \infty$$

showing $K|M^\perp \equiv 0$. Define P_0 to be orthogonal projection onto $M^\perp$. Then for $f \in H$,

$$f = P_0 f + (1 - P_0) f = P_0 f + \sum_{n=1}^{\infty} \langle f|\varphi_n\rangle \varphi_n$$

and

$$Kf = KP_0 f + K\sum_{n=1}^{\infty} \langle f|\varphi_n\rangle \varphi_n = \sum_{n=1}^{\infty} \lambda_n \langle f|\varphi_n\rangle \varphi_n$$

which proves Eq. (20.11).

Since $\{\lambda_n\}_{n=1}^{\infty} \subset \sigma(K)$ and $\sigma(K)$ is closed, it follows that $0 \in \sigma(K)$ and hence $\{\lambda_n\}_{n=1}^{\infty} \cup \{0\} \subset \sigma(K)$. Suppose that $z \notin \{\lambda_n\}_{n=1}^{\infty} \cup \{0\}$ and let d be the distance between z and $\{\lambda_n\}_{n=1}^{\infty} \cup \{0\}$. Notice that $d > 0$ because $\lim_{n\to\infty} \lambda_n = 0$.

A few simple computations show that:

$$(K - zI) f = \sum_{n=1}^{\infty} \langle f|\varphi_n\rangle (\lambda_n - z)\varphi_n - zP_0 f,$$

$(K - z)^{-1}$ exists,

$$(K - zI)^{-1} f = \sum_{n=1}^{\infty} \langle f|\varphi_n\rangle (\lambda_n - z)^{-1}\varphi_n - z^{-1}P_0 f,$$

and

$$\begin{aligned}\|(K - zI)^{-1} f\|^2 &= \sum_{n=1}^{\infty} |\langle f|\varphi_n\rangle|^2 \frac{1}{|\lambda_n - z|^2} + \frac{1}{|z|^2}\|P_0 f\|^2 \\ &\le \left(\frac{1}{d}\right)^2 \left(\sum_{n=1}^{\infty} |\langle f|\varphi_n\rangle|^2 + \|P_0 f\|^2\right) = \frac{1}{d^2}\|f\|^2.\end{aligned}$$

We have thus shown that $(K - zI)^{-1}$ exists, $\|(K - zI)^{-1}\| \le d^{-1} < \infty$ and hence $z \notin \sigma(K)$. ■

Theorem 20.12 (Structure of Compact Operators). *Let $K : H \to B$ be a compact operator. Then there exists $N \in \mathbb{N} \cup \{\infty\}$, orthonormal subsets $\{\varphi_n\}_{n=1}^N \subset H$ and $\{\psi_n\}_{n=1}^N \subset B$ and a sequence $\{\alpha_n\}_{n=1}^N \subset \mathbb{R}_+$ such that $\alpha_1 \ge \alpha_2 \ge \dots$ (with $\lim_{n\to\infty} \alpha_n = 0$ if $N = \infty$), $\|\psi_n\| \le 1$ for all n and*

$$Kf = \sum_{n=1}^{N} \alpha_n \langle f|\varphi_n\rangle \psi_n \text{ for all } f \in H. \tag{20.13}$$

Proof. Since K^*K is a self-adjoint compact operator, Theorem 20.11 implies there exists an orthonormal set $\{\varphi_n\}_{n=1}^N \subset H$ and positive numbers $\{\lambda_n\}_{n=1}^N$ such that

$$K^*K\psi = \sum_{n=1}^{N} \lambda_n \langle \psi|\varphi_n\rangle \varphi_n \text{ for all } \psi \in H.$$

Let A be the positive square root of K^*K defined by

$$A\psi := \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi|\varphi_n\rangle \varphi_n \text{ for all } \psi \in H.$$

A simple computation shows, $A^2 = K^*K$, and therefore,

$$\begin{aligned}\|A\psi\|^2 &= \langle A\psi|A\psi\rangle = \langle \psi|A^2\psi\rangle \\ &= \langle \psi|K^*K\psi\rangle = \langle K\psi|K\psi\rangle = \|K\psi\|^2\end{aligned}$$

for all $\psi \in H$. Hence we may define a unitary operator, $u : \overline{\operatorname{Ran}(A)} \to \overline{\operatorname{Ran}(K)}$ by the formula

$$uA\psi = K\psi \text{ for all } \psi \in H.$$

We then have

$$K\psi = uA\psi = \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi|\varphi_n\rangle u\varphi_n \tag{20.14}$$

which proves the result with $\psi_n := u\varphi_n$ and $\alpha_n = \sqrt{\lambda_n}$.

It is instructive to find ψ_n explicitly and to verify Eq. (20.14) by brute force. Since $\varphi_n = \lambda_n^{-1/2} A\varphi_n$,

$$\psi_n = \lambda_n^{-1/2} uA\varphi_n = \lambda_n^{-1/2} K\varphi_n$$

and

$$\langle K\varphi_n|K\varphi_m\rangle = \langle \varphi_n|K^*K\varphi_m\rangle = \lambda_n \delta_{mn}.$$

This verifies that $\{\psi_n\}_{n=1}^N$ is an orthonormal set. Moreover,

$$\sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi | \varphi_n \rangle \psi_n = \sum_{n=1}^{N} \sqrt{\lambda_n} \langle \psi | \varphi_n \rangle \lambda_n^{-1/2} K \varphi_n$$
$$= K \sum_{n=1}^{N} \langle \psi | \varphi_n \rangle \varphi_n = K\psi$$

since $\sum_{n=1}^{N} \langle \psi | \varphi_n \rangle \varphi_n = P\psi$ where P is orthogonal projection onto $\mathrm{Nul}(K)^{\perp}$.

Second Proof. Let $K = u\,|K|$ be the polar decomposition of K. Then $|K|$ is self-adjoint and compact, by Corollary **??** below, and hence by Theorem 20.11 there exists an orthonormal basis $\{\varphi_n\}_{n=1}^{N}$ for $\mathrm{Nul}(|K|)^{\perp} = \mathrm{Nul}(K)^{\perp}$ such that $|K|\,\varphi_n = \lambda_n \varphi_n$, $\lambda_1 \geq \lambda_2 \geq \dots$ and $\lim_{n\to\infty} \lambda_n = 0$ if $N = \infty$. For $f \in H$,

$$Kf = u\,|K| \sum_{n=1}^{N} \langle f | \varphi_n \rangle \varphi_n = \sum_{n=1}^{N} \langle f | \varphi_n \rangle u\,|K|\,\varphi_n = \sum_{n=1}^{N} \lambda_n \langle f | \varphi_n \rangle u \varphi_n$$

which is Eq. (20.13) with $\psi_n := u\varphi_n$. ■

Exercise 20.6 (Continuation of Example 14.12). Let $H := L^2([0,1], m)$, $k(x,y) := \min(x,y)$ for $x, y \in [0,1]$ and define $K : H \to H$ by

$$Kf(x) = \int_0^1 k(x,y)\, f(y)\, dy.$$

From Exercise 19.20 we know that K is a compact operator[2] on H. Since k is real and symmetric, it is easily seen that K is self-adjoint. Show:

1. If $g \in C^2([0,1])$ with $g(0) = 0 = g'(1)$, then $Kg'' = -g$. Use this to conclude $\langle Kf | g'' \rangle = -\langle f | g \rangle$ for all $g \in C_c^\infty((0,1))$ and consequently that $\mathrm{Nul}(K) = \{0\}$.
2. Now suppose that $f \in H$ is an eigenvector of K with eigenvalue $\lambda \neq 0$. Show that there is a version[3] of f which is in $C([0,1]) \cap C^2((0,1))$ and this version, still denoted by f, solves
$$\lambda f'' = -f \text{ with } f(0) = f'(1) = 0. \qquad (20.15)$$
where $f'(1) := \lim_{x\uparrow 1} f'(x)$.
3. Use Eq. (20.15) to find all the eigenvalues and eigenfunctions of K.
4. Use the results above along with the spectral Theorem 20.11, to show
$$\left\{ \sqrt{2} \sin\left(\left(n + \frac{1}{2} \right) \pi x \right) : n \in \mathbb{N}_0 \right\}$$
is an orthonormal basis for $L^2([0,1], m)$.

---

[2] See Exercise 19.9 from which it will follow that K is a Hilbert Schmidt operator and hence compact.

[3] A measurable function g is called a version of f iff $g = f$ a.e..

Part IV

More on Banach Spaces

21

The Hahn-Banach Theorem

Our goal here is to show that continuous dual, X^*, of a Banach space, X, is always sufficiently large. This will be the content of the Hahn-Banach Theorem 21.7 below.

Proposition 21.1. *Let X be a complex vector space over $\mathbb{C}$ and let $X_{\mathbb{R}}$ denote X thought of as a real vector space. If $f \in X^*$ and $u = \mathrm{Re} f \in X_{\mathbb{R}}^*$ then*

$$f(x) = u(x) - iu(ix). \tag{21.1}$$

Conversely if $u \in X_{\mathbb{R}}^*$ and f is defined by Eq. (21.1), then $f \in X^*$ and $\|u\|_{X_{\mathbb{R}}^*} = \|f\|_{X^*}$. More generally if p is a semi-norm (see Definition 4.24) on X, then

$$|f| \le p \text{ iff } u \le p.$$

Proof. Let $v(x) = \mathrm{Im}\ f(x)$, then

$$v(ix) = \mathrm{Im}\ f(ix) = \mathrm{Im}(if(x)) = \mathrm{Re} f(x) = u(x).$$

Therefore

$$f(x) = u(x) + iv(x) = u(x) + iu(-ix) = u(x) - iu(ix).$$

Conversely for $u \in X_{\mathbb{R}}^*$ let $f(x) = u(x) - iu(ix)$. Then

$$\begin{aligned} f((a+ib)x) &= u(ax + ibx) - iu(iax - bx) \\ &= au(x) + bu(ix) - i(au(ix) - bu(x)) \end{aligned}$$

while

$$(a+ib)f(x) = au(x) + bu(ix) + i(bu(x) - au(ix)).$$

So f is complex linear. Because $|u(x)| = |\mathrm{Re} f(x)| \le |f(x)|$, it follows that $\|u\| \le \|f\|$. For $x \in X$ choose $\lambda \in S^1 \subset \mathbb{C}$ such that $|f(x)| = \lambda f(x)$ so

$$|f(x)| = f(\lambda x) = u(\lambda x) \le \|u\|\ \|\lambda x\| = \|u\| \|x\|.$$

Since $x \in X$ is arbitrary, this shows that $\|f\| \le \|u\|$ so $\|f\| = \|u\|$.[1]

For the last assertion, it is clear that $|f| \le p$ implies that $u \le |u| \le |f| \le p$. Conversely if $u \le p$ and $x \in X$, choose $\lambda \in S^1 \subset \mathbb{C}$ such that $|f(x)| = \lambda f(x)$. Then

$$|f(x)| = \lambda f(x) = f(\lambda x) = u(\lambda x) \le p(\lambda x) = p(x)$$

holds for all $x \in X$. ■

Definition 21.2 (Minkowski functional). *A function $p : X \to [0, \infty)$ is a **Minkowski functional** if*

1. *$p(x + y) \le p(x) + p(y)$ for all $x, y \in X$ and*
2. *$p(cx) = cp(x)$ for all $c \ge 0$ and $x \in X$.*

Example 21.3. Suppose that $X = \mathbb{R}$ and

$$p(x) = \inf \{\lambda \ge 0 : x \in \lambda[-1, 2] = [-\lambda, 2\lambda]\}.$$

Notice that if $x \ge 0$, then $p(x) = x/2$ and if $x \le 0$ then $p(x) = -x$, i.e.

$$p(x) = \begin{cases} x/2 \text{ if } x \ge 0 \\ |x| \text{ if } x \le 0. \end{cases}$$

From this formula it is clear that $p(cx) = cp(x)$ for all $c \ge 0$ but not for $c < 0$. Moreover, p satisfies the triangle inequality, indeed if $p(x) = \lambda$ and $p(y) = \mu$, then $x \in \lambda[-1, 2]$ and $y \in \mu[-1, 2]$ so that

$$x + y \in \lambda[-1, 2] + \mu[-1, 2] \subset (\lambda + \mu) [-1, 2]$$

which shows that $p(x+y) \le \lambda + \mu = p(x) + p(y)$. To check the last set inclusion let $a, b \in [-1, 2]$, then

$$\lambda a + \mu b = (\lambda + \mu) \left(\frac{\lambda}{\lambda + \mu} a + \frac{\mu}{\lambda + \mu} b \right) \in (\lambda + \mu) [-1, 2]$$

[1] To understand better why $\|f\| = \|u\|$, notice that

$$\|f\|^2 = \sup_{\|x\|=1} |f(x)|^2 = \sup_{\|x\|=1} (|u(x)|^2 + |u(ix)|^2).$$

Suppose that $M = \sup_{\|x\|=1} |u(x)|$ and this supremum is attained at $x_0 \in X$ with $\|x_0\| = 1$. Replacing x_0 by $-x_0$ if necessary, we may assume that $u(x_0) = M$. Since u has a maximum at x_0,

$$\begin{aligned} 0 &= \frac{d}{dt}\Big|_0 u \left(\frac{x_0 + itx_0}{\|x_0 + itx_0\|} \right) \\ &= \frac{d}{dt}\Big|_0 \left\{ \frac{1}{|1 + it|} \ (u(x_0) + tu(ix_0)) \right\} = u(ix_0) \end{aligned}$$

since $\frac{d}{dt}|_0 |1 + it| = \frac{d}{dt}|_0 \sqrt{1 + t^2} = 0$.This explains why $\|f\| = \|u\|$.

since $[-1,2]$ is a convex set and $\frac{\lambda}{\lambda+\mu}+\frac{\mu}{\lambda+\mu}=1$.

In what follows X will always be a real or complex vector space at very least.

Definition 21.4. *Let $S\subset X$ be a set. Then*

1. *S is called* symmetric (or balanced or circled) *if $x\in S\Rightarrow \alpha x\in S$ whenever $|\alpha|=1$.*
2. *S is* absorbing *if for every $x\in X$, there exists $\alpha>0$ such that $x\in\alpha S$.*
3. *S is* linearly open *if for every $x_0\neq 0$, $\{\alpha:\alpha x_0\in S\}$ is open.*
4. *S is* star shaped *if $tx\in S$ for all $0\le t\le 1$ and $x\in X$.*

Remark 21.5. Let $S\subset X$ then

1. if S is absorbing or star shaped then $0\in S$.
2. If S is nonempty, convex and symmetric then $0\in S$.
3. The intersection of convex sets is convex.
4. If $0\in S$ and S is convex then S is star shaped.

Lemma 21.6. *Let X be a real or complex linear space and $S\subset X$ be absorbing and star shaped. Then*

$$N(x)=N_S(x):=\inf\{\lambda>0:x\in\lambda S\}=\left[\sup\{\mu>0:\mu x\in S\}\right]^{-1}$$

satisfies

1. *If $t\ge 0$ then $N(tx)=tN(x)$.*
2. *$\{N<1\}\subset S\subset\{N\le 1\}$.*
3. *If S is convex then $N(x+y)\le N(x)+N(y)$ for all $x,y\in X$.*
4. *If S is balanced then $N(\lambda x)=|\lambda|\,N(x)$ for all $\lambda\in\mathbb{F}$ and $x\in X$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$ depending on whether X is real or complex.*

Proof.

1. Since $0\in S$, $N(0)=0$ so it suffices to assume $t>0$ where
$$N(tx)=\inf\{\lambda:tx\in\lambda S\}=\inf\left\{\lambda:x\in\frac{\lambda}{t}S\right\}=\inf\{t\lambda:x\in\lambda S\}=tN(x).$$
2. If $N(x)<1$, then there exists $\lambda<1$ such that $x\in\lambda S$, i.e. $\lambda^{-1}x\in S$. Since S is star shaped, $x=\lambda(\lambda^{-1}x)\in S$ as well. If $x\in S$, then clearly $N(x)\le 1$, so we have proved item 2.
3. Now suppose S is convex and $\lambda>N(x)$ and $\mu>N(y)$ so that $x\in\lambda S$ and $y\in\mu S$. Then there exists $a,b\in S$ such that
$$x+y=\lambda a+\mu b=(\lambda+\mu)\left[\frac{\lambda}{\lambda+\mu}a+\frac{\mu}{\lambda+\mu}b\right].$$
Since S is convex, this implies $(x+y)\in(\lambda+\mu)\,S$ and therefore that $N(x+y)\le\lambda+\mu$.Letting $\lambda\downarrow N(x)$ and $\mu\downarrow N(y)$ gives the assertion in item 3.
4. Now suppose that S is balanced and $\lambda\in\mathbb{F}\setminus\{0\}$ (the case $\lambda=0$ being trivial) and $\mu>N(x)$. Then $x\in\mu S$ so $|\lambda|\,x\in|\lambda|\,\mu S$ and since S is balance, $\lambda x\in|\lambda|\,\mu S$. Therefore $N(\lambda x)\le|\lambda|\,\mu$ and so letting $\mu\downarrow N(x)$ we learn $N(\lambda x)\le|\lambda|\,N(x)$. This inequality with x replaced by $\lambda^{-1}x$ implies $N(x)\le|\lambda|\,N(\lambda^{-1}x)$ and then replacing λ by λ^{-1} shows $N(x)\le|\lambda|^{-1}N(\lambda x)$. This then implies $N(\lambda x)\ge|\lambda|\,N(x)$ which combined with the opposite inequality shows $N(\lambda x)=|\lambda|\,N(x)$.

■

BRUCE: Add in the relationship to convex sets and separation theorems, see Reed and Simon Vol. 1. for example. (See D:\Bruce\Classfil\Analysis\funct-anal\len.tex where this and more is already done!)

Theorem 21.7 (Hahn-Banach). *Let X be a real vector space, $p:X\to[0,\infty)$ be a Minikowski functional, $M\subset X$ be a subspace $f:M\to\mathbb{R}$ be a linear functional such that $f\le p$ on M. Then there exists a linear functional $F:X\to\mathbb{R}$ such that $F|_M=f$ and $F\le p$ on X.*

Proof. Step 1. We show for all $x\in X\setminus M$ there exists and extension F to $M\oplus\mathbb{R}x$ with the desired properties. If F exists and $\alpha=F(x)$, then for all $y\in M$ and $\lambda\in\mathbb{R}$ we must have

$$f(y)+\lambda\alpha=F(y+\lambda x)\le p(y+\lambda x). \tag{21.2}$$

Dividing this equation by $|\lambda|$ allows us to conclude that Eq. (21.2) is valid for all $y\in M$ and $\lambda\in\mathbb{R}$ iff

$$f(y)+\varepsilon\alpha\le p(y+\varepsilon x)\text{ for all }y\in M\text{ and }\varepsilon\in\{\pm1\}.$$

Equivalently put we must have, for all $y,z\in M$, that

$$\begin{aligned}\alpha&\le p(y+x)-f(y)\text{ and}\\ f(z)-p(z-x)&\le\alpha.\end{aligned}$$

Hence it is possible to find an $\alpha\in\mathbb{R}$ such that Eq. (21.2) holds iff

$$f(z)-p(z-x)\le p(y+x)-f(y)\text{ for all }y,z\in M. \tag{21.3}$$

(If Eq. (21.3) holds, then $\sup_{z\in M}[f(z)-p(z-x)] \le \inf_{y\in M}[p(y+x)-f(y)]$ and so we may choose $\alpha = \sup_{z\in M}[f(z)-p(z-x)]$ for example.) Now Equation (21.3) is equivalent to having

$$f(z)+f(y)=f(z+y)\le p(y+x)+p(z-x) \text{ for all } y,z\in M$$

and this last equation is valid because

$$f(z+y)\le p(z+y)=p(y+x+z-x)\le p(y+x)+p(z-x),$$

wherein we use $f\le p$ on M and the triangle inequality for p. In conclusion, if $\alpha := \sup_{z\in M}[f(z)-p(z-x)]$ and $F(y+\lambda x) := f(y)+\lambda\alpha$, then by following the above logic backwards, we have $F|_M = f$ and $F\le p$ on $M\oplus\mathbb{R}x$ showing F is the desired extension.

Step 2. Let us now write $F: X\to\mathbb{R}$ to mean F is defined on a linear subspace $D(F)\subset X$ and $F: D(F)\to\mathbb{R}$ is linear. For $F,G: X\to\mathbb{R}$ we will say $F\prec G$ if $D(F)\subset D(G)$ and $F=G|_{D(F)}$, that is G is an extension of F. Let

$$\mathcal{F}=\{F:X\to\mathbb{R}: f\prec F \text{ and } F\le p \text{ on } D(F)\}.$$

Then $(\mathcal{F},\prec)$ is a partially ordered set. If $\Phi\subset\mathcal{F}$ is a chain (i.e. a linearly ordered subset of $\mathcal{F}$) then Φ has an upper bound $G\in\mathcal{F}$ defined by $D(G)=\bigcup_{F\in\Phi} D(F)$ and $G(x)=F(x)$ for $x\in D(F)$. Then it is easily checked that $D(G)$ is a linear subspace, $G\in\mathcal{F}$, and $F\prec G$ for all $F\in\Phi$. We may now apply Zorn's Lemma[2] (see Theorem 2.14) to conclude there exists a maximal element $F\in\mathcal{F}$. Necessarily, $D(F)=X$ for otherwise we could extend F by step (1), violating the maximality of F. Thus F is the desired extension of f. ∎

Corollary 21.8. *Suppose that X is a complex vector space, $p: X\to[0,\infty)$ is a semi-norm, $M\subset X$ is a linear subspace, and $f: M\to\mathbb{C}$ is linear functional such that $|f(x)|\le p(x)$ for all $x\in M$. Then there exists $F\in X'$ (X' is the **algebraic** dual of X) such that $F|_M=f$ and $|F|\le p$.*

Proof. Let $u=\mathrm{Re}f$ then $u\le p$ on M and hence by Theorem 21.7, there exists $U\in X'_{\mathbb{R}}$ such that $U|_M=u$ and $U\le p$ on M. Define $F(x)=U(x)-iU(ix)$ then as in Proposition 21.1, $F=f$ on M and $|F|\le p$. ∎

Theorem 21.9. *Let X be a normed space $M\subset X$ be a closed subspace and $x\in X\setminus M$. Then there exists $f\in X^*$ such that $\|f\|=1$, $f(x)=\delta=d(x,M)$ and $f=0$ on M.*

[2] The use of Zorn's lemma in this step may be avoided in the case that $p(x)$ is a norm and X may be written as $\overline{M\oplus\mathrm{span}(\beta)}$ where $\beta:=\{x_n\}_{n=1}^\infty$ is a countable subset of X. In this case, by step (1) and induction, $f: M\to\mathbb{R}$ may be extended to a linear functional $F: M\oplus\mathrm{span}(\beta)\to\mathbb{R}$ with $F(x)\le p(x)$ for $x\in M\oplus\mathrm{span}(\beta)$. This function F then extends by continuity to X and gives the desired extension of f.

Proof. Define $h: M\oplus\mathbb{C}x\to\mathbb{C}$ by $h(m+\lambda x):=\lambda\delta$ for all $m\in M$ and $\lambda\in\mathbb{C}$. Then

$$\|h\| := \sup_{m\in M \text{ and } \lambda\ne 0}\frac{|\lambda|\,\delta}{\|m+\lambda x\|}=\sup_{m\in M \text{ and } \lambda\ne 0}\frac{\delta}{\|x+m/\lambda\|}=\frac{\delta}{\delta}=1$$

and by the Hahn – Banach theorem there exists $f\in X^*$ such that $f|_{M\oplus\mathbb{C}x}=h$ and $\|f\|\le 1$. Since $1=\|h\|\le\|f\|\le 1$, it follows that $\|f\|=1$. ∎

Corollary 21.10 (A density test). *Let M be a subspace of a normed space $(X,\|\cdot\|)$. The statement that M is dense in X is equivalent to the statement that the only $f\in X^*$ which vanishes on M is the zero linear functional.*

Proof. If $f\in X^*$ such that $f|_M=0$, then by continuity $f|_{\bar M}=0$. Thus if M is dense in X it follows that $f\equiv 0$. Conversely, if M is not dense, then $\bar M$ is a proper subspace of X and according to Theorem 21.9 there exists $f\in X^*\setminus\{0\}$ such that $f|_{\bar M}=0$. ∎

Corollary 21.11. *To each $x\in X$, let $\hat x\in X^{**}$ be defined by $\hat x(f)=f(x)$ for all $f\in X^*$. Then the map $x\in X\to\hat x\in X^{**}$ is a linear (injective) isometry of Banach spaces. In particular, for all $x\in X$ we have,*

$$\|x\|=\sup_{f\in X^*\setminus\{0\}}\frac{|f(x)|}{\|f\|_{X^*}}=\sup_{\|f\|_{X^*}=1}|f(x)|.$$

Proof. Since

$$|\hat x(f)|=|f(x)|\le\|f\|_{X^*}\|x\|_X \text{ for all } f\in X^*,$$

it follows that $\|\hat x\|_{X^{**}}\le\|x\|_X$. Now applying Theorem 21.9 with $M=\{0\}$, there exists $f\in X^*$ such that $\|f\|=1$ and $|\hat x(f)|=f(x)=\|x\|$, which shows that $\|\hat x\|_{X^{**}}\ge\|x\|_X$. This shows that $x\in X\to\hat x\in X^{**}$ is an isometry. Since isometries are necessarily injective, we are done. ∎

Definition 21.12. *A Banach space X is **reflexive** if the map $x\in X\to\hat x\in X^{**}$ is surjective.*

Example 21.13. Every Hilbert space H is reflexive. This is a consequence of the Riesz Theorem 18.17.

Exercise 21.1. Show all finite dimensional Banach spaces, X, are reflexive.

Definition 21.14. *Let X be a Banach space and $(\Omega,\mathcal{B},\mu)$ be a measure space. We say such a function $u:\Omega\to X$ is **weakly measurable** if $f\circ u:\Omega\to\mathbb{C}$ is measurable for all $f\in X^*$. A weakly measurable function $u:\Omega\to X$ is said to be **integrable** if there exists $U\in L^1(\Omega,\mathcal{B},\mu)$ such that $\|u(\omega)\|\le U(\omega)$ for all $\omega\in\Omega$.*

Exercise 21.2. Suppose that X is a separable Banach space. Show there exists $\varphi_n \in X^*$ such that

$$\|x\| = \sup_n |\varphi_n(x)| \text{ for all } x \in X. \tag{21.4}$$

Use this to conclude that Borel σ – algebra of X and the σ – algebra generated by $\varphi \in X^*$ are the same, i.e. $\sigma(X^*) = \mathcal{B}$. So if $(\Omega, \mathcal{B}, \mu)$ is a measure space and X is separable, a function $u : \Omega \to X$ is weakly integrable iff $u : \Omega \to X$ is $\mathcal{B}/\mathcal{B}_X$ – measurable and

$$\int_\Omega \|u(\omega)\| \, d\mu(\omega) < \infty.$$

Theorem 21.15 (Reflexive Integration Theory). *Suppose that X is a Banach space and $(\Omega, \mathcal{B}, \mu)$ is a measure space. The collection of weakly integrable functions, $u : \Omega \to X$, is a vector space and for each weakly integrable u we have $\tilde{u} \in X^{**}$ where*

$$\tilde{u}(f) := \int_\Omega [f \circ u] \, d\mu \text{ for all } f \in X^*. \tag{21.5}$$

If we further assume X is reflexive, there exists a unique $x_0 \in X$ such that $\tilde{u} = \hat{x}_0$ and we denote x_0 by $\int_\Omega u \, d\mu$. This integral in linear and $\int_\Omega u \, d\mu$ is the unique element of X such that

$$f\left(\int_\Omega u \, d\mu\right) = \int_\Omega [f \circ u] \, d\mu \text{ for all } f \in X^*. \tag{21.6}$$

Moreover if $\|u(\cdot)\| \le U \in L^1(\Omega, \mathcal{B}, \mu)$, then

$$\left\|\int_\Omega u \, d\mu\right\|_X \le \|U\|_1 = \int_\Omega U \, d\mu.$$

Proof. The collection of weakly integrable functions is easily seen to be a vector space. If $u : \Omega \to X$ is weakly integrable, it is easy to check that $\tilde{u}$ is linear. Moreover, since

$$|f \circ u| \le \|f\|_{X^*} \|u\|_X \le \|f\|_{X^*} \cdot U$$

it follows that

$$|\tilde{u}(f)| \le \int_\Omega |f \circ u| \, d\mu \le C \|f\|_{X^*}.$$

where $C := \int_\Omega U d\mu < \infty$. This shows that $\tilde{u} \in X^{**}$. The rest of the proof is now a simple consequence of this fact and the reflexivity assumption on X. ∎

Exercise 21.3. Suppose that X and Y are Banach spaces, $T \in L(X, U)$, $(\Omega, \mathcal{B}, \mu)$ is a measure space, and $u : \Omega \to X$ is an integrable function in the sense that;

1. For all $\lambda \in X^*$, $\lambda \circ u \in L^1(\mu)$ and
2. there exists a unique element $x \in X$ (denoted by $\int_\Omega u(\omega) \, d\mu(\omega)$) such that Eq. (21.6) holds.

Show $T \circ u : \Omega \to Y$ is an integrable function is the above sense and

$$\int_\Omega T[u(\omega)] \, d\mu(\omega) = T \int_\Omega u(\omega) \, d\mu(\omega).$$

[For situations where the hypothesis of this exercise hold see Theorems 17.15 and 21.15.]

Definition 21.16. *For subsets, $M \subset X$ and $N \subset X^*$, let*

$$M^0 := \{f \in X^* : f|_M = 0\} \text{ and}$$
$$N^\perp := \{x \in X : f(x) = 0 \text{ for all } f \in N\}.$$

We call M^0 the ***annihilator*** *of M and $N^\perp$ the* ***backwards annihilator*** *of N.*

Lemma 21.17. *Let $M \subset X$ and $N \subset X^*$, then*

1. *M^0 and $N^\perp$ are always closed subspace of X^* and X respectively.*
2. *If M is a subspace of X, then $\left(M^0\right)^\perp = \bar{M}$.*
3. *If N is a subspace, then $\bar{N} \subset \left(N^\perp\right)^0$ with equality if X is reflexive. Also see Exercise 21.4, Example 21.18, and Proposition 21.23 below.*

Proof. Since

$$M^0 = \cap_{x \in M} \text{Nul}(\hat{x}) \text{ and } N^\perp = \cap_{f \in M} \text{Nul}(f),$$

M^0 and $N^\perp$ are both formed as an intersection of closed subspaces and hence are themselves closed subspaces.

If $x \in M$, then $f(x) = 0$ for all $f \in M^0$ so that $x \in \left(M^0\right)^\perp$ and hence $\bar{M} \subset \left(M^0\right)^\perp$. If $x \notin \bar{M}$, then there exists (by Theorem 21.9) $f \in X^*$ such that $f|_M = 0$ while $f(x) \neq 0$, i.e. $f \in M^0$ yet $f(x) \neq 0$. This shows $x \notin \left(M^0\right)^\perp$ and we have shown $\left(M^0\right)^\perp \subset \bar{M}$. The proof of Item 3. is left to the reader in Exercise 21.4. ∎

Exercise 21.4. Prove Item 3. of Lemma 21.17, i.e. if If N is a subspace of X^*, then $\bar{N} \subset \left(N^\perp\right)^0$ with equality if X is reflexive.. Also show that it is possible that $\bar{N} \neq \left(N^\perp\right)^0$. **Hint:** let $X = Y^*$ where Y is a non-reflexive Banach space (see Theorem 14.21 and Theorem 21.20 below) and take $N = \hat{Y} \subset Y^{**} = X^*$.

Example 21.18 (Another exmaple where $\bar{N} \neq \left(N^{\perp}\right)^{0}$). As in Exercise 38.2, let (X, τ) be a compact Hausdorff space which supports a positive measure ν on $\mathcal{B} = \sigma(\tau)$ such that $\nu(X) \neq \sum_{x \in X} \nu(\{x\})$, i.e. ν is a not a counting type measure. Recall that $C(X)^*$ is isomorphic to the space of complex Radon measures on $(X, \mathcal{B})$ and let $\lambda \in C(X)^{**}$ be defined by

$$\lambda(\mu) = \sum_{x \in X} \mu(\{x\}).$$

Then take

$$N := \left\{\mu \in C(X)^* : \lambda(\mu) = \sum_{x \in X} \mu(\{x\}) = 0\right\}$$

which is a closed subspace $C(X)^*$. If $o \in X$ is a fixed point we will have $\mu_x := \delta_x - \delta_o \in N$ for all $x \in X$ and therefore if $f \in N^{\perp}$ we must have $0 = \mu_x(f) = f(x) - f(o)$ for all $x \in X$ which shows that $f = c$ is constant. We also know that $\mu = \nu - \sum_{x \in X} \nu(\{x\}) \delta_x \in N$ and therefore

$$0 = \mu(f) = \mu(c) = c \cdot \left[\nu(X) - \sum_{x \in X} \nu(\{x\})\right]$$

from which it follows that $c = 0$. Therefore we have shown $N^{\perp} = \{0\}$ and therefore $\left(N^{\perp}\right)^{0} = C(X)^*$ which properly contains N.

Proposition 21.19. *Suppose X is a Banach space, then $X^{***} = \widehat{(X^*)} \oplus \left(\hat{X}\right)^{0}$ where*

$$\left(\hat{X}\right)^{0} = \{\lambda \in X^{***} : \lambda(\hat{x}) = 0 \text{ for all } x \in X\}.$$

In particular X is reflexive iff X^* is reflexive.

Proof. Let $\psi \in X^{***}$ and define $f_\psi \in X^*$ by $f_\psi(x) := \psi(\hat{x})$ for all $x \in X$ and set $\psi' := \psi - \hat{f}_\psi$. For $x \in X$ (so $\hat{x} \in X^{**}$) we have

$$\psi'(\hat{x}) = \psi(\hat{x}) - \hat{f}_\psi(\hat{x}) = f_\psi(x) - \hat{x}(f_\psi) = f_\psi(x) - f_\psi(x) = 0.$$

This shows $\psi' \in \hat{X}^0$ and we have shown $X^{***} = \widehat{X^*} + \hat{X}^0$. If $\psi \in \widehat{X^*} \cap \hat{X}^0$, then $\psi = \hat{f}$ for some $f \in X^*$ and $0 = \hat{f}(\hat{x}) = \hat{x}(f) = f(x)$ for all $x \in X$, i.e. $f = 0$ so $\psi = 0$. Therefore $X^{***} = \widehat{X^*} \oplus \hat{X}^0$ as claimed.

If X is reflexive, then $\hat{X} = X^{**}$ and so $\hat{X}^0 = \{0\}$ showing $(X^*)^{**} = X^{***} = \widehat{(X^*)}$, i.e. X^* is reflexive. Conversely if X^* is reflexive we conclude that $\left(\hat{X}\right)^{0} = \{0\}$ and therefore

$$X^{**} = \{0\}^{\perp} = \left(\hat{X}^0\right)^{\perp} = \hat{X},$$

which shows $\hat{X}$ is reflexive. Here we have used

$$\left(\hat{X}^0\right)^{\perp} = \overline{\hat{X}} = \hat{X}$$

since $\hat{X}$ is a closed subspace of X^{**}. ■

Theorem 21.20 (Continuation of Theorem 14.21). *Let X be an **infinite** set, $\mu : X \to (0, \infty)$ be a function, $p \in [1, \infty]$, $q := p/(p-1)$ be the conjugate exponent and for $f \in \ell^q(\mu)$ define $\varphi_f : \ell^p(\mu) \to \mathbb{F}$ by*

$$\varphi_f(g) := \sum_{x \in X} f(x) g(x) \mu(x). \tag{21.7}$$

1. *$\ell^p(\mu)$ is reflexive for $p \in (1, \infty)$.*
2. *The map $\varphi : \ell^1(\mu) \to \ell^\infty(X)^*$ is not surjective.*
3. *$\ell^1(\mu)$ and $\ell^\infty(X)$ are **not** reflexive.*

See Lemma 21.21 below and Exercise 38.2 above for more examples of non-reflexive spaces.

Proof.

1. This basically follows from two applications of item 3 of Theorem 14.21. More precisely if $\lambda \in \ell^p(\mu)^{**}$, let $\tilde{\lambda} \in \ell^q(\mu)^*$ be defined by $\tilde{\lambda}(g) = \lambda(\varphi_g)$ for $g \in \ell^q(\mu)$. Then by item 3., there exists $f \in \ell^p(\mu)$ such that, for all $g \in \ell^q(\mu)$,
$$\lambda(\varphi_g) = \tilde{\lambda}(g) = \varphi_f(g) = \varphi_g(f) = \hat{f}(\varphi_g).$$
Since $\ell^p(\mu)^* = \{\varphi_g : g \in \ell^q(\mu)\}$, this implies that $\lambda = \hat{f}$ and so $\ell^p(\mu)$ is reflexive.
2. Recall $c_0(X)$ as defined in Notation 14.20 and is a closed subspace of $\ell^\infty(X)$, see Exercise 14.6. Let $\mathbf{1} \in \ell^\infty(X)$ denote the constant function 1 on X. Notice that $\|\mathbf{1} - f\|_\infty \geq 1$ for all $f \in c_0(X)$ and therefore, by the Hahn - Banach Theorem, there exists $\lambda \in \ell^\infty(X)^*$ such that $\lambda(\mathbf{1}) = 0$ while $\lambda|_{c_0(X)} \equiv 0$. Now if $\lambda = \varphi_f$ for some $f \in \ell^1(\mu)$, then $\mu(x) f(x) = \lambda(\delta_x) = 0$ for all x and f would have to be zero. This is absurd.
3. As we have seen $\ell^1(\mu)^* \cong \ell^\infty(X)$ while $\ell^\infty(X)^* \cong c_0(X)^* \neq \ell^1(\mu)$. Let $\lambda \in \ell^\infty(X)^*$ be the linear functional as described above. We view this as an element of $\ell^1(\mu)^{**}$ by using
$$\tilde{\lambda}(\varphi_g) := \lambda(g) \text{ for all } g \in \ell^\infty(X).$$
Suppose that $\tilde{\lambda} = \hat{f}$ for some $f \in \ell^1(\mu)$, then
$$\lambda(g) = \tilde{\lambda}(\varphi_g) = \hat{f}(\varphi_g) = \varphi_g(f) = \varphi_f(g).$$
But λ was constructed in such a way that $\lambda \neq \varphi_f$ for any $f \in \ell^1(\mu)$. It now follows from Proposition 21.19 that $\ell^1(\mu)^* \cong \ell^\infty(X)$ is also not reflexive.

■

Exercise 21.5. Suppose $p \in (1,\infty)$ and μ is a σ – finite measure on a measurable space $(X,\mathcal{M})$, then $L^p(X,\mathcal{M},\mu)$ is reflexive. **Hint:** model your proof on the proof of item 1. of Theorem 21.20 making use of Theorem 29.6.

Lemma 21.21. *Suppose that (X,o) is a pointed Hausdorff topological space (i.e. $o \in X$ is a fixed point) and ν is a finite measure on $\mathcal{B}_X$ such that*

1. $\mathrm{supp}(\nu) = X$ *while* $\nu(\{o\}) = 0$ *and*
2. *there exists* $f_n \in C(X)$ *such that* $f_n \to 1_{\{o\}}$ *boundedly as* $n \to \infty$.

(For example suppose $X = [0,1]$, $o = 0$, and $\mu = m$.)
Then the map
$$g \in L^1(\nu) \to \varphi_g \in L^\infty(\nu)^*$$
is not surjective and the Banach space $L^1(\nu)$ is not reflexive. (In other words, Theorem 29.6 may fail when $p = \infty$ and L^1 - spaces need not be reflexive.)

Proof. Since $\mathrm{supp}(\nu) = X$, if $f \in C(X)$ we have
$$\|f\|_{L^\infty(\nu)} = \sup\{|f(x)|\, x \in X\}$$
and we may view $C(X)$ as a closed subspace of $L^\infty(\nu)$. For $f \in C(X)$, let $\lambda(f) = f(o)$. Then $\|\lambda\|_{C(X)^*} = 1$, and therefore by Corollary 21.8 of the Hahn-Banach Theorem, there exists an extension $\Lambda \in (L^\infty(\nu))^*$ such that $\lambda = \Lambda|_{C(X)}$ and $\|\Lambda\| = 1$.

If $\Lambda = \varphi_g$ for some $g \in L^1(\nu)$ then we would have
$$f(o) = \lambda(f) = \Lambda(f) = \varphi_g(f) = \int_X f g d\nu \text{ for all } f \in C(X).$$
Applying this equality to the $\{f_n\}_{n=1}^\infty$ in item 2. of the statement of the lemma and then passing to the limit using the dominated convergence theorem, we arrive at the following contradiction;
$$1 = \lim_{n\to\infty} f_n(o) = \lim_{n\to\infty} \int_X f_n g d\nu = \int_X 1_{\{o\}} g d\nu = 0.$$
Hence we must conclude that $\Lambda \neq \varphi_g$ for any $g \in L^1(\nu)$.

Since, by Theorem 29.6, the map $f \in L^\infty(\nu) \to \varphi_f \in L^1(\nu)^*$ is an isometric isomorphism of Banach spaces we may define $L \in L^1(\nu)^{**}$ by
$$L(\varphi_f) := \Lambda(f) \text{ for all } f \in L^\infty(\nu).$$
If L were to equal $\hat{g}$ for some $g \in L^1(\nu)$, then
$$\Lambda(f) = L(\varphi_f) = \hat{g}(\varphi_f) = \varphi_f(g) = \int_X f g d\nu$$
for all $f \in C(X) \subset L^\infty(\nu)$. But we have just seen this is impossible and therefore $L \neq \hat{g}$ for any $g \in L^1(\nu)$ and thus $L^1(\nu)$ is not reflexive. ■

21.0.1 Hahn – Banach Theorem Problems

Exercise 21.6. Suppose that $f : [a,b] \to X$ is a continuous function such that $\dot{f}(t)$ exists for $t \in (a,b)$ and $\dot{f}$ extends to a continuous function on $[a,b]$. Then
$$\|f(b) - f(a)\| \leq \int_a^b \|\dot{f}(t)\| dt \leq (b-a) \cdot \left\|\dot{f}\right\|_\infty, \tag{21.8}$$
where $\left\|\dot{f}\right\|_\infty := \sup_{a<t<b} \left\|\dot{f}(t)\right\|_X$. **Hint:** the Hahn Banach Theorem 21.7 (or Corollary 21.8) implies
$$\|f(b) - f(a)\| = \sup_{\lambda \in X^*,\, \lambda \neq 0} \frac{|\lambda(f(b)) - \lambda(f(a))|}{\|\lambda\|}.$$

Exercise 21.7. Prove Theorem 32.43 using the following strategy.

1. Use the results from the proof in the text of Theorem 32.43 that
$$s \to \int_c^d f(s,t) dt \text{ and } t \to \int_a^b f(s,t) ds$$
are continuous maps.
2. For the moment take $X = \mathbb{R}$ and prove Eq. (32.31) holds by first proving it holds when $f(s,t) = s^m t^n$ with $m, n \in \mathbb{N}_0$. Then use this result along with Theorem 32.39 to show Eq. (32.31) holds for all $f \in C([a,b] \times [c,d], \mathbb{R})$.
3. For the general case, use the special case proved in item 2. along with Hahn Banach Theorem 21.7 (or Corollary 21.8).

Exercise 21.8 (Banach Valued Liouville's Theorem). (This exercise requires knowledge of complex variables.) Let X be a Banach space and $f : \mathbb{C} \to X$ be a function which is complex differentiable at all points $z \in \mathbb{C}$, i.e. $f'(z) := \lim_{h\to 0} (f(z+h) - f(z)/h$ exists for all $z \in \mathbb{C}$. If we further suppose that $M := \sup_{z\in\mathbb{C}} \|f(z)\| < \infty$, then f is constant. **Hint:** use the Hahn Banach Theorem 21.7 (or Corollary 21.8) and the fact the result holds if $X = \mathbb{C}$.

Exercise 21.9.]Let M be a finite dimensional subspace of a normed space, X. Show there exists a closed subspace, N, such that $X = M \oplus N$. **Hint:** let $\beta = \{x_1, \ldots, x_n\} \subset M$ be a basis for M and construct N making use of $\lambda_i \in X^*$ which you should construct to satisfy,
$$\lambda_i(x_j) = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j. \end{cases}$$

Exercise 21.10. Folland 5.21, p. 160.

Exercise 21.11. Let X be a Banach space such that X^* is separable. Show X is separable as well. (The converse is not true as can be seen by taking $X = \ell^1(\mathbb{N})$.) **Hint:** use the greedy algorithm, i.e. suppose $D \subset X^* \setminus \{0\}$ is a countable dense subset of X^*, for $\ell \in D$ choose $x_\ell \in X$ such that $\|x_\ell\| = 1$ and $|\ell(x_\ell)| \geq \frac{1}{2}\|\ell\|$.

Exercise 21.12. Show that any separable Banach space X for which X^* is not separable is **not** reflexive. In particular you may conclude that $L^1([0,1], m)$ is not reflexive where m is Lebesgue measure.

Exercise 21.13. Folland 5.26.

21.0.2 *Quotient spaces, adjoints, and more reflexivity

Definition 21.22. *Let X and Y be Banach spaces and $A : X \to Y$ be a linear operator. The **transpose** of A is the linear operator $A^\dagger : Y^* \to X^*$ defined by $(A^\dagger f)(x) = f(Ax)$ for $f \in Y^*$ and $x \in X$. The **null space** of A is the subspace $\mathrm{Nul}(A) := \{x \in X : Ax = 0\} \subset X$. For $M \subset X$ and $N \subset X^*$ let*

$$M^0 := \{f \in X^* : f|_M = 0\} \text{ and}$$
$$N^\perp := \{x \in X : f(x) = 0 \text{ for all } f \in N\}.$$

Proposition 21.23 (Basic properties of transposes and annihilators).

1. *$\|A\| = \|A^\dagger\|$ and $A^{\dagger\dagger}\hat{x} = \widehat{Ax}$ for all $x \in X$.*
2. *M^0 and $N^\perp$ are always closed subspaces of X^* and X respectively.*
3. *$(M^0)^\perp = \bar{M}$.*
4. *$\bar{N} \subset (N^\perp)^0$ with equality when X is reflexive. (See Exercise 21.4, Example 21.18 above which shows that $\bar{N} \neq (N^\perp)^0$ in general.)*
5. *$\mathrm{Nul}(A) = \mathrm{Ran}(A^\dagger)^\perp$ and $\mathrm{Nul}(A^\dagger) = \mathrm{Ran}(A)^0$. Moreover, $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^\dagger)^\perp$ and if X is reflexive, then $\overline{\mathrm{Ran}(A^\dagger)} = \mathrm{Nul}(A)^0$.*
6. *X is reflexive iff X^* is reflexive. More generally $X^{***} = \widehat{X^*} \oplus \hat{X}^0$ where*

$$\hat{X}^0 = \{\lambda \in X^{***} : \lambda(\hat{x}) = 0 \text{ for all } x \in X\}.$$

Proof.

1.

$$\|A\| = \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|=1} \sup_{\|f\|=1} |f(Ax)|$$
$$= \sup_{\|f\|=1} \sup_{\|x\|=1} \left|A^\dagger f(x)\right| = \sup_{\|f\|=1} \left\|A^\dagger f\right\| = \left\|A^\dagger\right\|.$$

2. This is an easy consequence of the assumed continuity off all linear functionals involved.
3. If $x \in M$, then $f(x) = 0$ for all $f \in M^0$ so that $x \in (M^0)^\perp$. Therefore $\bar{M} \subset (M^0)^\perp$. If $x \notin \bar{M}$, then there exists $f \in X^*$ such that $f|_M = 0$ while $f(x) \neq 0$, i.e. $f \in M^0$ yet $f(x) \neq 0$. This shows $x \notin (M^0)^\perp$ and we have shown $(M^0)^\perp \subset \bar{M}$.
4. It is again simple to show $N \subset (N^\perp)^0$ and therefore $\bar{N} \subset (N^\perp)^0$. Moreover, as above if $f \notin \bar{N}$ there exists $\psi \in X^{**}$ such that $\psi|_{\bar{N}} = 0$ while $\psi(f) \neq 0$. If X is reflexive, $\psi = \hat{x}$ for some $x \in X$ and since $g(x) = \psi(g) = 0$ for all $g \in \bar{N}$, we have $x \in N^\perp$. On the other hand, $f(x) = \psi(f) \neq 0$ so $f \notin (N^\perp)^0$. Thus again $(N^\perp)^0 \subset \bar{N}$.
5.

$$\begin{aligned}\mathrm{Nul}(A) &= \{x \in X : Ax = 0\} = \{x \in X : f(Ax) = 0 \ \forall\ f \in X^*\} \\ &= \left\{x \in X : A^\dagger f(x) = 0 \ \forall\ f \in X^*\right\} \\ &= \left\{x \in X : g(x) = 0 \ \forall\ g \in \mathrm{Ran}(A^\dagger)\right\} = \mathrm{Ran}(A^\dagger)^\perp.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathrm{Nul}(A^\dagger) &= \left\{f \in Y^* : A^\dagger f = 0\right\} = \left\{f \in Y^* : (A^\dagger f)(x) = 0 \ \forall\ x \in X\right\} \\ &= \{f \in Y^* : f(Ax) = 0 \ \forall\ x \in X\} \\ &= \left\{f \in Y^* : f|_{\mathrm{Ran}(A)} = 0\right\} = \mathrm{Ran}(A)^0.\end{aligned}$$

6. Let $\psi \in X^{***}$ and define $f_\psi \in X^*$ by $f_\psi(x) = \psi(\hat{x})$ for all $x \in X$ and set $\psi' := \psi - \hat{f}_\psi$. For $x \in X$ (so $\hat{x} \in X^{**}$) we have

$$\psi'(\hat{x}) = \psi(\hat{x}) - \hat{f}_\psi(\hat{x}) = f_\psi(x) - \hat{x}(f_\psi) = f_\psi(x) - f_\psi(x) = 0.$$

This shows $\psi' \in \hat{X}^0$ and we have shown $X^{***} = \widehat{X^*} + \hat{X}^0$. If $\psi \in \widehat{X^*} \cap \hat{X}^0$, then $\psi = \hat{f}$ for some $f \in X^*$ and $0 = \hat{f}(\hat{x}) = \hat{x}(f) = f(x)$ for all $x \in X$, i.e. $f = 0$ so $\psi = 0$. Therefore $X^{***} = \widehat{X^*} \oplus \hat{X}^0$ as claimed. If X is reflexive, then $\hat{X} = X^{**}$ and so $\hat{X}^0 = \{0\}$ showing $X^{***} = \widehat{X^*}$, i.e. X^* is reflexive. Conversely if X^* is reflexive we conclude that $\hat{X}^0 = \{0\}$ and therefore $X^{**} = \{0\}^\perp = \left(\hat{X}^0\right)^\perp = \hat{X}$, so that X is reflexive.

Alternative proof. Notice that $f_\psi = J^\dagger\psi$, where $J : X \to X^{**}$ is given by $Jx = \hat{x}$, and the composition

$$f \in X^* \xrightarrow{\hat{}} \hat{f} \in X^{***} \xrightarrow{J^\dagger} J^\dagger \hat{f} \in X^*$$

is the identity map since $\left(J^\dagger \hat{f}\right)(x) = \hat{f}(Jx) = \hat{f}(\hat{x}) = \hat{x}(f) = f(x)$ for all $x \in X$. Thus it follows that $X^* \xrightarrow{\hat{}} X^{***}$ is invertible iff $J^\dagger$ is its inverse which can happen iff $\operatorname{Nul}(J^\dagger) = \{0\}$. But as above $\operatorname{Nul}(J^\dagger) = \operatorname{Ran}(J)^0$ which will be zero iff $\overline{\operatorname{Ran}(J)} = X^{**}$ and since J is an isometry this is equivalent to saying $\operatorname{Ran}(J) = X^{**}$. So we have again shown X^* is reflexive iff X is reflexive.

■

Theorem 21.24 (Banach Space Factor Theorem). *Let X be a Banach space, $M \subset X$ be a proper closed subspace, X/M the quotient space, $\pi : X \to X/M$ the projection map $\pi(x) = x + M$ for $x \in X$ and define the quotient norm on X/M by*

$$\|\pi(x)\|_{X/M} = \|x+M\|_{X/M} = \inf_{m\in M} \|x+m\|_X .$$

Then:

1. *$\|\cdot\|_{X/M}$ is a norm on X/M.*
2. *The projection map $\pi : X \to X/M$ has norm 1, $\|\pi\| = 1$.*
3. *$(X/M, \|\cdot\|_{X/M})$ is a Banach space.*
4. *If Y is another normed space and $T : X \to Y$ is a bounded linear transformation such that $M \subset \operatorname{Nul}(T)$, then there exists a unique linear transformation $S : X/M \to Y$ such that $T = S \circ \pi$ and moreover $\|T\| = \|S\|$.*

Proof.

1. Clearly $\|x + M\| \ge 0$ and if $\|x + M\| = 0$, then there exists $m_n \in M$ such that $\|x + m_n\| \to 0$ as $n \to \infty$, i.e. $x = -\lim_{n\to\infty} m_n \in \bar{M} = M$. Since $x \in M$, $x + M = 0 \in X/M$. If $c \in \mathbb{C}\setminus\{0\}$, $x \in X$, then
$$\|cx + M\| = \inf_{m\in M} \|cx + m\| = |c| \inf_{m\in M} \|x + m/c\| = |c| \|x + M\|$$
because m/c runs through M as m runs through M. Let $x_1, x_2 \in X$ and $m_1, m_2 \in M$ then
$$\|x_1 + x_2 + M\| \le \|x_1 + x_2 + m_1 + m_2\| \le \|x_1 + m_1\| + \|x_2 + m_2\|.$$
Taking infimums over $m_1, m_2 \in M$ then implies
$$\|x_1 + x_2 + M\| \le \|x_1 + M\| + \|x_2 + M\|.$$
and we have completed the proof the $(X/M, \|\cdot\|)$ is a normed space.
2. Since $\|\pi(x)\| = \inf_{m\in M} \|x + m\| \le \|x\|$ for all $x \in X$, $\|\pi\| \le 1$. To see $\|\pi\| = 1$, let $x \in X \setminus M$ so that $\pi(x) \ne 0$. Given $\alpha \in (0, 1)$, there exists $m \in M$ such that
$$\|x + m\| \le \alpha^{-1} \|\pi(x)\| .$$
Therefore,
$$\frac{\|\pi(x+m)\|}{\|x+m\|} = \frac{\|\pi(x)\|}{\|x+m\|} \ge \frac{\alpha \|x+m\|}{\|x+m\|} = \alpha$$
which shows $\|\pi\| \ge \alpha$. Since $\alpha \in (0, 1)$ is arbitrary we conclude that $\|\pi(x)\| = 1$.
3. Let $\pi(x_n) \in X/M$ be a sequence such that $\sum \|\pi(x_n)\| < \infty$. As above there exists $m_n \in M$ such that $\|\pi(x_n)\| \ge \frac{1}{2}\|x_n + m_n\|$ and hence $\sum \|x_n + m_n\| \le 2\sum \|\pi(x_n)\| < \infty$. Since X is complete, $x := \sum_{n=1}^{\infty} (x_n + m_n)$ exists in X and therefore by the continuity of π,
$$\pi(x) = \sum_{n=1}^{\infty} \pi(x_n + m_n) = \sum_{n=1}^{\infty} \pi(x_n)$$
showing X/M is complete.
4. The existence of S is guaranteed by the "factor theorem" from linear algebra. Moreover $\|S\| = \|T\|$ because
$$\|T\| = \|S \circ \pi\| \le \|S\| \|\pi\| = \|S\|$$
and
$$\|S\| = \sup_{x\notin M} \frac{\|S(\pi(x))\|}{\|\pi(x)\|} = \sup_{x\notin M} \frac{\|Tx\|}{\|\pi(x)\|} \ge \sup_{x\notin M} \frac{\|Tx\|}{\|x\|} = \sup_{x\neq 0} \frac{\|Tx\|}{\|x\|} = \|T\| .$$

■

Theorem 21.25. *Let X be a Banach space. Then*

1. *Identifying X with $\hat{X} \subset X^{**}$, the weak $-*$ topology on X^{**} induces the weak topology on X. More explicitly, the map $x \in X \to \hat{x} \in \hat{X}$ is a homeomorphism when X is equipped with its weak topology and $\hat{X}$ with the relative topology coming from the weak-$*$ topology on X^{**}.*
2. *$\hat{X} \subset X^{**}$ is dense in the weak-$*$ topology on X^{**}.*
3. *Letting C and C^{**} be the closed unit balls in X and X^{**} respectively, then $\hat{C} := \{\hat{x} \in C^{**} : x \in C\}$ is dense in C^{**} in the weak $-*$ topology on X^{**}.*
4. *X is reflexive iff C is weakly compact.*

(See Definition 36.19 for the topologies being used here.)

Proof.

1. The weak $-*$ topology on X^{**} is generated by
$$\left\{\hat{f} : f \in X^*\right\} = \{\psi \in X^{**} \to \psi(f) : f \in X^*\}.$$
So the induced topology on X is generated by
$$\{x \in X \to \hat{x} \in X^{**} \to \hat{x}(f) = f(x) : f \in X^*\} = X^*$$
and so the induced topology on X is precisely the weak topology.
2. A basic weak - $*$ neighborhood of a point $\lambda \in X^{**}$ is of the form
$$\mathcal{N} := \cap_{k=1}^{n} \{\psi \in X^{**} : |\psi(f_k) - \lambda(f_k)| < \varepsilon\} \tag{21.9}$$
for some $\{f_k\}_{k=1}^n \subset X^*$ and $\varepsilon > 0$. be given. We must now find $x \in X$ such that $\hat{x} \in \mathcal{N}$, or equivalently so that
$$|\hat{x}(f_k) - \lambda(f_k)| = |f_k(x) - \lambda(f_k)| < \varepsilon \text{ for } k = 1, 2, \ldots, n. \tag{21.10}$$
In fact we will show there exists $x \in X$ such that $\lambda(f_k) = f_k(x)$ for $k = 1, 2, \ldots, n$. To prove this stronger assertion we may, by discarding some of the f_k's if necessary, assume that $\{f_k\}_{k=1}^n$ is a linearly independent set. Since the $\{f_k\}_{k=1}^n$ are linearly independent, the map $x \in X \to (f_1(x), \ldots, f_n(x)) \in \mathbb{C}^n$ is surjective (why) and hence there exists $x \in X$ such that
$$(f_1(x), \ldots, f_n(x)) = Tx = (\lambda(f_1), \ldots, \lambda(f_n)) \tag{21.11}$$
as desired.
3. Let $\lambda \in C^{**} \subset X^{**}$ and $\mathcal{N}$ be the weak - $*$ open neighborhood of λ as in Eq. (21.9). Working as before, given $\varepsilon > 0$, we need to find $x \in C$ such that Eq. (21.10). It will be left to the reader to verify that it suffices again to assume $\{f_k\}_{k=1}^n$ is a linearly independent set. (Hint: Suppose that $\{f_1, \ldots, f_m\}$ were a maximal linearly dependent subset of $\{f_k\}_{k=1}^n$, then each f_k with $k > m$ may be written as a linear combination $\{f_1, \ldots, f_m\}$.) As in the proof of item 2., there exists $x \in X$ such that Eq. (21.11) holds. The problem is that x may not be in C. To remedy this, let $N := \cap_{k=1}^n \mathrm{Nul}(f_k) = \mathrm{Nul}(T)$, $\pi : X \to X/N \cong \mathbb{C}^n$ be the projection map and $\bar{f}_k \in (X/N)^*$ be chosen so that $f_k = \bar{f}_k \circ \pi$ for $k = 1, 2, \ldots, n$. Then we have produced $x \in X$ such that
$$(\lambda(f_1), \ldots, \lambda(f_n)) = (f_1(x), \ldots, f_n(x)) = (\bar{f}_1(\pi(x)), \ldots, \bar{f}_n(\pi(x))).$$
Since $\left\{\bar{f}_1, \ldots, \bar{f}_n\right\}$ is a basis for $(X/N)^*$ we find
$$\begin{aligned}\|\pi(x)\| &= \sup_{\alpha \in \mathbb{C}^n \setminus \{0\}} \frac{\left|\sum_{i=1}^n \alpha_i \bar{f}_i(\pi(x))\right|}{\left\|\sum_{i=1}^n \alpha_i \bar{f}_i\right\|} = \sup_{\alpha \in \mathbb{C}^n \setminus \{0\}} \frac{\left|\sum_{i=1}^n \alpha_i \lambda(f_i)\right|}{\left\|\sum_{i=1}^n \alpha_i f_i\right\|} \\ &= \sup_{\alpha \in \mathbb{C}^n \setminus \{0\}} \frac{\left|\lambda(\sum_{i=1}^n \alpha_i f_i)\right|}{\left\|\sum_{i=1}^n \alpha_i f_i\right\|} \\ &\le \|\lambda\| \sup_{\alpha \in \mathbb{C}^n \setminus \{0\}} \frac{\left\|\sum_{i=1}^n \alpha_i f_i\right\|}{\left\|\sum_{i=1}^n \alpha_i f_i\right\|} = 1.\end{aligned}$$
Hence we have shown $\|\pi(x)\| \le 1$ and therefore for any $\alpha > 1$ there exists $y = x + n \in X$ such that $\|y\| < \alpha$ and $(\lambda(f_1), \ldots, \lambda(f_n)) = (f_1(y), \ldots, f_n(y))$. Hence
$$|\lambda(f_i) - f_i(y/\alpha)| \le \left|f_i(y) - \alpha^{-1} f_i(y)\right| \le (1 - \alpha^{-1}) |f_i(y)|$$
which can be arbitrarily small (i.e. less than ε) by choosing α sufficiently close to 1.
4. Let $\hat{C} := \{\hat{x} : x \in C\} \subset C^{**} \subset X^{**}$. If X is reflexive, $\hat{C} = C^{**}$ is weak - $*$ compact and hence by item 1., C is weakly compact in X. Conversely if C is weakly compact, then $\hat{C} \subset C^{**}$ is weak $-*$ compact being the continuous image of a continuous map. Since the weak $-*$ topology on X^{**} is Hausdorff, it follows that $\hat{C}$ is weak $-*$ closed and so by item 3, $C^{**} = \overline{\hat{C}}^{\text{weak}-*} = \hat{C}$. So if $\lambda \in X^{**}$, $\lambda / \|\lambda\| \in C^{**} = \hat{C}$, i.e. there exists $x \in C$ such that $\hat{x} = \lambda / \|\lambda\|$. This shows $\lambda = (\|\lambda\| x)\hat{}$ and therefore $\hat{X} = X^{**}$.

■

21.0.3 Hahn-Banach Theorem Problems

Exercise 21.14. Let X be a normed vector space. Show a linear functional, $f : X \to \mathbb{C}$, is bounded iff $M := f^{-1}(\{0\})$ is closed.

Exercise 21.15. Let M be a closed subspace of a normed space, X, and $x \in X \setminus M$. Show $M \oplus \mathbb{C}x$ is closed. **Hint**: make use of a $\lambda \in X^*$ which you should construct so that $\lambda(M) = 0$ while $\lambda(x) \neq 0$.

Exercise 21.16 (Uses quotient spaces). Let X be an infinite dimensional normed vector space. Show:

1. There exists a sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $\|x_n\| = 1$ for all n and $\|x_m - x_n\| \ge \frac{1}{2}$ for all $m \neq n$.
2. Show X is not locally compact.

22

Baire Category Theorem

Let (X,τ) be a topological space and $A \subset X$. Recall that the **interior of** A is defined as

$$A^o := \cup \{V : V \text{ is open and } V \subset A\}, \tag{22.1}$$

i.e. A^o is the largest open subset of A. Similarly the **closure of** A is defined as

$$\bar{A} = \cap \{F : F \text{ closed and } A \subset F\}, \tag{22.2}$$

i.e. $\bar{A}$ is the smallest closed set containing A.

Exercise 22.1. Suppose that (X,τ_X) and (Y,τ_Y) are two topological spaces and $\varphi : X \to Y$ is a homeomorphism, i.e. φ is continuous, invertible, and φ^{-1} is continuous. Show $\varphi(A^o) = [\varphi(A)]^o$ and $\varphi(\bar{A}) = \overline{\varphi(A)}$ for all $A \subset X$.

Proposition 22.1 (Interiors, Closures, and Complements). *Let (X,τ) be a topological space and $A \subset X$. Then*

1. $(A^o)^c = \overline{A^c}$ *and*
2. $(\bar{A})^c = (A^c)^o$.

Proof. Using Eqs. (22.1) and (22.2) we find,

$$\begin{aligned}(A^o)^c &= \cap \{V^c : V \text{ is open and } V \subset A\} \\ &= \cap \{V^c : V^c \text{ is closed and } A^c \subset V^c\} = \overline{A^c}\end{aligned}$$

and

$$\begin{aligned}(\bar{A})^c &= [\cap \{F : F \text{ closed and } A \subset F\}]^c \\ &= \cup \{F^c : F \text{ closed and } A \subset F\} \\ &= \cup \{F^c : F^c \text{ open and } F^c \subset A^c\} = (A^c)^o.\end{aligned}$$

■

Definition 22.2. *Let (X,τ) be a topological space. A set $E \subset X$ is said to be **nowhere dense** if $(\bar{E})^o = \emptyset$ i.e. $\bar{E}$ has empty interior.*

In other words E is nowhere dense if $\bar{E}$ contains no non-empty open subsets of X. In contrast recall that E is dense in X iff $\bar{E} = X$ which is equivalent to saying that $\bar{E}$ contains all open subsets of X. Here is a couple of simple remarks that we will use without comment in the future.

Exercise 22.2. Let (X,τ) be a topological space and E and G be subsets of X. Prove;

1. E is nowhere dense iff E^c has dense interior.
2. $G \subset X$ is dense iff $G \cap W \neq \emptyset$ for all $\emptyset \neq W \subset_o X$.

22.1 Metric Space Baire Category Theorem

Theorem 22.3 (Baire Category Theorem). *Let (X,ρ) be a complete metric space.*

1. *The countable intersection of dense open subsets of X is still dense in X, i.e. if $\{V_n\}_{n=1}^\infty$ is a sequence of dense open sets, then $G := \bigcap_{n=1}^\infty V_n$ is dense in X.*
2. *X is not the countable union of nowhere dense subsets of X, i.e. if $\{E_n\}_{n=1}^\infty$ is a sequence of nowhere dense sets, then $\bigcup_{n=1}^\infty E_n \subset \bigcup_{n=1}^\infty \bar{E}_n \subsetneqq X$ and in particular $X \neq \bigcup_{n=1}^\infty E_n$. In fact we have $\bigcup_{n=1}^\infty E_n$ has empty interior.*[1]

Proof. 1. From Exercise 22.2 we must show $W \cap G \neq \emptyset$ for all non-empty open sets $W \subset X$. Since V_1 is dense, $W \cap V_1 \neq \emptyset$ and hence there exists $x_1 \in X$ and $\varepsilon_1 > 0$ such that

$$\overline{B(x_1,\varepsilon_1)} \subset W \cap V_1.$$

Since V_2 is dense, $B(x_1,\varepsilon_1) \cap V_2 \neq \emptyset$ and hence there exists $x_2 \in X$ and $\varepsilon_2 > 0$ such that

$$\overline{B(x_2,\varepsilon_2)} \subset B(x_1,\varepsilon_1) \cap V_2.$$

Continuing this way inductively, we may choose $\{x_n \in X \text{ and } \varepsilon_n > 0\}_{n=1}^\infty$ such that

$$\overline{B(x_n,\varepsilon_n)} \subset B(x_{n-1},\varepsilon_{n-1}) \cap V_n \; \forall n.$$

Furthermore we can clearly do this construction in such a way that $\varepsilon_n \downarrow 0$ as $n \uparrow \infty$. Hence $\{x_n\}_{n=1}^\infty$ is Cauchy sequence and $x = \lim_{n\to\infty} x_n$ exists in X since

[1] Take $X = \mathbb{R}$ and $E_n = \{r_n\}_{n=1}^\infty$ so that $\cup E_n = \mathbb{Q}$. Thus $\cup E_n$ is dense in $\mathbb{R}$ so we certainly can not assert that the countable union of nowhere dense sets is nowhere dense.

X is complete. Since $\overline{B(x_n,\varepsilon_n)}$ is closed, $x \in \overline{B(x_n,\varepsilon_n)} \subset V_n$ so that $x \in V_n$ for all n and hence $x \in G$. Moreover, $x \in \overline{B(x_1,\varepsilon_1)} \subset W \cap V_1$ implies $x \in W$ and hence $x \in W \cap G$ showing $W \cap G \neq \emptyset$.

2. Now suppose that $\{E_n\}$ are nowhere dense. Let $F_n := \bar{E}_n$ so that $F_n^o = \emptyset$, i.e. $X = (F_n^o)^c = \overline{F_n^c}$. Thus F_n^c is a dense open set and therefore $G := \cap_{n=1}^\infty F_n^c$ is still dense in X, i.e. $X = \bar{G} = \overline{\cap_{n=1}^\infty F_n^c}$. This equivalent to

$$\emptyset = X^c = \left[\overline{\cap_{n=1}^\infty F_n^c}\right]^c = \left[(\cap_{n=1}^\infty F_n^c)^c\right]^o = (\cup_{n=1}^\infty F_n)^o = \left(\cup_{n=1}^\infty \bar{E}_n\right)^o,$$

which is to say $\cup_{n=1}^\infty E_n \subset \cup_{n=1}^\infty \bar{E}_n$ has empty interior and in particular $\cup_{n=1}^\infty E_n \subset \cup_{n=1}^\infty \bar{E}_n \neq X$. ■

Example 22.4. Suppose that X is a countable set and ρ is a metric on X for which no single point set is open. [Then (X,ρ) is **not** complete. Indeed we may assume $X = \mathbb{N}$ and let $E_n := \{n\} \subset \mathbb{N}$ for all $n \in \mathbb{N}$. Then E_n is closed and by assumption it has empty interior. Since $X = \cup_{n\in\mathbb{N}} E_n$, it follows from the Baire Category Theorem 22.3 that (X,ρ) can not be complete.

For example, take $X = \mathbb{Q}$ and $\rho(x,y) := |y - x|$. More generally, take any metric on $\mathbb{Q}$ which generates the same topology as the metric ρ, then this metric can not be complete.

22.2 Locally Compact Hausdorff Space Baire Category Theorem

Here is another version of the Baire Category theorem when X is a locally compact Hausdorff space.

Theorem 22.5. *Let X be a locally compact Hausdorff space.*

1. *If $\{V_n\}_{n=1}^\infty$ is a sequence of dense open sets, then $G := \bigcap_{n=1}^\infty V_n$ is dense in X.*
2. *If $\{E_n\}_{n=1}^\infty$ is a sequence of nowhere dense sets, then $\left[\bigcup_{n=1}^\infty E_n\right]^o = \emptyset$ and in particular $X \neq \bigcup_{n=1}^\infty E_n$.*

Proof. As in the proof of Theorem 22.3, the second assertion is a consequence of the first. To finish the proof, if suffices to show $G \cap W \neq \emptyset$ for all open sets $W \subset X$. Since V_1 is dense, there exists $x_1 \in V_1 \cap W$ and by Proposition 37.7 [below!] there exists $U_1 \subset_o X$ such that $x_1 \in U_1 \subset \bar{U}_1 \subset V_1 \cap W$ with $\bar{U}_1$ being compact. Similarly, there exists a non-empty open set U_2 such that $U_2 \subset \bar{U}_2 \subset U_1 \cap V_2$. Working inductively, we may find non-empty open sets $\{U_k\}_{k=1}^\infty$ such that $U_k \subset \bar{U}_k \subset U_{k-1} \cap V_k$. Since $\cap_{k=1}^n \bar{U}_k = \bar{U}_n \neq \emptyset$ for all n, the finite intersection characterization of $\bar{U}_1$ being compact implies that

$$\emptyset \neq \cap_{k=1}^\infty \bar{U}_k \subset G \cap W.$$

■

Definition 22.6. *A subset $E \subset X$ is **meager** or of the **first category** if $E = \bigcup_{n=1}^\infty E_n$ where each E_n is nowhere dense. And a set $R \subset X$ is called **residual**[2] if R^c is meager. See Exercise 22.3 below for an equivalent definition of residual.*

Remarks 22.7 *For those readers that already know some measure theory may want to think of meager as being the topological analogue of sets of measure 0 and residual as being the topological analogue of sets of full measure. (This analogy should not be taken too seriously, see Exercise **??**.)*

1. *R is residual iff R contains a countable intersection of dense open sets. Indeed if R is a residual set, then there exists nowhere dense sets $\{E_n\}$ such that*
$$R^c = \cup_{n=1}^\infty E_n \subset \cup_{n=1}^\infty \bar{E}_n.$$
Taking complements of this equation shows that
$$\cap_{n=1}^\infty \bar{E}_n^c \subset R,$$
i.e. R contains a set of the form $\cap_{n=1}^\infty V_n$ with each V_n $(= \bar{E}_n^c)$ being an open dense subset of X.
Conversely, if $\cap_{n=1}^\infty V_n \subset R$ with each V_n being an open dense subset of X, then $R^c \subset \cup_{n=1}^\infty V_n^c$ and hence $R^c = \cup_{n=1}^\infty E_n$ where each $E_n = R^c \cap V_n^c$, is a nowhere dense subset of X.
2. *A countable union of meager sets is meager and any subset of a meager set is meager. Indeed, if $E = \bigcup_{n=1}^\infty E_n$ is a meager set and $F \subset E$, then $F = \bigcup_{n=1}^\infty F \cap E_n$ is meager as well since $\left[\overline{F \cap E_n}\right]^o \subset \left[\overline{E_n}\right]^o = \emptyset$ for all n.*
3. *A countable intersection of residual sets is residual.*

Exercise 22.3. Recall that $R \subset X$ is a residual set if R^c is meager, i.e. R^c is the countable union of nowhere dense sets. Show R is residual iff $R = \cap_{n=1}^\infty A_n$ for some $\{A_n\}_{n=1}^\infty$ such that each A_n has dense interior, i.e. $\overline{A_n^o} = X$.

Remarks 22.8 *The Baire Category Theorems may now be stated as follows. If X is a complete metric space or X is a locally compact Hausdorff space, then*

1. *all residual sets are dense in X and*

[2] Dictionary definition: Relating to or indicating a remainder, "residual quantity."

2. *Meager sents have empty interior and in particular X is not meager.*

It should also be remarked that incomplete metric spaces may be meager. For example, let $X \subset C([0,1])$ be the subspace of polynomial functions on $[0,1]$ equipped with the supremum norm. Then $X = \cup_{n=1}^\infty E_n$ where $E_n \subset X$ denotes the subspace of polynomials of degree less than or equal to n. You are asked to show in Exercise 22.4 below that E_n is nowhere dense for all n. Hence X is meager and the empty set is residual in X.

Here is an application of Theorem 22.3.

Theorem 22.9. *Let $\mathcal{N} \subset C([0,1],\mathbb{R})$ be the set of nowhere differentiable functions. (Here a function f is said to be differentiable at 0 if $f'(0) := \lim_{t\downarrow 0} \frac{f(t)-f(0)}{t}$ exists and at 1 if $f'(1) := \lim_{t\uparrow 0} \frac{f(1)-f(t)}{1-t}$ exists.) Then $\mathcal{N}$ is a residual set so the "generic" continuous functions is nowhere differentiable.*

Proof. If $f \notin \mathcal{N}$, then $f'(x_0)$ exists for some $x_0 \in [0,1]$ and by the definition of the derivative and compactness of $[0,1]$, there exists $n \in \mathbb{N}$ such that $|f(x) - f(x_0)| \le n|x - x_0|\ \forall\ x \in [0,1]$. Thus if we define

$$E_n := \{f \in C([0,1]) : \exists\ x_0 \in [0,1]\ \ni\ |f(x) - f(x_0)| \le n|x - x_0|\ \forall\ x \in [0,1]\},$$

then we have just shown $\mathcal{N}^c \subset E := \cup_{n=1}^\infty E_n$. So to finish the proof it suffices to show (for each n) E_n is a closed subset of $C([0,1],\mathbb{R})$ with empty interior.

1. To prove E_n is closed, let $\{f_m\}_{m=1}^\infty \subset E_n$ be a sequence of functions such that there exists $f \in C([0,1],\mathbb{R})$ such that $\|f - f_m\|_\infty \to 0$ as $m \to \infty$. Since $f_m \in E_n$, there exists $x_m \in [0,1]$ such that

$$|f_m(x) - f_m(x_m)| \le n|x - x_m|\ \forall\ x \in [0,1]. \tag{22.3}$$

Since $[0,1]$ is a compact metric space, by passing to a subsequence if necessary, we may assume $x_0 = \lim_{m\to\infty} x_m \in [0,1]$ exists. Passing to the limit in Eq. (22.3), making use of the uniform convergence of $f_n \to f$ to show $\lim_{m\to\infty} f_m(x_m) = f(x_0)$, implies

$$|f(x) - f(x_0)| \le n|x - x_0|\ \forall\ x \in [0,1]$$

and therefore that $f \in E_n$. This shows E_n is a closed subset of $C([0,1],\mathbb{R})$.

2. To finish the proof, we will show $E_n^0 = \emptyset$ by showing for each $f \in E_n$ and $\varepsilon > 0$ given, there exists $g \in C([0,1],\mathbb{R}) \setminus E_n$ such that $\|f - g\|_\infty < \varepsilon$. We now construct g. Since $[0,1]$ is compact and f is continuous there exists $N \in \mathbb{N}$ such that $|f(x) - f(y)| < \varepsilon/2$ whenever $|y - x| < 1/N$. Let k denote the piecewise linear function on $[0,1]$ such that $k(\frac{m}{N}) = f(\frac{m}{N})$ for $m = 0,1,\dots,N$ and $k''(x) = 0$ for $x \notin \pi_N := \{m/N : m = 0,1,\dots,N\}$. Then it is easily seen that $\|f - k\|_u < \varepsilon/2$ and for $x \in (\frac{m}{N}, \frac{m+1}{N})$ that

$$|k'(x)| = \frac{|f(\frac{m+1}{N}) - f(\frac{m}{N})|}{\frac{1}{N}} < N\varepsilon/2.$$

We now make k "rougher" by adding a small wiggly function h which we define as follows. Let $M \in \mathbb{N}$ be chosen so that $4\varepsilon M > 2n$ and define h uniquely by $h(\frac{m}{M}) = (-1)^m \varepsilon/2$ for $m = 0,1,\dots,M$ and $h''(x) = 0$ for $x \notin \pi_M$. Then $\|h\|_\infty < \varepsilon$ and $|h'(x)| = 4\varepsilon M > 2n$ for $x \notin \pi_M$. See Figure 22.1 below. Finally

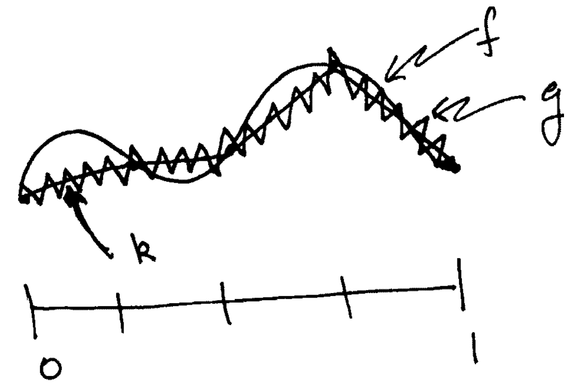


Fig. 22.1. Constgructing a rough approximation, g, to a continuous function f.

define $g := k + h$. Then

$$\|f - g\|_\infty \le \|f - k\|_\infty + \|h\|_\infty < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

and

$$|g'(x)| \ge |h'(x)| - |k'(x)| > 2n - n = n\ \forall x \notin \pi_M \cup \pi_N.$$

It now follows from this last equation and the mean value theorem that for any $x_0 \in [0,1]$,

$$\left|\frac{g(x) - g(x_0)}{x - x_0}\right| > n$$

for all $x \in [0,1]$ sufficiently close to x_0. This shows $g \notin E_n$ and so the proof is complete. ■

Here is an application of the Baire Category Theorem 22.5. For more applications along these lines, see [24] and the references therein.

Proposition 22.10. *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a function such that $f'(x)$ exists for all $x \in \mathbb{R}$. Let*

$$U := \bigcup_{\varepsilon>0}\left\{x\in\mathbb{R} : \sup_{|y|<\varepsilon}|f'(x+y)|<\infty\right\}.$$

Then U is a dense open set. (It is not true that $U=\mathbb{R}$ in general, see Example 25.27 below.)

Proof. It is easily seen from the definition of U that U is open. Let $W\subset_o\mathbb{R}$ be an open subset of $\mathbb{R}$. For $k\in\mathbb{N}$, let

$$\begin{aligned}E_k &:= \left\{x\in W : |f(y)-f(x)|\le k\,|y-x| \text{ when } |y-x|\le\frac{1}{k}\right\}\\ &= \bigcap_{z:|z|\le k^{-1}}\{x\in W : |f(x+z)-f(x)|\le k\,|z|\},\end{aligned}$$

which is a closed subset of $\mathbb{R}$ since f is continuous. Moreover, if $x\in W$ and $M=|f'(x)|$, then

$$\begin{aligned}|f(y)-f(x)| &= |f'(x)\,(y-x)+o\,(y-x)|\\ &\le (M+1)\,|y-x|\end{aligned}$$

for y close to x. (Here $o(y-x)$ denotes a function such that $\lim_{y\to x}o(y-x)/(y-x)=0$.) In particular, this shows that $x\in E_k$ for all k sufficiently large. Therefore $W=\cup_{k=1}^{\infty}E_k$ and since W is not meager by the Baire category Theorem 22.5, some E_k has non-empty interior. That is there exists $x_0\in E_k\subset W$ and $\varepsilon>0$ such that

$$J := (x_0-\varepsilon, x_0+\varepsilon)\subset E_k\subset W.$$

For $x\in J$, we have $|f(x+z)-f(x)|\le k\,|z|$ provided that $|z|\le k^{-1}$ and therefore that $|f'(x)|\le k$ for $x\in J$. Therefore $x_0\in U\cap W$ showing U is dense. ■

Remark 22.11. This proposition generalizes to functions $f:\mathbb{R}^n\to\mathbb{R}^m$ in an obvious way.

For our next application of Theorem 22.3, let $X := BC^{\infty}((-1,1))$ denote the set of smooth functions f on $(-1,1)$ such that f and all of its derivatives are bounded. In the metric

$$\rho(f,g) := \sum_{k=0}^{\infty}2^{-k}\frac{\left\|f^{(k)}-g^{(k)}\right\|_\infty}{1+\left\|f^{(k)}-g^{(k)}\right\|_\infty} \text{ for } f,g\in X,$$

X becomes a complete metric space.

Theorem 22.12. *Given an increasing sequence of positive numbers $\{M_n\}_{n=1}^{\infty}$, the set*

$$\mathcal{F} := \left\{f\in X : \limsup_{n\to\infty}\left|\frac{f^{(n)}(0)}{M_n}\right|\ge 1\right\}$$

is dense in X. In particular, there is a dense set of $f\in X$ such that the power series expansion of f at 0 has zero radius of convergence.

Proof. Step 1. Let $n\in\mathbb{N}$. Choose $g\in C_c^{\infty}((-1,1))$ such that $\|g\|_\infty<2^{-n}$ while $g'(0)=2M_n$ and define

$$f_n(x) := \int_0^x dt_{n-1}\int_0^{t_{n-1}}dt_{n-2}\dots\int_0^{t_2}dt_1 g(t_1).$$

Then for $k<n$,

$$f_n^{(k)}(x) = \int_0^x dt_{n-k-1}\int_0^{t_{n-k-1}}dt_{n-k-2}\dots\int_0^{t_2}dt_1 g(t_1),$$

$f^{(n)}(x)=g'(x)$, $f_n^{(n)}(0)=2M_n$ and $f_n^{(k)}$ satisfies

$$\left\|f_n^{(k)}\right\|_\infty \le \frac{2^{-n}}{(n-1-k)!}\le 2^{-n} \text{ for } k<n.$$

Consequently,

$$\begin{aligned}\rho(f_n,0) &= \sum_{k=0}^{\infty}2^{-k}\frac{\left\|f_n^{(k)}\right\|_\infty}{1+\left\|f_n^{(k)}\right\|_\infty}\\ &\le \sum_{k=0}^{n-1}2^{-k}2^{-n}+\sum_{k=n}^{\infty}2^{-k}\cdot 1\le 2\left(2^{-n}+2^{-n}\right)=4\cdot 2^{-n}.\end{aligned}$$

Thus we have constructed $f_n\in X$ such that $\lim_{n\to\infty}\rho(f_n,0)=0$ while $f_n^{(n)}(0)=2M_n$ for all n.

Step 2. The set

$$G_n := \cup_{m\ge n}\left\{f\in X : \left|f^{(m)}(0)\right|>M_m\right\}$$

is a dense open subset of X. The fact that G_n is open is clear. To see that G_n is dense, let $g\in X$ be given and define $g_m := g+\varepsilon_m f_m$ where $\varepsilon_m := sgn(g^{(m)}(0))$. Then

$$\left|g_m^{(m)}(0)\right| = \left|g^{(m)}(0)\right|+\left|f_m^{(m)}(0)\right|\ge 2M_m>M_m \text{ for all } m.$$

Therefore, $g_m\in G_n$ for all $m\ge n$ and since

$$\rho(g_m, g) = \rho(f_m, 0) \to 0 \text{ as } m \to \infty$$

it follows that $g \in \bar{G}_n$.

Step 3. By the Baire Category theorem, $\cap G_n$ is a dense subset of X. This completes the proof of the first assertion since

$$\mathcal{F} = \left\{ f \in X : \limsup_{n\to\infty} \left| \frac{f^{(n)}(0)}{M_n} \right| \geq 1 \right\}$$
$$= \cap_{n=1}^{\infty} \left\{ f \in X : \left| \frac{f^{(n)}(0)}{M_n} \right| \geq 1 \text{ for some } n \geq m \right\} \supset \cap_{n=1}^{\infty} G_n.$$

Step 4. Take $M_n = (n!)^2$ and recall that the power series expansion for f near 0 is given by $\sum_{n=0}^{\infty} \frac{f_n(0)}{n!} x^n$. This series can not converge for any $f \in \mathcal{F}$ and any $x \neq 0$ because

$$\limsup_{n\to\infty} \left| \frac{f_n(0)}{n!} x^n \right| = \limsup_{n\to\infty} \left| \frac{f_n(0)}{(n!)^2} n! x^n \right|$$
$$= \limsup_{n\to\infty} \left| \frac{f_n(0)}{(n!)^2} \right| \cdot \lim_{n\to\infty} n! \left| x^n \right| = \infty$$

where we have used $\lim_{n\to\infty} n! \left| x^n \right| = \infty$ and $\limsup_{n\to\infty} \left| \frac{f_n(0)}{(n!)^2} \right| \geq 1$. ■

Remark 22.13. Given a sequence of real number $\{a_n\}_{n=0}^{\infty}$ there always exists $f \in X$ such that $f^{(n)}(0) = a_n$. To construct such a function f, let $\varphi \in C_c^{\infty}(-1,1)$ be a function such that $\varphi = 1$ in a neighborhood of 0 and $\varepsilon_n \in (0,1)$ be chosen so that $\varepsilon_n \downarrow 0$ as $n \to \infty$ and $\sum_{n=0}^{\infty} |a_n| \varepsilon_n^n < \infty$. The desired function f can then be defined by

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n \varphi(x/\varepsilon_n) =: \sum_{n=0}^{\infty} g_n(x). \tag{22.4}$$

The fact that f is well defined and continuous follows from the estimate:

$$|g_n(x)| = \left| \frac{a_n}{n!} x^n \varphi(x/\varepsilon_n) \right| \leq \frac{\|\varphi\|_\infty}{n!} |a_n| \varepsilon_n^n$$

and the assumption that $\sum_{n=0}^{\infty} |a_n| \varepsilon_n^n < \infty$. The estimate

$$|g_n'(x)| = \left| \frac{a_n}{(n-1)!} x^{n-1} \varphi(x/\varepsilon_n) + \frac{a_n}{n! \varepsilon_n} x^n \varphi'(x/\varepsilon_n) \right|$$
$$\leq \frac{\|\varphi\|_\infty}{(n-1)!} |a_n| \varepsilon_n^{n-1} + \frac{\|\varphi'\|_\infty}{n!} |a_n| \varepsilon_n^n$$
$$\leq (\|\varphi\|_\infty + \|\varphi'\|_\infty) |a_n| \varepsilon_n^n$$

and the assumption that $\sum_{n=0}^{\infty} |a_n| \varepsilon_n^n < \infty$ shows $f \in C^1(-1,1)$ and $f'(x) = \sum_{n=0}^{\infty} g_n'(x)$. Similar arguments show $f \in C_c^k(-1,1)$ and $f^{(k)}(x) = \sum_{n=0}^{\infty} g_n^{(k)}(x)$ for all x and $k \in \mathbb{N}$. This completes the proof since, using $\varphi(x/\varepsilon_n) = 1$ for x in a neighborhood of 0, $g_n^{(k)}(0) = \delta_{k,n} a_k$ and hence

$$f^{(k)}(0) = \sum_{n=0}^{\infty} g_n^{(k)}(0) = a_k.$$

22.3 Baire Category Theorem Exercises

Exercise 22.4. Let $(X, \|\cdot\|)$ be a normed space and $E \subset X$ be a subspace.

1. If E is closed and proper subspace of X then E is nowhere dense.
2. If E is a proper finite dimensional subspace of X then E is nowhere dense.

Exercise 22.5. Now suppose that $(X, \|\cdot\|)$ is an infinite dimensional Banach space. Show that X can not have a countable **algebraic** basis. More explicitly, there is no countable subset $S \subset X$ such that every element $x \in X$ may be written as a **finite** linear combination of elements from S. **Hint:** make use of Exercise 22.4 and the Baire category theorem.

23

The Baire Category Theorem applied to Banach Sapces

In this chapter we will give some striking applications of the Baire Category theorem in the context of Banach spaces.

23.1 The Open Mapping Theorem

Theorem 23.1 (Open Mapping Theorem). *Let X, Y be Banach spaces, $T \in L(X,Y)$. If T is surjective then T is an open mapping, i.e. $T(V)$ is open in Y for all open subsets $V \subset X$.*

Proof. For all $r > 0$ let $B_r^X = \{x \in X : \|x\|_X < r\} \subset X$, $B_r^Y = \{y \in Y : \|y\|_Y < r\} \subset Y$ and $E_r = T(B_r^X) \subset Y$. The proof will be carried out by proving the following three assertions.

1. There exists $\alpha > 0$ such that $B_r^Y \subset \overline{E_{\alpha r}}$ for all $r > 0$.
2. For the same $\alpha > 0$, $B_r^Y \subset E_{\alpha r}$, i.e. we may remove the closure in assertion 1.
3. The last assertion implies T is an open mapping.

1. Since $Y = \bigcup_{n-1}^{\infty} E_n$, the Baire category Theorem 22.3 implies there exists n such that $\overline{E}_n^0 \neq \emptyset$, i.e. there exists $y \in \overline{E}_n$ and $\varepsilon > 0$ such that $\overline{B^Y(y,\varepsilon)} \subset \overline{E}_n$. Suppose $\|y'\| < \varepsilon$ then y and $y + y'$ are in $B^Y(y,\varepsilon) \subset \overline{E}_n$ hence there exists $\tilde{x}, x \in B_n^X$ such that $\|T\tilde{x} - (y+y')\|$ and $\|Tx - y\|$ may be made as small as we please, which we abbreviate as follows

$$\|T\tilde{x} - (y+y')\| \approx 0 \text{ and } \|Tx - y\| \approx 0.$$

Hence by the triangle inequality,

$$\begin{aligned}\|T(\tilde{x} - x) - y'\| &= \|T\tilde{x} - (y+y') - (Tx - y)\| \\ &\leq \|T\tilde{x} - (y+y')\| + \|Tx - y\| \approx 0\end{aligned}$$

with $\tilde{x} - x \in B_{2n}^X$. This shows that $y' \in \overline{E_{2n}}$ which implies $B_\varepsilon^Y \subset \overline{E_{2n}}$. For any $r > 0$ the map $\varphi_r : Y \to Y$ given by $\varphi_r(y) = \frac{r}{\varepsilon} y$ is a homeomorphism. Thus, with $\alpha = \frac{2n}{\varepsilon}$, $\varphi_r(E_{2n}) = E_{\frac{2n}{\varepsilon} r} = E_{\alpha r}$ and $\varphi_r(B_\varepsilon^Y) = B_r^Y$ and so it follows that

$$B_r^Y = \varphi_r(B_\varepsilon^Y) \subset \varphi_r\left(\bar{E}_{2n}\right) = \overline{\varphi_r\left(E_{2n}\right)} \subset \overline{E_{\alpha r}}.$$

2. Let $\alpha > 0$ be as in assertion 1. and $y \in B_r^Y$. Let $\{r_n\}_{n=0}^\infty$ be a strictly increasing sequence of numbers such that $r_0 = 0$, and $\|y\| < r_n < r$ for all $n > 0$, and $\lim_{n\to\infty} r_n = r$. Further let $\Delta_n := r_n - r_{n-1}$ for $n \in \mathbb{N}$ so that $r = \sum_{n=1}^\infty \Delta_n$. By assertion **1,** $y \in B_{r_1}^Y \subset \overline{E_{\alpha r_1}} = \overline{T\left(B_{\alpha r_1}^X\right)}$ so there exists $x_1 \in B_{\alpha r_1}^X$ such that $\|y - Tx_1\| < \Delta_2$. (Notice that $\|y - Tx_1\|$ can be made as small as we please.) Similarly, since $y - Tx_1 \in B_{\Delta_2}^Y \subset \bar{E}_{\alpha\Delta_2} = \overline{T\left(B_{\alpha\Delta_2}^X\right)}$ there exists $x_2 \in B_{\alpha\Delta_2}^X$ such that $\|y - Tx_1 - Tx_2\| < \Delta_3$. Continuing this way inductively, there exists $x_n \in B_{\alpha\Delta_n}^X$ such that

$$\|y - \sum_{k=1}^n Tx_k\| < \Delta_{n+1} \text{ for all } n \in \mathbb{N}. \tag{23.1}$$

Since $\sum_{n=1}^\infty \|x_n\| < \sum_{n=1}^\infty \alpha\Delta_n = \alpha r$, $x := \sum_{n=1}^\infty x_n$ exists and $\|x\| < \alpha r$, i.e. $x \in B_{\alpha r}^X$. Passing to the limit in Eq. (23.1) shows, $\|y - Tx\| = 0$ and hence $y \in T(B_{\alpha r}^X) = E_{\alpha r}$. Since $y \in B_r^Y$ was arbitrary we have shown $B_r^Y \subset T\left(B_{\alpha r}^X\right)$ for all $r > 0$.

3. If $x \in V \subset_o X$ and $y = Tx \in TV$ we must show that TV contains a ball $B^Y(y,r) = Tx + B_r^Y$ for some $r > 0$. Since $V - x$ is a neighborhood of $0 \in X$, there exists $r > 0$ such that $B_{\alpha r}^X \subset (V - x)$. Then by assertion 2.,

$$B_r^Y \subset TB_{\alpha r}^X \subset T(V - x) = T(V) - y$$

which is equivalent to $B^Y(y,r) = Tx + B_r^Y \subset TV$.

■

Corollary 23.2. *If X, Y are Banach spaces and $T \in L(X,Y)$ is invertible (i.e. a bijective linear transformation) then the inverse map, T^{-1}, is* ***bounded****, i.e. $T^{-1} \in L(Y,X)$. (Note that T^{-1} is automatically linear.)*

23.2 Closed Graph Theorem

Definition 23.3. *Let X and Y be normed spaces and $T : X \to Y$ be linear (not necessarily continuous) map.*

1. *Let $\Gamma : X \to X \times Y$ be the linear map defined by $\Gamma(x) := (x, T(x))$ for all $x \in X$ and let*
$$\Gamma(T) = \{(x, T(x)) : x \in X\}$$
*be the **graph** of T.*
2. *The operator T is said to be **closed** if $\Gamma(T)$ is closed subset of $X \times Y$.*

Exercise 23.1. Let $T : X \to Y$ be a linear map between normed vector spaces, show T is closed iff for all convergent sequences $\{x_n\}_{n=1}^{\infty} \subset X$ such that $\{Tx_n\}_{n=1}^{\infty} \subset Y$ is also convergent, we have $\lim_{n\to\infty} Tx_n = T(\lim_{n\to\infty} x_n)$. (Compare this with the statement that T is continuous iff for every convergent sequences $\{x_n\}_{n=1}^{\infty} \subset X$ we have $\{Tx_n\}_{n=1}^{\infty} \subset Y$ is **necessarily** convergent and $\lim_{n\to\infty} Tx_n = T(\lim_{n\to\infty} x_n)$.)

Theorem 23.4 (Closed Graph Theorem). *Let X and Y be Banach spaces and $T : X \to Y$ be linear map. Then T is continuous iff T is closed.*

Proof. If T is continuous and $(x_n, Tx_n) \to (x, y) \in X \times Y$ as $n \to \infty$ then $Tx_n \to Tx = y$ which implies $(x, y) = (x, Tx) \in \Gamma(T)$. Conversely suppose T is closed, i.e. $\Gamma(T)$ is a closed subspace of $X \times Y$ and is therefore a Banach space in its own right. The projection map $\pi_2 : X \times Y \to Y$ is continuous and $\pi_1|_{\Gamma(T)} : \Gamma(T) \to X$ is continuous bijection which implies $\pi_1|_{\Gamma(T)}^{-1}$ is bounded by the open mapping Theorem 23.1. Therefore $T = \pi_2 \circ \Gamma = \pi_2 \circ \pi_1|_{\Gamma(T)}^{-1}$ is bounded, being the composition of bounded operators since the following diagram commutes

$$\begin{array}{ccc} & \Gamma(T) & \\ \Gamma = \pi_1|_{\Gamma(T)}^{-1} \nearrow & & \searrow \pi_2 \\ X & \underset{T}{\longrightarrow} & Y \end{array}.$$

■

As an application we have the following proposition.

Proposition 23.5. *Let H be a Hilbert space. Suppose that $T : H \to H$ is a linear (not necessarily bounded) map such that there exists $T^* : H \to H$ such that*
$$\langle Tx|Y\rangle = \langle x|T^*Y\rangle \;\forall\, x, y \in H.$$
Then T is bounded.

Proof. It suffices to show T is closed. To prove this suppose that $x_n \in H$ such that $(x_n, Tx_n) \to (x, y) \in H \times H$. Then for any $z \in H$,
$$\langle Tx_n|z\rangle = \langle x_n|T^*z\rangle \longrightarrow \langle x|T^*z\rangle = \langle Tx|z\rangle \text{ as } n \to \infty.$$
On the other hand $\lim_{n\to\infty}\langle Tx_n|z\rangle = \langle y|z\rangle$ as well and therefore $\langle Tx|z\rangle = \langle y|z\rangle$ for all $z \in H$. This shows that $Tx = y$ and proves that T is closed. ■

Here is another example.

Example 23.6. Suppose that $\mathcal{M} \subset L^2([0,1], m)$ is a closed subspace such that each element of $\mathcal{M}$ has a representative in $C([0,1])$. We will abuse notation and simply write $\mathcal{M} \subset C([0,1])$. Then

1. There exists $A \in (0, \infty)$ such that $\|f\|_\infty \le A\|f\|_{L^2}$ for all $f \in \mathcal{M}$.
2. For all $x \in [0,1]$ there exists $g_x \in \mathcal{M}$ such that
$$f(x) = \langle f|g_x\rangle := \int_0^1 f(y)\, g_x(y)\, dy \text{ for all } f \in \mathcal{M}.$$
Moreover we have $\|g_x\| \le A$.
3. The subspace $\mathcal{M}$ is finite dimensional and $\dim(\mathcal{M}) \le A^2$.

Proof. 1) I will give a two proofs of part 1. Each proof requires that we first show that $(\mathcal{M}, \|\cdot\|_\infty)$ is a complete space. To prove this it suffices to show $\mathcal{M}$ is a closed subspace of $C([0,1])$. So let $\{f_n\} \subset \mathcal{M}$ and $f \in C([0,1])$ such that $\|f_n - f\|_\infty \to 0$ as $n \to \infty$. Then $\|f_n - f_m\|_{L^2} \le \|f_n - f_m\|_\infty \to 0$ as $m, n \to \infty$, and since $\mathcal{M}$ is closed in $L^2([0,1])$, $L^2 - \lim_{n\to\infty} f_n = g \in \mathcal{M}$. By passing to a subsequence if necessary we know that $g(x) = \lim_{n\to\infty} f_n(x) = f(x)$ for m - a.e. x. So $f = g \in \mathcal{M}$.

i) Let $i : (\mathcal{M}, \|\cdot\|_\infty) \to (\mathcal{M}, \|\cdot\|_2)$ be the identity map. Then i is bounded and bijective. By the open mapping theorem, $j = i^{-1}$ is bounded as well. Hence there exists $A < \infty$ such that $\|f\|_\infty = \|j(f)\| \le A\|f\|_2$ for all $f \in \mathcal{M}$.

ii) Let $j : (\mathcal{M}, \|\cdot\|_2) \to (\mathcal{M}, \|\cdot\|_\infty)$ be the identity map. We will shows that j is a closed operator and hence bounded by the closed graph Theorem 23.4. Suppose that $f_n \in \mathcal{M}$ such that $f_n \to f$ in L^2 and $f_n = j(f_n) \to g$ in $C([0,1])$. Then as in the first paragraph, we conclude that $g = f = j(f)$ a.e. showing j is closed. Now finish as in last line of proof i).

2) For $x \in [0,1]$, let $e_x : \mathcal{M} \to \mathbb{C}$ be the evaluation map $e_x(f) = f(x)$. Then
$$|e_x(f)| \le |f(x)| \le \|f\|_\infty \le A\|f\|_{L^2}$$
which shows that $e_x \in \mathcal{M}^*$. Hence there exists a unique element $g_x \in \mathcal{M}$ such that
$$f(x) = e_x(f) = \langle f, g_x\rangle \text{ for all } f \in \mathcal{M}.$$
Moreover $\|g_x\|_{L^2} = \|e_x\|_{\mathcal{M}^*} \le A$.

3) Let $\{f_j\}_{j=1}^n$ be an L^2 – orthonormal subset of $\mathcal{M}$. Then
$$A^2 \ge \|e_x\|_{\mathcal{M}^*}^2 = \|g_x\|_{L^2}^2 \ge \sum_{j=1}^n |\langle f_j, g_x\rangle|^2 = \sum_{j=1}^n |f_j(x)|^2$$
and integrating this equation over $x \in [0,1]$ implies that

$$A^2 \geq \sum_{j=1}^{n} \int_0^1 |f_j(x)|^2 dx = \sum_{j=1}^{n} 1 = n$$

which shows that $n \leq A^2$. Hence $\dim(\mathcal{M}) \leq A^2$. ∎

Remark 23.7. Keeping the notation in Example 23.6, $G(x,y) = g_x(y)$ for all $x, y \in [0,1]$. Then

$$f(x) = e_x(f) = \int_0^1 f(y)\overline{G(x,y)}dy \text{ for all } f \in \mathcal{M}.$$

The function G is called the reproducing kernel for $\mathcal{M}$.

The above example generalizes as follows.

Proposition 23.8. *Suppose that $(X, \mathcal{M}, \mu)$ is a finite measure space, $p \in [1, \infty)$ and W is a closed subspace of $L^p(\mu)$ such that $W \subset L^p(\mu) \cap L^\infty(\mu)$. Then* $\dim(W) < \infty$.

Proof. With out loss of generality we may assume that $\mu(X) = 1$. As in Example 23.6, we can show that W is a closed subspace of $L^\infty(\mu)$ and then (by the open mapping theorem) that there exists a constant $A < \infty$ such that $\|f\|_\infty \leq A \|f\|_p$ for all $f \in W$. Now if $1 \leq p \leq 2$, then

$$\|f\|_\infty \leq A \|f\|_p \leq A \|f\|_2$$

and if $p \in (2, \infty)$, then $\|f\|_p^p \leq \|f\|_2^2 \|f\|_\infty^{p-2}$ or equivalently,

$$\|f\|_p \leq \|f\|_2^{2/p} \|f\|_\infty^{1-2/p} \leq \|f\|_2^{2/p} \left(A \|f\|_p\right)^{1-2/p}$$

from which we learn that $\|f\|_p \leq A^{1-2/p} \|f\|_2$. Therefore $\|f\|_\infty \leq AA^{1-2/p} \|f\|_2$ and so for all $p \in [1, \infty)$ there exists a constant $B < \infty$ such that $\|f\|_\infty \leq B \|f\|_2$.

If $\{f_n\}_{n=1}^N$ be an orthonormal subset of W, $c = (c_1, \ldots, c_N) \in \mathbb{C}^N$, and $f_c = \sum_{n=1}^N c_n f_n$, then

$$\|f_c\|_\infty^2 \leq B^2 \|f_c\|_2^2 = B^2 \sum_{n=1}^{N} |c_n|^2 = B^2 |c|^2$$

where $|c|^2 := \sum_{n=1}^N |c_n|^2$. For each $c \in \mathbb{C}^N$, there is an exceptional set E_c such that for $x \notin E_c$,

$$|f_c(x)|^2 = \left|\sum_{n=1}^{N} c_n f_n(x)\right|^2 \leq B^2 |c|^2.$$

Let $\mathbb{D} := (\mathbb{Q} + i\mathbb{Q})^N$ and $E = \cap_{c \in \mathbb{D}} E_c$. Then $\mu(E) = 0$ and for $x \notin E$, $\left|\sum_{n=1}^N c_n f_n(x)\right| \leq B^2 |c|^2$ for all $c \in \mathbb{D}$. By continuity it then follows for $x \notin E$ that

$$\left|\sum_{n=1}^{N} c_n f_n(x)\right|^2 \leq B^2 |c|^2 \text{ for all } c \in \mathbb{C}^N.$$

Taking $c_n = f_n(x)$ in this inequality implies that

$$\left|\sum_{n=1}^{N} |f_n(x)|^2\right|^2 \leq B^2 \sum_{n=1}^{N} |f_n(x)|^2 \text{ for all } x \notin E$$

and therefore that

$$\sum_{n=1}^{N} |f_n(x)|^2 \leq B^2 \text{ for all } x \notin E.$$

Integrating this equation over x then implies that $N \leq B^2$, i.e. $\dim(W) \leq B^2$. ∎

23.3 Uniform Boundedness Principle

Theorem 23.9 (Uniform Boundedness Principle). *Let X and Y be normed vector spaces, $\mathcal{A} \subset L(X,Y)$ be a collection of bounded linear operators from X to Y,*

$$\begin{aligned} F = F_{\mathcal{A}} &= \{x \in X : \sup_{A \in \mathcal{A}} \|Ax\| < \infty\} \text{ and} \\ R = R_{\mathcal{A}} = F^c &= \{x \in X : \sup_{A \in \mathcal{A}} \|Ax\| = \infty\}. \end{aligned} \tag{23.2}$$

1. *If $\sup_{A \in \mathcal{A}} \|A\| < \infty$ then $F = X$.*
2. *If F is not meager, then $\sup_{A \in \mathcal{A}} \|A\| < \infty$.*
3. *If X is a Banach space then the following are equivalent;*
   a) *F is not meager,*
   b) *$\sup_{A \in \mathcal{A}} \|A\| < \infty$,*
   c) *$F = X$, i.e. $\sup_{A \in \mathcal{A}} \|Ax\| < \infty$ for all $x \in X$*

   In particular, when X is a Banach space,,

$$\sup_{A \in \mathcal{A}} \|Ax\| < \infty \text{ for all } x \in X \iff \sup_{A \in \mathcal{A}} \|A\| < \infty.$$

In words the collection of operators $\mathcal{A}$ is pointwise bounded iff $\mathcal{A}$ is uniformly bounded on the unit sphere in X.

4. *If X is a Banach space, then $\sup_{A\in\mathcal{A}}\|A\|=\infty$ iff R is residual. In particular if $\sup_{A\in\mathcal{A}}\|A\|=\infty$ then $\sup_{A\in\mathcal{A}}\|Ax\|=\infty$ for x in a dense subset of X.*

Proof. 1. If $M:=\sup_{A\in\mathcal{A}}\|A\|<\infty$, then $\sup_{A\in\mathcal{A}}\|Ax\|\le M\|x\|<\infty$ for all $x\in X$ showing $F=X$.

2. For each $n\in\mathbb{N}$, let $E_n\subset X$ be the closed sets given by

$$E_n=\{x:\sup_{A\in\mathcal{A}}\|Ax\|\le n\}=\bigcap_{A\in\mathcal{A}}\{x:\|Ax\|\le n\}.$$

Then $F=\cup_{n=1}^{\infty}E_n$ which is assumed to be non-meager and hence there exists an $n\in\mathbb{N}$ such that E_n has non-empty interior. Let $B_x(\delta)$ be a ball such that $\overline{B_x(\delta)}\subset E_n$. Then for $y\in X$ with $\|y\|=\delta$ we know $x-y\in\overline{B_x(\delta)}\subset E_n$, so that $Ay=Ax-A(x-y)$ and hence for any $A\in\mathcal{A}$,

$$\|Ay\|\le\|Ax\|+\|A(x-y)\|\le n+n=2n.$$

Hence it follows that $\|A\|\le 2n/\delta$ for all $A\in\mathcal{A}$, i.e. $\sup_{A\in\mathcal{A}}\|A\|\le 2n/\delta<\infty$.

3. $(a\implies b)$ follows from item 2, $(b\implies c)$ follows from item 1, and $(c\implies a)$ follows from the Baire Category Theorem 22.3, X is not meager.

4. Item 3. implies F is meager iff $\sup_{A\in\mathcal{A}}\|A\|=\infty$. Since $R=F^c$, R is residual iff F is meager, so R is residual iff $\sup_{A\in\mathcal{A}}\|A\|=\infty$. The last assertion follows from the fact the residual sets in a Banach space are dense. ∎

Remarks 23.10 *Let $S\subset X$ be the unit sphere in X, $f_A(x)=Ax$ for $x\in S$ and $A\in\mathcal{A}$.*

1. *The assertion $\sup_{A\in\mathcal{A}}\|Ax\|<\infty$ for all $x\in X$ implies $\sup_{A\in\mathcal{A}}\|A\|<\infty$ may be interpreted as follows. If $\sup_{A\in\mathcal{A}}\|f_A(x)\|<\infty$ for all $x\in S$, then $\sup_{A\in\mathcal{A}}\|f_A\|_\infty<\infty$ where $\|f_A\|_\infty:=\sup_{x\in S}\|f_A(x)\|=\|A\|$.*
2. *If $\dim(X)<\infty$ we may give a simple proof of this assertion. Indeed if $\{e_n\}_{n=1}^N\subset S$ is a basis for X there is a constant $\varepsilon>0$ such that $\left\|\sum_{n=1}^N\lambda_n e_n\right\|\ge\varepsilon\sum_{n=1}^N|\lambda_n|$ and so the assumption $\sup_{A\in\mathcal{A}}\|f_A(x)\|<\infty$ implies*

$$\sup_{A\in\mathcal{A}}\|A\|=\sup_{A\in\mathcal{A}}\sup_{\lambda\neq0}\frac{\left\|\sum_{n=1}^N\lambda_nAe_n\right\|}{\left\|\sum_{n=1}^N\lambda_ne_n\right\|}\le\sup_{A\in\mathcal{A}}\sup_{\lambda\neq0}\frac{\sum_{n=1}^N|\lambda_n|\,\|Ae_n\|}{\varepsilon\sum_{n=1}^N|\lambda_n|}$$
$$\le\varepsilon^{-1}\sup_{A\in\mathcal{A}}\sup_n\|Ae_n\|=\varepsilon^{-1}\sup_n\sup_{A\in\mathcal{A}}\|Ae_n\|<\infty.$$

Notice that we have used the linearity of each $A\in\mathcal{A}$ in a crucial way.

3. *If we drop the linearity assumption, so that $f_A\in C(S,Y)$ for all $A\in\mathcal{A}$ – some index set, then it is no longer true that $\sup_{A\in\mathcal{A}}\|f_A(x)\|<\infty$ for all $x\in S$, then $\sup_{A\in\mathcal{A}}\|f_A\|_\infty<\infty$. The reader is invited to construct a counterexample when $X=\mathbb{R}^2$ and $Y=\mathbb{R}$ by finding a sequence $\{f_n\}_{n=1}^\infty$ of continuous functions on S^1 such that $\lim_{n\to\infty}f_n(x)=0$ for all $x\in S^1$ while $\lim_{n\to\infty}\|f_n\|_{C(S^1)}=\infty$.*
4. *The assumption that X is a Banach space in item 3.of Theorem 23.9 can not be dropped. For example, let $X\subset C([0,1])$ be the polynomial functions on $[0,1]$ equipped with the uniform norm $\|\cdot\|_\infty$ and for $t\in(0,1]$, let $f_t(x):=(x(t)-x(0))/t$ for all $x\in X$. Then $\lim_{t\to0}f_t(x)=\frac{d}{dt}|_0x(t)$ and therefore $\sup_{t\in(0,1]}|f_t(x)|<\infty$ for all $x\in X$. If the conclusion of Theorem 23.9 (item 3.) were true we would have $M:=\sup_{t\in(0,1]}\|f_t\|<\infty$. This would then imply*

$$\left|\frac{x(t)-x(0)}{t}\right|\le M\|x\|_\infty\text{ for all }x\in X\text{ and }t\in(0,1].$$

Letting $t\downarrow0$ in this equation gives, $|\dot{x}(0)|\le M\|x\|_\infty$ for all $x\in X$. But taking $x(t)=t^n$ in this inequality shows $M=\infty$.

Example 23.11. Suppose that $\{c_n\}_{n=1}^\infty\subset\mathbb{C}$ is a sequence of numbers such that

$$\lim_{N\to\infty}\sum_{n=1}^N a_nc_n\text{ exists in }\mathbb{C}\text{ for all }a\in\ell^1.$$

Then $c\in\ell^\infty$.

Proof. Let $f_N\in\left(\ell^1\right)^*$ be given by $f_N(a)=\sum_{n=1}^N a_nc_n$ and set $M_N:=\max\{|c_n|:n=1,\dots,N\}$. Then

$$|f_N(a)|\le M_N\|a\|_{\ell^1}$$

and by taking $a=e_k$ with k such $M_N=|c_k|$, we learn that $\|f_N\|=M_N$. Now by assumption, $\lim_{N\to\infty}f_N(a)$ exists for all $a\in\ell^1$ and in particular,

$$\sup_N|f_N(a)|<\infty\text{ for all }a\in\ell^1.$$

So by the uniform boundedness principle, Theorem 23.9,

$$\infty>\sup_N\|f_N\|=\sup_N M_N=\sup\{|c_n|:n=1,2,3,\dots\}.$$

∎

23.3.1 Applications to Fourier Series

Let $T = S^1$ be the unit circle in S^1, $\varphi_n(z) := z^n$ for all $n \in \mathbb{Z}$, and m denote the normalized arc length measure on T, i.e. if $f : T \to [0,\infty)$ is measurable, then

$$\int_T f(w)dw := \int_T f dm := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta})d\theta.$$

From Section 40.1, we know $\{\varphi_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(T)$. For $n \in \mathbb{N}$ and $z \in T$, let

$$s_n(f,z) := \sum_{k=-n}^{n} \langle f|\varphi_n\rangle \varphi_k(z) = \int_T f(w)d_n(z\bar{w})dw$$

where

$$d_n(e^{i\theta}) := \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta},$$

see Eqs. (40.4) and (40.5).[1] By Theorem 40.3, for all $f \in L^2(T)$ we know

$$f = L^2(T) - \lim_{n\to\infty} s_n(f,\cdot).$$

On the other hand the next proposition shows; if we fix $z \in T$, then $\lim_{n\to\infty} s_n(f,z)$ does **not** even exist for the "typical" $f \in C(T) \subset L^2(T)$.

Proposition 23.12 (Lack of pointwise convergence). *For each $z \in T$, there exists a residual set $R_z \subset C(T)$ such that $\sup_n |s_n(f,z)| = \infty$ for all $f \in R_z$. Recall that $C(T)$ is a complete metric space, hence R_z is a dense subset of $C(T)$.*

Proof. By symmetry considerations, it suffices to assume $z = 1 \in T$. Let $\Lambda_n : C(T) \to \mathbb{C}$ be given by

$$\Lambda_n f := s_n(f,1) = \int_T f(w)d_n(\bar{w})dw.$$

An application of Corollary **??** below shows,

$$\|\Lambda_n\| = \|d_n\|_1 = \int_T |d_n(\bar{w})|\,dw = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|d_n(e^{-i\theta})\right|d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}\right|d\theta. \tag{23.3}$$

Of course we may prove this directly as follows. Since

$$|\Lambda_n f| = \left|\int_T f(w)d_n(\bar{w})dw\right| \le \int_T |f(w)d_n(\bar{w})|\,dw \le \|f\|_\infty \int_T |d_n(\bar{w})|\,dw,$$

we learn $\|\Lambda_n\| \le \int_T |d_n(\bar{w})|\,dw$. For all $\varepsilon > 0$, let

$$f_\varepsilon(z) := \frac{d_n(\bar{z})}{\sqrt{d_n^2(\bar{z}) + \varepsilon}}.$$

Then $\|f_\varepsilon\|_{C(T)} \le 1$ and hence

$$\|\Lambda_n\| \ge \lim_{\varepsilon\downarrow 0}|\Lambda_n f_\varepsilon| = \lim_{\varepsilon\downarrow 0}\int_T \frac{d_n^2(\bar{z})}{\sqrt{d_n^2(\bar{z})+\varepsilon}}dw = \int_T |d_n(\bar{z})|\,dw$$

and the verification of Eq. (23.3) is complete.

Using

$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\,dy\right| \le |x|$$

in Eq. (23.3) implies that

$$\|\Lambda_n\| \ge \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\sin(n+\frac{1}{2})\theta}{\frac{1}{2}\theta}\right|d\theta = \frac{2}{\pi}\int_0^{\pi}\left|\sin(n+\frac{1}{2})\theta\right|\frac{d\theta}{\theta}$$
$$= \frac{2}{\pi}\int_0^{\pi}\left|\sin(n+\frac{1}{2})\theta\right|\frac{d\theta}{\theta} = \int_0^{(n+\frac{1}{2})\pi}|\sin y|\frac{dy}{y} \to \infty \text{ as } n\to\infty \tag{23.4}$$

and hence $\sup_n \|\Lambda_n\| = \infty$. So by Theorem 23.9,

$$R_1 = \{f \in C(T) : \sup_n |\Lambda_n f| = \infty\}$$

is a residual set. ■

See Rudin Chapter 5 for more details.

[1] Letting $\alpha = e^{i\theta/2}$, we have

$$D_n(\theta) = \sum_{k=-n}^{n}\alpha^{2k} = \frac{\alpha^{2(n+1)} - \alpha^{-2n}}{\alpha^2 - 1} = \frac{\alpha^{2n+1} - \alpha^{-(2n+1)}}{\alpha - \alpha^{-1}} = \frac{2i\sin(n+\frac{1}{2})\theta}{2i\sin\frac{1}{2}\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}.$$

and therefore

$$D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}.$$

Lemma 23.13. *For $f \in L^1(T)$, let*

$$\tilde{f}(n) := \langle f, \varphi_n \rangle = \int_T f(w)\bar{w}^n dw.$$

*Then $\tilde{f} \in c_0 := C_0(\mathbb{Z})$ (i.e. $\lim_{n\to\infty} \tilde{f}(n) = 0$) and the map $f \in L^1(T) \to \tilde{f} \in c_0$ is a one to one bounded linear transformation into but **not** onto c_0.*

Proof. By Bessel's inequality, $\sum_{n\in\mathbb{Z}} \left|\tilde{f}(n)\right|^2 < \infty$ for all $f \in L^2(T)$ and in particular $\lim_{|n|\to\infty} \left|\tilde{f}(n)\right| = 0$. Given $f \in L^1(T)$ and $g \in L^2(T)$ we have

$$\left|\tilde{f}(n) - \hat{g}(n)\right| = \left|\int_T [f(w) - g(w)]\,\bar{w}^n dw\right| \le \|f - g\|_1$$

and hence

$$\limsup_{n\to\infty} \left|\tilde{f}(n)\right| = \limsup_{n\to\infty} \left|\tilde{f}(n) - \hat{g}(n)\right| \le \|f - g\|_1$$

for all $g \in L^2(T)$. Since $L^2(T)$ is dense in $L^1(T)$, it follows that $\limsup_{n\to\infty} \left|\tilde{f}(n)\right| = 0$ for all $f \in L^1$, i.e. $\tilde{f} \in c_0$. Since $\left|\tilde{f}(n)\right| \le \|f\|_1$, we have $\left\|\tilde{f}\right\|_{c_0} \le \|f\|_1$ showing that $\Lambda f := \tilde{f}$ is a bounded linear transformation from $L^1(T)$ to c_0. To see that Λ is injective, suppose $\tilde{f} = \Lambda f \equiv 0$, then $\int_T f(w)p(w,\bar{w})dw = 0$ for all polynomials p in w and $\bar{w}$. By the Stone - Wierestrass and the dominated convergence theorem, this implies that

$$\int_T f(w)g(w)dw = 0$$

for all $g \in C(T)$. Lemma 31.11 now implies $f = 0$ a.e. If Λ were surjective, the open mapping theorem would imply that $\Lambda^{-1} : c_0 \to L^1(T)$ is bounded. In particular this implies there exists $C < \infty$ such that

$$\|f\|_{L^1} \le C \left\|\tilde{f}\right\|_{c_0} \text{ for all } f \in L^1(T). \tag{23.5}$$

Taking $f = d_n$, we find (because $\tilde{d}_n(k) = 1_{|k|\le n}$) that $\left\|\tilde{d}_n\right\|_{c_0} = 1$ while (by Eq. (23.4)) $\lim_{n\to\infty} \|d_n\|_{L^1} = \infty$ contradicting Eq. (23.5). Therefore $\mathrm{Ran}(\Lambda) \ne c_0$. ■

23.4 Exercises

23.4.1 More Examples of Banach Spaces

Exercise 23.2. Folland 5.9, p. 155. (Drop this problem, or move to Chapter 15.)

Exercise 23.3. Folland 5.10, p. 155. (Drop this problem, or move later where it can be done.)

Exercise 23.4. Folland 5.11, p. 155. (Drop this problem, or move to Chapter 15.)

23.4.2 Open Mapping and Closed Operator Problems

Exercise 23.5. Let $X = \ell^1(\mathbb{N})$,

$$Y = \left\{ f \in X : \sum_{n=1}^{\infty} n\,|f(n)| < \infty \right\}$$

with Y being equipped with the $\ell^1(\mathbb{N})$ - norm, and $T : Y \to X$ be defined by $(Tf)(n) = nf(n)$. Show:

1. Y is a proper dense subspace of X and in particular Y is not complete
2. $T : Y \to X$ is a closed operator which is not bounded.
3. $T : Y \to X$ is algebraically invertible, $S := T^{-1} : X \to Y$ is bounded and surjective but not open.

Exercise 23.6. Let $X = C([0,1])$ and $Y = C^1([0,1]) \subset X$ with both X and Y being equipped with the uniform norm. Let $T : Y \to X$ be the linear map, $Tf = f'$. Here $C^1([0,1])$ denotes those functions, $f \in C^1((0,1)) \cap C([0,1])$ such that

$$f'(1) := \lim_{x\uparrow 1} f'(x) \text{ and } f'(0) := \lim_{x\downarrow 0} f'(x)$$

exist.

1. Y is a proper dense subspace of X and in particular Y is not complete.
2. $T : Y \to X$ is a closed operator which is not bounded.

Exercise 23.7. Folland 5.31, p. 164.

Exercise 23.8. Let X be a vector space equipped with two norms, $\|\cdot\|_1$ and $\|\cdot\|_2$ such that $\|\cdot\|_1 \le \|\cdot\|_2$ and X is complete relative to both norms. Show there is a constant $C < \infty$ such that $\|\cdot\|_2 \le C\|\cdot\|_1$.

Exercise 23.9 (No slowest decay rate). Show that it is impossible to find a sequence, $\{a_n\}_{n\in\mathbb{N}} \subset (0,\infty)$, with the following property: if $\{\lambda_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathbb{C}$, then $\sum_{n=1}^{\infty} |\lambda_n| < \infty$ iff $\sup a_n^{-1}|\lambda_n| < \infty$. (Poetically speaking, there is no "slowest rate" of decay for the summands of absolutely convergent series.)

Outline: For sake of contradiction suppose such a "magic" sequence $\{a_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ were to exists.

1. For $f \in \ell^\infty(\mathbb{N})$, let $(Tf)(n) := a_n f(n)$ for $n \in \mathbb{N}$. Verify that $Tf \in \ell^1(\mathbb{N})$ and $T : \ell^\infty(\mathbb{N}) \to \ell^1(\mathbb{N})$ is a bounded linear operator.
2. Show $T : \ell^\infty(\mathbb{N}) \to \ell^1(\mathbb{N})$ must be an invertible operator and that $T^{-1} : \ell^1(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ is necessarily bounded, i.e. $T : \ell^\infty(\mathbb{N}) \to \ell^1(\mathbb{N})$ is a homeomorphism.
3. Arrive at a contradiction by showing either that T^{-1} is not bounded or by using the fact that, D, the set of finitely supported sequences, is dense in $\ell^1(\mathbb{N})$ but not in $\ell^\infty(\mathbb{N})$.

Exercise 23.10. Folland 5.34, p. 164. (Not a very good problem, delete.)

Exercise 23.11. Folland 5.35, p. 164. (A quotient space exercise.)

Exercise 23.12. Folland 5.36, p. 164. (A quotient space exercise.)

Exercise 23.13. Suppose $T : X \to Y$ is a linear map between two Banach spaces such that $f \circ T \in X^*$ for all $f \in Y^*$. Show T is bounded.

Exercise 23.14. Suppose $T_n : X \to Y$ for $n \in \mathbb{N}$ is a sequence of bounded linear operators between two Banach spaces such that $\lim_{n\to\infty} T_n x$ exists for all $x \in X$. Show $Tx := \lim_{n\to\infty} T_n x$ defines a bounded linear operator from X to Y.

Exercise 23.15. Let X, Y and Z be Banach spaces and $B : X \times Y \to Z$ be a bilinear map such that $B(x,\cdot) \in L(Y,Z)$ and $B(\cdot,y) \in L(X,Z)$ for all $x \in X$ and $y \in Y$. Show there is a constant $M < \infty$ such that

$$\|B(x,y)\| \le M\,\|x\|\,\|y\| \text{ for all } (x,y) \in X \times Y$$

and conclude from this that $B : X \times Y \to Z$ is continuous.

Exercise 23.16. Folland 5.40, p. 165. (Condensation of singularities).

Exercise 23.17. Folland 5.41, p. 165. (Drop this exercise, it is 22.5.)

23.4.3 Weak Topology and Convergence Problems

Definition 23.14. *A sequence $\{x_n\}_{n=1}^\infty \subset X$ is **weakly Cauchy** if for all $V \in \tau_w$ such that $0 \in V$, $x_n - x_m \in V$ for all m,n sufficiently large. Similarly a sequence $\{f_n\}_{n=1}^\infty \subset X^*$ is **weak–∗ Cauchy** if for all $V \in \tau_{w^*}$ such that $0 \in V$, $f_n - f_m \in V$ for all m,n sufficiently large.*

Remark 23.15. These conditions are equivalent to $\{f(x_n)\}_{n=1}^\infty$ being Cauchy for all $f \in X^*$ and $\{f_n(x)\}_{n=1}^\infty$ being Cauchy for all $x \in X$ respectively.

Exercise 23.18. Let X and Y be Banach spaces. Show:

1. Every weakly Cauchy sequence in X is bounded.
2. Every weak-* Cauchy sequence in X^* is bounded.
3. If $\{T_n\}_{n=1}^\infty \subset L(X,Y)$ converges weakly (or strongly) then $\sup_n \|T_n\|_{L(X,Y)} < \infty$.

Exercise 23.19. Let X be a Banach space, $C := \{x \in X : \|x\| \le 1\}$ and $C^* := \{\lambda \in X^* : \|\lambda\|_{X^*} \le 1\}$ be the closed unit balls in X and X^* respectively.

1. Show C is weakly closed and C^* is weak-* closed in X and X^* respectively.
2. If $E \subset X$ is a norm-bounded set, then the weak closure, $\bar{E}^w \subset X$, is also norm bounded.
3. If $F \subset X^*$ is a norm-bounded set, then the weak-* closure, $\bar{E}^{w-*} \subset X^*$, is also norm bounded.
4. Every weak-* Cauchy sequence $\{f_n\} \subset X^*$ is weak-* convergent to some $f \in X^*$.

Exercise 23.20. Folland 5.49, p. 171.

Exercise 23.21. If X is a separable normed linear space, the weak-* topology on the closed unit ball in X^* is second countable and hence metrizable. (See Theorem 36.25.)

Exercise 23.22. Let X be a Banach space. Show every weakly compact subset of X is norm bounded and every weak–∗ compact subset of X^* is norm bounded.

Exercise 23.23. A vector subspace of a normed space X is norm-closed iff it is weakly closed. (If X is not reflexive, it is not necessarily true that a normed closed subspace of X^* need be weak* closed, see Exercise 23.25.) (**Hint:** this problem only uses the Hahn-Banach Theorem.)

Exercise 23.24. Let X be a Banach space, $\{T_n\}_{n=1}^\infty$ and $\{S_n\}_{n=1}^\infty$ be two sequences of bounded operators on X such that $T_n \to T$ and $S_n \to S$ strongly, and suppose $\{x_n\}_{n=1}^\infty \subset X$ such that $\lim_{n\to\infty} \|x_n - x\| = 0$. Show:

1. $\lim_{n\to\infty} \|T_n x_n - Tx\| = 0$ and that
2. $T_n S_n \to TS$ strongly as $n \to \infty$.

Exercise 23.25. Folland 5.52, p. 172.

Exercise 23.26. Let $H = \ell_2$ and $S_n(x_1, x_2, \dots) = (x_{n+1}, x_{n+2}, \dots)$. Show;

1. $S_n \xrightarrow{s} 0$ for all $x \in \ell_2$ while $\|S_n^* x\| = \|x\|$ for all $x \in \ell^2$ and therefore S_n^* is not strongly convergent to 0. This shows the adjoint operation is strongly discontinuous.

2. Observe that $S_n \overset{w}{\to} 0$ weakly and $S_n^* \overset{w}{\to} 0$ weakly, i.e. $\lim_{n\to\infty} \langle S_n x|y\rangle = 0 = \lim_{n\to\infty} \langle S_n^* x|y\rangle$. On the other hand verify that $S_n S_n^* = I$ for all n from which it follows that the map $(A, B) \to AB$ is not jointly continuous in the weak operator topology even though it is in the strong operator topology provided A is restricted to a bounded sets.

Exercise 23.27 (Inverse operation is strongly discontinuous). Let $H = \ell^2(\mathbb{N})$, $P_n : H \to H$ be orthogonal projection onto

$$H_n := \{x \in H : x_k = 0 \text{ for } k > n\},$$

and $Q_n = I - P_n$ be projection onto the remaining components. Let $\varepsilon_n \in (0, 1)$ be chosen so that $\varepsilon_n \downarrow 0$ and define $A_n = P_n + \varepsilon_n Q_n$. Show

1. $A_n \overset{s}{\to} I$ as $n \to \infty$ while
2. A_n^{-1} exists but $\left\{A_n^{-1}\right\}_{n=1}^{\infty}$ is not strongly convergent.

Part V

The Structure of Measures

Complex Measures, Radon-Nikodym Theorem

Definition 24.1. *A* ***signed measure*** ν *on a measurable space* $(X, \mathcal{M})$ *is a function* $\nu : \mathcal{M} \to \overline{\mathbb{R}}$ *such that*

1. *Either*
$$\nu(\mathcal{M}) := \{\nu(A) : A \in \mathcal{M}\} \subset (-\infty, \infty]$$
or $\nu(\mathcal{M}) \subset [-\infty, \infty)$.
2. ν *is countably additive, this is to say if* $E = \coprod_{j=1}^{\infty} E_j$ *with* $E_j \in \mathcal{M}$, *then*
$$\nu(E) = \sum_{j=1}^{\infty} \nu(E_j).$$
If $\nu(E) \in \mathbb{R}$ *then the series* $\sum_{j=1}^{\infty} \nu(E_j)$ *is absolutely convergent since it is independent of rearrangements.*
3. $\nu(\emptyset) = 0$.

If there exists $X_n \in \mathcal{M}$ *such that* $|\nu(X_n)| < \infty$ *and* $X = \cup_{n=1}^{\infty} X_n$, *then* ν *is said to be* σ – ***finite*** *and if* $\nu(\mathcal{M}) \subset \mathbb{R}$ *then* ν *is said to be a* ***finite signed measure.*** *Similarly, a countably additive set function* $\nu : \mathcal{M} \to \mathbb{C}$ *such that* $\nu(\emptyset) = 0$ *is called a* ***complex measure.***

Example 24.2. Suppose that μ_+ and μ_- are two positive measures on $\mathcal{M}$ such that either $\mu_+(X) < \infty$ or $\mu_-(X) < \infty$, then $\nu = \mu_+ - \mu_-$ is a signed measure. If both $\mu_+(X)$ and $\mu_-(X)$ are finite then ν is a finite signed measure and may also be considered to be a complex measure.

Example 24.3. Suppose that $g : X \to \overline{\mathbb{R}}$ is measurable and either $\int_E g^+ d\mu$ or $\int_E g^- d\mu < \infty$, then
$$\nu(A) = \int_A g d\mu \ \forall \ A \in \mathcal{M} \tag{24.1}$$
defines a signed measure. This is actually a special case of the last example with $\mu_\pm(A) := \int_A g^\pm d\mu$. Notice that the measure $\mu_\pm$ in this example have the property that they are concentrated on disjoint sets, namely μ_+ "lives" on $\{g > 0\}$ and μ_- "lives" on the set $\{g < 0\}$.

Example 24.4. Suppose that μ is a positive measure on $(X, \mathcal{M})$ and $g \in L^1(\mu)$, then ν given as in Eq. (24.1) is a complex measure on $(X, \mathcal{M})$. Also if $\{\mu_\pm^r, \mu_\pm^i\}$ is any collection of four positive finite measures on $(X, \mathcal{M})$, then
$$\nu := \mu_+^r - \mu_-^r + i\left(\mu_+^i - \mu_-^i\right) \tag{24.2}$$
is a complex measure.

If ν is given as in Eq. 24.1, then ν may be written as in Eq. (24.2) with $d\mu_\pm^r = (\operatorname{Re} g)_\pm d\mu$ and $d\mu_\pm^i = (\operatorname{Im} g)_\pm d\mu$.

24.1 The Radon-Nikodym Theorem

Definition 24.5. *Let* ν *be a complex or signed measure on* $(X, \mathcal{M})$. *A set* $E \in \mathcal{M}$ *is a* ***null*** *set or precisely a* ν – *null set if* $\nu(A) = 0$ *for all* $A \in \mathcal{M}$ *such that* $A \subset E$, *i.e.* $\nu|_{\mathcal{M}_E} = 0$. *Recall that* $\mathcal{M}_E := \{A \cap E : A \in \mathcal{M}\} = i_E^{-1}(\mathcal{M})$ *is the "trace of* M *on* E.

We will eventually show that every complex and σ – finite signed measure ν may be described as in Eq. (24.1). The next theorem is the first result in this direction.

Theorem 24.6 (A Baby Radon-Nikodym Theorem). *Suppose* $(X, \mathcal{M})$ *is a measurable space,* μ *is a positive finite measure on* $\mathcal{M}$ *and* ν *is a complex measure on* $\mathcal{M}$ *such that* $|\nu(A)| \le \mu(A)$ *for all* $A \in \mathcal{M}$. *Then* $d\nu = \rho d\mu$ *where* $|\rho| \le 1$. *Moreover if* ν *is a positive measure, then* $0 \le \rho \le 1$.

Proof. For a simple function, $f \in \mathcal{S}(X, \mathcal{M})$, let $\nu(f) := \sum_{a \in \mathbb{C}} a\nu(f = a)$. Then
$$|\nu(f)| \le \sum_{a \in \mathbb{C}} |a| \, |\nu(f = a)| \le \sum_{a \in \mathbb{C}} |a| \, \mu(f = a) = \int_X |f| \, d\mu.$$
So, by the B.L.T. Theorem 32.4, ν extends to a continuous linear functional on $L^1(\mu)$ satisfying the bounds
$$|\nu(f)| \le \int_X |f| \, d\mu \le \sqrt{\mu(X)} \, \|f\|_{L^2(\mu)} \text{ for all } f \in L^1(\mu).$$

The Riesz representation Theorem 18.17 then implies there exists a unique $\rho \in L^2(\mu)$ such that

$$\nu(f) = \int_X f \rho d\mu \text{ for all } f \in L^2(\mu).$$

Taking $A \in \mathcal{M}$ and $f = \overline{\operatorname{sgn}(\rho)} 1_A$ in this equation shows

$$\int_A |\rho| \, d\mu = \nu(\overline{\operatorname{sgn}(\rho)} 1_A) \leq \mu(A) = \int_A 1 d\mu$$

from which it follows that $|\rho| \leq 1$, $\mu -$ a.e. If ν is a positive measure, then for real f, $0 = \operatorname{Im}[\nu(f)] = \int_X \operatorname{Im} \rho f d\mu$ and taking $f = \operatorname{Im} \rho$ shows $0 = \int_X [\operatorname{Im} \rho]^2 \, d\mu$, i.e. $\operatorname{Im}(\rho(x)) = 0$ for $\mu -$ a.e. x and we have shown ρ is real a.e. Similarly,

$$0 \leq \nu(\operatorname{Re} \rho < 0) = \int_{\{\operatorname{Re} \rho < 0\}} \rho d\mu \leq 0,$$

shows $\rho \geq 0$ a.e. ■

Definition 24.7. *Let μ and ν be two signed or complex measures on $(X, \mathcal{M})$. Then:*

1. *μ and ν are **mutually singular** (written as $\mu \perp \nu$) if there exists $A \in \mathcal{M}$ such that A is a ν – null set and A^c is a μ – null set.*
2. *The measure ν **is absolutely continuous relative to** μ (written as $\nu \ll \mu$) provided $\nu(A) = 0$ whenever A is a μ – null set, i.e. all μ – null sets are ν – null sets as well.*

As an example, suppose that μ is a positive measure and $\rho \in L^1(\mu)$. Then the measure, $\nu := \rho\mu$ is absolutely continuous relative to μ. Indeed, if $\mu(A) = 0$ then

$$\rho(A) = \int_A \rho d\mu = 0$$

as well.

Lemma 24.8. *If μ_1, μ_2 and ν are signed measures on $(X, \mathcal{M})$ such that $\mu_1 \perp \nu$ and $\mu_2 \perp \nu$ and $\mu_1 + \mu_2$ is well defined, then $(\mu_1 + \mu_2) \perp \nu$. If $\{\mu_i\}_{i=1}^\infty$ is a sequence of positive measures such that $\mu_i \perp \nu$ for all i then $\mu = \sum_{i=1}^\infty \mu_i \perp \nu$ as well.*

Proof. In both cases, choose $A_i \in \mathcal{M}$ such that A_i is $\nu -$ null and A_i^c is μ_i-null for all i. Then by Lemma 24.15, $A := \cup_i A_i$ is still a ν –null set. Since

$$A^c = \cap_i A_i^c \subset A_m^c \text{ for all } m$$

we see that A^c is a μ_i - null set for all i and is therefore a null set for $\mu = \sum_{i=1}^\infty \mu_i$. This shows that $\mu \perp \nu$. ■

Throughout the remainder of this section μ will be always be a positive measure on $(X, \mathcal{M})$.

Definition 24.9 (Lebesgue Decomposition). *Suppose that ν is a signed (complex) measure and μ is a positive measure on $(X, \mathcal{M})$. Two signed (complex) measures ν_a and ν_s form a **Lebesgue decomposition** of ν relative to μ if*

1. *If $\nu(A) = \infty$ $(\nu(A) = -\infty)$ for some $A \in \mathcal{M}$ then $\nu_a(A) \neq -\infty$ $(\nu_a(A) \neq +\infty)$ and $\nu_s(A) \neq -\infty$ $(\nu_s(A) \neq +\infty)$.*
2. *$\nu = \nu_a + \nu_s$ which is well defined by assumption 1.*
3. *$\nu_a \ll \mu$ and $\nu_s \perp \mu$.*

Lemma 24.10. *Let ν is a signed (complex) measure and μ is a positive measure on $(X, \mathcal{M})$. If there exists a Lebesgue decomposition, $\nu = \nu_s + \nu_a$, of the measure ν relative to μ then it is unique. Moreover:*

1. *if ν is positive then ν_s and ν_a are positive.*
2. *If ν is a σ – finite measure then so are ν_s and ν_a.*

Proof. Since $\nu_s \perp \mu$, there exists $A \in \mathcal{M}$ such that $\mu(A) = 0$ and A^c is $\nu_s -$ null and because $\nu_a \ll \mu$, A is also a null set for ν_a. So for $C \in \mathcal{M}$, $\nu_a(C \cap A) = 0$ and $\nu_s(C \cap A^c) = 0$ from which it follows that

$$\nu(C) = \nu(C \cap A) + \nu(C \cap A^c) = \nu_s(C \cap A) + \nu_a(C \cap A^c)$$

and hence,

$$\begin{aligned} \nu_s(C) &= \nu_s(C \cap A) = \nu(C \cap A) \text{ and} \\ \nu_a(C) &= \nu_a(C \cap A^c) = \nu(C \cap A^c). \end{aligned} \tag{24.3}$$

Item 1. is now obvious from Eq. (24.3).

For Item 2., if ν is a $\sigma -$ finite measure then there exists $X_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty X_n$ and $|\nu(X_n)| < \infty$ for all n. Since $\nu(X_n) = \nu_a(X_n) + \nu_s(X_n)$, we must have $\nu_a(X_n) \in \mathbb{R}$ and $\nu_s(X_n) \in \mathbb{R}$ showing ν_a and ν_s are $\sigma -$ finite as well.

For the uniqueness assertion, if we have another decomposition $\nu = \tilde{\nu}_a + \tilde{\nu}_s$ with $\tilde{\nu}_s \perp \mu$ and $\tilde{\nu}_a \ll \mu$ we may choose $\tilde{A} \in \mathcal{M}$ such that $\mu(\tilde{A}) = 0$ and $\tilde{A}^c$ is $\tilde{\nu}_s -$ null. Then $B = A \cup \tilde{A}$ is still a μ - null set and $B^c = A^c \cap \tilde{A}^c$ is a null set for both ν_s and $\tilde{\nu}_s$. Therefore by the same arguments which proved Eq. (24.3),

$$\begin{aligned} \nu_s(C) &= \nu(C \cap B) = \tilde{\nu}_s(C) \text{ and} \\ \nu_a(C) &= \nu(C \cap B^c) = \tilde{\nu}_a(C) \text{ for all } C \in \mathcal{M}. \end{aligned}$$

■

Lemma 24.11. *Suppose μ is a positive measure on $(X,\mathcal{M})$ and $f,g : X \to \bar{\mathbb{R}}$ are extended integrable functions such that*

$$\int_A f d\mu = \int_A g d\mu \text{ for all } A \in \mathcal{M}, \tag{24.4}$$

$\int_X f_-\, d\mu < \infty$, $\int_X g_-\, d\mu < \infty$, and the measures $|f|\, d\mu$ and $|g|\, d\mu$ are σ – finite. Then $f(x) = g(x)$ for μ – a.e. x.

Proof. By assumption there exists $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\int_{X_n} |f|\, d\mu < \infty$ and $\int_{X_n} |g|\, d\mu < \infty$ for all n. Replacing A by $A \cap X_n$ in Eq. (24.4) implies

$$\int_A 1_{X_n} f d\mu = \int_{A\cap X_n} f d\mu = \int_{A\cap X_n} g d\mu = \int_A 1_{X_n} g d\mu$$

for all $A \in \mathcal{M}$. Since $1_{X_n} f$ and $1_{X_n} g$ are in $L^1(\mu)$ for all n, this equation implies $1_{X_n} f = 1_{X_n} g$, μ – a.e. Letting $n \to \infty$ then shows that $f = g$, μ – a.e. ∎

Remark 24.12. Suppose that f and g are two positive measurable functions on $(X,\mathcal{M},\mu)$ such that Eq. (24.4) holds. It is not in general true that $f = g$, μ – a.e. A trivial counterexample is to take $\mathcal{M} = 2^X$, $\mu(A) = \infty$ for all non-empty $A \in \mathcal{M}$, $f = 1_X$ and $g = 2 \cdot 1_X$. Then Eq. (24.4) holds yet $f \neq g$.

Theorem 24.13 (Radon Nikodym Theorem for Positive Measures). *Suppose that μ and ν are σ – finite positive measures on $(X,\mathcal{M})$. Then ν has a unique Lebesgue decomposition $\nu = \nu_a + \nu_s$ relative to μ and there exists a unique (modulo sets of μ – measure 0) function $\rho : X \to [0,\infty)$ such that $d\nu_a = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$.*

Proof. The uniqueness assertions follow directly from Lemmas 24.10 and 24.11.

Existence. (Von-Neumann's Proof.) First suppose that μ and ν are **finite** measures and let $\lambda = \mu + \nu$. By Theorem 24.6, $d\nu = h d\lambda$ with $0 \le h \le 1$ and this implies, for all non-negative measurable functions f, that

$$\nu(f) = \lambda(fh) = \mu(fh) + \nu(fh) \tag{24.5}$$

or equivalently

$$\nu(f(1-h)) = \mu(fh). \tag{24.6}$$

Taking $f = 1_{\{h=1\}}$ in Eq. (24.6) shows that

$$\mu(\{h=1\}) = \nu(1_{\{h=1\}}(1-h)) = 0,$$

i.e. $0 \le h(x) < 1$ for μ-a.e. x. Let

$$\rho := 1_{\{h<1\}} \frac{h}{1-h}$$

and then take $f = g 1_{\{h<1\}}(1-h)^{-1}$ with $g \ge 0$ in Eq. (24.6) to learn

$$\nu(g 1_{\{h<1\}}) = \mu(g 1_{\{h<1\}}(1-h)^{-1} h) = \mu(\rho g).$$

Hence if we define

$$\nu_a := 1_{\{h<1\}}\nu \text{ and } \nu_s := 1_{\{h=1\}}\nu,$$

we then have $\nu_s \perp \mu$ (since ν_s "lives" on $\{h=1\}$ while $\mu(h=1) = 0$) and $\nu_a = \rho\mu$ and in particular $\nu_a \ll \mu$. Hence $\nu = \nu_a + \nu_s$ is the desired Lebesgue decomposition of ν.[1]

If we further assume that $\nu \ll \mu$, then $\mu(h=1) = 0$ implies $\nu(h=1) = 0$ and hence that $\nu_s = 0$ and we conclude that $\nu = \nu_a = \rho\mu$.

For the σ – **finite case**, write $X = \coprod_{n=1}^\infty X_n$ where $X_n \in \mathcal{M}$ are chosen so that $\mu(X_n) < \infty$ and $\nu(X_n) < \infty$ for all n. Let $d\mu_n = 1_{X_n} d\mu$ and $d\nu_n = 1_{X_n} d\nu$. Then by what we have just proved there exists $\rho_n \in L^1(X,\mu_n) \subset L^1(X,\mu)$ and measure ν_n^s such that $d\nu_n = \rho_n d\mu_n + d\nu_n^s$ with $\nu_n^s \perp \mu_n$. Since μ_n and ν_n^s "live" on X_n (see Eq. (24.3) there exists $A_n \in \mathcal{M}_{X_n}$ such that $\mu(A_n) = \mu_n(A_n) = 0$ and

$$\nu_n^s(X \setminus A_n) = \nu_n^s(X_n \setminus A_n) = 0.$$

This shows that $\nu_n^s \perp \mu$ for all n and so by Lemma 24.8, $\nu_s := \sum_{n=1}^\infty \nu_n^s$ is singular relative to μ. Since

$$\nu = \sum_{n=1}^\infty \nu_n = \sum_{n=1}^\infty (\rho_n \mu_n + \nu_n^s) = \sum_{n=1}^\infty (\rho_n 1_{X_n} \mu + \nu_n^s) = \rho\mu + \nu_s,$$

where $\rho := \sum_{n=1}^\infty 1_{X_n} \rho_n$, it follows that $\nu = \nu_a + \nu_s$ with $\nu_a = \rho\mu$ is the Lebesgue decomposition of ν relative to μ. ∎

[1] Here is the motivation for this construction. Suppose that $d\nu = d\nu_s + \rho d\mu$ is the Radon-Nikodym decompostion and $X = A \coprod B$ such that $\nu_s(B) = 0$ and $\mu(A) = 0$. Then we find

$$\nu_s(f) + \mu(\rho f) = \nu(f) = \lambda(fg) = \nu(fg) + \mu(fg).$$

Letting $f \to 1_A f$ then implies that

$$\nu_s(1_A f) = \nu(1_A f g)$$

which show that $g = 1$ ν –a.e. on A. Also letting $f \to 1_B f$ implies that

$$\mu(\rho 1_B f(1-g)) = \nu(1_B f(1-g)) = \mu(1_B f g) = \mu(fg)$$

which shows that

$$\rho(1-g) = \rho 1_B(1-g) = g \ \mu - \text{a.e.}.$$

This shows that $\rho = \frac{g}{1-g}$ μ – a.e.

24.2 The Structure of Signed Measures

Definition 24.14. *Let ν be a signed measure on $(X,\mathcal{M})$ and $E\in\mathcal{M}$, then*

1. *E is **positive** if for all $A\in\mathcal{M}$ such that $A\subset E$, $\nu(A)\geq 0$, i.e. $\nu|_{\mathcal{M}_E}\geq 0$.*
2. *E is **negative** if for all $A\in\mathcal{M}$ such that $A\subset E$, $\nu(A)\leq 0$, i.e. $\nu|_{\mathcal{M}_E}\leq 0$.*

Lemma 24.15. *Suppose that ν is a signed measure on $(X,\mathcal{M})$. Then*

1. *Any subset of a positive set is positive.*
2. *The countable union of positive (negative or null) sets is still positive (negative or null).*
3. *Let us now further assume that $\nu(\mathcal{M})\subset[-\infty,\infty)$ and $E\in\mathcal{M}$ is a set such that $\nu(E)\in(0,\infty)$. Then there exists a positive set $P\subset E$ such that $\nu(P)\geq\nu(E)$.*

Proof. The first assertion is obvious. If $P_j\in\mathcal{M}$ are positive sets, let $P=\bigcup_{n=1}^{\infty}P_n$. By replacing P_n by the positive set $P_n\setminus\left(\bigcup_{j=1}^{n-1}P_j\right)$ we may assume that the $\{P_n\}_{n=1}^{\infty}$ are pairwise disjoint so that $P=\coprod_{n=1}^{\infty}P_n$. Now if $E\subset P$ and $E\in\mathcal{M}$, $E=\coprod_{n=1}^{\infty}(E\cap P_n)$ so $\nu(E)=\sum_{n=1}^{\infty}\nu(E\cap P_n)\geq 0$.which shows that P is positive. The proof for the negative and the null case is analogous.

The idea for proving the third assertion is to keep removing "big" sets of negative measure from E. The set remaining from this procedure will be P. We now proceed to the formal proof. For all $A\in\mathcal{M}$ let

$$n(A)=1\wedge\sup\{-\nu(B):B\subset A\}.$$

Since $\nu(\emptyset)=0$, $n(A)\geq 0$ and $n(A)=0$ iff A is positive. Choose $A_0\subset E$ such that $-\nu(A_0)\geq\frac{1}{2}n(E)$ and set $E_1=E\setminus A_0$, then choose $A_1\subset E_1$ such that $-\nu(A_1)\geq\frac{1}{2}n(E_1)$ and set $E_2=E\setminus(A_0\cup A_1)$. Continue this procedure inductively, namely if $A_0,\dots,A_{k-1}$ have been chosen let $E_k=E\setminus\left(\coprod_{i=0}^{k-1}A_i\right)$ and choose $A_k\subset E_k$ such that $-\nu(A_k)\geq\frac{1}{2}n(E_k)$. Let $P:=E\setminus\coprod_{k=0}^{\infty}A_k=\bigcap_{k=0}^{\infty}E_k$, then $E=P\cup\coprod_{k=0}^{\infty}A_k$ and hence

$$(0,\infty)\ni\nu(E)=\nu(P)+\sum_{k=0}^{\infty}\nu(A_k)=\nu(P)-\sum_{k=0}^{\infty}-\nu(A_k)\leq\nu(P). \tag{24.7}$$

From Eq. (24.7) we learn that$\sum_{k=0}^{\infty}-\nu(A_k)<\infty$ and in particular that $\lim_{k\to\infty}(-\nu(A_k))=0$. Since $0\leq\frac{1}{2}n(E_k)\leq-\nu(A_k)$, this also implies $\lim_{k\to\infty}n(E_k)=0$. If $A\in\mathcal{M}$ with $A\subset P$, then $A\subset E_k$ for all k and so, for k large so that $n(E_k)<1$, we find $-\nu(A)\leq n(E_k)$. Letting $k\to\infty$ in this estimate shows $-\nu(A)\leq 0$ or equivalently $\nu(A)\geq 0$. Since $A\subset P$ was arbitrary, we conclude that P is a positive set such that $\nu(P)\geq\nu(E)$. ∎

24.2.1 Hahn Decomposition Theorem

Definition 24.16. *Suppose that ν is a signed measure on $(X,\mathcal{M})$. A **Hahn decomposition** for ν is a partition $\{P,N=P^c\}$ of X such that P is positive and N is negative.*

Theorem 24.17 (Hahn Decomposition Theorem). *Every signed measure space $(X,\mathcal{M},\nu)$ has a Hahn decomposition, $\{P,N\}$. Moreover, if $\{\tilde{P},\tilde{N}\}$ is another Hahn decomposition, then $P\Delta\tilde{P}=N\Delta\tilde{N}$ is a null set, so the decomposition is unique modulo null sets.*

Proof. With out loss of generality we may assume that $\nu(\mathcal{M})\subset[-\infty,\infty)$. If not just consider $-\nu$ instead.

Uniqueness. For any $A\in\mathcal{M}$, we have

$$\nu(A)=\nu(A\cap P)+\nu(A\cap N)\leq\nu(A\cap P)\leq\nu(P).$$

In particular, taking $A=P\cup\tilde{P}$, we learn

$$\nu(P)\leq\nu(P\cup\tilde{P})\leq\nu(P)$$

or equivalently that $\nu(P)=\nu\left(P\cup\tilde{P}\right)$. Of course by symmetry we also have

$$\nu(P)=\nu\left(P\cup\tilde{P}\right)=\nu\left(\tilde{P}\right)=:s.$$

Since also,

$$s=\nu(P\cup\tilde{P})=\nu(P)+\nu(\tilde{P})-\nu(P\cap\tilde{P})=2s-\nu(P\cap\tilde{P}),$$

we also have $\nu(P\cap\tilde{P})=s$. Finally using $P\cup\tilde{P}=\left[P\cap\tilde{P}\right]\coprod\left(\tilde{P}\Delta P\right)$, we conclude that

$$s=\nu(P\cup\tilde{P})=\nu(P\cap\tilde{P})+\nu(\tilde{P}\Delta P)=s+\nu(\tilde{P}\Delta P)$$

which shows $\nu(\tilde{P}\Delta P)=0$. Thus $N\Delta\tilde{N}=\tilde{P}\Delta P$ is a positive set with zero measure, i.e. $N\Delta\tilde{N}=\tilde{P}\Delta P$ is a null set and this proves the uniqueness assertion.

Existence. Let

$$s:=\sup\{\nu(A):A\in\mathcal{M}\}$$

which is non-negative since $\nu(\emptyset) = 0$. If $s = 0$, we are done since $P = \emptyset$ and $N = X$ is the desired decomposition. So assume $s > 0$ and choose $A_n \in \mathcal{M}$ such that $\nu(A_n) > 0$ and $\lim_{n\to\infty} \nu(A_n) = s$. By Lemma 24.15 there exists positive sets $P_n \subset A_n$ such that $\nu(P_n) \ge \nu(A_n)$. Then $s \ge \nu(P_n) \ge \nu(A_n) \to s$ as $n \to \infty$ implies that $s = \lim_{n\to\infty} \nu(P_n)$. The set $P := \cup_{n=1}^\infty P_n$ is a positive set being the union of positive sets and since $P_n \subset P$ for all n,

$$\nu(P) \ge \nu(P_n) \to s \text{ as } n \to \infty.$$

This shows that $\nu(P) \ge s$ and hence by the definition of s, $s = \nu(P) < \infty$.

I now claim that $N = P^c$ is a negative set and therefore, $\{P, N\}$ is the desired Hahn decomposition. If N were not negative, we could find $E \subset N = P^c$ such that $\nu(E) > 0$. We then would have

$$\nu(P \cup E) = \nu(P) + \nu(E) = s + \nu(E) > s$$

which contradicts the definition of s. ■

24.2.2 Jordan Decomposition

Theorem 24.18 (Jordan Decomposition). *If ν is a signed measure on $(X, \mathcal{M})$, there exist unique positive measure $\nu_\pm$ on $(X, \mathcal{M})$ such that $\nu_+ \perp \nu_-$ and $\nu = \nu_+ - \nu_-$. This decomposition is called the **Jordan decomposition** of ν.*

Proof. Let $\{P, N\}$ be a Hahn decomposition for ν and define

$$\nu_+(E) := \nu(P \cap E) \text{ and } \nu_-(E) := -\nu(N \cap E) \ \forall\ E \in \mathcal{M}.$$

Then it is easily verified that $\nu = \nu_+ - \nu_-$ is a Jordan decomposition of ν. The reader is asked to prove the uniqueness of this decomposition in Exercise 24.12. ■

Definition 24.19. *The measure, $|\nu| := \nu_+ + \nu_-$ is called the total variation of ν. A signed measure is called σ – **finite** provided that $\nu_\pm$ (or equivalently $|\nu| := \nu_+ + \nu_-$) are σ -finite measures.*

Lemma 24.20. *Let ν be a signed measure on $(X, \mathcal{M})$ and $A \in \mathcal{M}$. If $\nu(A) \in \mathbb{R}$ then $\nu(B) \in \mathbb{R}$ for all $B \subset A$. Moreover, $\nu(A) \in \mathbb{R}$ iff $|\nu|(A) < \infty$. In particular, ν is σ finite iff $|\nu|$ is σ – finite. Furthermore if $P, N \in \mathcal{M}$ is a Hahn decomposition for ν and $g = 1_P - 1_N$, then $d\nu = gd|\nu|$, i.e.*

$$\nu(A) = \int_A gd|\nu| \text{ for all } A \in \mathcal{M}.$$

Proof. Suppose that $B \subset A$ and $|\nu(B)| = \infty$ then since $\nu(A) = \nu(B) + \nu(A \setminus B)$ we must have $|\nu(A)| = \infty$. Let $P, N \in \mathcal{M}$ be a Hahn decomposition for ν, then

$$\begin{aligned}\nu(A) &= \nu(A \cap P) + \nu(A \cap N) = |\nu(A \cap P)| - |\nu(A \cap N)| \text{ and} \\ |\nu|(A) &= \nu(A \cap P) - \nu(A \cap N) = |\nu(A \cap P)| + |\nu(A \cap N)|. \end{aligned} \tag{24.8}$$

Therefore $\nu(A) \in \mathbb{R}$ iff $\nu(A \cap P) \in \mathbb{R}$ and $\nu(A \cap N) \in \mathbb{R}$ iff $|\nu|(A) < \infty$. Finally,

$$\begin{aligned}\nu(A) &= \nu(A \cap P) + \nu(A \cap N) \\ &= |\nu|(A \cap P) - |\nu|(A \cap N) \\ &= \int_A (1_P - 1_N) d|\nu|\end{aligned}$$

which shows that $d\nu = gd|\nu|$. ■

Lemma 24.21. *Suppose that μ is a positive measure on $(X, \mathcal{M})$ and $g : X \to \mathbb{R}$ is an extended μ-integrable function. If ν is the signed measure $d\nu = gd\mu$, then $d\nu_\pm = g_\pm d\mu$ and $d|\nu| = |g|d\mu$. We also have*

$$|\nu|(A) = \sup\{\int_A f\ d\nu : |f| \le 1\} \text{ for all } A \in \mathcal{M}. \tag{24.9}$$

Proof. The pair, $P = \{g > 0\}$ and $N = \{g \le 0\} = P^c$ is a Hahn decomposition for ν. Therefore

$$\nu_+(A) = \nu(A \cap P) = \int_{A\cap P} gd\mu = \int_A 1_{\{g>0\}} gd\mu = \int_A g_+ d\mu,$$

$$\nu_-(A) = -\nu(A \cap N) = -\int_{A\cap N} gd\mu = -\int_A 1_{\{g\le 0\}} gd\mu = -\int_A g_- d\mu.$$

and

$$\begin{aligned}|\nu|(A) = \nu_+(A) + \nu_-(A) &= \int_A g_+ d\mu - \int_A g_- d\mu \\ &= \int_A (g_+ - g_-)\, d\mu = \int_A |g|\, d\mu.\end{aligned}$$

If $A \in \mathcal{M}$ and $|f| \le 1$, then

$$\begin{aligned}\left|\int_A f\ d\nu\right| &= \left|\int_A f\ d\nu_+ - \int_A f\ d\nu_-\right| \le \left|\int_A f\ d\nu_+\right| + \left|\int_A f\ d\nu_-\right| \\ &\le \int_A |f|\, d\nu_+ + \int_A |f|\, d\nu_- = \int_A |f|\ d|\nu| \le |\nu|(A).\end{aligned}$$

For the reverse inequality, let $f := 1_P - 1_N$ then

$$\int_A f \, d\nu = \nu(A \cap P) - \nu(A \cap N) = \nu_+(A) + \nu_-(A) = |\nu|(A).$$

■

Definition 24.22. *Let ν be a signed measure on $(X, \mathcal{M})$, let*

$$L^1(\nu) := L^1(\nu_+) \cap L^1(\nu_-) = L^1(|\nu|)$$

and for $f \in L^1(\nu)$ we define

$$\int_X f d\nu := \int_X f d\nu_+ - \int_X f d\nu_-.$$

Lemma 24.23. *Let μ be a positive measure on $(X, \mathcal{M})$, g be an extended integrable function on $(X, \mathcal{M}, \mu)$ and $d\nu = g d\mu$. Then $L^1(\nu) = L^1(|g| \, d\mu)$ and for $f \in L^1(\nu)$,*

$$\int_X f d\nu = \int_X f g d\mu.$$

Proof. By Lemma 24.21, $d\nu_+ = g_+ d\mu$, $d\nu_- = g_- d\mu$, and $d|\nu| = |g| \, d\mu$ so that $L^1(\nu) = L^1(|\nu|) = L^1(|g| \, d\mu)$ and for $f \in L^1(\nu)$,

$$\begin{aligned} \int_X f d\nu &= \int_X f d\nu_+ - \int_X f d\nu_- = \int_X f g_+ d\mu - \int_X f g_- d\mu \\ &= \int_X f \left(g_+ - g_-\right) d\mu = \int_X f g d\mu. \end{aligned}$$

■

Exercise 24.1 (Obsolete now?). Let ν be a $\sigma -$ finite signed measure, $f \in L^1(|\nu|)$ and define

$$\int_X f d\nu := \int_X f d\nu_+ - \int_X f d\nu_-.$$

Suppose that μ is a $\sigma -$ finite measure and $\nu \ll \mu$. Show

$$\int_X f d\nu = \int_X f \frac{d\nu}{d\mu} d\mu. \tag{24.10}$$

BRUCE: this seems to already be done in Lemma 24.23.

Lemma 24.24. *Suppose ν is a signed measure, μ is a positive measure and $\nu = \nu_a + \nu_s$ is a Lebesgue decomposition (see Definition 24.9) of ν relative to μ, then $|\nu| = |\nu_a| + |\nu_s|$.*

Proof. Let $A \in \mathcal{M}$ be chosen so that A is a null set for ν_a and A^c is a null set for ν_s. Let $A = P' \coprod N'$ be a Hahn decomposition of $\nu_s|_{\mathcal{M}_A}$ and $A^c = \tilde{P} \coprod \tilde{N}$ be a Hahn decomposition of $\nu_a|_{\mathcal{M}_{A^c}}$. Let $P = P' \cup \tilde{P}$ and $N = N' \cup \tilde{N}$. Since for $C \in \mathcal{M}$,

$$\begin{aligned} \nu(C \cap P) &= \nu(C \cap P') + \nu(C \cap \tilde{P}) \\ &= \nu_s(C \cap P') + \nu_a(C \cap \tilde{P}) \geq 0 \end{aligned}$$

and

$$\begin{aligned} \nu(C \cap N) &= \nu(C \cap N') + \nu(C \cap \tilde{N}) \\ &= \nu_s(C \cap N') + \nu_a(C \cap \tilde{N}) \leq 0 \end{aligned}$$

we see that $\{P, N\}$ is a Hahn decomposition for ν. It also easy to see that $\{P, N\}$ is a Hahn decomposition for both ν_s and ν_a as well. Therefore,

$$\begin{aligned} |\nu|(C) &= \nu(C \cap P) - \nu(C \cap N) \\ &= \nu_s(C \cap P) - \nu_s(C \cap N) + \nu_a(C \cap P) - \nu_a(C \cap N) \\ &= |\nu_s|(C) + |\nu_a|(C). \end{aligned}$$

■

Lemma 24.25.

1. *Let ν be a signed measure and μ be a positive measure on $(X, \mathcal{M})$ such that $\nu \ll \mu$ and $\nu \perp \mu$, then $\nu \equiv 0$.*
2. *Suppose that $\nu = \sum_{i=1}^{\infty} \nu_i$ where ν_i are positive measures on $(X, \mathcal{M})$ such that $\nu_i \ll \mu$, then $\nu \ll \mu$.*
3. *Also if ν_1 and ν_2 are two signed measure such that $\nu_i \ll \mu$ for $i = 1, 2$ and $\nu = \nu_1 + \nu_2$ is well defined, then $\nu \ll \mu$.*

Proof. 1. Because $\nu \perp \mu$, there exists $A \in \mathcal{M}$ such that A is a $\nu -$ null set and $B = A^c$ is a μ - null set. Since B is $\mu -$ null and $\nu \ll \mu$, B is also $\nu -$ null. This shows by Lemma 24.15 that $X = A \cup B$ is also $\nu -$ null, i.e. ν is the zero measure. The proof of items 2. and 3. are easy and will be left to the reader. ■

Theorem 24.26 (Radon Nikodym Theorem for Signed Measures). *Let ν be a σ – finite signed measure and μ be a σ – finite positive measure on $(X, \mathcal{M})$. Then ν has a unique Lebesgue decomposition $\nu = \nu_a + \nu_s$ relative to μ and there exists a unique (modulo sets of μ – measure 0) extended integrable function $\rho : X \to \mathbb{R}$ such that $d\nu_a = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$, i.e. $d\nu = \rho d\mu$ iff $\nu \ll \mu$.*

Proof. Uniqueness. Is a direct consequence of Lemmas 24.10 and 24.11. **Existence.** Let $\nu = \nu_+ - \nu_-$ be the Jordan decomposition of ν. Assume, without loss of generality, that $\nu_+(X) < \infty$, i.e. $\nu(A) < \infty$ for all $A \in \mathcal{M}$. By the Radon Nikodym Theorem 24.13 for positive measures there exist functions $f_\pm : X \to [0, \infty)$ and measures $\lambda_\pm$ such that $\nu_\pm = \mu_{f_\pm} + \lambda_\pm$ with $\lambda_\pm \perp \mu$. Since

$$\infty > \nu_+(X) = \mu_{f_+}(X) + \lambda_+(X),$$

$f_+ \in L^1(\mu)$ and $\lambda_+(X) < \infty$ so that $f = f_+ - f_-$ is an extended integrable function, $d\nu_a := f d\mu$ and $\nu_s = \lambda_+ - \lambda_-$ are signed measures. This finishes the existence proof since

$$\nu = \nu_+ - \nu_- = \mu_{f_+} + \lambda_+ - \left(\mu_{f_-} + \lambda_-\right) = \nu_a + \nu_s$$

and $\nu_s = (\lambda_+ - \lambda_-) \perp \mu$ by Lemma 24.8. For the final statement, if $\nu_s = 0$, then $d\nu = \rho d\mu$ and hence $\nu \ll \mu$. Conversely if $\nu \ll \mu$, then $d\nu_s = d\nu - \rho d\mu \ll \mu$, so by Lemma 24.15, $\nu_s = 0$. Alternatively just use the uniqueness of the Lebesgue decomposition to conclude $\nu_a = \nu$ and $\nu_s = 0$. Or more directly, choose $B \in \mathcal{M}$ such that $\mu(B^c) = 0$ and B is a ν_s – null set. Since $\nu \ll \mu$, B^c is also a ν – null set so that, for $A \in \mathcal{M}$,

$$\nu(A) = \nu(A \cap B) = \nu_a(A \cap B) + \nu_s(A \cap B) = \nu_a(A \cap B).$$

■

Notation 24.27 *The function f is called the Radon-Nikodym derivative of ν relative to μ and we will denote this function by $\frac{d\nu}{d\mu}$.*

24.3 Complex Measures

Suppose that ν is a complex measure on $(X, \mathcal{M})$, let $\nu_r := \operatorname{Re}\nu$, $\nu_i := \operatorname{Im}\nu$ and $\mu := |\nu_r| + |\nu_i|$. Then μ is a finite positive measure on $\mathcal{M}$ such that $\nu_r \ll \mu$ and $\nu_i \ll \mu$. By the Radon-Nikodym Theorem 24.26, there exists real functions $h, k \in L^1(\mu)$ such that $d\nu_r = h\, d\mu$ and $d\nu_i = k\, d\mu$. So letting $g := h + ik \in L^1(\mu)$,

$$d\nu = (h + ik)d\mu = g d\mu$$

showing every complex measure may be written as in Eq. (24.1).

Lemma 24.28. *Suppose that ν is a complex measure on $(X, \mathcal{M})$, and for $i = 1, 2$ let μ_i be a finite positive measure on $(X, \mathcal{M})$ such that $d\nu = g_i d\mu_i$ with $g_i \in L^1(\mu_i)$. Then*

$$\int_A |g_1|\, d\mu_1 = \int_A |g_2|\, d\mu_2 \text{ for all } A \in \mathcal{M}.$$

In particular, we may define a positive measure $|\nu|$ on $(X, \mathcal{M})$ by

$$|\nu|(A) = \int_A |g_1|\, d\mu_1 \text{ for all } A \in \mathcal{M}.$$

The finite positive measure $|\nu|$ is called the ***total variation measure*** *of ν.*

Proof. Let $\lambda = \mu_1 + \mu_2$ so that $\mu_i \ll \lambda$. Let $\rho_i = d\mu_i/d\lambda \geq 0$ and $h_i = \rho_i g_i$. Since

$$\nu(A) = \int_A g_i d\mu_i = \int_A g_i \rho_i d\lambda = \int_A h_i d\lambda \text{ for all } A \in \mathcal{M},$$

$h_1 = h_2$, λ –a.e. Therefore

$$\begin{aligned}\int_A |g_1|\, d\mu_1 &= \int_A |g_1|\, \rho_1 d\lambda = \int_A |h_1|\, d\lambda \\ &= \int_A |h_2|\, d\lambda = \int_A |g_2|\, \rho_2 d\lambda = \int_A |g_2|\, d\mu_2.\end{aligned}$$

■

Definition 24.29. *Given a complex measure ν, let $\nu_r = \operatorname{Re}\nu$ and $\nu_i = \operatorname{Im}\nu$ so that ν_r and ν_i are finite signed measures such that*

$$\nu(A) = \nu_r(A) + i\nu_i(A) \text{ for all } A \in \mathcal{M}.$$

Let $L^1(\nu) := L^1(\nu_r) \cap L^1(\nu_i)$ and for $f \in L^1(\nu)$ define

$$\int_X f d\nu := \int_X f d\nu_r + i \int_X f d\nu_i.$$

Example 24.30. Suppose that μ is a positive measure on $(X, \mathcal{M})$, $g \in L^1(\mu)$ and $\nu(A) = \int_A g d\mu$ as in Example 24.4, then $L^1(\nu) = L^1(|g|\, d\mu)$ and for $f \in L^1(\nu)$

$$\int_X f d\nu = \int_X f g d\mu. \tag{24.11}$$

To check Eq. (24.11), notice that $d\nu_r = \operatorname{Re} g\; d\mu$ and $d\nu_i = \operatorname{Im} g\; d\mu$ so that (using Lemma 24.23)

$$L^1(\nu) = L^1(\operatorname{Re} g d\mu) \cap L^1(\operatorname{Im} g d\mu) = L^1(|\operatorname{Re} g|\, d\mu) \cap L^1(|\operatorname{Im} g|\, d\mu) = L^1(|g|\, d\mu).$$

If $f \in L^1(\nu)$, then

$$\int_X f d\nu := \int_X f \operatorname{Re} g d\mu + i \int_X f \operatorname{Im} g d\mu = \int_X f g d\mu.$$

Remark 24.31. Suppose that ν is a complex measure on $(X,\mathcal{M})$ such that $d\nu = gd\mu$ and as above $d|\nu| = |g|\,d\mu$. Letting

$$\rho = \operatorname{sgn}(\rho) := \begin{cases} \frac{g}{|g|} & \text{if } |g| \neq 0 \\ 1 & \text{if } |g| = 0 \end{cases}$$

we see that

$$d\nu = gd\mu = \rho\,|g|\,d\mu = \rho d\,|\nu|$$

and $|\rho| = 1$ and ρ is uniquely defined modulo $|\nu|$ – null sets. We will denote ρ by $d\nu/d\,|\nu|$. With this notation, it follows from Example 24.30 that $L^1(\nu) := L^1(|\nu|)$ and for $f \in L^1(\nu)$,

$$\int_X f d\nu = \int_X f \frac{d\nu}{d\,|\nu|} d\,|\nu|\,.$$

We now give a number of methods for computing the total variation, $|\nu|$, of a complex or signed measure ν.

Proposition 24.32 (Total Variation). *Suppose $\mathcal{A} \subset 2^X$ is an algebra, $\mathcal{M} = \sigma(\mathcal{A})$, ν is a complex (or a signed measure which is σ – finite on $\mathcal{A}$) on $(X,\mathcal{M})$ and for $E \in \mathcal{M}$ let*

$$\mu_0(E) = \sup\left\{\sum_1^n |\nu(E_j)| : E_j \in \mathcal{A}_E \ni E_i \cap E_j = \delta_{ij}E_i,\ n = 1,2,\dots\right\}$$

$$\mu_1(E) = \sup\left\{\sum_1^n |\nu(E_j)| : E_j \in \mathcal{M}_E \ni E_i \cap E_j = \delta_{ij}E_i,\ n = 1,2,\dots\right\}$$

$$\mu_2(E) = \sup\left\{\sum_1^\infty |\nu(E_j)| : E_j \in \mathcal{M}_E \ni E_i \cap E_j = \delta_{ij}E_i\right\}$$

$$\mu_3(E) = \sup\left\{\left|\int_E f d\nu\right| : f \text{ is measurable with } |f| \leq 1\right\}$$

$$\mu_4(E) = \sup\left\{\left|\int_E f d\nu\right| : f \in \mathbb{S}_f(\mathcal{A},|\nu|) \text{ with } |f| \leq 1\right\}.$$

then $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = |\nu|$.

Proof. Let $\rho = d\nu/d\,|\nu|$ and recall that $|\rho| = 1$, $|\nu|$ – a.e.
Step 1. $\mu_4 \leq |\nu| = \mu_3$. If f is measurable with $|f| \leq 1$ then

$$\left|\int_E f\ d\nu\right| = \left|\int_E f\ \rho d\,|\nu|\right| \leq \int_E |f|\ d\,|\nu| \leq \int_E 1 d\,|\nu| = |\nu|\,(E)$$

from which we conclude that $\mu_4 \leq \mu_3 \leq |\nu|$. Taking $f = \bar\rho$ above shows

$$\left|\int_E f\ d\nu\right| = \int_E \bar\rho\ \rho\ d\,|\nu| = \int_E 1\ d\,|\nu| = |\nu|\,(E)$$

which shows that $|\nu| \leq \mu_3$ and hence $|\nu| = \mu_3$.

Step 2. $\mu_4 \geq |\nu|$. Let $X_m \in \mathcal{A}$ be chosen so that $|\nu|\,(X_m) < \infty$ and $X_m \uparrow X$ as $m \to \infty$. By Theorem 31.15 (or Remark **??** or Corollary **??** below), there exists $\rho_n \in \mathbb{S}_f(\mathcal{A},\mu)$ such that $\rho_n \to \rho 1_{X_m}$ in $L^1(|\nu|)$ and each ρ_n may be written in the form

$$\rho_n = \sum_{k=1}^N z_k 1_{A_k} \tag{24.12}$$

where $z_k \in \mathbb{C}$ and $A_k \in \mathcal{A}$ and $A_k \cap A_j = \emptyset$ if $k \neq j$. I claim that we may assume that $|z_k| \leq 1$ in Eq. (24.12) for if $|z_k| > 1$ and $x \in A_k$,

$$|\rho(x) - z_k| \geq \left|\rho(x) - |z_k|^{-1} z_k\right|.$$

This is evident from Figure 24.1 and formally follows from the fact that

$$\frac{d}{dt}\left|\rho(x) - t\,|z_k|^{-1} z_k\right|^2 = 2\left[t - \operatorname{Re}(|z_k|^{-1} z_k\overline{\rho(x)})\right] \geq 0$$

when $t \geq 1$. ■

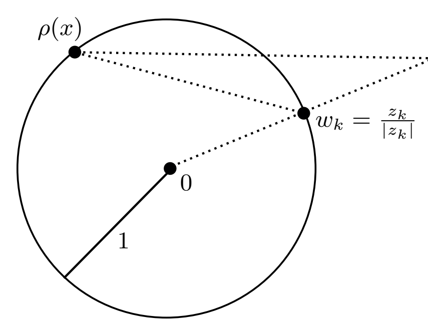


Fig. 24.1. Sliding points to the unit circle.

Therefore if we define

$$w_k := \begin{cases} |z_k|^{-1} z_k & \text{if } |z_k| > 1 \\ z_k & \text{if } |z_k| \leq 1 \end{cases}$$

and $\tilde\rho_n = \sum_{k=1}^N w_k 1_{A_k}$ then

$$|\rho(x) - \rho_n(x)| \geq |\rho(x) - \tilde\rho_n(x)|$$

and therefore $\tilde{\rho}_n \to \rho 1_{X_m}$ in $L^1(|\nu|)$. So we now assume that ρ_n is as in Eq. (24.12) with $|z_k| \le 1$. Now

$$\left|\int_E \bar{\rho}_n d\nu - \int_E \bar{\rho} 1_{X_m} d\nu\right| \le \left|\int_E (\bar{\rho}_n d\nu - \bar{\rho} 1_{X_m}) \rho d|\nu|\right| \le \int_E |\bar{\rho}_n - \bar{\rho} 1_{X_m}| \, d|\nu| \to 0 \text{ as } n \to \infty$$

and hence

$$\mu_4(E) \ge \left|\int_E \bar{\rho} 1_{X_m} d\nu\right| = |\nu| (E \cap X_m) \text{ for all } m.$$

Letting $m \uparrow \infty$ in this equation shows $\mu_4 \ge |\nu|$ which combined with step 1. shows $\mu_3 = \mu_4 = |\nu|$.

Proof. Step 3. $\mu_0 = \mu_1 = \mu_2 = |\nu|$. Clearly $\mu_0 \le \mu_1 \le \mu_2$. Suppose $\{E_j\}_{j=1}^\infty \subset \mathcal{M}_E$ be a collection of pairwise disjoint sets, then

$$\sum_{j=1}^\infty |\nu(E_j)| = \sum_{j=1}^\infty \int_{E_j} \rho d|\nu| \le \sum_{j=1}^\infty |\nu|(E_j) = |\nu|(\cup E_j) \le |\nu|(E)$$

which shows that $\mu_2 \le |\nu| = \mu_4$. So it suffices to show $\mu_4 \le \mu_0$. But if $f \in \mathbb{S}_f(\mathcal{A}, |\nu|)$ with $|f| \le 1$, then f may be expressed as $f = \sum_{k=1}^N z_k 1_{A_k}$ with $|z_k| \le 1$ and $A_k \cap A_j = \delta_{ij} A_k$. Therefore,

$$\left|\int_E f d\nu\right| = \left|\sum_{k=1}^N z_k \nu(A_k \cap E)\right| \le \sum_{k=1}^N |z_k| \, |\nu(A_k \cap E)| \le \sum_{k=1}^N |\nu(A_k \cap E)| \le \mu_0(A).$$

Since this equation holds for all $f \in \mathbb{S}_f(\mathcal{A}, |\nu|)$ with $|f| \le 1$, $\mu_4 \le \mu_0$ as claimed. ∎

Theorem 24.33 (Radon Nikodym Theorem for Complex Measures). *Let ν be a complex measure and μ be a σ – finite positive measure on $(X, \mathcal{M})$. Then ν has a unique Lebesgue decomposition $\nu = \nu_a + \nu_s$ relative to μ and there exists a unique element $\rho \in L^1(\mu)$ such that such that $d\nu_a = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$, i.e. $d\nu = \rho d\mu$ iff $\nu \ll \mu$.*

Proof. Uniqueness. Is a direct consequence of Lemmas 24.10 and 24.11. **Existence.** Let $g : X \to S^1 \subset \mathbb{C}$ be a function such that $d\nu = g d|\nu|$. By Theorem 24.13, there exists $h \in L^1(\mu)$ and a positive measure $|\nu|_s$ such that $|\nu|_s \perp \mu$ and $d|\nu| = h d\mu + d|\nu|_s$. Hence we have $d\nu = \rho d\mu + d\nu_s$ with $\rho := gh \in L^1(\mu)$ and $d\nu_s := g d|\nu|_s$. This finishes the proof since, as is easily verified, $\nu_s \perp \mu$. ∎

24.4 Absolute Continuity on an Algebra

The following results will be needed in Section 25.4 below.

Exercise 24.2. Let $\nu = \nu^r + i\nu^i$ is a complex measure on a measurable space, $(X, \mathcal{M})$, then $|\nu^r| \le |\nu|$, $|\nu^i| \le |\nu|$ and $|\nu| \le |\nu^r| + |\nu^i|$.

Exercise 24.3. Let ν be a signed measure on a measurable space, $(X, \mathcal{M})$. If $A \in \mathcal{M}$ is set such that there exists $M < \infty$ such that $|\nu(B)| \le M$ for all $B \in \mathcal{M}_A = \{C \cap A : C \in \mathcal{M}\}$, then $|\nu|(A) \le 2M$. If ν is complex measure with $A \in \mathcal{M}$ and $M < \infty$ as above, then $|\nu|(A) \le 4M$.

Lemma 24.34. *Let ν be a complex or a signed measure on $(X, \mathcal{M})$. Then $A \in \mathcal{M}$ is a ν – null set iff $|\nu|(A) = 0$. In particular if μ is a positive measure on $(X, \mathcal{M})$, $\nu \ll \mu$ iff $|\nu| \ll \mu$.*

Proof. In all cases we have $|\nu(A)| \le |\nu|(A)$ for all $A \in \mathcal{M}$ which clearly shows that $|\nu|(A) = 0$ implies A is a ν – null set. Conversely if A is a ν – null set, then, by definition, $\nu|_{\mathcal{M}_A} \equiv 0$ so by Proposition 24.32

$$|\nu|(A) = \sup\left\{\sum_1^\infty |\nu(E_j)| : E_j \in \mathcal{M}_A \ni E_i \cap E_j = \delta_{ij} E_i\right\} = 0.$$

since $E_j \subset A$ implies $\mu(E_j) = 0$ and hence $\nu(E_j) = 0$.

Alternate Proofs that A is ν – null implies $|\nu|(A) = 0$.

1) Suppose ν is a signed measure and $\{P, N = P^c\} \subset \mathcal{M}$ is a Hahn decomposition for ν. Then

$$|\nu|(A) = \nu(A \cap P) - \nu(A \cap N) = 0.$$

Now suppose that ν is a complex measure. Then A is a null set for both $\nu_r := \operatorname{Re} \nu$ and $\nu_i := \operatorname{Im} \nu$. Therefore $|\nu|(A) \le |\nu_r|(A) + |\nu_i|(A) = 0$.

2) Here is another proof in the complex case. Let $\rho = \frac{d\nu}{d|\nu|}$, then by assumption of A being ν – null,

$$0 = \nu(B) = \int_B \rho d|\nu| \text{ for all } B \in \mathcal{M}_A.$$

This shows that $\rho 1_A = 0$, $|\nu|$ – a.e. and hence

$$|\nu|(A) = \int_A |\rho| \, d|\nu| = \int_X 1_A |\rho| \, d|\nu| = 0.$$

∎

Theorem 24.35 ($\epsilon - \delta$ Definition of Absolute Continuity). *Let ν be a complex measure and μ be a positive measure on $(X, \mathcal{M})$. Then $\nu \ll \mu$ iff for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $|\nu(A)| < \varepsilon$ whenever $A \in \mathcal{M}$ and $\mu(A) < \delta$.*

Proof. ($\Longleftarrow$) If $\mu(A) = 0$ then $|\nu(A)| < \varepsilon$ for all $\varepsilon > 0$ which shows that $\nu(A) = 0$, i.e. $\nu \ll \mu$.

($\Longrightarrow$) Since $\nu \ll \mu$ iff $|\nu| \ll \mu$ and $|\nu(A)| \leq |\nu|(A)$ for all $A \in \mathcal{M}$, it suffices to assume $\nu \geq 0$ with $\nu(X) < \infty$. Suppose for the sake of contradiction there exists $\varepsilon > 0$ and $A_n \in \mathcal{M}$ such that $\nu(A_n) \geq \varepsilon > 0$ while $\mu(A_n) \leq \frac{1}{2^n}$. Let

$$A = \{A_n \text{ i.o.}\} = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n$$

so that

$$\mu(A) = \lim_{N\to\infty} \mu\left(\cup_{n\geq N} A_n\right) \leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \mu(A_n) \leq \lim_{N\to\infty} 2^{-(N-1)} = 0.$$

On the other hand,

$$\nu(A) = \lim_{N\to\infty} \nu\left(\cup_{n\geq N} A_n\right) \geq \lim_{n\to\infty} \inf \nu(A_n) \geq \varepsilon > 0$$

showing that ν is not absolutely continuous relative to μ. ■

Corollary 24.36. *Let μ be a positive measure on $(X, \mathcal{M})$ and $f \in L^1(d\mu)$. Then for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\left|\int_A f \, d\mu\right| < \varepsilon$ for all $A \in \mathcal{M}$ such that $\mu(A) < \delta$.*

Proof. Apply theorem 24.35 to the signed measure $\nu(A) = \int_A f \, d\mu$ for all $A \in \mathcal{M}$.

Alternative proof. If the statement in the corollary were false, there would exists $\varepsilon > 0$ and $A_n \in \mathcal{M}$ such that $\mu(A_n) \downarrow 0$ while $\left|\int_{A_n} f \, d\mu\right| \geq \varepsilon$ for all n. On the other hand $|1_{A_n} f| \leq |f| \in L^1(\mu)$ and $1_{A_n} f \overset{\mu}{\to} 0$ as $n \to \infty$ and so by the dominated convergence theorem in Corollary 16.21 we may conclude,

$$\lim_{n\to\infty} \int_{A_n} f \, d\mu = \lim_{n\to\infty} \int_X 1_{A_n} f \, d\mu = 0$$

which leads to the desired contradiction. ■

Theorem 24.37 (Absolute Continuity on an Algebra). *Let ν be a complex measure and μ be a positive measure on $(X, \mathcal{M})$. Suppose that $\mathcal{A} \subset \mathcal{M}$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and that μ is σ – finite on $\mathcal{A}$. Then $\nu \ll \mu$ iff for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $|\nu(A)| < \varepsilon$ for all $A \in \mathcal{A}$ which satisfy $\mu(A) < \delta$.*

Proof. ($\Longrightarrow$) This implication is a consequence of Theorem 24.35.

($\Longleftarrow$) If $|\nu(A)| < \varepsilon$ for all $A \in \mathcal{A}$ with $\mu(A) < \delta$, then by Exercise 24.3, $|\nu|(A) \leq 4\varepsilon$ for all $A \in \mathcal{A}$ with $\mu(A) < \delta$. Because of this argument, we may now replace ν by $|\nu|$ and hence we may assume that ν is a positive finite measure.

Let $\varepsilon > 0$ and $\delta > 0$ be such that $\nu(A) < \varepsilon$ for all $A \in \mathcal{A}$ with $\mu(A) < \delta$. Suppose that $B \in \mathcal{M}$ with $\mu(B) < \delta$ and $\alpha \in (0, \delta - \mu(B))$. By Corollary 31.18, there exists $A \in \mathcal{A}$ such that

$$\mu(A \Delta B) + \nu(A \Delta B) = (\mu + \nu)(A \Delta B) < \alpha.$$

In particular it follows that $\mu(A) \leq \mu(B) + \mu(A \Delta B) < \delta$ and hence by assumption $\nu(A) < \varepsilon$. Therefore,

$$\nu(B) \leq \nu(A) + \nu(A \Delta B) < \varepsilon + \alpha$$

and letting $\alpha \downarrow 0$ in this inequality shows $\nu(B) \leq \varepsilon$.

Alternative Proof. Let $\varepsilon > 0$ and $\delta > 0$ be such that $\nu(A) < \varepsilon$ for all $A \in \mathcal{A}$ with $\mu(A) < \delta$. Suppose that $B \in \mathcal{M}$ with $\mu(B) < \delta$. Use the regularity Theorem **??** below (or see Theorem **??** or Corollary **??**) to find $A \in \mathcal{A}_\sigma$ such that $B \subset A$ and $\mu(B) \leq \mu(A) < \delta$. Write $A = \cup_n A_n$ with $A_n \in \mathcal{A}$. By replacing A_n by $\cup_{j=1}^n A_j$ if necessary we may assume that A_n is increasing in n. Then $\mu(A_n) \leq \mu(A) < \delta$ for each n and hence by assumption $\nu(A_n) < \varepsilon$. Since $B \subset A = \cup_n A_n$ it follows that $\nu(B) \leq \nu(A) = \lim_{n\to\infty} \nu(A_n) \leq \varepsilon$. Thus we have shown that $\nu(B) \leq \varepsilon$ for all $B \in \mathcal{M}$ such that $\mu(B) < \delta$. ■

24.5 Exercises

Exercise 24.4. Prove Theorem 29.6 for $p \in [1, 2]$ by directly applying the Riesz theorem to $\varphi|_{L^2(\mu)}$.

Exercise 24.5. Show $|\nu|$ be defined as in Eq. (29.9) is a positive measure. Here is an outline.

1. Show
$$|\nu|(A) + |\nu|(B) \leq |\nu|(A \cup B). \tag{24.13}$$
when A, B are disjoint sets in $\mathcal{M}$.

2. If $A = \coprod_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{M}$ then
$$|\nu|(A) \le \sum_{n=1}^{\infty} |\nu|(A_n). \tag{24.14}$$
3. From Eqs. (24.13) and (24.14) it follows that $|\nu|$ is finitely additive, and hence
$$|\nu|(A) = \sum_{n=1}^{N} |\nu|(A_n) + |\nu|(\cup_{n>N} A_n) \ge \sum_{n=1}^{N} |\nu|(A_n).$$
Letting $N \to \infty$ in this inequality shows $|\nu|(A) \ge \sum_{n=1}^{\infty} |\nu|(A_n)$ which combined with Eq. (24.14) shows $|\nu|$ is countably additive.

Exercise 24.6. Suppose X is a set, $\mathcal{A} \subset 2^X$ is an algebra, and $\nu : \mathcal{A} \to \mathbb{C}$ is a finitely additive measure. For any $A \in \mathcal{A}$, let
$$|\nu|(A) := \sup\left\{ \sum_{i=1}^{n} |\nu(A_i)| : A = \coprod_{i=1}^{n} A_i \text{ with } A_i \in \mathcal{A} \text{ and } n \in \mathbb{N} \right\}.$$

1. Suppose $\mathbb{P} := \{A_i\}_{i=1}^n \subset \mathcal{A}$ is a partition of $A \in \mathcal{A}$ and $\{B_j\}_{j=1}^m \subset \mathcal{A}$ is partition of A which refines $\mathbb{P}$ (i.e. for each j there exists an i such that $B_j \subset A_i$), then
$$\sum_{i=1}^{n} |\nu(A_i)| \le \sum_{j=1}^{m} |\nu(B_j)|. \tag{24.15}$$
2. The **total variation**,$|\nu| : \mathcal{A} \to [0, \infty]$, of ν is a finitely additive measure on $\mathcal{A}$.

Exercise 24.7. Suppose that $\{\nu_n\}_{n=1}^{\infty}$ are complex measures on a measurable space, $(X, \mathcal{M})$.

1. If $\sum_{n=1}^{\infty} |\nu_n|(X) < \infty$, then $\nu := \sum_{n=1}^{\infty} \nu_n$ is a complex measure.
2. If there is a finite positive measure, $\mu : \mathcal{M} \to [0, \infty)$ such that $|\nu_n(A)| \le \mu(A)$ for all $A \in \mathcal{M}$ and $\nu(A) := \lim_{n\to\infty} \nu_n(A)$ exists for all $A \in \mathcal{M}$, then ν is also a complex measure.

Exercise 24.8. Let $(X, \mathcal{M})$ be a measurable space and $M(X)$ denote the space of complex measures on $(X, \mathcal{M})$ and for $\mu \in M(X)$ let $\|\mu\| := |\mu|(X)$. Show $(M(X), \|\cdot\|)$ is a Banach space.

Exercise 24.9. Suppose μ_i, ν_i are σ − finite positive measures on measurable spaces, $(X_i, \mathcal{M}_i)$, for $i = 1, 2$. If $\nu_i \ll \mu_i$ for $i = 1, 2$ then $\nu_1 \otimes \nu_2 \ll \mu_1 \otimes \mu_2$ and in fact
$$\frac{d(\nu_1 \otimes \nu_2)}{d(\mu_1 \otimes \mu_2)}(x_1, x_2) = \rho_1 \otimes \rho_2(x_1, x_2) := \rho_1(x_1)\rho_2(x_2)$$
where $\rho_i := d\nu_i/d\mu_i$ for $i = 1, 2$.

Exercise 24.10. Let $X = [0, 1]$, $\mathcal{M} := \mathcal{B}_{[0,1]}$, m be Lebesgue measure and μ be counting measure on X. Show

1. $m \ll \mu$ yet there is not function ρ such that $dm = \rho d\mu$.
2. Counting measure μ has no Lebesgue decomposition relative to m.

Exercise 24.11. Suppose that ν is a signed or complex measure on $(X, \mathcal{M})$ and $A_n \in \mathcal{M}$ such that either $A_n \uparrow A$ or $A_n \downarrow A$ and $\nu(A_1) \in \mathbb{R}$, then show $\nu(A) = \lim_{n\to\infty} \nu(A_n)$.

Exercise 24.12. Let $(X, \mathcal{M})$ be a measurable space, $\nu : \mathcal{M} \to [-\infty, \infty)$ be a signed measure, and $\nu = \nu_+ - \nu_-$ be a Jordan decomposition of ν. If $\nu := \alpha - \beta$ with α and β being positive measures and $\alpha(X) < \infty$, show $\nu_+ \le \alpha$ and $\nu_- \le \beta$. Us this result to prove the uniqueness of Jordan decompositions stated in Theorem 24.18.

Exercise 24.13. Let ν_1 and ν_2 be two signed measures on $(X, \mathcal{M})$ which are assumed to be valued in $[-\infty, \infty)$. Show, $|\nu_1 + \nu_2| \le |\nu_1| + |\nu_2|$. **Hint:** use Exercise 24.12 along with the observation that $\nu_1 + \nu_2 = (\nu_1^+ + \nu_2^+) - (\nu_1^- + \nu_2^-)$, where $\nu_i^{\pm} := (\nu_i)_{\pm}$.

Exercise 24.14. Folland Exercise 3.7a on p. 88.

Exercise 24.15. Show Theorem 24.35 may fail if ν is not finite. (For a hint, see problem 3.10 on p. 92 of Folland.)

Exercise 24.16. Folland 3.14 on p. 92.

Exercise 24.17. Folland 3.15 on p. 92.

Exercise 24.18. If ν is a complex measure on $(X, \mathcal{M})$ such that $|\nu|(X) = \nu(X)$, then $\nu = |\nu|$.

Exercise 24.19. Suppose ν is a complex or a signed measure on a measurable space, $(X, \mathcal{M})$. Show $A \in \mathcal{M}$ is a ν - null set iff $|\nu|(A) = 0$. Use this to conclude that if μ is a positive measure, then $\nu \perp \mu$ iff $|\nu| \perp \mu$.

25

Lebesgue Differentiation and the Fundamental Theorem of Calculus

BRUCE: replace $\mathbb{R}^n$ by $\mathbb{R}^d$ in this section?

Notation 25.1 *In this chapter, let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n}$ denote the Borel σ – algebra on $\mathbb{R}^n$ and m be Lebesgue measure on $\mathcal{B}$. If V is an open subset of $\mathbb{R}^n$, let $L^1_{loc}(V) := L^1_{loc}(V, m)$ and simply write L^1_{loc} for $L^1_{loc}(\mathbb{R}^n)$. We will also write $|A|$ for $m(A)$ when $A \in \mathcal{B}$.*

Definition 25.2. *A collection of measurable sets $\{E\}_{r>0} \subset \mathcal{B}$ is said to shrink nicely to $x \in \mathbb{R}^n$ if (i) $E_r \subset \overline{B(x,r)}$ for all $r > 0$ and (ii) there exists $\alpha > 0$ such that $m(E_r) \geq \alpha m(B(x,r))$. We will abbreviate this by writing $E_r \downarrow \{x\}$ nicely. (Notice that it is not required that $x \in E_r$ for any $r > 0$.*

The main result of this chapter is the following theorem.

Theorem 25.3. *Suppose that ν is a complex measure on $(\mathbb{R}^n, \mathcal{B})$, then there exists $g \in L^1(\mathbb{R}^n, m)$ and a complex measure ν_s such that $\nu_s \perp m$, $d\nu = g dm + d\nu_s$, and for m - a.e. x,*

$$g(x) = \lim_{r \downarrow 0} \frac{\nu(E_r)}{m(E_r)} \text{ and} \tag{25.1}$$

$$0 = \lim_{r \downarrow 0} \frac{\nu_s(E_r)}{m(E_r)} \tag{25.2}$$

for any collection of $\{E_r\}_{r>0} \subset \mathcal{B}$ which shrink nicely to $\{x\}$. (Eq. (25.1) holds for all $x \in \mathcal{L}(g)$ – the Lebesgue set of g, see Definition 25.12 and Theorem 25.13 below.)

Proof. The existence of g and ν_s such that $\nu_s \perp m$ and $d\nu = g dm + d\nu_s$ is a consequence of the Radon-Nikodym Theorem 24.33. Since

$$\frac{\nu(E_r)}{m(E_r)} = \frac{1}{m(E_r)} \int_{E_r} g(x) dm(x) + \frac{\nu_s(E_r)}{m(E_r)}$$

Eq. (25.1) is a consequence of Theorem 25.14 and Corollary 25.16 below. ■

The rest of this chapter will be devoted to filling in the details of the proof of this theorem.

25.1 A Covering Lemma and Averaging Operators

Lemma 25.4 (Covering Lemma). *Let $\mathcal{E}$ be a collection of open balls in $\mathbb{R}^n$ and $U = \cup_{B \in \mathcal{E}} B$. If $c < m(U)$, then there exists disjoint balls $B_1, \ldots, B_k \in \mathcal{E}$ such that $c < 3^n \sum_{j=1}^{k} m(B_j)$.*

Proof. Choose a compact set $K \subset U$ such that $m(K) > c$ and then let $\mathcal{E}_1 \subset \mathcal{E}$ be a finite subcover of K. Choose $B_1 \in \mathcal{E}_1$ to be a ball with largest diameter in $\mathcal{E}_1$. Let $\mathcal{E}_2 = \{A \in \mathcal{E}_1 : A \cap B_1 = \emptyset\}$. If $\mathcal{E}_2$ is not empty, choose $B_2 \in \mathcal{E}_2$ to be a ball with largest diameter in $\mathcal{E}_2$. Similarly let $\mathcal{E}_3 = \{A \in \mathcal{E}_2 : A \cap B_2 = \emptyset\}$ and if $\mathcal{E}_3$ is not empty, choose $B_3 \in \mathcal{E}_3$ to be a ball with largest diameter in $\mathcal{E}_3$. Continue choosing $B_i \in \mathcal{E}$ for $i = 1, 2, \ldots, k$ this way until $\mathcal{E}_{k+1}$ is empty, see Figure 25.1 below. If $B = B(x_0, r) \subset \mathbb{R}^n$, let $B^* = B(x_0, 3r) \subset \mathbb{R}^n$, that is B^*

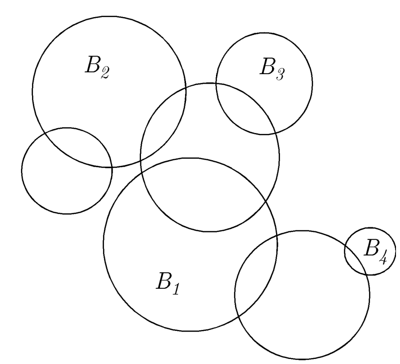


Fig. 25.1. Picking out the large disjoint balls via the "greedy algorithm."

is the ball concentric with B which has three times the radius of B. We will now show $K \subset \cup_{i=1}^k B_i^*$. For each $A \in \mathcal{E}_1$ there exists a first i such that $B_i \cap A \neq \emptyset$. In this case $\text{diam}(A) \leq \text{diam}(B_i)$ and $A \subset B_i^*$. Therefore $A \subset \cup_{i=1}^k B_i^*$ and hence $K \subset \cup\{A : A \in \mathcal{E}_1\} \subset \cup_{i=1}^k B_i^*$. Hence by sub-additivity,

$$c < m(K) \le \sum_{i=1}^{k} m(B_i^*) \le 3^n \sum_{i=1}^{k} m(B_i).$$

■

Definition 25.5. *For $f \in L^1_{loc}$, $x \in \mathbb{R}^n$ and $r > 0$ let*

$$(A_r f)(x) = \frac{1}{|B(x,r)|} \int\limits_{B(x,r)} f dm \tag{25.3}$$

where $B(x,r) = B(x,r) \subset \mathbb{R}^n$, and $|A| := m(A)$.

Lemma 25.6. *Let $f \in L^1_{loc}$, then for each $x \in \mathbb{R}^n$, $(0,\infty) \ni r \to (A_r f)(x) \in \mathbb{C}$ is continuous and for each $r > 0$, $\mathbb{R}^n \ni x \to (A_r f)(x) \in \mathbb{C}$ is measurable.*

Proof. Recall that $|B(x,r)| = m(E_1)r^n$ which is continuous in r. Also $\lim_{r\to r_0} 1_{B(x,r)}(y) = 1_{B(x,r_0)}(y)$ if $|y| \ne r_0$ and since $m(\{y : |y| \ne r_0\}) = 0$ (you prove!), $\lim_{r\to r_0} 1_{B(x,r)}(y) = 1_{B(x,r_0)}(y)$ for m -a.e. y. So by the dominated convergence theorem,

$$\lim_{r\to r_0} \int\limits_{B(x,r)} f dm = \int\limits_{B(x,r_0)} f dm$$

and therefore

$$(A_r f)(x) = \frac{1}{m(E_1)r^n} \int\limits_{B(x,r)} f dm$$

is continuous in r. Let $g_r(x,y) := 1_{B(x,r)}(y) = 1_{|x-y|<r}$. Then g_r is $\mathcal{B} \otimes \mathcal{B}$ – measurable (for example write it as a limit of continuous functions or just notice that $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by $F(x,y) := |x-y|$ is continuous) and so that by Fubini's theorem

$$x \to \int\limits_{B(x,r)} f dm = \int\limits_{B(x,r)} g_r(x,y) f(y) dm(y)$$

is $\mathcal{B}$ – measurable and hence so is $x \to (A_r f)(x)$. ■

25.2 Maximal Functions

Definition 25.7. *For $f \in L^1(m)$, the Hardy - Littlewood maximal function Hf is defined by*

$$(Hf)(x) = \sup_{r>0} A_r |f|(x).$$

Lemma 25.6 allows us to write

$$(Hf)(x) = \sup_{r\in\mathbb{Q},\, r>0} A_r |f|(x)$$

from which it follows that Hf is measurable.

Theorem 25.8 (Maximal Inequality). *If $f \in L^1(m)$ and $\alpha > 0$, then*

$$m(Hf > \alpha) \le \frac{3^n}{\alpha} \|f\|_{L^1}.$$

*(**Remark:** this theorem extends to $f \in L^1(m; X)$ where X is a separable Banach space – just replace $|\cdot|$ in the definition and proofs by $\|\cdot\|_X$.)*

This should be compared with Chebyshev's inequality which states that

$$m(|f| > \alpha) \le \frac{\|f\|_{L^1}}{\alpha}.$$

Proof. Let $E_\alpha := \{Hf > \alpha\}$. For all $x \in E_\alpha$ there exists r_x such that $A_{r_x} |f|(x) > \alpha$, i.e.

$$|B_x(r_x)| < \frac{1}{\alpha} \int_{B_x(r_x)} |f| \, dm.$$

Since $E_\alpha \subset \cup_{x\in E_\alpha} B_x(r_x)$, if $c < m(E_\alpha) \le m(\cup_{x\in E_\alpha} B_x(r_x))$ then, using Lemma 25.4, there exists $x_1, \dots, x_k \in E_\alpha$ and disjoint balls $B_i = B_{x_i}(r_{x_i})$ for $i = 1, 2, \dots, k$ such that

$$c < \sum_{i=1}^{k} 3^n |B_i| < \sum \frac{3^n}{\alpha} \int_{B_i} |f| \, dm \le \frac{3^n}{\alpha} \int_{\mathbb{R}^n} |f| \, dm = \frac{3^n}{\alpha} \|f\|_{L^1}.$$

This shows that $c < 3^n \alpha^{-1} \|f\|_{L^1}$ for all $c < m(E_\alpha)$ which proves $m(E_\alpha) \le 3^n \alpha^{-1} \|f\|$. ■

Theorem 25.9. *If $f \in L^1_{loc}$ then $\lim_{r\downarrow 0} (A_r f)(x) = f(x)$ for m – a.e. $x \in \mathbb{R}^n$.*

Proof. With out loss of generality we may assume $f \in L^1(m)$. We now begin with the special case where $f = g \in L^1(m)$ is also continuous. In this case we find:

$$\begin{aligned}|(A_r g)(x) - g(x)| &\le \frac{1}{|B(x,r)|} \int_{B(x,r)} |g(y) - g(x)| dm(y) \\ &\le \sup_{y\in B(x,r)} |g(y) - g(x)| \to 0 \text{ as } r \to 0.\end{aligned}$$

In fact we have shown that $(A_r g)(x) \to g(x)$ as $r \to 0$ uniformly for x in compact subsets of $\mathbb{R}^n$. For general $f \in L^1(m)$,

$$\begin{aligned}|A_r f(x) - f(x)| &\le |A_r f(x) - A_r g(x)| + |A_r g(x) - g(x)| + |g(x) - f(x)| \\ &= |A_r(f-g)(x)| + |A_r g(x) - g(x)| + |g(x) - f(x)| \\ &\le H(f-g)(x) + |A_r g(x) - g(x)| + |g(x) - f(x)|\end{aligned}$$

and therefore,

$$\overline{\lim_{r\downarrow 0}} |A_r f(x) - f(x)| \le H(f-g)(x) + |g(x) - f(x)|.$$

So if $\alpha > 0$, then

$$E_\alpha := \left\{\overline{\lim_{r\downarrow 0}} |A_r f(x) - f(x)| > \alpha\right\} \subset \left\{H(f-g) > \frac{\alpha}{2}\right\} \cup \left\{|g-f| > \frac{\alpha}{2}\right\}$$

and thus

$$\begin{aligned}m(E_\alpha) &\le m\left(H(f-g) > \frac{\alpha}{2}\right) + m\left(|g-f| > \frac{\alpha}{2}\right) \\ &\le \frac{3^n}{\alpha/2}\|f-g\|_{L^1} + \frac{1}{\alpha/2}\|f-g\|_{L^1} \\ &\le 2(3^n+1)\alpha^{-1}\|f-g\|_{L^1},\end{aligned}$$

where in the second inequality we have used the Maximal inequality (Theorem 25.8) and Chebyshev's inequality. Since this is true for all continuous $g \in C(\mathbb{R}^n) \cap L^1(m)$ and this set is dense in $L^1(m)$, we may make $\|f-g\|_{L^1}$ as small as we please. This shows that

$$m\left(\left\{x : \overline{\lim_{r\downarrow 0}} |A_r f(x) - f(x)| > 0\right\}\right) = m(\cup_{n=1}^\infty E_{1/n}) \le \sum_{n=1}^\infty m(E_{1/n}) = 0.$$

■

Remark 25.10. Theorem 25.9 also holds for $f \in L^1(m; X)$ where X is a separable Banach space. The only point is to observe that $C_c(\mathbb{R}^n; X)$ are still dense in $L^1(m; X)$. To prove we use the fact that X – valued L^1 – simple functions are dense in $L^1(m; X)$ and so it suffices to show that $1_A \cdot x$ may be approximated by $g \in C_c(\mathbb{R}^n; X)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$ with $m(A) < \infty$ and $x \in X$. But this is easy to do by taking $g = \varphi \cdot x$ where $\|\varphi - 1_A\|_{L^1(m)}$ is small and $\varphi \in C_c(\mathbb{R}^n, \mathbb{R})$.

Corollary 25.11. *If $d\mu = g dm$ with $g \in L^1_{loc}$ then*

$$\frac{\mu(B(x,r))}{|B(x,r)|} = A_r g(x) \to g(x) \text{ for } m \text{ – a.e. } x.$$

25.3 Lebesque Set

Definition 25.12. *For $f \in L^1_{loc}(m)$, the **Lebesgue set** of f is*

$$\begin{aligned}\mathcal{L}(f) &:= \left\{x \in \mathbb{R}^n : \lim_{r\downarrow 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)| dy = 0\right\} \\ &= \left\{x \in \mathbb{R}^n : \lim_{r\downarrow 0} (A_r |f(\cdot) - f(x)|)(x) = 0\right\}.\end{aligned}$$

More generally, if $p \in [1,\infty)$ and $f \in L^p_{loc}(m)$, let

$$\mathcal{L}_p(f) := \left\{x \in \mathbb{R}^n : \lim_{r\downarrow 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|^p dy = 0\right\}$$

Theorem 25.13. *Suppose $1 \le p < \infty$ and $f \in L^p_{loc}(m)$, then $m(\mathbb{R}^d \setminus \mathcal{L}_p(f)) = 0$. (This result also holds for $f \in L^p_{loc}(m; X)$ where X is a separable Banach space. One need only replace $\mathbb{Q} + i\mathbb{Q}$ by a countable dense subset of X and $|\cdot|$ by $\|\cdot\|_X$ in the proof below.)*

Proof. For $w \in \mathbb{C}$ define $g_w(x) = |f(x) - w|^p$ and

$$E_w := \left\{x : \lim_{r\downarrow 0} (A_r g_w)(x) \ne g_w(x)\right\}$$

and further let

$$E = \bigcup_{w \in \mathbb{Q}+i\mathbb{Q}} E_w.$$

Then by Theorem 25.9 $m(E_w) = 0$ for all $w \in \mathbb{C}$ and therefore $m(E) = 0$. By definition of E, if $x \notin E$ then.

$$\lim_{r\downarrow 0} (A_r |f(\cdot) - w|^p)(x) = |f(x) - w|^p$$

for all $w \in \mathbb{Q} + i\mathbb{Q}$. Letting $q := \frac{p}{p-1}$ (so that $p/q = p-1$) we have

$$\begin{aligned}|f(\cdot) - f(x)|^p &\le (|f(\cdot) - w| + |w - f(x)|)^p \\ &\le 2^{p/q}(|f(\cdot) - w|^p + |w - f(x)|^p) = 2^{p-1}(|f(\cdot) - w|^p + |w - f(x)|^p),\end{aligned}$$

$$(A_r |f(\cdot) - f(x)|^p)(x) \le 2^{p-1}(A_r |f(\cdot) - w|^p)(x) + 2^{p-1}|w - f(x)|$$

and hence for $x \notin E$,

$$\overline{\lim_{r\downarrow 0}}(A_r|f(\cdot)-f(x)|^p)(x) \le 2^{p-1}|f(x)-w|^p + 2^{p-1}|w-f(x)|^p = 2^p|f(x)-w|^p.$$

Since this is true for all $w \in \mathbb{Q}+i\mathbb{Q}$, we see that

$$\overline{\lim_{r\downarrow 0}}(A_r|f(\cdot)-f(x)|^p)(x) = 0 \text{ for all } x \notin E,$$

i.e. $E^c \subset \mathcal{L}_p(f)$ or equivalently $(\mathcal{L}_p(f))^c \subset E$. So $m(\mathbb{R}^n \setminus \mathcal{L}_p(f)) \le m(E) = 0$. ■

Theorem 25.14 (Lebesque Differentiation Theorem). *If $f \in L^p_{loc}$ and $x \in \mathcal{L}_p(f)$ (so in particular for m – a.e. x), then*

$$\lim_{r\downarrow 0} \frac{1}{m(E_r)} \int_{E_r} |f(y)-f(x)|^p dy = 0$$

and

$$\lim_{r\downarrow 0} \frac{1}{m(E_r)} \int_{E_r} f(y)dy = f(x)$$

when $E_r \downarrow \{x\}$ nicely, see Definition 25.2.

Proof. For $x \in \mathcal{L}_p(f)$, by Hölder's inequality (Theorem 16.1) or Jensen's inequality (Theorem 28.8), we have

$$\begin{aligned}\left|\frac{1}{m(E_r)}\int_{E_r} f(y)dy - f(x)\right|^p &= \left|\frac{1}{m(E_r)}\int_{E_r}(f(y)-f(x))\,dy\right|^p \\ &\le \frac{1}{m(E_r)}\int_{E_r}|f(y)-f(x)|^p dy \\ &\le \frac{1}{\alpha m(B(x,r))}\int_{B(x,r)}|f(y)-f(x)|^p dy\end{aligned}$$

which tends to zero as $r \downarrow 0$ by Theorem 25.13. In the second inequality we have used the fact that $m(\overline{B(x,r)} \setminus B(x,r)) = 0$. ■

Lemma 25.15. *Suppose λ is positive K – finite measure on $\mathcal{B} := \mathcal{B}_{\mathbb{R}^n}$ such that $\lambda \perp m$. Then for m – a.e. x,*

$$\lim_{r\downarrow 0} \frac{\lambda(B(x,r))}{m(B(x,r))} = 0.$$

Proof. Let $A \in \mathcal{B}$ such that $\lambda(A) = 0$ and $m(A^c) = 0$. By the regularity theorem (see Theorem 38.16, Corollary **??** or Exercise **??**), for all $\varepsilon > 0$ there exists an open set $V_\varepsilon \subset \mathbb{R}^n$ such that $A \subset V_\varepsilon$ and $\lambda(V_\varepsilon) < \varepsilon$. For the rest of this argument, we will assume m has been extended to the Lebesgue measurable sets, $\mathcal{L} := \bar{\mathcal{B}}^m$. Let

$$F_k := \left\{x \in A : \overline{\lim_{r\downarrow 0}} \frac{\lambda(B(x,r))}{m(B(x,r))} > \frac{1}{k}\right\}$$

the for $x \in F_k$ choose $r_x > 0$ such that $B_x(r_x) \subset V_\varepsilon$ (see Figure 25.2) and $\frac{\lambda(B(x,r_x))}{m(B(x,r_x))} > \frac{1}{k}$, i.e.

$$m(B(x,r_x)) < k\lambda(B(x,r_x)).$$

Let $\mathcal{E} = \{B(x,r_x)\}_{x\in F_k}$ and $U := \bigcup_{x\in F_k} B(x,r_x) \subset V_\varepsilon$. Heuristically if all the

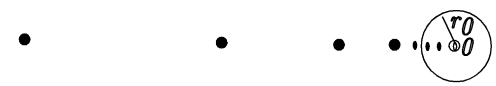


Fig. 25.2. In this picture we imagine that $\lambda = \sum_{n=1}^\infty n^{-2}\delta_{1/n}$ and $A = \mathbb{R}^2 \setminus \{(-1/n,0) : n \in \mathbb{N}\}$. We may approximate A by the open sets, $V_N := \mathbb{R}^2 \setminus \{(-1/n,0) : 1 \le n \le N\}$, since $\lambda(V_N) = \sum_{n=N+1}^\infty n^{-2} \to 0$ as $N \to \infty$. (Of course we could simplify matters in this setting by choosing $A = V := \mathbb{R}^2 \setminus (\{(-1/n,0) : 1 \le n \le N\} \cup \{0\})$, but this would not be very enlightening.)

balls in $\mathcal{E}$ were disjoint and $\mathcal{E}$ were countable, then

$$\begin{aligned}m(F_k) &\le \sum_{x\in F_k} m(B(x,r_x)) < k\sum_{x\in F_k}\lambda(B(x,r_x)) \\ &= k\lambda(U) \le k\ \lambda(V_\varepsilon) \le k\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary this would imply that $F_k \in \mathcal{L}$ and $m(F_k) = 0$. To fix the above argument, suppose that $c < m(U)$ and use the covering lemma to find disjoint balls $B_1, \dots, B_N \in \mathcal{E}$ such that

$$\begin{aligned}c &< 3^n \sum_{i=1}^N m(B_i) < k3^n \sum_{i=1}^N \lambda(B_i) \\ &\le k3^n\lambda(U) \le k3^n\lambda(V_\varepsilon) \le k3^n\varepsilon.\end{aligned}$$

Since $c < m(U)$ is arbitrary we learn that $m(U) \le k3^n\varepsilon$. This argument shows open sets U_ε such that $F_k \subset U_\varepsilon$ and $m(U_\varepsilon) \le k3^n\varepsilon$ for all $\varepsilon > 0$. Therefore $F_k \subset G := \cap_{l=1}^\infty U_{1/l} \in \mathcal{B}$ with $m(G) = 0$ which shows $F_k \in \mathcal{L}$ and $m(F_k) = 0$. Since

$$F_\infty := \left\{ x \in A : \overline{\lim_{r \downarrow 0}} \frac{\lambda(B(x,r))}{m(B(x,r))} > 0 \right\} = \cup_{k=1}^\infty F_k \in \mathcal{L},$$

it also follows that $F_\infty \in \mathcal{L}$ and $m(F_\infty) = 0$. Since

$$\left\{ x \in \mathbb{R}^n : \overline{\lim_{r \downarrow 0}} \frac{\lambda(B(x,r))}{m(B(x,r))} > 0 \right\} \subset F_\infty \cup A^c$$

and $m(A^c) = 0$, we have shown

$$m\left(\left\{ x \in \mathbb{R}^n : \overline{\lim_{r \downarrow 0}} \frac{\lambda(B(x,r))}{m(B(x,r))} > 0 \right\} \right) = 0.$$

■

Corollary 25.16. *Let λ be a complex or a K – finite signed measure (i.e. $\nu(K) \in \mathbb{R}$ for all $K \sqsubset\sqsubset \mathbb{R}^n$) such that $\lambda \perp m$. Then for m – a.e. x,*

$$\lim_{r \downarrow 0} \frac{\lambda(E_r)}{m(E_r)} = 0$$

whenever $E_r \downarrow \{x\}$ nicely.

Proof. By Exercise 24.19, $\lambda \perp m$ implies $|\lambda| \perp m$. Hence the result follows from Lemma 25.15 and the inequalities,

$$\frac{|\lambda(E_r)|}{m(E_r)} \le \frac{|\lambda|(E_r)}{\alpha m(B(x,r))} \le \frac{|\lambda|(\overline{B(x,r)})}{\alpha m(B(x,r))} \le \frac{|\lambda|(B(x,2r))}{\alpha 2^{-n} m(B(x,2r))}.$$

■

25.4 The Fundamental Theorem of Calculus

In this section we will restrict the results above to the one dimensional setting. The following notation will be in force for the rest of this chapter. (BRUCE: make sure this notation agrees with the notation in Notation 25.21.)

Notation 25.17 *Let*

1. *m be one dimensional Lebesgue measure on $\mathcal{B} := \mathcal{B}_\mathbb{R}$,*
2. *α, β be numbers in $\bar{\mathbb{R}}$ such that $-\infty \le \alpha < \beta \le \infty$,*
3. *$\mathcal{A} = \mathcal{A}_{[\alpha,\beta]}$ be the algebra generated by sets of the form $(a,b] \cap [\alpha,\beta]$ with $-\infty \le a < b \le \infty$,*
4. *$\mathcal{A}^b$ denote those sets in $\mathcal{A}$ which are bounded,*
5. *and $\mathcal{B}_{[\alpha,\beta]}$ be the Borel σ – algebra on $[\alpha,\beta] \cap \mathbb{R}$.*

Notation 25.18 *Given a function $F : \mathbb{R} \to \bar{\mathbb{R}}$ or $F : \mathbb{R} \to \mathbb{C}$, let $F(x-) = \lim_{y \uparrow x} F(y)$, $F(x+) = \lim_{y \downarrow x} F(y)$ and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x)$ whenever the limits exist. Notice that if F is a monotone functions then $F(\pm\infty)$ and $F(x\pm)$ exist for all x.*

25.4.1 Increasing Functions

Theorem 25.19 (Monotone functions). *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and define $G(x) = F(x+)$. Then:*

1. *The function G is increasing and right continuous.*
2. *For $x \in \mathbb{R}$, $G(x) = \lim_{y \downarrow x} F(y-)$.*
3. *The set of discontinuities of F, $\{x \in \mathbb{R} : F(x+) > F(x-)\}$, is countable. Moreover for each $N > 0$,*
$$\sum_{x \in (-N,N]} [F(x+) - F(x-)] \le F(N) - F(-N) < \infty. \tag{25.4}$$
4. *There exists a unique measure, ν_G on $\mathcal{B} = \mathcal{B}_\mathbb{R}$ such that*
$$\nu_G((a,b]) = G(b) - G(a) \text{ for all } a < b.$$
5. *For m – a.e. x, $F'(x)$ and $G'(x)$ exists and $F'(x) = G'(x)$. (Notice that $F'(x)$ and $G'(x)$ are non-negative when they exist.)*
6. *The function F' ($= G'$ a.e.) is in $L^1_{loc}(m)$ and there exists a unique positive measure ν_s on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that*
$$F(b+) - F(a+) = \int_a^b F' dm + \nu_s((a,b]) \text{ for all } -\infty < a < b < \infty.$$
Furthermore, the measure ν_s is singular relative to m and $F' \in L^1(\mathbb{R}, m)$ if F is bounded.

Proof.

1. The following observation shows G is increasing: if $x < y$ then
$$F(x-) \le F(x) \le F(x+) = G(x) \le F(y-) \le F(y) \le F(y+) = G(y). \tag{25.5}$$
Since G is increasing, $G(x) \le G(x+)$. If $y > x$ then $G(x+) \le F(y)$ and hence $G(x+) \le F(x+) = G(x)$, i.e. $G(x+) = G(x)$ which is to say G is right continuous. (For another proof, see Eq. (38.28) of Theorem 38.29 below.)
2. Since $G(x) \le F(y-) \le F(y)$ for all $y > x$, it follows that
$$G(x) \le \lim_{y \downarrow x} F(y-) \le \lim_{y \downarrow x} F(y) = G(x)$$
showing $G(x) = \lim_{y \downarrow x} F(y-)$.

3. By Eq. (25.5), if $x \neq y$ then
$$(F(x-), F(x+)] \cap (F(y-), F(y+)] = \emptyset.$$
Therefore, $\{(F(x-), F(x+)]\}_{x\in\mathbb{R}}$ are disjoint possible empty intervals in $\mathbb{R}$. Let $N \in \mathbb{N}$ and $\alpha \subset_f (-N, N)$ be a finite set, then
$$\coprod_{x\in\alpha}(F(x-), F(x+)] \subset (F(-N), F(N)]$$
and therefore,
$$\sum_{x\in\alpha}[F(x+) - F(x-)] \le F(N) - F(-N) < \infty.$$
Since this is true for all $\alpha \subset_f (-N, N]$, Eq. (25.4) holds. Eq. (25.4) shows
$$\Gamma_N := \{x \in (-N, N) | F(x+) - F(x-) > 0\}$$
is countable and hence so is
$$\Gamma := \{x \in \mathbb{R} | F(x+) - F(x-) > 0\} = \cup_{N=1}^{\infty}\Gamma_N.$$
4. Item 4. is a direct consequence of Theorem 8.33 (or Theorem 38.26 below). Notice that ν_G is a finite measure when F and hence G is bounded.
5. Theorem 25.3 now asserts that ν_G decomposes as;
$$d\nu_G = g dm + d\nu_s,$$
where $\nu_s \perp m$, $g \in L^1_{\text{loc}}(\mathbb{R}, m)$ with $g \in L^1(\mathbb{R}, m)$ if F is bounded. Moreover Theorem 25.3 implies, for m - a.e. x,
$$g(x) = \lim_{r\downarrow 0}(\nu_G(E_r)/m(E_r)),$$
where $\{E_r\}_{r>0}$ is any collection of sets shrink nicely to $\{x\}$. Since $(x, x+r] \downarrow \{x\}$ and $(x-r, x] \downarrow \{x\}$ nicely,
$$g(x) = \lim_{r\downarrow 0}\frac{\nu_G(x, x+r])}{m((x, x+r])} = \lim_{r\downarrow 0}\frac{G(x+r) - G(x)}{r} = \frac{d}{dx^+}G(x) \qquad (25.6)$$
and
$$g(x) = \lim_{r\downarrow 0}\frac{\nu_G((x-r, x])}{m((x-r, x])} = \lim_{r\downarrow 0}\frac{G(x) - G(x-r)}{r} = \lim_{r\downarrow 0}\frac{G(x-r) - G(x)}{-r} = \frac{d}{dx^-}G(x) \qquad (25.7)$$
exist and are equal for m - a.e. x, i.e. $G'(x) = g(x)$ exists for m -a.e. x. For $x \in \mathbb{R}$, let
$$H(x) := G(x) - F(x) = F(x+) - F(x) \ge 0.$$
Since $F(x) = G(x) - H(x)$, the proof of **5.** will be complete once we show $H'(x) = 0$ for m – a.e. x. From Item 3.,
$$\Lambda := \{x \in \mathbb{R} : F(x+) > F(x)\} \subset \{x \in \mathbb{R} : F(x+) > F(x-)\}$$
is a countable set and
$$\sum_{x\in(-N,N)} H(x) = \sum_{x\in(-N,N)} (F(x+){-}F(x)) \le \sum_{x\in(-N,N)} (F(x+){-}F(x-)) < \infty$$
for all $N < \infty$. Therefore $\lambda := \sum_{x\in\mathbb{R}} H(x)\delta_x$ (i.e. $\lambda(A) := \sum_{x\in A} H(x)$ for all $A \in \mathcal{B}_{\mathbb{R}}$) defines a Radon measure on $\mathcal{B}_{\mathbb{R}}$. Since $\lambda(\Lambda^c) = 0$ and $m(\Lambda) = 0$, the measure $\lambda \perp m$. By Corollary 25.16 for m - a.e. x,
$$\begin{aligned}\left|\frac{H(x+r) - H(x)}{r}\right| &\le \frac{|H(x+r)| + |H(x)|}{|r|} \\ &\le \frac{H(x+|r|) + H(x-|r|) + H(x)}{|r|} \\ &\le 2\frac{\lambda([x-|r|, x+|r|])}{2|r|}\end{aligned}$$
and the last term goes to zero as $r \to 0$ because $\{[x-r, x+r]\}_{r>0}$ shrinks nicely to $\{x\}$ as $r \downarrow 0$ and $m([x-|r|, x+|r|]) = 2|r|$. Hence we conclude for m – a.e. x that $H'(x) = 0$.
6. From Theorem 25.3, item 5. and Eqs. (25.6) and (25.7), $F' = G' \in L^1_{loc}(m)$ and $d\nu_G = F'dm + d\nu_s$ where ν_s is a positive measure such that $\nu_s \perp m$. Applying this equation to an interval of the form $(a, b]$ gives
$$F(b+) - F(a+) = \nu_G((a, b]) = \int_a^b F'dm + \nu_s((a, b]). \qquad (25.8)$$
The uniqueness of ν_s such that this equation holds is a consequence of Theorem **??**. As we have already mentioned, when F is bounded then $F' \in L^1(\mathbb{R}, m)$. This can also be seen directly by letting $a \to -\infty$ and $b \to +\infty$ in Eq. (25.8).

■

Example 25.20. Let $C \subset [0, 1]$ denote the Cantor set constructed as follows. Let $C_1 = [0, 1] \setminus (1/3, 2/3)$, $C_2 := C_1 \setminus [(1/9, 2/9) \cup (7/9, 8/9)]$, etc., so that we keep removing the middle thirds at each stage in the construction. Letting $C_0 := [0,1]$, we have $m(C_{n+1}) = \frac{2}{3} m(C_n)$ for $n \geq 0$ and hence $m(C_n) = (2/3)^n \to 0$ as $n \to \infty$. We now let

$$C := \cap_{n=1}^{\infty} C_n = \left\{ x = \sum_{j=0}^{\infty} a_j 3^{-j} : a_j \in \{0,2\} \right\}$$

and since $C_n \downarrow C$ it follows that $m(C) = \lim_{n\to\infty} m(C_n) = 0$. Associated to this set is the so called Cantor function $F(x) := \lim_{n\to\infty} f_n(x)$ where the $\{f_n\}_{n=1}^{\infty}$ are continuous non-decreasing functions such that $f_n(0) = 0$, $f_n(1) = 1$ with the f_n pictured in Figure 25.3 below.From the pictures one sees that

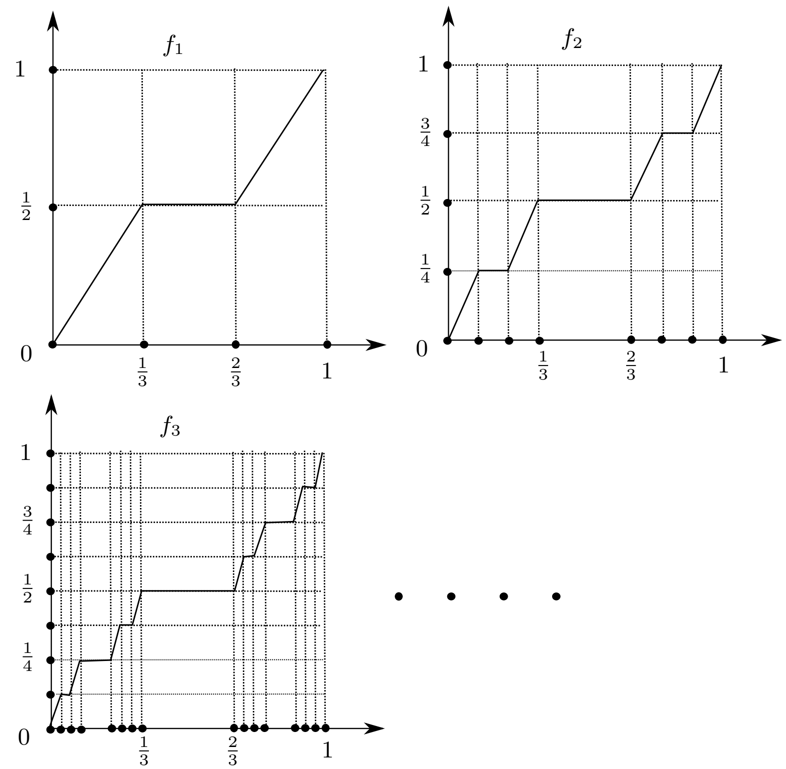


Fig. 25.3. Constructing a Cantor function.

$\{f_n\}$ are uniformly Cauchy, hence there exists $F \in C([0,1])$ such that $F(x) := \lim_{n\to\infty} f_n(x)$. The function F has the following properties,

1. F is continuous and non-decreasing.
2. $F'(x) = 0$ for m – a.e. $x \in [0,1]$ because F is flat on all of the middle third open intervals used to construct the Cantor set C and the total measure of these intervals is 1 as proved above.
3. The measure on $\mathcal{B}_{[0,1]}$ associated to F, namely $\nu([0,b]) = F(b)$ is singular relative to Lebesgue measure and $\nu(\{x\}) = 0$ for all $x \in [0,1]$. Notice that $\nu([0,1]) = 1$. In particular, the function F certainly does not satisfy the fundamental theorem of calculus despite the fact that $F'(x) = 0$ for a.e. x.
4. There are in fact many known examples of continuous increasing functions whose derivative is zero almost everywhere, see [15,17,22] and the references therein and also see Problem 3.5.40 on p. 109 of Folland for a simple example. Regarding the fact that this behavior is "typical" among the continuous increasing functions, see [24].

25.4.2 Functions of Bounded Variation

Our next goal is to prove an analogue of Theorem 25.19 for complex valued F. Let $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$ be fixed. The following notation will be used throughout this section.

Notation 25.21 *Let $(X, \mathcal{B})$ denote one of the following four measure spaces: $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, $\left((-\infty,\beta], \mathcal{B}_{(-\infty,\beta)}\right)$, $\left((\alpha,\infty), \mathcal{B}_{(\alpha,\infty)}\right)$ or $\left((\alpha,\beta], \mathcal{B}_{(\alpha,\beta]}\right)$ and let $\bar{X}$ denote the closure of X in $\mathbb{R}$ and $\bar{X}_\infty$ denote the closure of X in $\bar{\mathbb{R}} := [-\infty,\infty]$.*[1] *We further let $\mathcal{A}$ denote the algebra of half open intervals in X, i.e. the algebra generated by the sets, $\{(a,b] \cap X : -\infty \leq a \leq b \leq \infty\}$. Also let $\mathcal{A}_b$ be those $A \in \mathcal{A}$ which are bounded.*

Definition 25.22. *For $-\infty \leq a < b \leq \infty$, a partition $\mathbb{P}$ of $[a,b] \cap \bar{X}$ is a finite subset of $[a,b] \cap \bar{X}$ such that $\{a,b\} \cap \bar{X} \subset \mathbb{P}$. For $x \in [\min \mathbb{P}, \max \mathbb{P}]$, let*

$$x_+ = x_+^{\mathbb{P}} = \min\{y \in \mathbb{P} : y > x\} \wedge \max \mathbb{P}$$

where $\min \emptyset := \infty$.

For example, if $X = (\alpha,\infty)$, then a partition of $\bar{X} = [\alpha,\infty)$ is a finite subset, $\mathbb{P}$, of $[\alpha,\infty)$ such that $\alpha \in \mathbb{P}$ and if $\alpha \leq a < b < \infty$, then a partition of $[a,b]$ is a finite subset, $\mathbb{P}$, of $[a,b]$ such that $a, b \in \mathbb{P}$, see Figure 25.4.

The following proposition will help motivate a number of concepts which will need to introduce.

Proposition 25.23. *Suppose ν is a complex measure on $(X,\mathcal{B})$ and $F : \bar{X} \to \mathbb{C}$ is a function*

$$\nu((a,b]) = F(b) - F(a)$$

[1] So $\bar{X}$ is either $\mathbb{R}$, $(-\infty,\beta]$, $[\alpha,\infty)$, or $[\alpha,\beta]$ respectively and $\bar{X}_\infty$ is either $[-\infty,\infty]$, $[-\infty,\beta]$, $[\alpha,\infty]$, or $[\alpha,\beta]$ respectively.

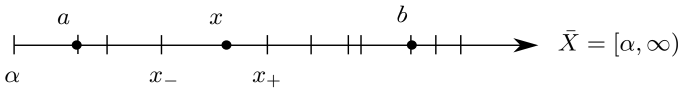


Fig. 25.4. In this figure, $X = (\alpha, \infty)$ and partitions of X and $[a, b]$ with $[a, b] \subset \bar{X}$ have been shown by the vertical lines. The meaing of x_+ is also depicted.

for all $a, b \in \bar{X}$ with $a < b$.(For example one may let $F(x) := \nu((-\infty, x] \cap X))$.) Then

1. *$F : \bar{X} \to \mathbb{C}$ is a right continuous function,*
2. *For all $a, b \in \bar{X}$ with $a < b$,*
$$|\nu|(a, b] = \sup_{\mathbb{P}} \sum_{x \in \mathbb{P}} |\nu(x, x_+]| = \sup_{\mathbb{P}} \sum_{x \in \mathbb{P}} |F(x_+) - F(x)| \tag{25.9}$$
where supremum is over all partitions $\mathbb{P}$ of $[a, b]$.
3. *If $\inf X = -\infty$ then Eq. (25.9) remains valid for $a = -\infty$ and moreover,*
$$|\nu|((-\infty, b]) = \lim_{a \to -\infty} |\nu|(a, b]. \tag{25.10}$$
Similar statements hold in case $\sup X = +\infty$ in which case we may take $b = \infty$ above. In particular if $X = \mathbb{R}$, then
$$|\nu|(\mathbb{R}) = \sup \left\{ \sum_{x \in \mathbb{P}} |F(x_+) - F(x)| : \mathbb{P} \text{ is a partition of } \mathbb{R} \right\}$$
$$= \lim_{\substack{a \to -\infty \\ b \to \infty}} \sup \left\{ \sum_{x \in \mathbb{P}} |F(x_+) - F(x)| : \mathbb{P} \text{ is a partition of } [a, b] \right\}.$$
4. *$\nu \ll m$ on X iff for all $\varepsilon > 0$ there exists $\delta > 0$ such that*
$$\sum_{i=1}^{n} |\nu((a_i, b_i])| = \sum_{i=1}^{n} |F(b_i) - F(a_i)| < \varepsilon \tag{25.11}$$
whenever $\{(a_i, b_i]\}_{i=1}^n$ are disjoint subintervals of X such that $\sum_{i=1}^{n} (b_i - a_i) < \delta$.

Proof. 1. The right continuity of F is a consequence of the continuity of ν under decreasing limits of sets.

2 and 3. When $a, b \in \bar{X}$, Eq. (25.9) follows from Proposition 24.32 and the fact that $\mathcal{B} = \sigma(\mathcal{A})$. The verification of item 3. is left for Exercise 25.1.

4. Equation (25.11) is a consequence of Theorem 24.37 and the following remarks:

a) $\{(a_i, b_i) \cap X\}_{i=1}^n$ are disjoint intervals iff $\{(a_i, b_i] \cap X\}_{i=1}^n$ are disjoint intervals,
b) $m(X \cap (\cup_{i=1}^n (a_i, b_i])) \le \sum_{i=1}^{n} (b_i - a_i)$, and
c) the general element $A \in \mathcal{A}_b$ is of the form $A = X \cap \left(\coprod_{i=1}^n (a_i, b_i]\right)$.

■

Exercise 25.1. Prove Item 3. of Proposition 25.23.

Definition 25.24 (Total variation of a function). *The **total variation** of a function $F : \bar{X} \to \mathbb{C}$ on $(a, b] \cap X \subset \bar{X}_\infty$ ($b = \infty$ is allowed here) is defined by*
$$T_F((a, b] \cap X) = \sup_{\mathbb{P}} \sum_{x \in \mathbb{P}} |F(x_+) - F(x)|$$
where supremum is over all partitions $\mathbb{P}$ of $[a, b] \cap X$. Also let
$$T_F(b) := T_F((\inf X, b]) \text{ for all } b \in X.$$
*The function F is said to have **bounded variation on** $(a, b] \cap X$ if $T_F((a, b] \cap X) < \infty$ and F is said to be of **bounded variation,** and we write $F \in BV(X)$, if $T_F(X) < \infty$.*

Definition 25.25 (Absolute continuity). *A function $F : \bar{X} \to \mathbb{C}$ is **absolutely continuous** if for all $\varepsilon > 0$ there exists $\delta > 0$ such that*
$$\sum_{i=1}^{n} |F(b_i) - F(a_i)| < \varepsilon \tag{25.12}$$
whenever $\{(a_i, b_i]\}_{i=1}^n$ are disjoint subintervals of X such that $\sum_{i=1}^{n} (b_i - a_i) < \delta$.

Exercise 25.2. Let $F, G : \bar{X} \to \mathbb{C}$ be and $\lambda \in \mathbb{C}$ be given. Show

1. $T_{F+G} \le T_F + T_G$ and $T_{\lambda F} = |\lambda| T_F$. Conclude from this that $BV(X)$ is a vector space.
2. $T_{\operatorname{Re} F} \le T_F$, $T_{\operatorname{Im} F} \le T_F$, and $T_F \le T_{\operatorname{Re} F} + T_{\operatorname{Im} F}$. In particular $F \in BV(X)$ iff $\operatorname{Re} F$ and $\operatorname{Im} F$ are in $BV(X)$.
3. If $F : \bar{X} \to \mathbb{C}$ is absolutely continuous then $F : \bar{X} \to \mathbb{C}$ is continuous and in fact is uniformly continuous.

Lemma 25.26 (Examples). *Let $F : \bar{X} \to \mathbb{F}$ be given, where $\mathbb{F}$ is either $\mathbb{R}$ of $\mathbb{C}$.*

1. *If $F:\bar{X}\to\mathbb{R}$ is a monotone function, then $T_F((a,b])=|F(b)-F(a)|$ for all $a,b\in\bar{X}$ with $a<b$. So $F\in BV(X)$ iff F is bounded (which will be the case if $X=[\alpha,\beta]$).*
2. *If $F:[\alpha,\beta]\to\mathbb{C}$ is absolutely continuous then $F\in BV((\alpha,\beta])$.*
3. *If $F\in C([\alpha,\beta]\to\mathbb{R})$, $F'(x)$ is differentiable for* ***all*** *$x\in(\alpha,\beta)$, and $\sup_{x\in(\alpha,\beta)}|F'(x)|=M<\infty$, then F is absolutely continuous[2] and*
$$T_F((a,b])\le M(b-a)\ \forall\ \alpha\le a<b\le\beta.$$
4. *Let $f\in L^1(X,m)$ and set*
$$F(x)=\int_{(-\infty,x]\cap\bar{X}}fdm \text{ for all } x\in\bar{X}. \tag{25.13}$$
Then $F:\bar{X}\to\mathbb{C}$ is absolutely continuous.

Proof.

1. If F is monotone increasing and $\mathbb{P}$ is a partition of $(a,b]$ then
$$\sum_{x\in\mathbb{P}}|F(x_+)-F(x)|=\sum_{x\in\mathbb{P}}(F(x_+)-F(x))=F(b)-F(a)$$
so that $T_F((a,b])=F(b)-F(a)$. Similarly, one shows
$$T_F((a,b])=F(a)-F(b)=|F(b)-F(a)|$$
if F is monotone decreasing. Also note that $F\in BV(\mathbb{R})$ iff $|F(\infty)-F(-\infty)|<\infty$, where $F(\pm\infty)=\lim_{x\to\pm\infty}F(x)$.
2. Since F is absolutely continuous, there exists $\delta>0$ such that whenever $a,b\in\bar{X}$ with $a<b$ and $b-a<\delta$, then
$$\sum_{x\in\mathbb{P}}|F(x_+)-F(x)|\le 1$$
for all partitions, $\mathbb{P}$, of $[a,b]$. This shows that $T_F((a,b])\le 1$ for all $a<b$ with $b-a<\delta$. Thus using Eq. (25.15), it follows that $T_F((a,b])\le N<\infty$ provided $N\in\mathbb{N}$ is chosen so that $b-a<N\delta$.
3. Suppose that $\{(a_i,b_i]\}_{i=1}^n$ are disjoint subintervals of $(a,b]$, then by the mean value theorem,
$$\begin{aligned}\sum_{i=1}^n|F(b_i)-F(a_i)|&\le\sum_{i=1}^n|F'(c_i)|(b_i-a_i)\le M\cdot m\left(\cup_{i=1}^n(a_i,b_i)\right)\\&\le M\sum_{i=1}^n(b_i-a_i)\le M(b-a)\end{aligned}$$
form which it easily follows that F is absolutely continuous. Moreover we may conclude that $T_F((a,b])\le M(b-a)$.
4. Let ν be the positive measure $d\nu=|f|\,dm$ on $(a,b]$. Again let $\{(a_i,b_i]\}_{i=1}^n$ be disjoint subintervals of $(a,b]$, then
$$\begin{aligned}\sum_{i=1}^n|F(b_i)-F(a_i)|&=\sum_{i=1}^n\left|\int_{(a_i,b_i]}fdm\right|\\&\le\sum_{i=1}^n\int_{(a_i,b_i]}|f|\,dm\\&=\int_{\cup_{i=1}^n(a_i,b_i]}|f|\,dm=\nu(\cup_{i=1}^n(a_i,b_i]).\end{aligned}\tag{25.14}$$
Since ν is absolutely continuous relative to m, by Theorem 24.35 (or Corollary 24.36 or Theorem24.37), for all $\varepsilon>0$ there exist $\delta>0$ such that $\nu(A)<\varepsilon$ if $m(A)<\delta$. Applying this result with $A=\cup_{i=1}^n(a_i,b_i]$, it follows from Eq. (25.14) that F satisfies the definition of being absolutely continuous. Furthermore, Eq. (25.14) also may be used to show
$$T_F((a,b])\le\int_{(a,b]}|f|\,dm.$$

■

[2] It is proved in Natanson or in Rudin that this is also true if $F\in C([\alpha,\beta])$ such that $F'(x)$ exists for **all** $x\in(\alpha,\beta)$ and $F'\in L^1([\alpha,\beta],m)$.

Example 25.27 (See I. P. Natanson,"Theory of functions of a real variable," p.269.). In each of the two examples below, $f\in C([-1,1])$.

1. Let $f(x)=|x|^{3/2}\sin\frac{1}{x}$ with $f(0)=0$, then f is everywhere differentiable but f' is not bounded near zero. However, f' is in $L^1([-1,1])$.
2. Let $f(x)=x^2\cos\frac{\pi}{x^2}$ with $f(0)=0$, then f is everywhere differentiable but $f'\notin L^1(-\varepsilon,\varepsilon)$ for any $\varepsilon\in(0,1)$. Indeed, if $0\notin(\alpha,\beta)$ then
$$\int_\alpha^\beta f'(x)dx=f(\beta)-f(\alpha)=\beta^2\cos\frac{\pi}{\beta^2}-\alpha^2\cos\frac{\pi}{\alpha^2}.$$
Now take $\alpha_n:=\sqrt{\frac{2}{4n+1}}$ and $\beta_n=1/\sqrt{2n}$. Then
$$\int_{\alpha_n}^{\beta_n}f'(x)dx=\frac{2}{4n+1}\cos\frac{\pi(4n+1)}{2}-\frac{1}{2n}\cos 2n\pi=\frac{1}{2n}$$
and noting that $\{(\alpha_n,\beta_n)\}_{n=1}^\infty$ are all disjoint, we find $\int_0^\varepsilon|f'(x)|\,dx=\infty$.

Theorem 25.28. *Let $F:\mathbb{R}\to\mathbb{C}$ be any function.*

1. *For $a < b < c$,*
$$T_F((a,c]) = T_F((a,b]) + T_F((b,c]). \tag{25.15}$$
Letting $a = \alpha$ in this expression implies
$$T_F(c) = T_F(b) + T_F((b,c]) \tag{25.16}$$
and in particular T_F is monotone increasing.
2. *Now suppose $F : \mathbb{R} \to \mathbb{R}$ and $F \in BV(\mathbb{R})$. Then the functions $F_\pm := (T_F \pm F)/2$ are bounded and increasing functions.*
3. *A function $F : \mathbb{R} \to \mathbb{R}$ is in BV iff $F = F_+ - F_-$ where $F_\pm$ are bounded increasing functions. In particular if $F \in BV(\mathbb{R})$, then $F(a+) := \lim_{y \downarrow a} F(y)$ exists for all $a \in \mathbb{R}$.*
4. *(Optional) If $F \in BV(\mathbb{R})$ and $a \in \mathbb{R}$, then*
$$T_F(a+) - T_F(a) \le \liminf_{y \downarrow a} |F(y) - F(a)|. \tag{25.17}$$

Proof.

1. (Item 1. is a special case of Exercise 24.6. Nevertheless we will give a proof here.) By the triangle inequality, if $\mathbb{P}$ and $\mathbb{P}'$ are partition of $[a,c]$ such that $\mathbb{P} \subset \mathbb{P}'$, then
$$\sum_{x \in \mathbb{P}} |F(x_+) - F(x)| \le \sum_{x \in \mathbb{P}'} |F(x_+) - F(x)|.$$
So if $\mathbb{P}$ is a partition of $[a,c]$, then $\mathbb{P} \subset \mathbb{P}' := \mathbb{P} \cup \{b\}$ implies
$$\begin{aligned}\sum_{x \in \mathbb{P}} |F(x_+) - F(x)| &\le \sum_{x \in \mathbb{P}'} |F(x_+) - F(x)| \\ &= \sum_{x \in \mathbb{P}' \cap (a,b]} |F(x_+) - F(x)| + \sum_{x \in \mathbb{P}' \cap [b,c]} |F(x_+) - F(x)| \\ &\le T_F((a,b]) + T_F((b,c]).\end{aligned}$$
Thus we see that
$$T_F((a,c]) \le T_F((a,b]) + T_F((b,c]).$$
Similarly if $\mathbb{P}_1$ is a partition of $[a,b]$ and $\mathbb{P}_2$ is a partition of $[b,c]$, then $\mathbb{P} = \mathbb{P}_1 \cup \mathbb{P}_2$ is a partition of $[a,c]$ and
$$\sum_{x \in \mathbb{P}_1} |F(x_+) - F(x)| + \sum_{x \in \mathbb{P}_2} |F(x_+) - F(x)| = \sum_{x \in \mathbb{P}} |F(x_+) - F(x)| \le T_F((a,c]).$$
From this we conclude
$$T_F((a,b]) + T_F((b,c]) \le T_F((a,c])$$
which finishes the proof of Eqs. (25.15) and (25.16).
2. By Item 1., for all $a < b$,
$$T_F(b) - T_F(a) = T_F((a,b]) \ge |F(b) - F(a)| \tag{25.18}$$
and therefore
$$T_F(b) \pm F(b) \ge T_F(a) \pm F(a)$$
which shows that $F_\pm$ are increasing. Moreover from Eq. (25.18), for $b \ge 0$ and $a \le 0$,
$$\begin{aligned}|F(b)| &\le |F(b) - F(0)| + |F(0)| \le T_F(0,b] + |F(0)| \\ &\le T_F(0,\infty) + |F(0)|\end{aligned}$$
and similarly
$$|F(a)| \le |F(0)| + T_F(-\infty, 0)$$
which shows that F is bounded by $|F(0)| + T_F(\mathbb{R})$. Therefore the functions, F_+ and F_- are bounded as well.
3. By Exercise 25.2 if $F = F_+ - F_-$, then
$$\begin{aligned}T_F((a,b]) &\le T_{F_+}((a,b]) + T_{F_-}((a,b]) \\ &= |F_+(b) - F_+(a)| + |F_-(b) - F_-(a)|\end{aligned}$$
which is bounded showing that $F \in BV$. Conversely if F is bounded variation, then $F = F_+ - F_-$ where $F_\pm$ are defined as in Item 1.
4. Choose some $b > a$. Then for any $\varepsilon > 0$ we may choose a partition $\mathbb{P}$ of $[a,b]$ such that
$$T_F(b) - T_F(a) = T_F((a,b]) \le \sum_{x \in \mathbb{P}} |F(x_+) - F(x)| + \varepsilon. \tag{25.19}$$
Let $y \in (a, a_+)$, then
$$\begin{aligned}\sum_{x \in \mathbb{P}} |F(x_+) - F(x)| + \varepsilon &\le \sum_{x \in \mathbb{P} \cup \{y\}} |F(x_+) - F(x)| + \varepsilon \\ &= |F(y) - F(a)| + \sum_{x \in \mathbb{P} \setminus \{y\}} |F(x_+) - F(x)| + \varepsilon \\ &\le |F(y) - F(a)| + T_F((y,b]) + \varepsilon. \end{aligned} \tag{25.20}$$
Combining Eqs. (25.19) and (25.20) shows
$$T_F((a,b]) = T_F(b) - T_F(a) \le |F(y) - F(a)| + T_F((y,b]) + \varepsilon$$
or equivalently that

$$T_F(y) - T_F(a) = T_F((a, y]) \le |F(y) - F(a)| + \varepsilon.$$

Since $y \in (a, a_+)$ is arbitrary we conclude that

$$T_F(a+) - T_F(a) = \liminf_{y \downarrow a} T_F(y) - T_F(a) \le \liminf_{y \downarrow a} |F(y) - F(a)| + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary this proves Eq. (25.17). ∎

Theorem 25.29 (Bounded variation functions). *Suppose $F : \bar{X} \to \mathbb{C}$ is in $BV(X)$, then*

1. *$F(x+) := \lim_{y \downarrow x} F(y)$ and $F(x-) := \lim_{y \uparrow x} F(y)$ exist for all $x \in \bar{X}$. By convention, if $X \subset (\alpha, \infty]$ then $F(\alpha-) = F(\alpha)$ and if $X \subset (-\infty, \beta]$ then $F(\beta+) := F(\beta)$. Let $G(x) := F(x+)$ and $G(\pm\infty) = F(\pm\infty)$ where appropriate.*
2. *If $\inf X = -\infty$, then $F(-\infty) := \lim_{x \to -\infty} F(x)$ exists and if $\sup X = +\infty$ then $F(\infty) := \lim_{x \to \infty} F(x)$ exists.*
3. *The set of points of discontinuity, $\{x \in X : \lim_{y \to x} F(y) \ne F(x)\}$, of F is at most countable and in particular $G(x) = F(x+)$ for all but a countable number $x \in X$.*
4. *For m – a.e. x, $F'(x)$ and $G'(x)$ exist and $F'(x) = G'(x)$.*
5. *The function G is right continuous on X. Moreover, there exists a unique complex measure, $\nu = \nu_F$, on $(X, \mathcal{B})$ such that, for all $a, b \in \bar{X}$ with $a < b$,*
$$\nu((a, b]) = G(b) - G(a) = F(b+) - F(a+). \tag{25.21}$$
6. *$F' \in L^1(X, m)$ and the Lebesgue decomposition of ν may be written as*
$$d\nu_F = F' dm + d\nu_s \tag{25.22}$$
where ν_s is a measure singular to m. In particular,
$$G(b) - G(a) = F(b+) - F(a+) = \int_a^b F' dm + \nu_s((a, b]) \tag{25.23}$$
whenever $a, b \in \bar{X}$ with $a < b$.
7. *$\nu_s = 0$ iff G is absolutely continuous[3] on $\bar{X}$.*

Proof. If $X \ne \mathbb{R}$, extend F to all of $\mathbb{R}$ by requiring F be constant on each of the connected components of $\mathbb{R} \setminus X^o$. For example if $X = [\alpha, \beta]$, extend F to $\mathbb{R}$ by setting $F(x) := F(\alpha)$ for $x \le \alpha$ and $F(x) = F(\beta)$ for $x \ge \beta$. With this extension it is easily seen that $T_F(\mathbb{R}) = T_F(X)$ and $T_F(x)$ is constant on the connected components of $\mathbb{R} \setminus X^o$. Thus we may now assume $X = \mathbb{R}$ and $T_F(\mathbb{R}) < \infty$. Moreover, by considering the real and imaginary parts of F separately we may assume F is real valued. So we now assume $X = \mathbb{R}$ and $F : \mathbb{R} \to \mathbb{R}$ is in $BV := BV(\mathbb{R})$.

1. – 4. By Theorem 25.28, the functions $F_\pm := (T_F \pm F)/2$ are bounded and increasing functions. Since $F = F_+ - F_-$, items 1. – 4. are now easy consequences of Theorem 25.19 applies to F_+ and F_-.

5. Let $G_\pm(x) := F_\pm(x+)$ and $G_\pm(\infty) = F_\pm(\infty)$ and $G_\pm(-\infty) = F_\pm(-\infty)$, then

$$G(x) = F(x+) = G_+(x) - G_-(x)$$

and as in Theorem 8.33 (or Theorems 25.19 38.26), there exists unique positive finite measures, $\nu_\pm$, such that

$$\nu_\pm((a, b]) = G_\pm(b) - G_\pm(a) \text{ for all } a < b.$$

Then $\nu := \nu_+ - \nu_-$ is a finite signed measure with the property that

$$\nu((a, b]) = G(b) - G(a) = F(b+) - F(a+) \text{ for all } a < b.$$

We will prove the uniqueness of the measure ν below.

6. Since $\nu_\pm$ have Lebesgue decompositions given by

$$d\nu_\pm = F'_\pm dm + d(\nu_\pm)_s$$

with $F'_\pm \in L^1(m)$ and $(\nu_\pm)_s \perp m$, it follows that

$$d\nu = \left(F'_+ - F'_-\right) dm + d\nu_s = F' dm + d\nu_s$$

with $F' = F'_+ - F'_-$ (m -a.e.), $F' \in L^1(\mathbb{R}, m)$ and $\nu_s \perp m$, where

$$\nu_s := (\nu_+)_s - (\nu_-)_s.$$

7. This is a consequence of Theorems 24.26 (or 24.33) and 24.37. Alternatively, if $\nu_s = 0$ then G is absolutely continuous from Eq. (25.23) and item 4. of Lemma 25.26. For the converse direction assume that G is absolutely continuous and $A \in \mathcal{B}_\mathbb{R}$ such that $m(A) = 0$. By regularity of m and $|\nu|$ we can find a decreasing sequence of open sets $\{U_j\}_{j=1}^\infty$ such that $m(U_j \setminus A) \to 0$ and $|\nu|(U_j \setminus A) \to 0$ as $j \to \infty$ and therefore $m(U_j) \to m(A) = 0$ as $j \to \infty$ and

$$|\nu(U_j) - \nu(A)| = |\nu(U_j \setminus A)| \le |\nu|(U_j \setminus A) \to 0 \text{ as } j \to \infty.$$

For given j, $U_j = \cup_{n=1}^N J_n$ where $N \in \mathbb{N} \cup \{\infty\}$ and $\{J_n = (a_n, b_n)\}_{n=1}^N$ are disjoint open intervals.[4] For $K \le N$ with $K < \infty$ we have $m\left(\cup_{n=1}^K J_n\right) \le m(U_j) =: \delta_j$ and so from the Definition 25.25 of absolute continuity of G

[3] We can not say that F is absolutely continuous here as can be seen by taking $F(x) = 1_{\{0\}}(x)$.

[4] This is the content of Exercise 35.22. For completeness let me sketch the proof here.

$$\left|\nu\left(\cup_{n=1}^K J_n\right)\right| = \left|\sum_{n=1}^K \left[G\left(b_n\right) - G\left(a_n\right)\right]\right| \le \sum_{n=1}^K \left|G\left(b_n\right) - G\left(a_n\right)\right| < \varepsilon_j \quad (25.24)$$

where ε_j is some sequence of positive numbers such that $\varepsilon_j \downarrow 0$. Letting $K \to \infty$ in Eq. (25.24) shows $|\nu(U_j)| \le \varepsilon_j$ for each j and then letting $j \to \infty$ shows $|\nu(A)| = 0$. Thus we have shown $\nu \ll m$ which implies $\nu_s = 0$ as $\nu_s \perp m$.

Uniqueness of ν. It now only remains to prove that ν satisfying Eq. (25.21) is unique. Suppose that $\tilde{\nu}$ is another such measure such that Eq. (25.21) holds with ν replaced by $\tilde{\nu}$. Then for $(a,b] \subset \mathbb{R}$,

$$|\nu|((a,b]) = \sup_{\mathbb{P}} \sum_{x\in\mathbb{P}} |G(x_+) - G(x)| = |\tilde{\nu}|((a,b])$$

where the supremum is over all partition of $[a,b]$. This shows that $|\nu| = |\tilde{\nu}|$ on $\mathcal{A} \subset \mathcal{B}$ and so by the measure uniqueness Theorem **??**, $|\nu| = |\tilde{\nu}|$ on $\mathcal{B}$. It now follows that $|\nu| + \nu$ and $|\tilde{\nu}| + \tilde{\nu}$ are finite positive measure on $\mathcal{B}$ such that, for all $a < b$,

$$\begin{aligned}(|\nu| + \nu)((a,b]) &= |\nu|((a,b]) + (G(b) - G(a)) \\ &= |\tilde{\nu}|((a,b]) + (G(b) - G(a)) \\ &= (|\tilde{\nu}| + \tilde{\nu})((a,b]).\end{aligned}$$

Hence another application of Theorem **??** shows

$$|\nu| + \nu = |\tilde{\nu}| + \tilde{\nu} = |\nu| + \tilde{\nu} \text{ on } \mathcal{B},$$

and hence $\nu = \tilde{\nu}$ on $\mathcal{B}$.

Alternative proofs of uniqueness of ν. The uniqueness may be proved by any number of other means. For example one may apply the multiplicative system Theorem 11.2 as follows. Let $-\infty < \alpha < \beta < \infty$ be given and take $\mathcal{H}$ to be the collection of bounded real measurable functions on $(\alpha,\beta]$ such that $\int_{(\alpha,\beta]} f d\nu = \int_{(\alpha,\beta]} f d\tilde{\nu}$ and M being the multiplicative system,

$$M := \left\{1_{(a,b]} : \alpha \le a < b \le \beta\right\}.$$

Then it follows from Theorem 11.2 that $\int_{(\alpha,\beta]} f d\nu = \int_{(\alpha,\beta]} f d\tilde{\nu}$ for all bounded measurable functions on $(\alpha,\beta]$ so that $\nu = \tilde{\nu}$ on $\mathcal{B}_{(\alpha,\beta]}$. As simple limiting argument then shows that $\nu = \tilde{\nu}$ on $\mathcal{B}_{\mathbb{R}}$.

Alternatively one could apply the monotone class Theorem (Lemma **??**) with $\mathcal{C} := \{A \in \mathcal{B} : \nu(A) = \tilde{\nu}(A)\}$ and $\mathcal{A}$ the algebra of half open intervals. Or one could use the $\pi - \lambda$ Theorem **??**, with $\mathcal{D} = \{A \in \mathcal{B} : \nu(A) = \tilde{\nu}(A)\}$ and $\mathcal{C} := \{(a,b] : a,b \in \mathbb{R} \text{ with } a < b\}$. ■

Corollary 25.30. *If $F \in BV(X)$ then $\nu_F \perp m$ iff $F' = 0$ m-a.e.*

Proof. This is a consequence of Eq. (25.22) and the uniqueness of the Lebesgue decomposition. In more detail, if $F'(x) = 0$ for m-a.e. x, then by Eq. (25.22), $\nu_F = \nu_s \perp m$. If $\nu_F \perp m$, then by Eq. (25.22), $F'dm = d\nu_F - d\nu_s \perp dm$ and by Lemma 24.8 $F'dm = 0$, i.e. $F' = 0$ m -a.e. ■

Corollary 25.31. *Let $F : \bar{X} \to \mathbb{C}$ be a right continuous function in $BV(X)$, ν_F be the associated complex measure and*

$$d\nu_F = F'dm + d\nu_s \quad (25.25)$$

be the its Lebesgue decomposition. Then the following are equivalent,

1. *F is absolutely continuous,*
2. *$\nu_F \ll m$,*
3. *$\nu_s = 0$, and*
4. *for all $a,b \in X$ with $a < b$,*

$$F(b) - F(a) = \int_{(a,b]} F'(x)dm(x). \quad (25.26)$$

Proof. The equivalence of **1.** and **2.** was established in Proposition 25.23 and the equivalence of **2.** and **3.** is trivial. (If $\nu_F \ll m$, then $d\nu_s = d\nu_F - F'dm \ll dm$ which implies, by Lemma 24.25, that $\nu_s = 0$.) If $\nu_F \ll m$ and $G(x) := F(x+)$, then the identity,

$$F(b) - F(a) = F(b+) - F(a-) = \int_a^b F'(x)\, dm(x),$$

implies F is continuous.

(The equivalence of **4.** and **1.**, **2.**, and **3.)** If F is absolutely continuous, then $\nu_s = 0$ and Eq. (25.26) follows from Eq. (25.25). Conversely let

$$\rho(A) := \int_A F'(x)dm(x) \text{ for all } A \in \mathcal{B}.$$

Recall by the Radon - Nikodym theorem that $\int_{\mathbb{R}} |F'(x)|\, dm(x) < \infty$ so that ρ is a complex measure on $\mathcal{B}$. So if Eq. (25.26) holds, then $\rho = \nu_F$ on the algebra generated by half open intervals. Therefore $\rho = \nu_F$ as in the uniqueness part of the proof of Theorem 25.29. Therefore $d\nu_F = F'dm \ll dm$. ■

For $x \in V$, let $a_x := \inf\{a : (a,x] \subset V\}$ and $b_x := \sup\{b : [x,b) \subset V\}$. Since V is open, $a_x < x < b_x$ and it is easily seen that $J_x := (a_x, b_x) \subset V$. Moreover if $y \in V$ and $J_x \cap J_y \ne \emptyset$, then $J_x = J_y$. The collection, $\{J_x : x \in V\}$, is at most countable since we may label each $J \in \{J_x : x \in V\}$ by choosing a rational number $r \in J$. Letting $\{J_n : n < N\}$, with $N = \infty$ allowed, be an enumeration of $\{J_x : x \in V\}$, we have $V = \coprod_{n<N} J_n$ as desired.

Theorem 25.32 (The fundamental theorem of calculus). *Suppose that $F : [\alpha, \beta] \to \mathbb{C}$ is a measurable function. Then the following are equivalent:*

1. *F is absolutely continuous on $[\alpha, \beta]$.*
2. *There exists $f \in L^1([\alpha, \beta]), dm)$ such that*

$$F(x) - F(\alpha) = \int_\alpha^x f dm \ \forall x \in [\alpha, \beta] \tag{25.27}$$

3. *F' exists a.e., $F' \in L^1([\alpha, \beta], dm)$ and*

$$F(x) - F(\alpha) = \int_\alpha^x F' dm \, \forall x \in [\alpha, \beta]. \tag{25.28}$$

Moreover if F is given as in Eq. (25.27), then $F' = f$ a.e.

Proof. 1. $\Longrightarrow$ 3. If F is absolutely continuous then $F \in BV((\alpha, \beta])$ and F is continuous on $[\alpha, \beta]$. Hence Eq. (25.28) holds by Corollary 25.31. The assertion 3. $\Longrightarrow$ 2. is trivial and we have already seen in Lemma 25.26 that 2. implies 1. The last assertion follows from Theorem 25.3 or can be seen comparing Eqs. (25.27) with (25.28) in which case,

$$\int_\alpha^x f dm = \int_\alpha^x F' dm \, \forall x \in [\alpha, \beta].$$

This only happens if $f = F'$ a.e. by a simple multiplicative systems theorem argument. ■

Corollary 25.33 (Integration by parts). *Suppose $-\infty < \alpha < \beta < \infty$ and $F, G : [\alpha, \beta] \to \mathbb{C}$ are two absolutely continuous functions. Then*

$$\int_\alpha^\beta F' G dm = -\int_\alpha^\beta F G' dm + FG|_\alpha^\beta.$$

Proof. Suppose that $\{(a_i, b_i]\}_{i=1}^n$ is a sequence of disjoint intervals in $[\alpha, \beta]$, then

$$\begin{aligned}\sum_{i=1}^n &|F(b_i)G(b_i) - F(a_i)G(a_i)| \\ &\le \sum_{i=1}^n |F(b_i)| \, |G(b_i) - G(a_i)| + \sum_{i=1}^n |F(b_i) - F(a_i)| \, |G(a_i)| \\ &\le \|F\|_\infty \sum_{i=1}^n |G(b_i) - G(a_i)| + \|G\|_\infty \sum_{i=1}^n |F(b_i) - F(a_i)| \, .\end{aligned}$$

From this inequality, one easily deduces the absolutely continuity of the product FG from the absolutely continuity of F and G. Therefore,

$$FG|_\alpha^\beta = \int_\alpha^\beta (FG)' dm = \int_\alpha^\beta (F'G + FG') dm.$$

■

25.4.3 Alternative method to Proving Theorem 25.29

For simplicity assume that $\alpha = -\infty$, $\beta = \infty$, $F \in BV$,

$$\mathcal{A}^b := \{A \in \mathcal{A} : A \text{ is bounded}\},$$

and $\mathcal{S}_c(\mathcal{A})$ denote simple functions of the form $f = \sum_{i=1}^n \lambda_i 1_{A_i}$ with $A_i \in \mathcal{A}^b$. Let $\nu^0 = \nu_F^0$ be the finitely additive set function on such that $\nu^0((a, b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$. As in the case of an increasing function F (see Lemma **??** and the text preceding it) we may define a linear functional, $I_F : \mathcal{S}_c(\mathcal{A}) \to \mathbb{C}$, by

$$I_F(f) = \sum_{\lambda \in \mathbb{C}} \lambda \nu^0(f = \lambda).$$

If we write $f = \sum_{i=1}^N \lambda_i 1_{(a_i, b_i]}$ with $\{(a_i, b_i]\}_{i=1}^N$ pairwise disjoint subsets of $\mathcal{A}^b$ inside $(a, b]$ we learn

$$|I_F(f)| = \left| \sum_{i=1}^N \lambda_i (F(b_i) - F(a_i) \right| \le \sum_{i=1}^N |\lambda_i| \, |F(b_i) - F(a_i)| \le \|f\|_\infty T_F((a, b]). \tag{25.29}$$

In the usual way this estimate allows us to extend I_F to the those compactly supported functions, $\overline{\mathcal{S}_c(\mathcal{A})}$, in the closure of $\mathcal{S}_c(\mathcal{A})$. As usual we will still denote the extension of I_F to $\overline{\mathcal{S}_c(\mathcal{A})}$ by I_F and recall that $\overline{\mathcal{S}_c(\mathcal{A})}$ contains $C_c(\mathbb{R}, \mathbb{C})$. The estimate in Eq. (25.29) still holds for this extension and in particular we have

$$|I(f)| \le T_F(\infty) \cdot \|f\|_\infty \text{ for all } f \in C_c(\mathbb{R}, \mathbb{C}).$$

Therefore I extends uniquely by continuity to an element of $C_0(\mathbb{R}, \mathbb{C})^*$. So by appealing to the complex Riesz Theorem (Corollary **??**) there exists a unique complex measure $\nu = \nu_F$ such that

$$I_F(f) = \int_{\mathbb{R}} f d\nu \text{ for all } f \in C_c(\mathbb{R}). \tag{25.30}$$

This leads to the following theorem.

Theorem 25.34. *To each function $F \in BV$ there exists a unique measure $\nu = \nu_F$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that Eq. (25.30) holds. Moreover, $F(x+) = \lim_{y \downarrow x} F(y)$ exists for all $x \in \mathbb{R}$ and the measure ν satisfies*

$$\nu((a,b]) = F(b+) - F(a+) \text{ for all } -\infty < a < b < \infty. \tag{25.31}$$

Remark 25.35. By applying Theorem 25.34 to the function $x \to F(-x)$ one shows every $F \in BV$ has left hand limits as well, i.e. $F(x-) = \lim_{y \uparrow x} F(y)$ exists for all $x \in \mathbb{R}$.

Proof. We must still prove $F(x+)$ exists for all $x \in \mathbb{R}$ and Eq. (25.31) holds. To prove let ψ_b and φ_ε be the functions shown in Figure 25.5 below. The reader should check that $\psi_b \in \overline{\mathcal{S}_c(\mathcal{A})}$. Notice that

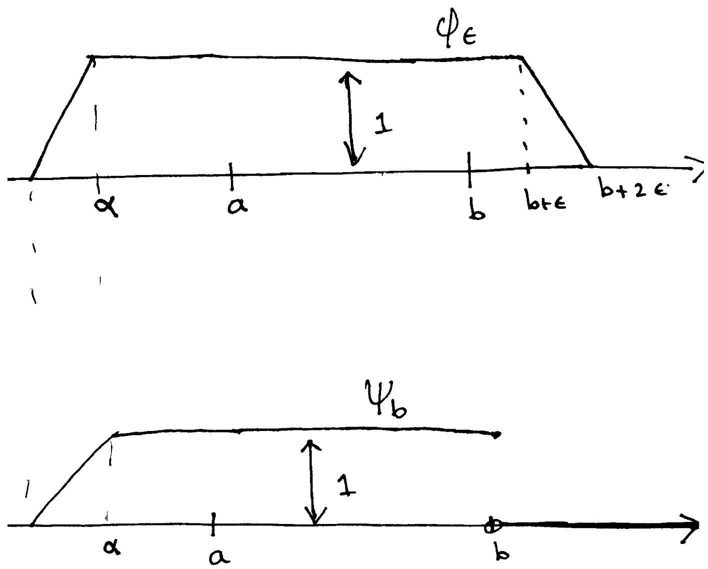


Fig. 25.5. A couple of functions in $\overline{\mathcal{S}_c(\mathcal{A})}$.

$$I_F(\psi_{b+\varepsilon}) = I_F(\psi_\alpha + 1_{(\alpha, b+\varepsilon]}) = I_F(\psi_\alpha) + F(b+\varepsilon) - F(\alpha)$$

and since $\|\varphi_\varepsilon - \psi_{b+\varepsilon}\|_\infty = 1$,

$$\begin{aligned}|I(\varphi_\varepsilon) - I_F(\psi_{b+\varepsilon})| &= |I_F(\varphi_\varepsilon - \psi_{b+\varepsilon})| \\ &\le T_F([b+\varepsilon, b+2\varepsilon]) = T_F(b+2\varepsilon) - T_F(b+\varepsilon),\end{aligned}$$

which implies $O(\varepsilon) := I(\varphi_\varepsilon) - I_F(\psi_{b+\varepsilon}) \to 0$ as $\varepsilon \downarrow 0$ because T_F is monotonic. Therefore,

$$\begin{aligned}I(\varphi_\varepsilon) &= I_F(\psi_{b+\varepsilon}) + I(\varphi_\varepsilon) - I_F(\psi_{b+\varepsilon}) \\ &= I_F(\psi_\alpha) + F(b+\varepsilon) - F(\alpha) + O(\varepsilon).\end{aligned} \tag{25.32}$$

Because φ_ε converges boundedly to ψ_b as $\varepsilon \downarrow 0$, the dominated convergence theorem implies

$$\lim_{\varepsilon \downarrow 0} I(\varphi_\varepsilon) = \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \varphi_\varepsilon d\nu = \int_{\mathbb{R}} \psi_b d\nu = \int_{\mathbb{R}} \psi_\alpha d\nu + \nu((\alpha, b]).$$

So we may let $\varepsilon \downarrow 0$ in Eq. (25.32) to learn $F(b+)$ exists and

$$\int_{\mathbb{R}} \psi_\alpha d\nu + \nu((\alpha, b]) = I_F(\psi_\alpha) + F(b+) - F(\alpha).$$

Similarly this equation holds with b replaced by a, i.e.

$$\int_{\mathbb{R}} \psi_\alpha d\nu + \nu((\alpha, a]) = I_F(\psi_\alpha) + F(a+) - F(\alpha).$$

Subtracting the last two equations proves Eq. (25.31). ■

Remark 25.36. Given Theorem 25.34 we may now prove Theorem 25.29 in the same we proved Theorem 25.19.

25.5 The connection of Weak and pointwise derivatives

Theorem 25.37. *Suppose Let $\Omega \subset \mathbb{R}$ be an open interval and $f \in L^1_{loc}(\Omega)$. Then there exists a complex measure μ on $\mathcal{B}_\Omega$ such that*

$$-\langle f, \varphi' \rangle = \mu(\varphi) := \int_\Omega \varphi d\mu \text{ for all } \varphi \in C_c^\infty(\Omega) \tag{25.33}$$

iff there exists a right continuous function F of bounded variation such that $F = f$ a.e. In this case $\mu = \mu_F$, i.e. $\mu((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.

Proof. Suppose $f = F$ a.e. where F is as above and let $\mu = \mu_F$ be the associated measure on $\mathcal{B}_\Omega$. Let $G(t) = F(t) - F(-\infty) = \mu((-\infty, t])$, then using Fubini's theorem and the fundamental theorem of calculus,

$$\begin{aligned}-\langle f, \varphi' \rangle = -\langle F, \varphi' \rangle = -\langle G, \varphi' \rangle &= -\int_\Omega \varphi'(t) \left[\int_\Omega 1_{(-\infty,t]}(s) d\mu(s)\right] dt \\ &= -\int_\Omega \int_\Omega \varphi'(t) 1_{(-\infty,t]}(s) dt d\mu(s) = \int_\Omega \varphi(s) d\mu(s) = \mu(\varphi).\end{aligned}$$

Conversely if Eq. (25.33) holds for some measure μ, let $F(t) := \mu((-\infty, t])$ then working backwards from above,

$$-\langle f, \varphi' \rangle = \mu(\varphi) = \int_\Omega \varphi(s) d\mu(s) = -\int_\Omega \int_\Omega \varphi'(t) 1_{(-\infty,t]}(s) dt d\mu(s)$$
$$= -\int_\Omega \varphi'(t) F(t) dt.$$

This shows $\partial^{(w)}(f - F) = 0$ and therefore by Proposition 41.25, $f = F + c$ a.e. for some constant $c \in \mathbb{C}$. Since $F + c$ is right continuous with bounded variation, the proof is complete. ■

Proposition 25.38. *Let $\Omega \subset \mathbb{R}$ be an open interval and $f \in L^1_{loc}(\Omega)$. Then $\partial^w f$ exists in $L^1_{loc}(\Omega)$ iff f has a continuous version $\tilde{f}$ which is absolutely continuous on all compact subintervals of Ω. Moreover, $\partial^w f = \tilde{f}'$ a.e., where $\tilde{f}'(x)$ is the usual pointwise derivative.*

Proof. If f is locally absolutely continuous and $\varphi \in C_c^\infty(\Omega)$ with $\operatorname{supp}(\varphi) \subset [a, b] \subset \Omega$, then by integration by parts, Corollary 25.33,

$$\int_\Omega f' \varphi dm = \int_a^b f' \varphi dm = -\int_a^b f \varphi' dm + f\varphi|_a^b = -\int_\Omega f \varphi' dm.$$

This shows $\partial^w f$ exists and $\partial^w f = f' \in L^1_{loc}(\Omega)$. Now suppose that $\partial^w f$ exists in $L^1_{loc}(\Omega)$ and $a \in \Omega$. Define $F \in C(\Omega)$ by $F(x) := \int_a^x \partial^w f(y) dy$. Then F is absolutely continuous on compacts and therefore by fundamental theorem of calculus for absolutely continuous functions (Theorem 25.32), $F'(x)$ exists and is equal to $\partial^w f(x)$ for a.e. $x \in \Omega$. Moreover, by the first part of the argument, $\partial^w F$ exists and $\partial^w F = \partial^w f$, and so by Proposition 41.25 there is a constant c such that

$$\tilde{f}(x) := F(x) + c = f(x) \text{ for a.e. } x \in \Omega.$$

■

Definition 25.39. *Let X and Y be metric spaces. A function $u : X \to Y$ is said to be* ***Lipschitz*** *if there exists $C < \infty$ such that*

$$d^Y(u(x), u(x')) \le C d^X(x, x') \text{ for all } x, x' \in X$$

and said to be ***locally Lipschitz*** *if for all compact subsets $K \subset X$ there exists $C_K < \infty$ such that*

$$d^Y(u(x), u(x')) \le C_K d^X(x, x') \text{ for all } x, x' \in K.$$

Proposition 25.40 (Rademacher's theorem). *Let $u \in L^1_{loc}(\Omega)$. Then there exists a locally Lipschitz function $\tilde{u} : \Omega \to \mathbb{C}$ such that $\tilde{u} = u$ a.e. iff $(weak - \partial_i u) \in L^1_{loc}(\Omega)$ exists and is locally (essentially) bounded for $i = 1, 2, \ldots, d$.*

Proof. Suppose $u = \tilde{u}$ a.e. and $\tilde{u}$ is Lipschitz and let $p \in (1, \infty)$ and V be a precompact open set such that $\bar{V} \subset W$ and let $V_\varepsilon := \left\{x \in \Omega : \operatorname{dist}(x, \bar{V}) \le \varepsilon\right\}$. Then for $\varepsilon < \operatorname{dist}(\bar{V}, \Omega^c)$, $V_\varepsilon \subset \Omega$ and therefore there is constant $C(V, \varepsilon) < \infty$ such that $|\tilde{u}(y) - \tilde{u}(x)| \le C(V, \varepsilon) |y - x|$ for all $x, y \in V_\varepsilon$. So for $0 < |h| \le 1$ and $v \in \mathbb{R}^d$ with $|v| = 1$,

$$\int_V \left| \frac{u(x + hv) - u(x)}{h} \right|^p dx = \int_V \left| \frac{\tilde{u}(x + hv) - \tilde{u}(x)}{h} \right|^p dx \le C(V, \varepsilon) |v|^p.$$

Therefore Theorem 41.18 may be applied to conclude $\partial_v u$ exists in L^p and moreover,

$$\lim_{h \to 0} \frac{\tilde{u}(x + hv) - \tilde{u}(x)}{h} = \partial_v u(x) \text{ for } m - \text{a.e. } x \in V.$$

Since there exists $\{h_n\}_{n=1}^\infty \subset \mathbb{R} \setminus \{0\}$ such that $\lim_{n \to \infty} h_n = 0$ and

$$|\partial_v u(x)| = \lim_{n \to \infty} \left| \frac{\tilde{u}(x + h_n v) - \tilde{u}(x)}{h_n} \right| \le C(V) \text{ for a.e. } x \in V,$$

it follows that $\|\partial_v u\|_\infty \le C(V)$ where $C(V) := \lim_{\varepsilon \downarrow 0} C(V, \varepsilon)$.

Conversely, let $\Omega_\varepsilon := \{x \in \Omega : \operatorname{dist}(x, \Omega^c) > \varepsilon\}$ and $\eta \in C_c^\infty(B(0, 1), [0, \infty))$ such that $\int_{\mathbb{R}^n} \eta(x) dx = 1$, $\eta_m(x) = m^n \eta(mx)$ and $u_m := u * \eta_m$ as in the proof of Theorem 41.18. Suppose $V \subset_o \Omega$ with $\bar{V} \subset \Omega$ and ε is sufficiently small. Then $u_m \in C^\infty(\Omega_\varepsilon)$, $\partial_v u_m = \partial_v u * \eta_m$, $|\partial_v u_m(x)| \le \|\partial_v u\|_{L^\infty(V_{m^{-1}})} =: C(V, m) < \infty$ and therefore for $x, y \in \bar{V}$ with $|y - x| \le \varepsilon$,

$$\begin{aligned} |u_m(y) - u_m(x)| &= \left| \int_0^1 \frac{d}{dt} u_m(x + t(y - x)) dt \right| \\ &= \left| \int_0^1 (y - x) \cdot \nabla u_m(x + t(y - x)) dt \right| \\ &\le \int_0^1 |y - x| \cdot |\nabla u_m(x + t(y - x))| dt \le C(V, m) |y - x|. \end{aligned} \tag{25.34}$$

By passing to a subsequence if necessary, we may assume that $\lim_{m \to \infty} u_m(x) = u(x)$ for m − a.e. $x \in \bar{V}$ and then letting $m \to \infty$ in Eq. (25.34) implies

$$|u(y) - u(x)| \le C(V) |y - x| \text{ for all } x, y \in \bar{V} \setminus E \text{ and } |y - x| \le \varepsilon \tag{25.35}$$

where $E \subset \bar{V}$ is a $m-$ null set. Define $\tilde{u}_V : \bar{V} \to \mathbb{C}$ by $\tilde{u}_V = u$ on $\bar{V} \setminus E$ and $\tilde{u}_V(x) = \lim_{y \to x, y \notin E} u(y)$ if $x \in E$. Then clearly $\tilde{u}_V = u$ a.e. on $\bar{V}$ and it is easy to show $\tilde{u}_V$ is well defined and $\tilde{u}_V : \bar{V} \to \mathbb{C}$ is continuous and still satisfies

$$|\tilde{u}_V(y) - \tilde{u}_V(x)| \le C_V |y-x| \text{ for } x, y \in \bar{V} \text{ with } |y-x| \le \varepsilon.$$

Since $\tilde{u}_V$ is continuous on $\bar{V}$ there exists $M_V < \infty$ such that $|\tilde{u}_V| \le M_V$ on $\bar{V}$. Hence if $x, y \in \bar{V}$ with $|x-y| \ge \varepsilon$, we find

$$\frac{|\tilde{u}_V(y) - \tilde{u}_V(x)|}{|y-x|} \le \frac{2M}{\varepsilon}$$

and hence

$$|\tilde{u}_V(y) - \tilde{u}_V(x)| \le \max\left\{C_V, \frac{2M_V}{\varepsilon}\right\} |y-x| \text{ for } x, y \in \bar{V}$$

showing $\tilde{u}_V$ is Lipschitz on $\bar{V}$. To complete the proof, choose precompact open sets V_n such that $V_n \subset \bar{V}_n \subset V_{n+1} \subset \Omega$ for all n and for $x \in V_n$ let $\tilde{u}(x) := \tilde{u}_{V_n}(x)$.

Alternative way to construct the function $\tilde{u}_V$. For $x \in V \setminus E$,

$$\begin{aligned} |u_m(x) - u(x)| &= \left|\int_V u(x-y)\eta(my)m^n dy - u(x)\right| = \left|\int_V [u(x-y/m) - u(x)]\,\eta(y)dy\right| \\ &\le \int_V |u(x-y/m) - u(x)|\,\eta(y)dy \le \frac{C}{m}\int_V |y|\,\eta(y)dy \end{aligned}$$

wherein the last equality we have used Eq. (25.35) with V replaced by V_ε for some small $\varepsilon > 0$. Letting $K := C\int_V |y|\,\eta(y)dy < \infty$ we have shown

$$\|u_m - u\|_\infty \le K/m \to 0 \text{ as } m \to \infty$$

and consequently

$$\|u_m - u_n\|_\infty = \|u_m - u_n\|_\infty \le 2K/m \to 0 \text{ as } m \to \infty.$$

Therefore, u_n converges uniformly to a continuous function $\tilde{u}_V$. ■

The next theorem is from Chapter 1. of Maz'ja [16].

Theorem 25.41. *Let $p \ge 1$ and Ω be an open subset of $\mathbb{R}^d$, $x \in \mathbb{R}^d$ be written as $x = (y,t) \in \mathbb{R}^{d-1} \times \mathbb{R}$,*

$$Y := \left\{y \in \mathbb{R}^{d-1} : (\{y\} \times \mathbb{R}) \cap \Omega \ne \emptyset\right\}$$

and $u \in L^p(\Omega)$. Then $\partial_t u$ exists weakly in $L^p(\Omega)$ iff there is a version $\tilde{u}$ of u such that for a.e. $y \in Y$ the function $t \to \tilde{u}(y,t)$ is absolutely continuous, $\partial_t u(y,t) = \frac{\partial \tilde{u}(y,t)}{\partial t}$ a.e., and $\left\|\frac{\partial \tilde{u}}{\partial t}\right\|_{L^p(\Omega)} < \infty$.

Proof. For the proof of Theorem 25.41, it suffices to consider the case where $\Omega = (0,1)^d$. Write $x \in \Omega$ as $x = (y,t) \in Y \times (0,1) = (0,1)^{d-1} \times (0,1)$ and $\partial_t u$ for the weak derivative $\partial_{e_d} u$. By assumption

$$\int_\Omega |\partial_t u(y,t)|\,dydt = \|\partial_t u\|_1 \le \|\partial_t u\|_p < \infty$$

and so by Fubini's theorem there exists a set of full measure, $Y_0 \subset Y$, such that

$$\int_0^1 |\partial_t u(y,t)|\,dt < \infty \text{ for } y \in Y_0.$$

So for $y \in Y_0$, the function $v(y,t) := \int_0^t \partial_t u(y,\tau)d\tau$ is well defined and absolutely continuous in t with $\frac{\partial}{\partial t}v(y,t) = \partial_t u(y,t)$ for a.e. $t \in (0,1)$. Let $\xi \in C_c^\infty(Y)$ and $\eta \in C_c^\infty((0,1))$, then integration by parts for absolutely functions implies

$$\int_0^1 v(y,t)\dot{\eta}(t)dt = -\int_0^1 \frac{\partial}{\partial t}v(y,t)\eta(t)dt \text{ for all } y \in Y_0.$$

Multiplying both sides of this equation by $\xi(y)$ and integrating in y shows

$$\begin{aligned} \int_\Omega v(x)\dot{\eta}(t)\xi(y)dydt &= -\int_\Omega \frac{\partial}{\partial t}v(y,t)\eta(t)\xi(y)dydt \\ &= -\int_\Omega \partial_t u(y,t)\eta(t)\xi(y)dydt. \end{aligned}$$

Using the definition of the weak derivative, this equation may be written as

$$\int_\Omega u(x)\dot{\eta}(t)\xi(y)dydt = -\int_\Omega \partial_t u(x)\eta(t)\xi(y)dydt$$

and comparing the last two equations shows

$$\int_\Omega [v(x) - u(x)]\,\dot{\eta}(t)\xi(y)dydt = 0.$$

Since $\xi \in C_c^\infty(Y)$ is arbitrary, this implies there exists a set $Y_1 \subset Y_0$ of full measure such that

$$\int_\Omega [v(y,t) - u(y,t)]\,\dot{\eta}(t)dt = 0 \text{ for all } y \in Y_1$$

from which we conclude, using Proposition 41.25, that $u(y,t) = v(y,t) + C(y)$ for $t \in J_y$ where $m_{d-1}(J_y) = 1$, here m_k denotes $k-$ dimensional Lebesgue measure. In conclusion we have shown that

$$u(y,t) = \tilde{u}(y,t) := \int_0^t \partial_t u(y,\tau)d\tau + C(y) \text{ for all } y \in Y_1 \text{ and } t \in J_y. \quad (25.36)$$

We can be more precise about the formula for $\tilde{u}(y,t)$ by integrating both sides of Eq. (25.36) on t we learn

$$\begin{aligned} C(y) &= \int_0^1 dt \int_0^t \partial_\tau u(y,\tau)d\tau - \int_0^1 u(y,t)dt \\ &= \int_0^1 (1-\tau)\,\partial_\tau u(y,\tau)d\tau - \int_0^1 u(y,t)dt \\ &= \int_0^1 \left[(1-t)\,\partial_t u(y,t) - u(y,t)\right] dt \end{aligned}$$

and hence

$$\tilde{u}(y,t) := \int_0^t \partial_\tau u(y,\tau)d\tau + \int_0^1 \left[(1-\tau)\,\partial_\tau u(y,\tau) - u(y,\tau)\right] d\tau$$

which is well defined for $y \in Y_0$. For the converse suppose that such a $\tilde{u}$ exists, then for $\varphi \in C_c^\infty(\Omega)$,

$$\begin{aligned} \int_\Omega u(y,t)\partial_t\varphi(y,t)dydt &= \int_\Omega \tilde{u}(y,t)\partial_t\varphi(y,t)dtdy \\ &= -\int_\Omega \frac{\partial\tilde{u}(y,t)}{\partial t}\varphi(y,t)dtdy \end{aligned}$$

wherein we have used integration by parts for absolutely continuous functions. From this equation we learn the weak derivative $\partial_t u(y,t)$ exists and is given by $\frac{\partial\tilde{u}(y,t)}{\partial t}$ a.e. ■

25.6 Exercises

Exercise 25.3. Folland 3.22 on p. 100.

Exercise 25.4. Folland 3.24 on p. 100.

Exercise 25.5. Folland 3.25 on p. 100.

Exercise 25.6. Folland 3.27 on p. 107.

Exercise 25.7 (Look at but do not hand in). Folland 3.28 on p. 107.

Exercise 25.8 (Look at but do not hand in). Folland 3.29 on p. 107.

Exercise 25.9. Folland 3.30 on p. 107.

Exercise 25.10. Folland 3.33 on p. 108.

Exercise 25.11. Folland 3.35 on p. 108.

Exercise 25.12. Folland 3.37 on p. 108.

Exercise 25.13. Folland 3.39 on p. 108.

Exercise 25.14 (Look at but do not hand in.). Folland 3.40 on p. 108.

Exercise 25.15 (Folland 8.4 on p. 239). If $f \in L^\infty(\mathbb{R}^n, m)$ and $\|\tau_y f - f\|_\infty \to 0$ as $|y| \to \infty$, then f agrees a.e. with a uniformly continous function. (See hints in the book.)

Exercise 25.16 (Global Integration by Parts Formula). Suppose that $f, g : \mathbb{R} \to \mathbb{R}$ are locally absolutely continuous functions[5] such that $f'g$, fg', and fg are all Lebesgue integrable functions on $\mathbb{R}$. Prove the following integration by parts formula;

$$\int_{\mathbb{R}} f'(x) \cdot g(x)\,dx = -\int_{\mathbb{R}} f(x) \cdot g'(x)\,dx. \quad (25.37)$$

Similarly show that; if $f, g : [0,\infty) \to [0,\infty)$ are locally absolutely continuous functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $[0,\infty)$, then

$$\int_0^\infty f'(x) \cdot g(x)\,dx = -f(0)\,g(0) - \int_0^\infty f(x) \cdot g'(x)\,dx. \quad (25.38)$$

Outline: 1. First use the theory developed to see that Eq. (25.37) holds if $f(x) = 0$ for $|x| \geq N$ for some $N < \infty$.

2. Let $\psi : \mathbb{R} \to [0,1]$ be a continuously differentiable function such that $\psi(x) = 1$ if $|x| \leq 1$ and $\psi(x) = 0$ if $|x| \geq 2$.[6] For any $\varepsilon > 0$ let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ Write out the identity in Eq. (25.37) with $f(x)$ being replaced by $f(x)\,\psi_\varepsilon(x)$.

3. Now use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in the identity you found in step 2.

4. A similar outline works to prove Eq. (25.38).

[5] This means that f and g restricted to any bounded interval in $\mathbb{R}$ are absolutely continuous on that interval.

[6] You may assume the existence of such a ψ, we will deal with this later.

.

25.7 Summary of B.V. and A.C. Functions

We take for granted Theorem 25.44 below in this summary.

Notation 25.42 *In this chapter, let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n}$ denote the Borel σ – algebra on $\mathbb{R}^n$ and m be Lebesgue measure on $\mathcal{B}$. If V is an open subset of $\mathbb{R}^n$, let $L^1_{loc}(V) := L^1_{loc}(V, m)$ and simply write L^1_{loc} for $L^1_{loc}(\mathbb{R}^n)$. We will also write $|A|$ for $m(A)$ when $A \in \mathcal{B}$.*

Definition 25.43. *A collection of measurable sets $\{E\}_{r>0} \subset \mathcal{B}$ is said to shrink nicely to $x \in \mathbb{R}^n$ if (i) $E_r \subset \overline{B(x,r)}$ for all $r > 0$ and (ii) there exists $\alpha > 0$ such that $m(E_r) \geq \alpha m(B(x,r))$. We will abbreviate this by writing $E_r \downarrow \{x\}$ nicely. (Notice that it is not required that $x \in E_r$ for any $r > 0$.*

Theorem 25.44. *Suppose that ν is a complex measure on $(\mathbb{R}^n, \mathcal{B})$, then there exists $g \in L^1(\mathbb{R}^n, m)$ and a complex measure ν_s such that $\nu_s \perp m$, $d\nu = gdm + d\nu_s$, and for m - a.e. x,*

$$g(x) = \lim_{r \downarrow 0} \frac{\nu(E_r)}{m(E_r)} \text{ and} \tag{25.39}$$

$$0 = \lim_{r \downarrow 0} \frac{\nu_s(E_r)}{m(E_r)} \tag{25.40}$$

for any collection of $\{E_r\}_{r>0} \subset \mathcal{B}$ which shrink nicely to $\{x\}$. (Eq. (25.39) holds for all $x \in \mathcal{L}(g)$ – the Lebesgue set of g, see Definition 25.12 and Theorem 25.13.)

Exercise 25.17. Suppose that $(X, \mathcal{M})$ is a measurable space and μ and ν are complex measures on $\mathcal{M}$ which are singular to one another. Then $|\mu| \perp |\nu|$ and $|\mu + \nu| = |\mu| + |\nu|$.

Theorem 25.45 (Bounded variation functions). *Suppose [7] $F = U + iV : \mathbb{R} \to \mathbb{C}$ is in $BV(\mathbb{R})$. Then $U = U_+ - U_-$ and $V = V_+ - V_-$ where $U_\pm$ and $V_\pm$ are increasing functions.[8] Consequently $F(x+)$ and $F(x-)$ exists for all $x \in \mathbb{R}$. Let $G(x) = F(x+)$ be the right continuous version of F.*

1. *The set of points of discontinuity, $\{x \in X : \lim_{y \to x} F(y) \neq F(x)\}$, of F is at most countable and in particular $G(x) = F(x+)$ for all but a countable number $x \in X$.*
2. *Let $H(x) := G(x) - F(x) = F(x+) - F(x)$, then $\sum_x |H(x)| \leq T_F(\infty) < \infty$ and $H'(x) = 0$ for m – a.e. x.*

---

[7] For our purposes it suffices to assume there is some (large) $M > 0$ where $F|_{(-\infty,-M]}$ and $F|_{[M,\infty)}$ are constant functions.

[8] In fact, $U_\pm = \frac{T_U \pm U}{2}$ and $V_\pm = \frac{T_V \pm V}{2}$.

3. For m – a.e. x, $F'(x)$ and $G'(x)$ exist and $F'(x) = G'(x)$.

4. There exists a unique complex measure, $\nu = \nu_F$, on $(\mathbb{R}, \mathcal{B})$ such that, for all $a, b \in \mathbb{R}$ with $a < b$,

$$\nu((a, b]) = G(b) - G(a) = F(b+) - F(a+). \tag{25.41}$$

5. $F' \in L^1(\mathbb{R}, m)$ and the Lebesgue decomposition of ν may be written as

$$d\nu_F = F' dm + d\nu_s \tag{25.42}$$

where ν_s is a measure singular to m. In particular,

$$G(b) - G(a) = F(b+) - F(a+) = \int_a^b F' dm + \nu_s((a, b]) \tag{25.43}$$

whenever $a, b \in \mathbb{R}$ with $a < b$.
Remark: *from Exercise 25.17 we know that*

$$\infty > T_G(\infty) = |\nu_G|(\mathbb{R}) = \int_{\mathbb{R}} |F'| \, dm + |\nu_s|(\mathbb{R}).$$

Proof. 1. We may assume that F is real and increasing in which case one easily sees for any finite subset, $\Lambda \subset \mathbb{R}$, that

$$\sum_{x \in \Lambda} |F(x+) - F(x-)| = \lim_{\varepsilon \downarrow 0} \sum_{x \in \Lambda} |F(x + \varepsilon) - F(x - \varepsilon)| \leq T_F(\infty) < \infty$$

which implies,

$$\sum_{x \in \mathbb{R}} |F(x+) - F(x-)| \leq T_F(\infty) < \infty.$$

2. Similarly one sees that

$$\sum_{x \in \Lambda} |H(x)| = \lim_{\varepsilon \downarrow 0} \sum_{x \in \Lambda} |F(x + \varepsilon) - F(x)| \leq T_F(\infty) < \infty$$

so that $\sum_x |H(x)| \leq T_F(\infty) < \infty$. Because of this fact, $\lambda := \sum_{x \in \mathbb{R}} |H(x)| \delta_x$, i.e.

$$\lambda(A) := \sum_{x \in A} |H(x)| \text{ for all } A \in \mathcal{B}_{\mathbb{R}}$$

defines a positive Radon measure on $\mathcal{B}_{\mathbb{R}}$. Since $\lambda(\Lambda^c) = 0$ and $m(\Lambda) = 0$, the measure $\lambda \perp m$. By Corollary 25.16 for m - a.e. x,

$$\begin{aligned} \left| \frac{H(x+r) - H(x)}{r} \right| &\leq \frac{|H(x+r)| + |H(x)|}{|r|} \\ &\leq \frac{H(x+|r|) + H(x-|r|) + H(x)}{|r|} \\ &\leq 2 \frac{\lambda([x - |r|, x + |r|])}{2|r|} \end{aligned}$$

and the last term goes to zero as $r \to 0$ because $\{[x - r, x + r]\}_{r>0}$ shrinks nicely to $\{x\}$ as $r \downarrow 0$ and $m([x - |r|, x + |r|]) = 2|r|$. Hence we conclude for m – a.e. x that $H'(x) = 0$.

3.-5. The existence of ν follows from the increasing function case and the splitting of F already described. We then have by Theorem 25.44 that

$$d\nu = g dm + d\nu_s,$$

where $\nu_s \perp m$, $g \in L^1_{\text{loc}}(\mathbb{R}, m)$ with $g \in L^1(\mathbb{R}, m)$ if F is bounded. Moreover Theorem 25.44, for m - a.e. x,

$$g(x) = \lim_{r \downarrow 0} (\nu_G(E_r)/m(E_r)),$$

where $\{E_r\}_{r>0}$ is any collection of sets shrink nicely to $\{x\}$. Since $(x, x+r] \downarrow \{x\}$ and $(x - r, x] \downarrow \{x\}$ nicely,

$$g(x) = \lim_{r \downarrow 0} \frac{\nu_G(x, x + r])}{m((x, x + r])} = \lim_{r \downarrow 0} \frac{G(x + r) - G(x)}{r} = \frac{d}{dx^+} G(x) \tag{25.44}$$

and

$$\begin{aligned} g(x) &= \lim_{r \downarrow 0} \frac{\nu_G((x - r, x])}{m((x - r, x])} = \lim_{r \downarrow 0} \frac{G(x) - G(x - r)}{r} \\ &= \lim_{r \downarrow 0} \frac{G(x - r) - G(x)}{-r} = \frac{d}{dx^-} G(x) \end{aligned} \tag{25.45}$$

exist and are equal for m - a.e. x, i.e. $G'(x) = g(x)$ exists for m -a.e. x. Since $F = G - H$ with $H' = 0$ a.e. it follows that $F' = G'$ a.e. and Eqs. (25.42) and (25.43) are proved. ■

25.7.1 Absolute Continuity

Proposition 25.46. *Let μ be a positive measure on $(X, \mathcal{M})$ and $f \in L^1(d\mu)$ and $\nu(A) := \int_A f d\mu$. Then for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|\nu|(A) < \varepsilon$ for all $A \in \mathcal{M}$ such that $\mu(A) < \delta$. In particular if $(X, \mathcal{M}, \mu) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, m)$ and $G(x) := \int_{-\infty}^x f dm$ we have for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\sum_{i=1}^n |G(b_i) - G(a_i)| < \varepsilon$$

where ever $\{(a_i, b_i)\}_{i=1}^n$ are disjoint open intervals such that

$$m(\cup_{i=1}^n (a_i, b_i)) = \sum_{i=1}^n (b_i - a_i) < \delta.$$

Proof. If the statement in the corollary were false, there would exists $\varepsilon > 0$ and $A_n \in \mathcal{M}$ such that $\mu(A_n) \downarrow 0$ while $\left|\int_{A_n} f\, d\mu\right| \geq \varepsilon$ for all n. On the other hand $|1_{A_n} f| \leq |f| \in L^1(\mu)$ and $1_{A_n} f \overset{\mu}{\to} 0$ as $n \to \infty$ and so by the dominated convergence theorem in Corollary 16.21 we may conclude,

$$\lim_{n\to\infty} \int_{A_n} f\, d\mu = \lim_{n\to\infty} \int_X 1_{A_n} f\, d\mu = 0$$

which leads to the desired contradiction.

For the second assertion let $\mu = m$ and suppose that A is of the form, $A := \sum_{i=1}^n (a_i, b_i)$,[9] where $\{(a_i, b_i)\}_{i=1}^n$ are disjoint open bounded intervals of $\mathbb{R}$. Then if $\nu = \nu_G$ and $d\nu_G = f dm$, then (with ε and δ as above)

$$\sum_{i=1}^n |G(b_i) - G(a_i)| = \sum_{i=1}^n |\nu_G(a_i, b_i)| \leq \sum_{i=1}^n |\nu_G|(a_i, b_i) = |\nu_G|(A) < \varepsilon$$

provided

$$\sum_{i=1}^n (b_i - a_i) = m(A) < \delta.$$

■

Definition 25.47 (Absolute continuity). *A function $F : \bar{X} \to \mathbb{C}$ is* ***absolutely continuous*** *if for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\sum_{i=1}^n |F(b_i) - F(a_i)| < \varepsilon \tag{25.46}$$

whenever $\{(a_i, b_i]\}_{i=1}^n$ are disjoint subintervals of X such that $\sum_{i=1}^n (b_i - a_i) < \delta$.

Proposition 25.48. *If G is an absolutely continuous function (necessarily continuous with bounded variation), then $\nu_G \ll m$ where $\nu = \nu_G$ is the complex measure associated to G.*

Proof. Assume that $A \in \mathcal{B}_{\mathbb{R}}$ such that $m(A) = 0$. By regularity of m and $|\nu|$ we can find a decreasing sequence of open sets $\{U_j\}_{j=1}^\infty$ such that $m(U_j \setminus A) \to 0$ and $|\nu|(U_j \setminus A) \to 0$ as $j \to \infty$ and therefore $m(U_j) \to m(A) = 0$ as $j \to \infty$ and

$$|\nu(U_j) - \nu(A)| = |\nu(U_j \setminus A)| \leq |\nu|(U_j \setminus A) \to 0 \text{ as } j \to \infty.$$

For given j, $U_j = \cup_{n=1}^N J_n$ where $N \in \mathbb{N} \cup \{\infty\}$ and $\{J_n = (a_n, b_n)\}_{n=1}^N$ are disjoint open intervals.[10] For $K \leq N$ with $K < \infty$ we have $m\left(\cup_{n=1}^K J_n\right) \leq m(U_j) =: \delta_j$ and so from the Definition 25.47 of absolute continuity of G

$$\left|\nu\left(\cup_{n=1}^K J_n\right)\right| = \left|\sum_{n=1}^K [G(b_n) - G(a_n)]\right| \leq \sum_{n=1}^K |G(b_n) - G(a_n)| < \varepsilon_j \tag{25.47}$$

where ε_j is some sequence of positive numbers such that $\varepsilon_j \downarrow 0$. Letting $K \to \infty$ in Eq. (25.47) shows $|\nu(U_j)| \leq \varepsilon_j$ for each j and then letting $j \to \infty$ shows $|\nu(A)| = 0$. ■

Exercise 25.18. Let $F, G : \bar{X} \to \mathbb{C}$ be and $\lambda \in \mathbb{C}$ be given. Show

1. $T_{F+G} \leq T_F + T_G$ and $T_{\lambda F} = |\lambda| T_F$. Conclude from this that $BV(X)$ is a vector space.
2. $T_{\mathrm{Re}\, F} \leq T_F$, $T_{\mathrm{Im}\, F} \leq T_F$, and $T_F \leq T_{\mathrm{Re}\, F} + T_{\mathrm{Im}\, F}$. In particular $F \in BV(X)$ iff $\mathrm{Re}\, F$ and $\mathrm{Im}\, F$ are in $BV(X)$.
3. If $F : \bar{X} \to \mathbb{C}$ is absolutely continuous then $F : \bar{X} \to \mathbb{C}$ is continuous and in fact is uniformly continuous.

Lemma 25.49 (Examples). *Let $F : \bar{X} \to \mathbb{F}$ be given, where $\mathbb{F}$ is either $\mathbb{R}$ of $\mathbb{C}$.*

1. *If $F : \bar{X} \to \mathbb{R}$ is a monotone function, then $T_F((a, b]) = |F(b) - F(a)|$ for all $a, b \in \bar{X}$ with $a < b$. So $F \in BV(X)$ iff F is bounded (which will be the case if $X = [\alpha, \beta]$).*
2. *If $F : [\alpha, \beta] \to \mathbb{C}$ is absolutely continuous then $F \in BV((\alpha, \beta])$.*
3. *If $F \in C([\alpha, \beta] \to \mathbb{R})$, $F'(x)$ is differentiable for* ***all*** *$x \in (\alpha, \beta)$, and $\sup_{x \in (\alpha, \beta)} |F'(x)| = M < \infty$, then F is absolutely continuous*[11] *and*
$$T_F((a, b]) \leq M(b - a) \;\forall\; \alpha \leq a < b \leq \beta.$$
4. *Let $f \in L^1(X, m)$ and set*
$$F(x) = \int_{(-\infty, x] \cap \bar{X}} f dm \text{ for all } x \in \bar{X}. \tag{25.48}$$
Then $F : \bar{X} \to \mathbb{C}$ is absolutely continuous.

---

[9] The notation, $\sum_{i=1}^n A_i$, is short hand for $\cup_{i=1}^n A_i$ with the additional assumption that $A_i \cap A_j = \emptyset$ for all $i \neq j$.

[10] This is the content of Exercise 35.22. For completeness let me sketch the proof here. For $x \in V$, let $a_x := \inf\{a : (a, x] \subset V\}$ and $b_x := \sup\{b : [x, b) \subset V\}$. Since V is open, $a_x < x < b_x$ and it is easily seen that $J_x := (a_x, b_x) \subset V$. Moreover if $y \in V$ and $J_x \cap J_y \neq \emptyset$, then $J_x = J_y$. The collection, $\{J_x : x \in V\}$, is at most countable since we may label each $J \in \{J_x : x \in V\}$ by choosing a rational number $r \in J$. Letting $\{J_n : n < N\}$, with $N = \infty$ allowed, be an enumeration of $\{J_x : x \in V\}$, we have $V = \coprod_{n<N} J_n$ as desired.

[11] It is proved in Natanson or in Rudin that this is also true if $F \in C([\alpha, \beta])$ such that $F'(x)$ exists for **all** $x \in (\alpha, \beta)$ and $F' \in L^1([\alpha, \beta], m)$.

Proof.

1. If F is monotone increasing and $\mathbb{P}$ is a partition of $(a,b]$ then
$$\sum_{x\in\mathbb{P}} |F(x_+) - F(x)| = \sum_{x\in\mathbb{P}} (F(x_+) - F(x)) = F(b) - F(a)$$
so that $T_F((a,b]) = F(b) - F(a)$. Similarly, one shows
$$T_F((a,b]) = F(a) - F(b) = |F(b) - F(a)|$$
if F is monotone decreasing. Also note that $F \in BV(\mathbb{R})$ iff $|F(\infty) - F(-\infty)| < \infty$, where $F(\pm\infty) = \lim_{x\to\pm\infty} F(x)$.
2. Since F is absolutely continuous, there exists $\delta > 0$ such that whenever $a, b \in \bar{X}$ with $a < b$ and $b - a < \delta$, then
$$\sum_{x\in\mathbb{P}} |F(x_+) - F(x)| \le 1$$
for all partitions, $\mathbb{P}$, of $[a,b]$. This shows that $T_F((a,b]) \le 1$ for all $a < b$ with $b - a < \delta$. Thus using Eq. (25.15)[12], it follows that $T_F((a,b]) \le N < \infty$ provided $N \in \mathbb{N}$ is chosen so that $b - a < N\delta$.
3. Suppose that $\{(a_i, b_i]\}_{i=1}^n$ are disjoint subintervals of $(a,b]$, then by the mean value theorem,
$$\begin{aligned}\sum_{i=1}^n |F(b_i) - F(a_i)| &\le \sum_{i=1}^n |F'(c_i)| (b_i - a_i) \le M \cdot m\left(\cup_{i=1}^n (a_i, b_i)\right) \\ &\le M \sum_{i=1}^n (b_i - a_i) \le M(b-a)\end{aligned}$$
form which it easily follows that F is absolutely continuous. Moreover we may conclude that $T_F((a,b]) \le M(b-a)$.
4. Let ν be the positive measure $d\nu = |f|\, dm$ on $(a,b]$. Again let $\{(a_i, b_i]\}_{i=1}^n$ be disjoint subintervals of $(a,b]$, then
$$\begin{aligned}\sum_{i=1}^n |F(b_i) - F(a_i)| &= \sum_{i=1}^n \left|\int_{(a_i,b_i]} f dm\right| \\ &\le \sum_{i=1}^n \int_{(a_i,b_i]} |f|\, dm \\ &= \int_{\cup_{i=1}^n (a_i,b_i]} |f|\, dm = \nu(\cup_{i=1}^n (a_i, b_i]). \end{aligned} \tag{25.49}$$
Since ν is absolutely continuous relative to m, by Theorem 24.35 (or Corollary 24.36 or Theorem24.37), for all $\varepsilon > 0$ there exist $\delta > 0$ such that $\nu(A) < \varepsilon$ if $m(A) < \delta$. Applying this result with $A = \cup_{i=1}^n (a_i, b_i]$, it follows from Eq. (25.49) that F satisfies the definition of being absolutely continuous. Furthermore, Eq. (25.49) also may be used to show
$$T_F((a,b]) \le \int_{(a,b]} |f|\, dm.$$

■

Theorem 25.50 (Absolute Continuity). *Suppose $F \in BV(\mathbb{R})$, $G = F(x+)$, and ν_F be the associated complex measure and*
$$d\nu_F = F' dm + d\nu_s \tag{25.50}$$
be the its Lebesgue decomposition. Then the following are equivalent,

1. *G is absolutely continuous*[13]*,*
2. *$\nu_F \ll m$,*
3. *$\nu_s = 0$, and*
4. *for all $a, b \in X$ with $a < b$,*
$$G(b) - G(a) = \int_{(a,b]} F'(x) dm(x). \tag{25.51}$$

*(**Moral:** absolutely continuous functions are those functions which satisfy the fundamental theorem of calculus.)*

Proof. The equivalence of **1.** and **2.** was established in Proposition 25.23 or by Propositions 25.46 and 25.48. The equivalence of **2.** and **3.** is trivial. (If $\nu_F \ll m$, then $d\nu_s = d\nu_F - F' dm \ll dm$ which implies, by Lemma 24.25, that $\nu_s = 0$.)

[12] This equation states
$$T_F((a,c]) = T_F((a,b]) + T_F((b,c])$$
for $-\infty < a < b < c < \infty$.

[13] We can not say that F is absolutely continuous here as can be seen by taking $F(x) = 1_{\{0\}}(x)$.

(The equivalence of **4.** and **1.**, **2.**, and **3.**) If $\nu_F \ll m$ (i.e. $\nu_s = 0$), then the identity in Eq. (25.51) holds which also implies G is continuous. Conversely if Eq. (25.51) holds, then $\nu_G(A) = \int_A F'(x)dm(x)$ whenever A is a half open interval and we have seen (see Example 11.6)

$$\nu_G(A) = \int_A F'(x)dm(x) \text{ for all } A \in \mathcal{B}$$

which certainly implies $\nu_G \ll m$.

■

Theorem 25.51 (The fundamental theorem of calculus). *Suppose that $F : [\alpha, \beta] \to \mathbb{C}$ is a measurable function. Then the following are equivalent:*

1. *F is absolutely continuous on $[\alpha, \beta]$.*
2. *There exists $f \in L^1([\alpha, \beta]), dm)$ such that*

$$F(x) - F(\alpha) = \int_\alpha^x f dm \; \forall x \in [\alpha, \beta] \tag{25.52}$$

3. *F' exists a.e., $F' \in L^1([\alpha, \beta], dm)$ and*

$$F(x) - F(\alpha) = \int_\alpha^x F' dm \, \forall x \in [\alpha, \beta]. \tag{25.53}$$

Proof. 1. $\Longrightarrow$ 3. If F is absolutely continuous then $F \in BV((\alpha, \beta])$ and F is continuous on $[\alpha, \beta]$. Hence Eq. (25.53) holds by Corollary 25.31. The assertion 3. $\Longrightarrow$ 2. is trivial and we have already seen in Lemma 25.49 that 2. implies 1. ■

Corollary 25.52 (Integration by parts). *Suppose $-\infty < \alpha < \beta < \infty$ and $F, G : [\alpha, \beta] \to \mathbb{C}$ are two absolutely continuous functions. Then*

$$\int_\alpha^\beta F'Gdm = -\int_\alpha^\beta FG'dm + FG|_\alpha^\beta.$$

Proof. Suppose that $\{(a_i, b_i]\}_{i=1}^n$ is a sequence of disjoint intervals in $[\alpha, \beta]$, then

$$\begin{aligned}
&\sum_{i=1}^n |F(b_i)G(b_i) - F(a_i)G(a_i)| \\
&\quad \le \sum_{i=1}^n |F(b_i)| \, |G(b_i) - G(a_i)| + \sum_{i=1}^n |F(b_i) - F(a_i)| \, |G(a_i)| \\
&\quad \le \|F\|_\infty \sum_{i=1}^n |G(b_i) - G(a_i)| + \|G\|_\infty \sum_{i=1}^n |F(b_i) - F(a_i)| \, .
\end{aligned}$$

From this inequality, one easily deduces the absolutely continuity of the product FG from the absolutely continuity of F and G. Therefore,

$$FG|_\alpha^\beta = \int_\alpha^\beta (FG)'dm = \int_\alpha^\beta (F'G + FG')dm.$$

■

Geometric Integration

Definition 26.1. *A subset $M \subset \mathbb{R}^n$ is a $n-1$ dimensional C^k-**Hypersurface** if for all $x_0 \in M$ there exists $\varepsilon > 0$ an open set $0 \in D \subset \mathbb{R}^n$ and a C^k-diffeomorphism $\psi : D \to B(x_0, \varepsilon)$ such that $\psi(D \cap \{x_n = 0\}) = B(x_0, \varepsilon) \cap M$. See Figure 26.1 below.*

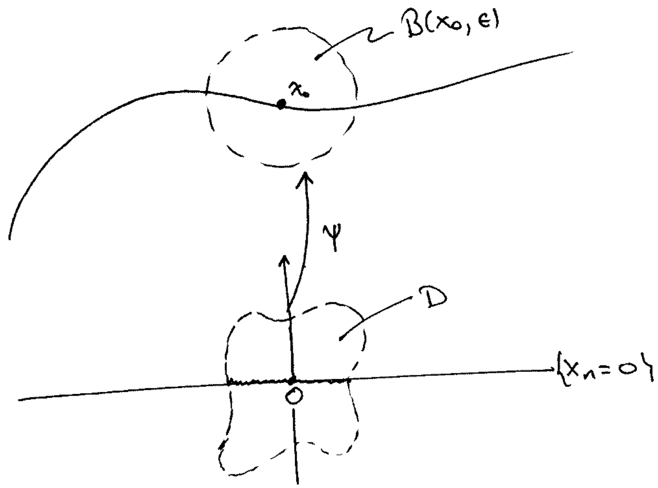


Fig. 26.1. An embedded submanifold of $\mathbb{R}^2$.

Example 26.2. Suppose $V \subset_0 \mathbb{R}^{n-1}$ and $g : V \xrightarrow{C^k} \mathbb{R}$. Then $M := \Gamma(g) = \{(y, g(y)) : y \in V\}$ is a C^k hypersurface. To verify this assertion, given $x_0 = (y_0, g(y_0)) \in \Gamma(g)$ define

$$\psi(y, z) := (y + y_0, g(y + y_0) - z).$$

Then $\psi : \{V - y_0\} \times \mathbb{R} \xrightarrow{C^k} V \times \mathbb{R}$ diffeomorphism

$$\psi((V - y_0) \times \{0\}) = \{(y + y_0, g(y + y_0)) : y \in V - y_0\} = \Gamma(g).$$

Proposition 26.3 (Parametrized Surfaces). *Let $k \geq 1$, $D \subset_0 \mathbb{R}^{n-1}$ and $\Sigma \in C^k(D, \mathbb{R}^n)$ satisfy*

1. *$\Sigma : D \to M := \Sigma(D)$ is a homeomorphism and*
2. *$\Sigma'(y) : \mathbb{R}^{n-1} \to \mathbb{R}^n$ is injective for all $y \in D$. (We will call M a C^k – **parametrized surface** and $\Sigma : D \to M$ a **parametrization of** M.)*

Then M is a C^k-hypersurface in $\mathbb{R}^n$. Moreover if $f \in C(W \subset_0 \mathbb{R}^d, \mathbb{R}^n)$ is a continuous function such that $f(W) \subset M$, then $f \in C^k(W, \mathbb{R}^n)$ iff $\Sigma^{-1} \circ f \in C^k(U, D)$.

Proof. Let $y_0 \in D$ and $x_0 = \Sigma(y_0)$ and n_0 be a normal vector to M at x_0, i.e. $n_0 \perp \operatorname{Ran}(\Sigma'(y_0))$, and let

$$\psi(t, y) := \Sigma(y_0 + y) + t\, n_0 \text{ for } t \in \mathbb{R} \text{ and } y \in D - y_0,$$

see Figure 26.2 below. Since $D_y\psi(0,0) = \Sigma'(y_0)$ and $\frac{\partial \psi}{\partial t}(0,0) = n_0 \notin$

Fig. 26.2. Showing a parametrized surface is an embedded hyper-surface.

$\operatorname{Ran}(\Sigma'(y_0))$, $\psi'(0,0)$ is invertible. so by the inverse function theorem there exists a neighborhood V of $(0,0) \in \mathbb{R}^n$ such that $\psi|_V : V \to \mathbb{R}^n$ is a C^k – diffeomorphism. Choose an $\varepsilon > 0$ such that $B(x_0, \varepsilon) \cap M \subset \Sigma(V \cap \{t = 0\})$ and $B(x_0, \varepsilon) \subset \psi(V)$. Then set $U := \psi^{-1}(B(x_0, \varepsilon))$. One finds $\psi|_U : U \to B(x_0, \varepsilon)$ has the desired properties. Now suppose $f \in C(W \subset_0 \mathbb{R}^d, \mathbb{R}^n)$ such that $f(W) \subset M$, $a \in W$ and $x_0 = f(a) \in M$. By shrinking W if necessary we may assume $f(W) \subset B(x_0, \varepsilon)$ where $B(x_0, \varepsilon)$ is the ball used previously. (This is where we used the continuity of f.) Then

$$\Sigma^{-1} \circ f = \pi \circ \psi^{-1} \circ f$$

where π is projection onto $\{t = 0\}$. Form this identity it clearly follows $\Sigma^{-1} \circ f$ is C^k if f is C^k. The converse is easier since if $\Sigma^{-1} \circ f$ is C^k then $f = \Sigma \circ (\Sigma^{-1} \circ f)$ is C^k as well. ■

26.1 Surface Integrals

Definition 26.4. *Suppose $\Sigma : D \subset_0 \mathbb{R}^{n-1} \to M \subset \mathbb{R}^n$ is a C^1- parameterized hypersurface of $\mathbb{R}^n$ and $f \in C_c(M, \mathbb{R})$. Then the surface integral of f over M, $\int_M f \, d\sigma$, is defined by*

$$\int_M f \, d\sigma = \int_D f \circ \Sigma(y) \left| \det[\frac{\partial \Sigma(y)}{\partial y_1}|, \dots, \frac{\partial \Sigma(y)}{\partial y_{n-1}}|n(y)] \right| dy$$
$$= \int_D f \circ \Sigma(y) \left| \det[\Sigma'(y) e_1 | \dots | \Sigma'(y) e_{n-1} | n(y)] \right| dy$$

where $n(y) \in \mathbb{R}^n$ is a unit normal vector perpendicular of $ran(\Sigma'(y))$ for each $y \in D$. We will abbreviate this formula by writing

$$d\sigma = \left| \det[\frac{\partial \Sigma(y)}{\partial y_1}|, \dots, \frac{\partial \Sigma(y)}{\partial y_{n-1}}|n(y)] \right| dy, \tag{26.1}$$

see Figure 26.3 below for the motivation.

Remark 26.5. Let $A = A(y) := [\Sigma'(y) e_1, \dots, \Sigma'(y) e_{n-1}, n(y)]$. Then

$$A^{\mathrm{tr}} A = \begin{bmatrix} \partial_1 \Sigma^t \\ \partial_2 \Sigma^t \\ \vdots \\ \partial_{n-1} \Sigma^t \\ n^t \end{bmatrix} [\partial_1 \Sigma | \dots | \partial_{n-1} \Sigma | n]$$
$$= \begin{bmatrix} \partial_1 \Sigma \cdot \partial_1 \Sigma & \partial_1 \Sigma \cdot \partial_2 \Sigma & \dots & \partial_1 \Sigma \cdot \partial_{n-1} \Sigma & 0 \\ \partial_2 \Sigma \cdot \partial_1 \Sigma & \partial_2 \Sigma \cdot \partial_2 \Sigma & \dots & \partial_2 \Sigma \cdot \partial_{n-1} \Sigma & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \partial_{n-1} \Sigma \cdot \partial_1 \Sigma & \partial_{n-1} \Sigma \cdot \partial_2 \Sigma & \dots & \partial_{n-1} \Sigma \cdot \partial_{n-1} \Sigma & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

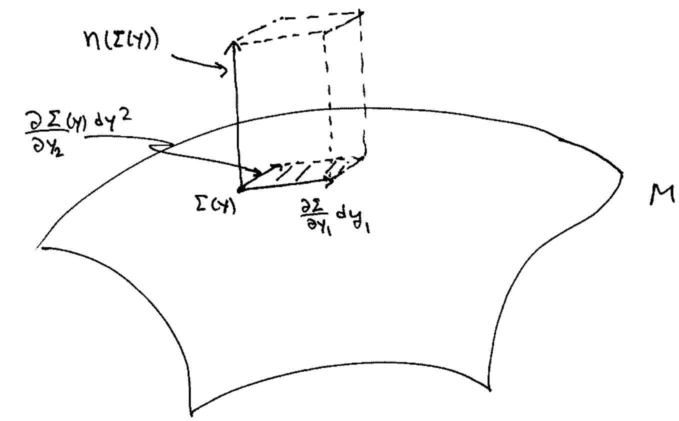


Fig. 26.3. The approximate area spanned by $\Sigma([y, y+dy])$ should be equal to the area spaced by $\frac{\partial \Sigma(y)}{\partial y_1} dy_1$ and $\frac{\partial \Sigma(y)}{\partial y_2} dy_2$ which is equal to the volume of the parallelepiped spanned by $\frac{\partial \Sigma(y)}{\partial y_1} dy_1$, $\frac{\partial \Sigma(y)}{\partial y_2} dy_2$ and $n(\Sigma(y))$ and hence the formula in Eq. (26.1).

and therefore

$$\left| \det[\frac{\partial \Sigma(y)}{\partial y_1}|, \dots, \frac{\partial \Sigma(y)}{\partial y_{n-1}}|n(y)] \right| = |\det(A)| \, dy = \sqrt{\det(A^{\mathrm{tr}} A)} dy$$
$$= \sqrt{\det\left[(\partial_i \Sigma \cdot \partial_j \Sigma)_{i,j=1}^{n-1}\right]}$$
$$= \sqrt{\det\left[(\Sigma')^{\mathrm{tr}} \Sigma'\right]}.$$

This implies $d\sigma = \rho^{\Sigma}(y) dy$ or more precisely that

$$\int_M f \, d\sigma = \int_D f \circ \Sigma(y) \rho^{\Sigma}(y) dy$$

where

$$\rho^{\Sigma}(y) := \sqrt{\det\left[(\partial_i \Sigma \cdot \partial_j \Sigma)_{i,j=1}^{n-1}\right]} = \sqrt{\det\left[(\Sigma')^{\mathrm{tr}} \Sigma'\right]}.$$

The next lemma shows that $\int_M f \, d\sigma$ is well defined, i.e. independent of how M is parametrized.

Example 26.6. Suppose $V \subset_0 \mathbb{R}^{n-1}$ and $g : V \xrightarrow{C^k} \mathbb{R}$ and $M := \Gamma(g) = \{(y, g(y)) : y \in V\}$ as in Example 26.2. We now compute $d\sigma$ in the parametrization $\Sigma : V \to M$ defined by $\Sigma(y) = (y, g(y))$. To simplify notation, let

$$\nabla g(y) := (\partial_1 g(y), \ldots, \partial_{n-1} g(y)).$$

As is standard from multivariable calculus (and is easily verified),

$$n(y) := \frac{(\nabla g(y), -1)}{\sqrt{1 + |\nabla g(y)|^2}}$$

is a normal vector to M at $\Sigma(y)$, i.e. $n(y) \cdot \partial_k \Sigma(y) = 0$ for all $k = 1, 2 \ldots, n-1$. Therefore,

$$\begin{aligned} d\sigma &= |\det[\partial_1 \Sigma| \ldots |\partial_{n-1}\Sigma|n]|\, dy \\ &= \frac{1}{\sqrt{1+|\nabla g(y)|^2}} \left| \det \begin{bmatrix} I_{n-1} & (\nabla g)^{tr} \\ \nabla g & -1 \end{bmatrix} \right| dy \\ &= \frac{1}{\sqrt{1+|\nabla g(y)|^2}} \left| \det \begin{bmatrix} I_{n-1} & 0 \\ \nabla g & -1-|\nabla g|^2 \end{bmatrix} \right| dy \\ &= \frac{1}{\sqrt{1+|\nabla g(y)|^2}} \left(1 + |\nabla g(y)|^2\right) dy = \sqrt{1+|\nabla g(y)|^2} dy. \end{aligned}$$

Hence if $g : M \to \mathbb{R}$, we have

$$\int_M g d\sigma = \int_V g(\Sigma(y)) \sqrt{1+|\nabla g(y)|^2} dy.$$

Example 26.7. Keeping the same notation as in Example 26.6, but now taking $V := B(0,r) \subset \mathbb{R}^{n-1}$ and $g(y) := \sqrt{r^2 - |y|^2}$. In this case $M = S_+^{n-1}$, the upper-hemisphere of S^{n-1}, $\nabla g(y) = -y/g(y)$,

$$d\sigma = \sqrt{1 + |y|^2 / g^2(y)} dy = \frac{r}{g(y)} dy$$

and so

$$\int_{S_+^{n-1}} g d\sigma = \int_{|y|<r} g(y, \sqrt{r^2 - |y|^2}) \frac{r}{\sqrt{r^2 - |y|^2}} dy.$$

A similar computation shows, with S_-^{n-1} being the lower hemisphere, that

$$\int_{S_-^{n-1}} g d\sigma = \int_{|y|<r} g(y, -\sqrt{r^2 - |y|^2}) \frac{r}{\sqrt{r^2 - |y|^2}} dy.$$

Lemma 26.8. *If $\tilde{\Sigma} : \tilde{D} \to M$ is another C^k – parametrization of M, then*

$$\int_D f \circ \Sigma(y) \rho^{\Sigma}(y) dy = \int_{\tilde{D}} f \circ \tilde{\Sigma}(y) \rho^{\tilde{\Sigma}}(y) dy.$$

Proof. By Proposition 26.3, $\varphi := \Sigma^{-1} \circ \tilde{\Sigma} : \tilde{D} \to D$ is a C^k – diffeomorphism. By the change of variables theorem on $\mathbb{R}^{n-1}$ with $y = \varphi(\tilde{y})$ (using $\tilde{\Sigma} = \Sigma \circ \varphi$, see Figure 26.4) we find

$$\begin{aligned} \int_{\tilde{D}} f \circ \tilde{\Sigma}(\tilde{y}) \rho^{\tilde{\Sigma}}(\tilde{y}) d\tilde{y} &= \int_{\tilde{D}} f \circ \tilde{\Sigma} \sqrt{\det\left(\tilde{\Sigma}'\right)^{\mathrm{tr}} \tilde{\Sigma}'} d\tilde{y} \\ &= \int_{\tilde{D}} f \circ \Sigma \circ \varphi \sqrt{\det (\Sigma \circ \varphi)'^{\mathrm{tr}} (\Sigma \circ \varphi)'} d\tilde{y} \\ &= \int_{\tilde{D}} f \circ \Sigma \circ \varphi \sqrt{\det\left[(\Sigma'(\varphi)\varphi')^{\mathrm{tr}} \Sigma'(\varphi)\varphi'\right]} d\tilde{y} \\ &= \int_{\tilde{D}} f \circ \Sigma \circ \varphi \sqrt{\det[\varphi'^{\mathrm{tr}} \left[\Sigma'(\varphi)^{\mathrm{tr}} \Sigma'(\varphi)\right] \varphi'} d\tilde{y} \\ &= \int_{\tilde{D}} (f \circ \Sigma \circ \varphi) \cdot \left(\sqrt{\det \Sigma'^{\mathrm{tr}} \Sigma'}\right) \circ \varphi \cdot |\det \varphi'|\, d\tilde{y} \\ &= \int_D f \circ \Sigma \sqrt{\det \Sigma'^{\mathrm{tr}} \Sigma'} dy. \end{aligned}$$

■

Definition 26.9. *Let M be a C^1-embedded hypersurface and $f \in C_c(M)$. Then we define the* ***surface integral of*** f ***over*** M *as*

$$\int_M f\, d\sigma = \sum_{i=1}^n \int_{M_i} \varphi_i f\, d\sigma$$

where $\varphi_i \in C_c^1(M, [0,1])$ are chosen so that $\sum_i \varphi_i \le 1$ with equality on $\mathrm{supp}(f)$ and the $\mathrm{supp}(\varphi_i f) \subset M_i \subset M$ where M_i is a subregion of M which may be viewed as a parametrized surface.

Remark 26.10. The integral $\int_M f\, d\sigma$ is well defined for if $\psi_j \in C_c^1(M, [0,1])$ is another sequence satisfying the properties of $\{\varphi_i\}$ with $\mathrm{supp}(\psi_j) \subset M_j' \subset M$ then (using Lemma 26.8 implicitly)

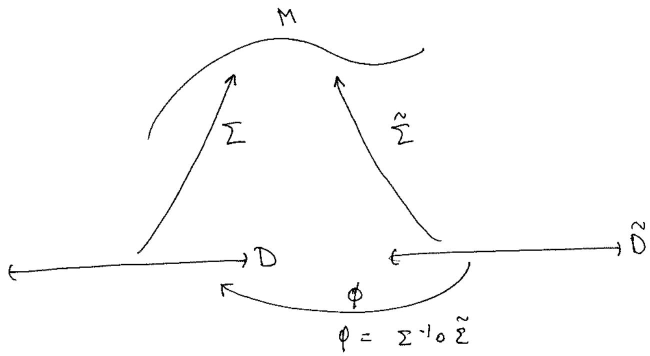


Fig. 26.4. Verifying surface integrals are independent of parametrization.

$$\sum_i \int_{M_i} \varphi_i f \, d\sigma = \sum_i \int_{M_i} \sum_j \psi_j \varphi_i f \, d\sigma = \sum_{ij} \int_{M_i \cap M_j'} \psi_j \varphi_i f \, d\sigma$$

with a similar computation showing

$$\sum_j \int_{M_j'} \psi_i f \, d\sigma = \sum_{ji} \int_{M_i \cap M_j'} \psi_j \varphi_i f \, d\sigma = \sum_{ij} \int_{M_i \cap M_j'} \psi_j \varphi_i f \, d\sigma.$$

Remark 26.11. By the Reisz theorem, there exists a unique Radon measure σ on M such that

$$\int_M f \, d\sigma = \int_M f \, d\sigma.$$

This σ is called surface measure on M. BRUCE: do we really need to appeal to the Reisz – Markov theorem here?

Lemma 26.12 (Surface Measure). *Let M be a C^2 – embedded hypersurface in $\mathbb{R}^n$ and $B \subset M$ be a measurable set such that $\bar{B}$ is compact and contained inside $\Sigma(D)$ where $\Sigma : D \to M \subset \mathbb{R}^n$ is a parametrization. Then*

$$\sigma(B) = \lim_{\varepsilon \downarrow 0} m(B^\varepsilon) = \frac{d}{d\varepsilon}|_{0+} m(B^\varepsilon)$$

where

$$B^\varepsilon := \{x + t\, n(x) : x \in B, 0 \le t \le \varepsilon\}$$

and $n(x)$ is a unit normal to M at $x \in M$, see Figure 26.5.

Proof. Let $A := \Sigma^{-1}(B)$ and $\nu(y) := n(\Sigma(y))$ so that $\nu \in C^{k-1}(D, \mathbb{R}^n)$ if $\Sigma \in C^k(D, \mathbb{R}^n)$. Define

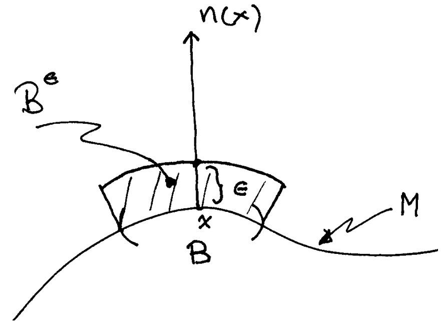


Fig. 26.5. Computing the surface area of B as the volume of an ϵ - fattened neighborhood of B.

$$\psi(y,t) = \Sigma(y) + tn(\Sigma(y)) = \Sigma(y) + t\nu(y)$$

so that $B^\varepsilon = \psi(A \times [0, \varepsilon])$. Hence by the change of variables formula

$$m(B^\varepsilon) = \int_{A \times [0,\varepsilon]} |\det \psi'(y,t)| dy \, dt = \int_0^\varepsilon dt \int_A dy |\det \psi'(y,t)| \tag{26.2}$$

so that by the fundamental theorem of calculus,

$$\frac{d}{d\varepsilon}|_{0+} m(B^\varepsilon) = \frac{d}{d\varepsilon}|_{0+} \int_0^\varepsilon dt \int_A dy |\det \psi'(y,t)| = \int_A |\det \psi'(y,0)| dy.$$

But

$$|\det \psi'(y,0)| = |\det[\Sigma'(y)|n(\Sigma(y))]| = \rho_\Sigma(y)$$

which shows

$$\frac{d}{d\varepsilon}|_{0+} m(B^\varepsilon) = \int_A \rho_\Sigma(y) dy = \int_D 1_B(\Sigma(y)) \rho_\Sigma(y) dy =: \sigma(B).$$

■

Example 26.13. Let $\Sigma = rS^{n-1}$ be the sphere of radius $r > 0$ contained in $\mathbb{R}^n$ and for $B \subset \Sigma$ and $\alpha > 0$ let

$$B_\alpha := \{t\omega : \omega \in B \text{ and } 0 \le t \le \alpha\} = \alpha B_1.$$

Assuming $N(\omega) = \omega / r$ is the outward pointing normal to rS^{n-1}, we have

$$B^\varepsilon = B_{(1+\varepsilon/r)} \setminus B_1 = [(1 + \varepsilon/r) B_1] \setminus B_1$$

and hence

$$\begin{aligned} m(B^{\varepsilon}) &= m\left([(1+\varepsilon/r)B_1] \setminus B_1\right) = m\left([(1+\varepsilon/r)B_1]\right) - m(B_1) \\ &= \left[(1+\varepsilon/r)^n - 1\right] m(B_1). \end{aligned}$$

Therefore,

$$\begin{aligned} \sigma(B) &= \frac{d}{d\varepsilon}|_0 \left[(1+\varepsilon/r)^n - 1\right] m(B_1) = \frac{n}{r} m(B_1) \\ &= n r^{n-1} m\left(r^{-1} B_1\right) = r^{n-1} \sigma(r^{-1} B), \end{aligned}$$

i.e.

$$\sigma(B) = \frac{n}{r} m(B_1) = n r^{n-1} m\left(r^{-1} B_1\right) = r^{n-1} \sigma(r^{-1} B).$$

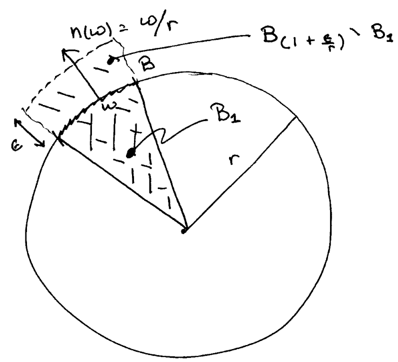


Fig. 26.6. Computing the area of region B on the surface of the sphere of radiur r.

Theorem 26.14. *If $f : \mathbb{R}^n \to [0,\infty]$ is a $(\mathcal{B}_{\mathbb{R}^n}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^n} f(x) dm(x) = \int_{[0,\infty)\times S^{n-1}} f(r\ \omega)\ r^{n-1} dr d\sigma(\omega). \tag{26.3}$$

In particular if $f : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable then

$$\int_{\mathbb{R}^n} f(|x|) dx = \int_0^{\infty} f(r) dV(r) \tag{26.4}$$

where $V(r) = m\left(B(0,r)\right) = r^n m\left(B(0,1)\right) = n^{-1}\sigma\left(S^{n-1}\right) r^n$.

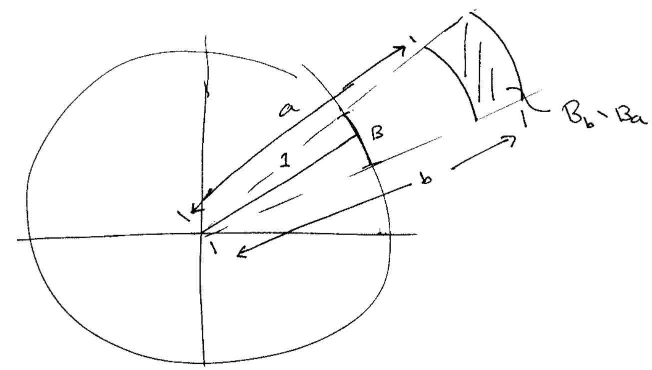


Fig. 26.7. The region $B_b \setminus B_a$.

Proof. Let $B \subset S^{n-1}$, $0 < a < b$ and let $f(x) = 1_{B_b \setminus B_a}(x)$, see Figure 26.7. Then

$$\begin{aligned} \int_{[0,\infty)\times S^{n-1}} f(r\ \omega)\ r^{n-1} dr d\sigma(\omega) &= \int_{[0,\infty)\times S^{n-1}} 1_B(\omega) 1_{[a,b]}(r)\ r^{n-1} dr d\sigma(\omega) \\ &= \sigma(B) \int_a^b r^{n-1} dr = n^{-1}\sigma(B)\left(b^n - a^n\right) \\ &= m(B_1)\left(b^n - a^n\right) = m\left(B_b \setminus B_a\right) \\ &= \int_{\mathbb{R}^n} f(x) dm(x). \end{aligned}$$

Since sets of the form $B_b \setminus B_a$ generate $\mathcal{B}_{\mathbb{R}^n}$ and are closed under intersections, this suffices to prove the theorem. Alternatively one may show that any $f \in C_c(\mathbb{R}^n)$ may be uniformly approximated by linear combinations of characteristic functions of the form $1_{B_b \setminus B_a}$. Indeed, let $S^{n-1} = \bigcup_{i=1}^{K} B_i$ be a partition of S^{n-1} with B_i small and choose $w_i \in B_i$. Let $0 < r_1 < r_2 < r_3 < \cdots < r_n = R < \infty$. Assume $\operatorname{supp}(f) \subset B(0,R)$. Then $\{(B_i)_{r_{j+1}} \setminus (B_i)_{r_j}\}_{i,j}$ partitions $\mathbb{R}^n$ into small regions. Therefore

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)dx &\cong \sum f(r_j\omega_i)m((B_i)_{r_{j+1}} \setminus (B_i)_{r_j}) \\ &= \sum f\left(r_j\omega_i\right)(r_{j+1}^n - r_j^n)\, m((B_i)_1) \\ &= \sum f(r_j\omega_i)\int_{r_j}^{r_{j+1}} r^{n-1}dr\ nm((B_i)_1) \\ &= \sum \int_{r_j}^{r_{j+1}} f(r_j\omega_i)r^{n-1}dr\ \sigma(B_i) \\ &\cong \sum_{ij}\int_{r_j}^{r_{j+1}}\left(\int_{S^{n-1}} f(r_j\omega)d\sigma(\omega)\right) r^{n-1}dr \\ &\cong \int_0^\infty \left(\int_{S^{n-1}} f(r\omega)d\sigma(\omega)\right) r^{n-1}dr.\end{aligned}$$

Eq. (26.4) is a simple special case of Eq. (26.3). It can also be proved directly as follows. Suppose first $f \in C_c^1([0,\infty))$ then

$$\begin{aligned}\int_{\mathbb{R}^n} f(|x|)dx &= -\int_{\mathbb{R}^n} dx \int_{|x|}^\infty dr f'(r) = -\int_{\mathbb{R}^n} dx \int_{\mathbb{R}} 1_{|x|\le r} f'(r)dr \\ &= -\int_0^\infty V(r)f'(r)dr = \int_0^\infty V'(r)f(r)dr.\end{aligned}$$

The result now extends to general f by a density argument. ■

We are now going to work out some integrals using Eq. (26.3). The first we leave as an exercise.

Exercise 26.1. Use the results of Example 26.7 and Theorem 26.14 to show,

$$\sigma(S^{n-1}) = 2\sigma(S^{n-2})\int_0^1 \frac{1}{\sqrt{1-\rho^2}}\rho^{n-2}d\rho.$$

The result in Exercise 26.1 may be used to compute the volume of spheres in any dimension. This method will be left to the reader. We will do this in another way. The first step will be to directly compute the following Gaussian integrals. The result will also be needed for later purposes.

Lemma 26.15. *Let $a > 0$ and*

$$I_n(a) := \int_{\mathbb{R}^n} e^{-a|x|^2}dm(x). \tag{26.5}$$

Then $I_n(a) = (\pi/a)^{n/2}$.

Proof. By Tonelli's theorem and induction,

$$\begin{aligned}I_n(a) &= \int_{\mathbb{R}^{n-1}\times\mathbb{R}} e^{-a|y|^2}e^{-at^2}m_{n-1}(dy)\ dt \\ &= I_{n-1}(a)I_1(a) = I_1^n(a).\end{aligned} \tag{26.6}$$

So it suffices to compute:

$$I_2(a) = \int_{\mathbb{R}^2} e^{-a|x|^2}dm(x) = \int_{\mathbb{R}^2\setminus\{0\}} e^{-a(x_1^2+x_2^2)}dx_1dx_2.$$

Writing this integral in polar coordinates (see Example **??**) gives

$$\begin{aligned}I_2(a) &= \int_0^\infty dr\ r\int_0^{2\pi} d\theta\ e^{-ar^2} = 2\pi\int_0^\infty re^{-ar^2}dr \\ &= 2\pi\lim_{M\to\infty}\int_0^M re^{-ar^2}dr = 2\pi\lim_{M\to\infty}\frac{e^{-ar^2}}{-2a}\int_0^M = \frac{2\pi}{2a} = \pi/a.\end{aligned}$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (26.6). ■

Corollary 26.16. *Let $S^{n-1} \subset \mathbb{R}^n$ be the unit sphere in $\mathbb{R}^n$ and*

$$\Gamma(x) := \int_0^\infty u^{x-1}e^{-u}du \text{ for } x > 0$$

*be the **gamma function.** Then*

1. *The surface area $\sigma(S^{n-1})$ of the unit sphere $S^{n-1} \subset \mathbb{R}^n$ is*
$$\sigma(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}. \tag{26.7}$$
2. *The Γ – function satisfies*
   a) *$\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ and $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$.*
   b) *For $n \in \mathbb{N}$,*
$$\Gamma(n+1) = n! \text{ and } \Gamma(n+1/2) = \frac{(2n-1)!!}{2^n}\cdot\sqrt{\pi}. \tag{26.8}$$
3. *For $n \in \mathbb{N}$,*
$$\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{n!} \text{ and } \sigma(S^{2n}) = \frac{2\,(2\pi)^n}{(2n-1)!!}. \tag{26.9}$$

Proof. Let I_n be as in Lemma 26.15. Using Theorem 26.14 we may alternatively compute $\pi^{n/2} = I_n(1)$ as

$$\pi^{n/2} = I_n(1) = \int_0^\infty dr\ r^{n-1}e^{-r^2} \int_{S^{n-1}} d\sigma = \sigma(S^{n-1}) \int_0^\infty r^{n-1}e^{-r^2}dr.$$

We simplify this last integral by making the change of variables $u = r^2$ so that $r = u^{1/2}$ and $dr = \frac{1}{2}u^{-1/2}du$. The result is

$$\begin{aligned}\int_0^\infty r^{n-1}e^{-r^2}dr &= \int_0^\infty u^{\frac{n-1}{2}}e^{-u}\frac{1}{2}u^{-1/2}du \\ &= \frac{1}{2}\int_0^\infty u^{\frac{n}{2}-1}e^{-u}du = \frac{1}{2}\Gamma(n/2). \qquad (26.10)\end{aligned}$$

Collecting these observations implies that

$$\pi^{n/2} = I_n(1) = \frac{1}{2}\sigma(S^{n-1})\Gamma(n/2)$$

which proves Eq. (26.7). The computation of $\Gamma(1)$ is easy and is left to the reader. By Eq. (26.10),

$$\begin{aligned}\Gamma(1/2) &= 2\int_0^\infty e^{-r^2}dr = \int_{-\infty}^\infty e^{-r^2}dr \\ &= I_1(1) = \sqrt{\pi}.\end{aligned}$$

The relation, $\Gamma(x+1) = x\Gamma(x)$ is the consequence of integration by parts:

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty e^{-u}\ u^{x+1}\ \frac{du}{u} = \int_0^\infty u^x\left(-\frac{d}{du}\ e^{-u}\right)du \\ &= x\int_0^\infty u^{x-1}\ e^{-u}\ du = x\ \Gamma(x).\end{aligned}$$

Eq. (26.8) follows by induction from the relations just proved. Eq. (26.9) is a consequence of items 1. and 2. as follows:

$$\sigma(S^{2n+1}) = \frac{2\pi^{(2n+2)/2}}{\Gamma((2n+2)/2)} = \frac{2\pi^{n+1}}{\Gamma(n+1)} = \frac{2\pi^{n+1}}{n!}$$

and

$$\begin{aligned}\sigma(S^{2n}) &= \frac{2\pi^{(2n+1)/2}}{\Gamma((2n+1)/2)} = \frac{2\pi^{n+1/2}}{\Gamma(n+1/2)} = \frac{2\pi^{n+1/2}}{\frac{(2n-1)!!}{2^n}\cdot\sqrt{\pi}} \\ &= \frac{2\,(2\pi)^n}{(2n-1)!!}.\end{aligned}$$

■

26.2 More spherical coordinates

In this section we will define spherical coordinates in all dimensions. Along the way we will develop an explicit method for computing surface integrals on spheres. As usual when $n = 2$ define spherical coordinates $(r, \theta) \in (0, \infty) \times [0, 2\pi)$ so that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = \psi_2(\theta, r).$$

For $n = 3$ we let $x_3 = r\cos\varphi_1$ and then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \psi_2(\theta, r\sin\varphi_1),$$

as can be seen from Figure 26.8, so that

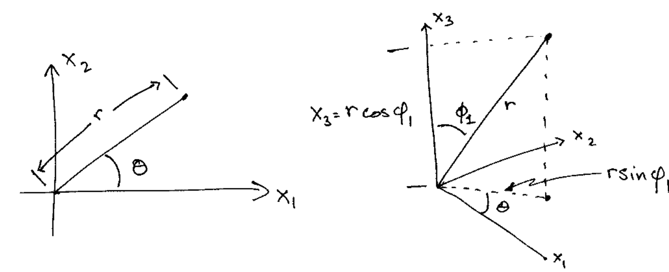


Fig. 26.8. Setting up polar coordinates in two and three dimensions.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \psi_2(\theta, r\sin\varphi_1) \\ r\cos\varphi_1 \end{pmatrix} = \begin{pmatrix} r\sin\varphi_1\cos\theta \\ r\sin\varphi_1\sin\theta \\ r\cos\varphi_1 \end{pmatrix} =: \psi_3(\theta, \varphi_1, r,).$$

We continue to work inductively this way to define

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} \psi_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r\sin\varphi_{n-1},) \\ r\cos\varphi_{n-1} \end{pmatrix} = \psi_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r).$$

So for example,

$$\begin{aligned} x_1 &= r\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_2\cos\varphi_1 \\ x_4 &= r\cos\varphi_2 \end{aligned}$$

and more generally,

$$\begin{aligned} x_1 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\cos\varphi_1 \\ &\vdots \\ x_{n-2} &= r\sin\varphi_{n-2}\sin\varphi_{n-3}\cos\varphi_{n-4} \\ x_{n-1} &= r\sin\varphi_{n-2}\cos\varphi_{n-3} \\ x_n &= r\cos\varphi_{n-2}. \end{aligned} \tag{26.11}$$

By the change of variables formula,

$$\begin{aligned} &\int_{\mathbb{R}^n} f(x)dm(x) \\ &= \int_0^\infty dr \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} d\varphi_1\dots d\varphi_{n-2}d\theta \Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) f(\psi_n(\theta,\varphi_1,\dots,\varphi_{n-2},r)) \end{aligned} \tag{26.12}$$

where

$$\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) := \left|\det\psi_n'(\theta,\varphi_1,\dots,\varphi_{n-2},r)\right|.$$

Proposition 26.17. *The Jacobian, Δ_n is given by*

$$\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) = r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1. \tag{26.13}$$

If f is a function on rS^{n-1} – the sphere of radius r centered at 0 inside of $\mathbb{R}^n$, then

$$\begin{aligned} &\int_{rS^{n-1}} f(x)d\sigma(x) = r^{n-1}\int_{S^{n-1}} f(r\omega)d\sigma(\omega) \\ &= \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} f(\psi_n(\theta,\varphi_1,\dots,\varphi_{n-2},r))\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r)d\varphi_1\dots d\varphi_{n-2}d\theta \end{aligned} \tag{26.14}$$

Proof. We are going to compute Δ_n inductively. Letting $\rho := r\sin\varphi_{n-1}$ and writing $\frac{\partial\psi_n}{\partial\xi}$ for $\frac{\partial\psi_n}{\partial\xi}(\theta,\varphi_1,\dots,\varphi_{n-2},\rho)$ we have

$$\begin{aligned} &\Delta_{n+1}(\theta,\varphi_1,\dots,\varphi_{n-2},\varphi_{n-1},r) \\ &\quad = \left|\begin{bmatrix} \frac{\partial\psi_n}{\partial\theta} & \frac{\partial\psi_n}{\partial\varphi_1} & \dots & \frac{\partial\psi_n}{\partial\varphi_{n-2}} & \frac{\partial\psi_n}{\partial\rho} r\cos\varphi_{n-1} & \frac{\partial\psi_n}{\partial\rho}\sin\varphi_{n-1} \\ 0 & 0 & \dots & 0 & -r\sin\varphi_{n-1} & \cos\varphi_{n-1} \end{bmatrix}\right| \\ &\quad = r\left(\cos^2\varphi_{n-1} + \sin^2\varphi_{n-1}\right)\Delta_n(,\theta,\varphi_1,\dots,\varphi_{n-2},\rho) \\ &\quad = r\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r\sin\varphi_{n-1}), \end{aligned}$$

i.e.

$$\Delta_{n+1}(\theta,\varphi_1,\dots,\varphi_{n-2},\varphi_{n-1},r) = r\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r\sin\varphi_{n-1}). \tag{26.15}$$

To arrive at this result we have expanded the determinant along the bottom row. Staring with the well known and easy to compute fact that $\Delta_2(\theta,r) = r$, Eq. (26.15) implies

$$\begin{aligned} \Delta_3(\theta,\varphi_1,r) &= r\Delta_2(\theta,r\sin\varphi_1) = r^2\sin\varphi_1 \\ \Delta_4(\theta,\varphi_1,\varphi_2,r) &= r\Delta_3(\theta,\varphi_1,r\sin\varphi_2) = r^3\sin^2\varphi_2\sin\varphi_1 \\ &\vdots \\ \Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) &= r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1 \end{aligned}$$

which proves Eq. (26.13). Eq. (26.14) now follows from Eqs. (26.3), (26.12) and (26.13). ■

As a simple application, Eq. (26.14) implies

$$\begin{aligned} \sigma(S^{n-1}) &= \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} \sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1 d\varphi_1\dots d\varphi_{n-2}d\theta \\ &= 2\pi\prod_{k=1}^{n-2}\gamma_k = \sigma(S^{n-2})\gamma_{n-2} \end{aligned} \tag{26.16}$$

where $\gamma_k := \int_0^\pi \sin^k\varphi d\varphi$. If $k \ge 1$, we have by integration by parts that,

$$\begin{aligned} \gamma_k &= \int_0^\pi \sin^k\varphi d\varphi = -\int_0^\pi \sin^{k-1}\varphi\ d\cos\varphi = 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\varphi\ \cos^2\varphi d\varphi \\ &= 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\varphi\ \left(1-\sin^2\varphi\right)d\varphi = 2\delta_{k,1} + (k-1)\left[\gamma_{k-2} - \gamma_k\right] \end{aligned}$$

and hence γ_k satisfies $\gamma_0 = \pi$, $\gamma_1 = 2$ and the recursion relation

$$\gamma_k = \frac{k-1}{k}\gamma_{k-2} \text{ for } k \ge 2.$$

Hence we may conclude

$$\gamma_0 = \pi,\ \gamma_1 = 2,\ \gamma_2 = \frac{1}{2}\pi,\ \gamma_3 = \frac{2}{3}2,\ \gamma_4 = \frac{3}{4}\frac{1}{2}\pi,\ \gamma_5 = \frac{4}{5}\frac{2}{3}2,\ \gamma_6 = \frac{5}{6}\frac{3}{4}\frac{1}{2}\pi$$

and more generally by induction that

$$\gamma_{2k} = \pi\frac{(2k-1)!!}{(2k)!!} \text{ and } \gamma_{2k+1} = 2\frac{(2k)!!}{(2k+1)!!}.$$

Indeed,

$$\gamma_{2(k+1)+1} = \frac{2k+2}{2k+3}\gamma_{2k+1} = \frac{2k+2}{2k+3}2\frac{(2k)!!}{(2k+1)!!} = 2\frac{[2(k+1)]!!}{(2(k+1)+1)!!}$$

and

$$\gamma_{2(k+1)} = \frac{2k+1}{2k+1}\gamma_{2k} = \frac{2k+1}{2k+2}\pi\frac{(2k-1)!!}{(2k)!!} = \pi\frac{(2k+1)!!}{(2k+2)!!}.$$

The recursion relation in Eq. (26.16) may be written as

$$\sigma(S^n) = \sigma\left(S^{n-1}\right)\gamma_{n-1} \tag{26.17}$$

which combined with $\sigma\left(S^1\right) = 2\pi$ implies

$$\begin{aligned}
\sigma\left(S^1\right) &= 2\pi, \\
\sigma(S^2) &= 2\pi \cdot \gamma_1 = 2\pi \cdot 2, \\
\sigma(S^3) &= 2\pi \cdot 2 \cdot \gamma_2 = 2\pi \cdot 2 \cdot \frac{1}{2}\pi = \frac{2^2\pi^2}{2!!}, \\
\sigma(S^4) &= \frac{2^2\pi^2}{2!!} \cdot \gamma_3 = \frac{2^2\pi^2}{2!!} \cdot 2\frac{2}{3} = \frac{2^3\pi^2}{3!!} \\
\sigma(S^5) &= 2\pi \cdot 2 \cdot \frac{1}{2}\pi \cdot \frac{2}{3}2 \cdot \frac{3}{4}\frac{1}{2}\pi = \frac{2^3\pi^3}{4!!}, \\
\sigma(S^6) &= 2\pi \cdot 2 \cdot \frac{1}{2}\pi \cdot \frac{2}{3}2 \cdot \frac{3}{4}\frac{1}{2}\pi \cdot \frac{4}{5}\frac{2}{3}2 = \frac{2^4\pi^3}{5!!}
\end{aligned}$$

and more generally that

$$\sigma(S^{2n}) = \frac{2\,(2\pi)^n}{(2n-1)!!} \text{ and } \sigma(S^{2n+1}) = \frac{(2\pi)^{n+1}}{(2n)!!} \tag{26.18}$$

which is verified inductively using Eq. (26.17). Indeed,

$$\sigma(S^{2n+1}) = \sigma(S^{2n})\gamma_{2n} = \frac{2\,(2\pi)^n}{(2n-1)!!}\pi\frac{(2n-1)!!}{(2n)!!} = \frac{(2\pi)^{n+1}}{(2n)!!}$$

and

$$\sigma(S^{(n+1)}) = \sigma(S^{2n+2}) = \sigma(S^{2n+1})\gamma_{2n+1} = \frac{(2\pi)^{n+1}}{(2n)!!}2\frac{(2n)!!}{(2n+1)!!} = \frac{2\,(2\pi)^{n+1}}{(2n+1)!!}.$$

Using

$$(2n)!! = 2n\,(2(n-1))\ldots(2\cdot 1) = 2^n n!$$

we may write $\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (26.9) and (26.18 in agreement. We may also write the formula in Eq. (26.18) as

$$\sigma(S^n) = \begin{cases} \frac{2(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ even} \\ \frac{(2\pi)^{\frac{n+1}{2}}}{(n-1)!!} & \text{for } n \text{ odd.} \end{cases}$$

26.3 n – dimensional manifolds with boundaries

Definition 26.18. *A set $\Omega \subset \mathbb{R}^n$ is said to be a C^k –* ***manifold with boundary*** *if for each $x_0 \in \partial\Omega := \Omega \setminus \Omega^o$ (here Ω^o is the interior of Ω) there exists $\varepsilon > 0$ an open set $0 \in D \subset \mathbb{R}^n$ and a C^k-diffeomorphism $\psi : D \to B(x_0, \varepsilon)$ such that $\psi(D \cap \{y_n \geq 0\}) = B(x_0, \varepsilon) \cap \Omega$. See Figure 26.9 below. We call $\partial\Omega$ the* ***manifold boundary*** *of Ω.*

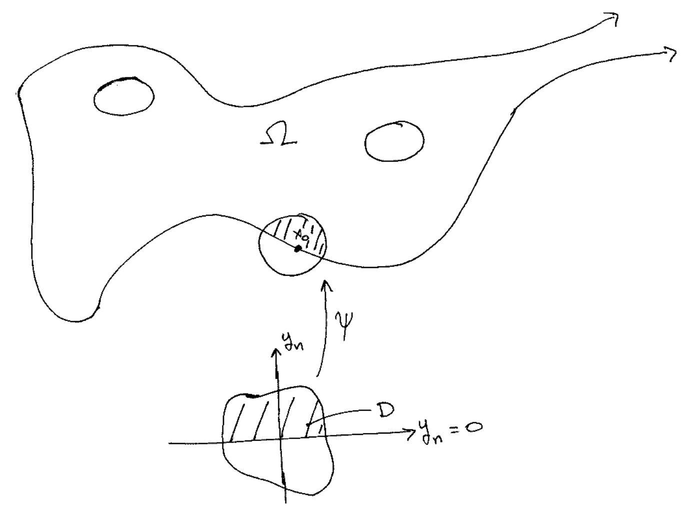


Fig. 26.9. Flattening out a neighborhood of a boundary point.

Remarks 26.19 *1. In Definition 26.18 we have defined $\partial\Omega = \Omega \setminus \Omega^o$ which is* ***not*** *the topological boundary of Ω, defined by $\text{bd}(\Omega) := \bar{\Omega} \setminus \Omega^0$. Clearly we always have $\partial\Omega \subset \text{bd}(\Omega)$ with equality iff Ω is closed.*
2. It is easily checked that if $\Omega \subset \mathbb{R}^n$ is a C^k – manifold with boundary, then $\partial\Omega$ is a C^k – hypersurface in $\mathbb{R}^n$.

The reader is left to verify the following examples.

Example 26.20. Let $\mathbb{H}^n = \{x \in \mathbb{R}^n : x_n > 0\}$.

1. $\bar{\mathbb{H}}^n$ is a C^∞ – manifold with boundary and

$$\partial\bar{\mathbb{H}}^n = \text{bd}\left(\bar{\mathbb{H}}^n\right) = \mathbb{R}^{n-1} \times \{0\}.$$

2. $\Omega = \overline{B(\xi, r)}$ is a C^∞ – manifold with boundary and $\partial\Omega = \mathrm{bd}(B(\xi, r))$, as the reader should verify. See Exercise 26.2 for a general result containing this statement.
3. Let U be the open unit ball in $\mathbb{R}^{n-1}$, then $\Omega = \mathbb{H}^n \cup (U \times \{0\})$ is a C^∞ – manifold with boundary and $\partial\Omega = U \times \{0\}$ while $\mathrm{bd}(\Omega) = \mathbb{R}^{n-1} \times \{0\}$.
4. Now let $\Omega = \mathbb{H}^n \cup (\bar{U} \times \{0\})$, then Ω is **not** a C^1 – manifold with boundary. The bad points are $\mathrm{bd}(U) \times \{0\}$.
5. Suppose V is an open subset of $\mathbb{R}^{n-1}$ and $g : V \to \mathbb{R}$ is a C^k – function and set
$$\Omega := \{(y, z) \in V \times \mathbb{R} \subset \mathbb{R}^n : z \geq g(y)\},$$
then Ω is a C^k – manifold with boundary and $\partial\Omega = \Gamma(g)$ – the graph of g. Again the reader should check this statement.
6. Let
$$\Omega = [(0,1) \times (0,1)] \cup [(-1,0) \times (-1,0)] \cup [(-1,1) \times \{0\}]$$
in which case
$$\Omega^o = [(0,1) \times (0,1)] \cup [(-1,0) \times (-1,0)]$$
and hence $\partial\Omega = (-1,1) \times \{0\}$ is a C^k – hypersurface in $\mathbb{R}^2$. Nevertheless Ω is not a C^k – manifold with boundary as can be seen by looking at the point $(0,0) \in \partial\Omega$.
7. If $\Omega = S^{n-1} \subset \mathbb{R}^n$, then $\partial\Omega = \Omega$ is a C^∞ - hypersurface. However, as in the previous example Ω is not an n – dimensional C^k – manifold with boundary despite the fact that Ω is now closed. (**Warning:** there is a clash of notation here with that of the more general theory of manifolds where $\partial S^{n-1} = \emptyset$ when viewing S^{n-1} as a manifold in its own right.)

Lemma 26.21. *Suppose $\Omega \subset_o \mathbb{R}^n$ such that $\mathrm{bd}(\Omega)$ is a C^k – hypersurface, then $\bar{\Omega}$ is C^k – manifold with boundary. (It is not necessarily true that $\partial\bar{\Omega} = \overline{\mathrm{bd}(\Omega)}$. For example, let $\Omega := B(0,1) \cup \{x \in \mathbb{R}^n : 1 < |x| < 2\}$. In this case $\bar{\Omega} = \overline{B(0,2)}$ so $\partial\bar{\Omega} = \{x \in \mathbb{R}^n : |x| = 2\}$ while $\mathrm{bd}(\Omega) = \{x \in \mathbb{R}^n : |x| = 2$ or $|x| = 1\}$.)*

Proof. Claim: Suppose $U = (-1,1)^n \subset_o \mathbb{R}^n$ and $V \subset_o U$ such that $\mathrm{bd}(V) \cap U = \partial\mathbb{H}^n \cap U$. Then V is either, $U_+ := U \cap \mathbb{H}^n = U \cap \{x_n > 0\}$ or $U_- := U \cap \{x_n < 0\}$ or $U \setminus \partial\mathbb{H}^n = U_+ \cup U_-$. To prove the claim, first observe that $V \subset U \setminus \partial\mathbb{H}^n$ and V is not empty, so either $V \cap U_+$ or $V \cap U_-$ is not empty. Suppose for example there exists $\xi \in V \cap U_+$. Let $\sigma : [0,1) \to U \cap \mathbb{H}^n$ be a continuous path such that $\sigma(0) = \xi$ and
$$T = \sup\{t < 1 : \sigma([0,t]) \subset V\}.$$
If $T \neq 1$, then $\eta := \sigma(T)$ is a point in U_+ which is also in $bd(V) = \bar{V} \setminus V$. But this contradicts $bd(V) \cap U = \partial\mathbb{H}^n \cap U$ and hence $T = 1$. Because U_+ is path connected, we have shown $U_+ \subset V$. Similarly if $V \cap U_- \neq \emptyset$, then $U_- \subset V$ as well and this completes the proof of the claim. We are now ready to show $\bar{\Omega}$ is a C^k – manifold with boundary. To this end, suppose
$$\xi \in \partial\bar{\Omega} = \mathrm{bd}(\bar{\Omega}) = \bar{\Omega} \setminus \bar{\Omega}^o \subset \bar{\Omega} \setminus \Omega = \mathrm{bd}(\Omega).$$
Since $\mathrm{bd}(\Omega)$ is a C^k – hypersurface, we may find an open neighborhood O of ξ such that there exists a C^k – diffeomorphism $\psi : U \to O$ such that $\psi(O \cap \mathrm{bd}(\Omega)) = U \cap \mathbb{H}^n$. Recall that
$$O \cap \mathrm{bd}(\Omega) = O \cap \bar{\Omega} \cap \Omega^c = \overline{\Omega \cap O}^O \setminus (O \setminus \Omega) = \mathrm{bd}_O(\Omega \cap O)$$
where $\overline{A}^O$ and $\mathrm{bd}_O(A)$ denotes the closure and boundary of a set $A \subset O$ in the relative topology on A. Since ψ is a C^k – diffeomorphism, it follows that $V := \psi(O \cap \Omega)$ is an open set such that
$$\mathrm{bd}(V) \cap U = \mathrm{bd}_U(V) = \psi(\mathrm{bd}_O(\Omega \cap O)) = \psi(O \cap \mathrm{bd}(\Omega)) = U \cap \mathbb{H}^n.$$
Therefore by the claim, we learn either $V = U_+$ of U_- or $U_+ \cup U_-$. However the latter case can not occur because in this case ξ would be in the interior of $\bar{\Omega}$ and hence not in $\mathrm{bd}(\bar{\Omega})$. This completes the proof, since by changing the sign on the n^{th} coordinate of ψ if necessary, we may arrange it so that $\psi(\bar{\Omega} \cap O) = U_+$. ■

Exercise 26.2. Suppose $F : \mathbb{R}^n \to \mathbb{R}$ is a C^k – function such that
$$\{F < 0\} := \{x \in \mathbb{R}^n : F(x) < 0\} \neq \emptyset$$
and $F'(\xi) : \mathbb{R}^n \to \mathbb{R}$ is surjective (or equivalently $\nabla F(\xi) \neq 0$) for all
$$\xi \in \{F = 0\} := \{x \in \mathbb{R}^n : F(x) = 0\}.$$
Then $\Omega := \{F \leq 0\}$ is a C^k – manifold with boundary and $\partial\Omega = \{F = 0\}$.

Hint: For $\xi \in \{F = 0\}$, let $A : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be a linear transformation such that $A|_{\mathrm{Nul}(F'(\xi))} : \mathrm{Nul}(F'(\xi)) \to \mathbb{R}^{n-1}$ is invertible and $A|_{\mathrm{Nul}(F'(\xi))^\perp} \equiv 0$ and then define
$$\varphi(x) := (A(x - \xi), -F(x)) \in \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R}^n.$$
Now use the inverse function theorem to construct ψ.

Definition 26.22 (Outward pointing unit normal vector). *Let Ω be a C^1 – manifold with boundary, the **outward pointing unit normal** to $\partial\Omega$ is the unique function $n : \partial\Omega \to \mathbb{R}^n$ satisfying the following requirements.*

1. *(Unit length.) $|n(x)| = 1$ for all $x \in \partial\Omega$.*
2. *(Orthogonality to $\partial\Omega$.) If $x_0 \in \partial\Omega$ and $\psi : D \to B(x_0, \varepsilon)$ is as in the Definition 26.18, then $n(x_o) \perp \psi'(0)(\partial\mathbb{H}^n)$, i.e. $n(x_0)$ is perpendicular of $\partial\Omega$.*
3. *(Outward Pointing.) If $\varphi := \psi^{-1}$, then $\varphi'(0)n(x_o) \cdot e_n < 0$ or equivalently put $\psi'(0)e_n \cdot n(x_0) < 0$, see Figure 26.11 below.*

26.4 Divergence Theorem

Theorem 26.23 (Divergence Theorem). *Let $\Omega \subset \mathbb{R}^n$ be a manifold with C^2 – boundary and $n : \partial\Omega \to \mathbb{R}^n$ be the unit outward pointing normal to Ω. If $Z \in C_c(\Omega, \mathbb{R}^n) \cap C^1(\Omega^o, \mathbb{R}^n)$ and*

$$\int_\Omega |\nabla \cdot Z| dm < \infty \tag{26.19}$$

then

$$\int_{\partial\Omega} Z(x) \cdot n(x) d\sigma(x) = \int_\Omega \nabla \cdot Z(x)\ dx. \tag{26.20}$$

The proof of Theorem 26.23 will be given after stating a few corollaries and then a number preliminary results.

Example 26.24. Let

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{on } [0,1], \\ 0 & \text{if } x = 0 \end{cases}$$

then $f \in C([0,1]) \cap C^\infty((0,1))$ and $f'(x) = \sin\left(\frac{1}{x}\right) - \frac{1}{x}\sin\left(\frac{1}{x}\right)$ for $x > 0$. Since

$$\int_0^1 \frac{1}{x}\left|\sin\left(\frac{1}{x}\right)\right| dx = \int_1^\infty u|\sin(u)|\frac{1}{u^2}du = \int_1^\infty \frac{|\sin(u)|}{u}du = \infty,$$

$\int_0^1 |f'(x)|\, dx = \infty$ and the integrability assumption, $\int_\Omega |\nabla \cdot Z| dx < \infty$, in Theorem 26.23 is necessary.

Corollary 26.25. *Let $\Omega \subset \mathbb{R}^n$ be a **closed** manifold with C^2 – boundary and $n : \partial\Omega \to \mathbb{R}^n$ be the outward pointing unit normal to Ω. If $Z \in C(\Omega, \mathbb{R}^n) \cap C^1(\Omega^o, \mathbb{R}^n)$ and*

$$\int_\Omega \{|Z| + |\nabla \cdot Z|\}\, dm + \int_{\partial\Omega} |Z \cdot n|\, d\sigma < \infty \tag{26.21}$$

then Eq. (26.20) is valid, i.e.

$$\int_{\partial\Omega} Z(x) \cdot n(x) d\sigma(x) = \int_\Omega \nabla \cdot Z(x)\ dx.$$

Proof. Let $\psi \in C_c^\infty(\mathbb{R}^n, [0,1])$ such that $\psi = 1$ in a neighborhood of 0 and set $\psi_k(x) := \psi(x/k)$ and $Z_k := \psi_k Z$. We have $\text{supp}(Z_k) \subset \text{supp}(\psi_k) \cap \Omega$ – which is a compact set since Ω is closed. Since $\nabla\psi_k(x) = \frac{1}{k}(\nabla\psi)(x/k)$ is bounded,

$$\begin{aligned} \int_\Omega |\nabla \cdot Z_k|\, dm &= \int_\Omega |\nabla\psi_k \cdot Z + \psi_k \nabla \cdot Z|\, dm \\ &\le C\int_\Omega |Z|\, dm + \int_\Omega |\nabla \cdot Z|\, dm < \infty. \end{aligned}$$

Hence Theorem 26.23 implies

$$\int_\Omega \nabla \cdot Z_k dm = \int_{\partial\Omega} Z_k \cdot n d\sigma. \tag{26.22}$$

By the D.C.T.,

$$\begin{aligned} \int_\Omega \nabla \cdot Z_k dm &= \int_\Omega \left[\frac{1}{k}(\nabla\psi)(x/k) \cdot Z(x) + \psi(x/k)\nabla \cdot Z(x)\right] dx \\ &\to \int_\Omega \nabla \cdot Z dm \end{aligned}$$

and

$$\int_{\partial\Omega} Z_k \cdot n\ d\sigma = \int_{\partial\Omega} \psi_k Z \cdot n\ d\sigma \to \int_{\partial\Omega} Z \cdot n\ d\sigma,$$

which completes the proof by passing the limit in Eq. (26.22). ∎

Corollary 26.26 (Integration by parts I). *Let $\Omega \subset \mathbb{R}^n$ be a closed manifold with C^2 – boundary, $n : \partial\Omega \to \mathbb{R}^n$ be the outward pointing normal to Ω, $Z \in C(\Omega, \mathbb{R}^n) \cap C^1(\Omega^o, \mathbb{R}^n)$ and $f \in C(\Omega, \mathbb{R}) \cap C^1(\Omega^o, \mathbb{R})$ such that*

$$\int_\Omega \{|f|\,[|Z| + |\nabla \cdot Z|] + |\nabla f|\,|Z|\}\, dm + \int_{\partial\Omega} |f|\,|Z \cdot n|\, d\sigma < \infty$$

then

$$\int_\Omega f(x)\nabla \cdot Z(x)\ dx = -\int_\Omega \nabla f(x) \cdot Z(x)\ dx + \int_{\partial\Omega} f(x)\, Z(x) \cdot n(x) d\sigma(x).$$

Proof. Apply Corollary 26.25 with Z replaced by fZ. ∎

Corollary 26.27 (Integration by parts II). *Let $\Omega \subset \mathbb{R}^n$ be a closed manifold with C^2 – boundary , $n : \partial\Omega \to \mathbb{R}^n$ be the outward pointing normal to Ω and $f, g \in C(\Omega, \mathbb{R}) \cap C^1(\Omega^o, \mathbb{R})$ such that*

$$\int_\Omega \{|f|\,|g| + |\partial_i f|\,|g| + |f|\,|\partial_i g|\}\, dm + \int_{\partial\Omega} |fgn_i|\, d\sigma < \infty$$

then

$$\int_\Omega f(x)\partial_i g(x)\ dm = -\int_\Omega \partial_i f(x)\cdot g(x)\ dm + \int_{\partial\Omega} f(x)g(x)n_i(x)d\sigma(x).$$

Proof. Apply Corollary 26.26 with Z chosen so that $Z_j = 0$ if $j \neq i$ and $Z_i = g$, (i.e. $Z = (0,\dots,g,0\dots,0)$). ■

Proposition 26.28. *Let Ω be as in Corollary 26.25 and suppose $u, v \in C^2(\Omega^o)\cap C^1(\Omega)$ such that $u, v, \nabla u, \nabla v, \Delta u, \Delta v \in L^2(\Omega)$ and $u, v, \frac{\partial u}{\partial n}, \frac{\partial v}{\partial n} \in L^2(\partial\Omega, d\sigma)$ then*

$$\int_\Omega \triangle u\cdot v dm = -\int_\Omega \nabla u\cdot\nabla v dm + \int_{\partial\Omega} v\frac{\partial u}{\partial n}d\sigma \tag{26.23}$$

and

$$\int_\Omega (\triangle u v - \triangle v\, u)dm = \int_{\partial\Omega}\left(v\frac{\partial u}{\partial n} - \frac{\partial v}{\partial n}u\right)d\sigma. \tag{26.24}$$

Proof. Eq. (26.23) follows by applying Corollary 26.26 with $f = v$ and $Z = \nabla u$. Similarly applying Corollary 26.26 with $f = u$ and $Z = \nabla v$ implies

$$\int_\Omega \triangle v\cdot u dm = -\int_\Omega \nabla u\cdot\nabla v dm + \int_{\partial\Omega} u\frac{\partial v}{\partial n}d\sigma$$

and subtracting this equation from Eq. (26.23) implies Eq. (26.24). ■

Lemma 26.29. *Let $\Omega_t = \varphi_t(\Omega)$ be a smoothly varying domain and $f:\mathbb{R}^n\to\mathbb{R}$. Then*

$$\frac{d}{dt}\int_{\Omega_t} f\ dx = \int_{\partial\Omega_t} f\,(Y_t\cdot n)\ d\sigma$$

where $Y_t(x) = \frac{d}{d\varepsilon}\big|_0 \varphi_{t+\varepsilon}(\varphi_t^{-1}(x))$ as in Figure 26.10.

Proof. With out loss of generality we may compute the derivative at $t = 0$ and replace Ω by $\varphi_0(\Omega)$ and φ_t by $\varphi_t\circ\varphi_0^{-1}$ if necessary so that $\varphi_0(x) = x$ and $Y(x) = \frac{d}{dt}\big|_0\varphi_t(x)$. By the change of variables theorem,

$$\int_{\Omega_t} f\, dx = \int f\, dx = \int_\Omega f\circ\varphi_t(x)\det[\varphi_t'(x)]dx$$

and hence

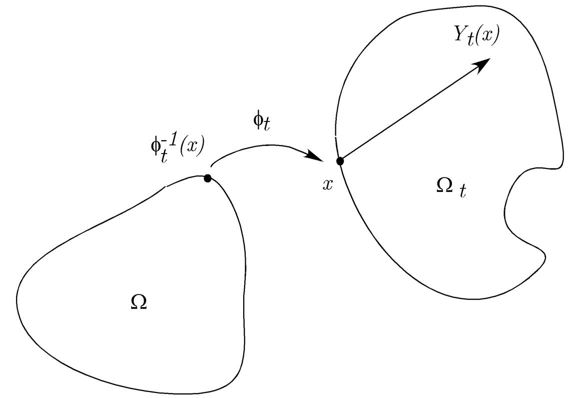


Fig. 26.10. The vector-field $Y_t(x)$ measures the velocity of the boundary point x at time t.

$$\begin{aligned}\frac{d}{dt}\Big|_0\int_{\Omega_t} f\ dx &= \int_{\Omega_0}[Yf(x) + \frac{d}{dt}\Big|_0\det[\varphi_t'(x)]f(x)]dx\\ &= \int_{\Omega_0}[Yf(x) + (\nabla\cdot Y(x))\,f(x)]dx\\ &= \int_{\Omega_0}\nabla\cdot(fY)(x)dx = \int_{\partial\Omega_0} f(x)Y(x)\cdot n(x)\ d\sigma(x).\end{aligned}$$

In the second equality we have used the fact that

$$\frac{d}{dt}\Big|_0\det[\varphi_t'(x)] = tr\left[\frac{d}{dt}\Big|_0\varphi_t'(x)\right] = tr\,[Y'(x)] = \nabla\cdot Y(x).$$

■

26.5 The proof of the Divergence Theorem

Lemma 26.30. *Suppose $\Omega\subset_o\mathbb{R}^n$ and $Z\in C^1(\Omega,\mathbb{R}^n)$ and $f\in C_c^1(\Omega,\mathbb{R})$, then*

$$\int_\Omega f\nabla\cdot Z\, dx = -\int_\Omega \nabla f\cdot Z\, dx.$$

Proof. Let $W := fZ$ on Ω and $W = 0$ on Ω^c, then $W\in C_c(\mathbb{R}^n,\mathbb{R}^n)$. By Fubini's theorem and the fundamental theorem of calculus,

$$\int_{\Omega} \nabla \cdot (fZ)\, dx = \int_{\mathbb{R}^n} (\nabla \cdot W) dx = \sum_{i=1}^{n} \int_{\mathbb{R}^n} \frac{\partial W^i}{\partial x^i}\, dx_1 \dots dx_n = 0.$$

This completes the proof because $\nabla \cdot (fZ) = \nabla f \cdot Z + f \nabla \cdot Z$. ■

Corollary 26.31. *If $\Omega \subset \mathbb{R}^n$, $Z \in C^1(\Omega, \mathbb{R}^n)$ and $g \in C(\Omega, \mathbb{R})$ then $g = \nabla \cdot Z$ iff*

$$\int_{\Omega} gf\, dx = -\int_{\Omega} Z \cdot \nabla f\, dx \text{ for all } f \in C_c^1(\Omega). \tag{26.25}$$

Proof. By Lemma 26.30, Eq. (26.25) holds iff

$$\int_{\Omega} gf\, dx = \int_{\Omega} \nabla \cdot Z\, f\, dx \text{ for all } f \in C_c^1(\Omega)$$

which happens iff $g = \nabla \cdot Z$. ■

Proposition 26.32 (Behavior of ∇ under coordinate transformations). *Let $\psi : W \to \Omega$ is a C^2 – diffeomorphism where W and Ω and open subsets of $\mathbb{R}^n$. Given $f \in C^1(\Omega, \mathbb{R})$ and $Z \in C^1(\Omega, \mathbb{R}^n)$ let $f^\psi = f \circ \psi \in C^1(W, \mathbb{R})$ and $Z^\psi \in C^1(W, \mathbb{R}^n)$ be defined by $Z^\psi(y) = \psi'(y)^{-1} Z(\psi(y))$. Then*

1. $\nabla f^\psi = \nabla(f \circ \psi) = (\psi')^{\text{tr}} (\nabla f) \circ \psi$ *and*
2. $\nabla \cdot [\det \psi'\ Z^\psi] = (\nabla \cdot Z) \circ \psi \cdot \det \psi'$. *(Notice that we use ψ is C^2 at this point.)*

Proof. 1. Let $v \in \mathbb{R}^n$, then by definition of the gradient and using the chain rule,

$$\nabla(f \circ \psi) \cdot v = \partial_v (f \circ \psi) = \nabla f(\psi) \cdot \psi' v = (\psi')^{\text{tr}} \nabla f(\psi) \cdot v.$$

2. Let $f \in C_c^1(\Omega)$. By the change of variables formula,

$$\begin{aligned}\int_{\Omega} f \nabla \cdot Z dm &= \int_W f \circ \psi\, (\nabla \cdot Z) \circ \psi\, |\det \psi'|\, dm \\ &= \int_W f^\psi\, (\nabla \cdot Z) \circ \psi\, |\det \psi'|\, dm.\end{aligned} \tag{26.26}$$

On the other hand

$$\begin{aligned}\int_{\Omega} f \nabla \cdot Z dm &= -\int_{\Omega} \nabla f \cdot Z dm = -\int_W \nabla f(\psi) \cdot Z(\psi)\, |\det \psi'|\, dm \\ &= -\int_W \left[(\psi')^{\text{tr}}\right]^{-1} \nabla f^\psi \cdot Z(\psi)\, |\det \psi'|\, dm \\ &= -\int_W \nabla f^\psi \cdot (\psi')^{-1} Z(\psi)\, |\det \psi'|\, dm \\ &= -\int_W \left(\nabla f^\psi \cdot Z^\psi\right) |\det \psi'|\, dm \\ &= \int_W f^\psi \nabla \cdot \left(|\det \psi'|\ Z^\psi\right) dm.\end{aligned} \tag{26.27}$$

Since Eqs. (26.26) and (26.27) hold for all $f \in C_c^1(\Omega)$ we may conclude

$$\nabla \cdot \left(|\det \psi'|\ Z^\psi\right) = (\nabla \cdot Z) \circ \psi\, |\det \psi'|$$

and by linearity this proves item 2. ■

Lemma 26.33. *Eq. (26.20 of the divergence Theorem 26.23 holds when $\Omega = \bar{\mathbb{H}}^n = \{x \in \mathbb{R}^n : x_n \geq 0\}$ and $Z \in C_c(\bar{\mathbb{H}}^n, \mathbb{R}^n) \cap C^1(\mathbb{H}^n, \mathbb{R}^n)$ satisfies*

$$\int_{\mathbb{H}^n} |\nabla \cdot Z|\ dx < \infty$$

Proof. In this case $\partial \Omega = \mathbb{R}^{n-1} \times \{0\}$ and $n(x) = -e_n$ for $x \in \partial \Omega$ is the outward pointing normal to Ω. By Fubini's theorem and the fundamental theorem of calculus,

$$\sum_{i=1}^{n-1} \int_{x_n > \delta} \frac{\partial Z^i}{\partial x^i}\, dx = 0$$

and

$$\int_{x_n > \delta} \frac{\partial Z_n}{\partial x_n}\, dx = -\int_{\mathbb{R}^{n-1}} Z_n(y, \delta) dy.$$

Therefore. using the dominated convergence theorem,

$$\begin{aligned}\int_{\mathbb{H}^n} \nabla \cdot Z\ dx &= \lim_{\delta \downarrow 0} \int_{x_n > \delta} \nabla \cdot Z\, dx = -\lim_{\delta \downarrow 0} \int_{\mathbb{R}^{n-1}} Z_n(y, \delta) dy \\ &= -\int_{\mathbb{R}^{n-1}} Z_n(y, 0) dy = \int_{\partial \mathbb{H}^n} Z(x) \cdot n(x)\ d\sigma(x).\end{aligned}$$

■

Remark 26.34. The same argument used in the proof of Lemma 26.33 shows Theorem 26.23 holds when

$$\Omega = \bar{\mathbb{R}}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i\}.$$

Notice that $\bar{\mathbb{R}}^n_+$ has a corners and edges, etc. and so $\partial\Omega$ is not smooth in this case.

26.5.1 The Proof of the Divergence Theorem 26.23

Proof. First suppose that $\operatorname{supp}(Z)$ is a compact subset of $B(x_0,\varepsilon)\cap\Omega$ for some $x_0 \in \partial\Omega$ and $\varepsilon > 0$ is sufficiently small so that there exists $V \subset_o \mathbb{R}^n$ and C^2 – diffeomorphism $\psi : V \longrightarrow B(x_0,\varepsilon)$ (see Figure 26.11) such that $\psi(V\cap\{y_n>0\}) = B(x_0,\varepsilon)\cap\Omega^o$ and

$$\psi(V\cap\{y_n=0\}) = B(x_0,\varepsilon)\cap\partial\Omega.$$

Because n is the outward pointing normal, $n(\psi(y))\cdot\psi'(y)e_n < 0$ on $y_n = 0$.

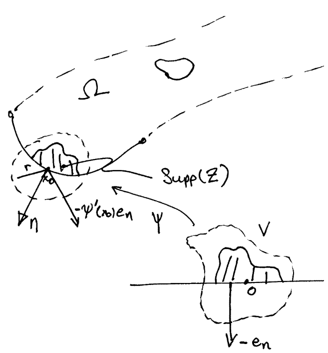


Fig. 26.11. Reducing the divergence theorem for general Ω to $\Omega = \mathbb{H}^n$.

Since V is connected and $\det\psi'(y)$ is never zero on V, $\varsigma := \operatorname{sgn}(\det\psi'(y)) \in \{\pm1\}$ is constant independent of $y \in V$. For $y \in \partial\bar{\mathbb{H}}^n$,

$$\begin{aligned}(Z\cdot n)(\psi(y))\,&|\det[\psi'(y)e_1|\dots|\psi'(y)e_{n-1}|n(\psi(y))]| \\ &= -\varsigma(Z\cdot n)(\psi(y))\det[\psi'(y)e_1|\dots|\psi'(y)e_{n-1}|n(\psi(y))] \\ &= -\varsigma\det[\psi'(y)e_1|\dots|\psi'(y)e_{n-1}|Z(\psi(y))] \\ &= -\varsigma\det[\psi'(y)e_1|\dots|\psi'(y)e_{n-1}|\psi'(y)Z^\psi(y)] \\ &= -\varsigma\det\psi'(y)\cdot\det[e_1|\dots|e_{n-1}|Z^\psi(y)] \\ &= -\left|\det\psi'(y)\right| Z^\psi(y)\cdot e_n,\end{aligned}$$

wherein the second equality we used the linearity properties of the determinant and the identity

$$Z(\psi(y)) = Z\cdot n(\psi(y)) + \sum_{i=1}^{n-1}\alpha_i\psi'(y)e_i \text{ for some } \alpha_i.$$

Starting with the definition of the surface integral we find

$$\begin{aligned}\int_{\partial\Omega} Z\cdot n\,d\sigma &= \int_{\partial\bar{\mathbb{H}}^n}(Z\cdot n)(\psi(y))\,|\det[\psi'(y)e_1|\dots|\psi'(y)e_{n-1}|n(\psi(y))]|\;dy \\ &= \int_{\partial\bar{\mathbb{H}}^n}\det\psi'(y)Z^\psi(y)\cdot(-e_n)\;dy \\ &= \int_{\mathbb{H}^n}\nabla\cdot\left[\det\psi' Z^\psi\right]dm \text{ (by Lemma 26.33)} \\ &= \int_{\mathbb{H}^n}[(\nabla\cdot Z)\circ\psi]\det\psi' dm \text{ (by Proposition 26.32)} \\ &= \int_\Omega(\nabla\cdot Z)\,dm \text{ (by the Change of variables theorem).}\end{aligned}$$

2) We now prove the general case where $Z \in C_c(\Omega,\mathbb{R}^n)\cap C^1(\Omega^o,\mathbb{R}^n)$ and $\int_\Omega|\nabla\cdot Z|dm < \infty$. Using Theorem **??**, we may choose $\varphi_i \in C_c^\infty(\mathbb{R}^n)$ such that

1. $\sum_{i=1}^N\varphi_i \leq 1$ with equality in a neighborhood of $K = \operatorname{Supp}(Z)$.
2. For all i either $\operatorname{supp}(\varphi_i)\subset\Omega$ or $\operatorname{supp}(\varphi_i)\subset B(x_0,\varepsilon)$ where $x_0 \in \partial\Omega$ and $\varepsilon > 0$ are as in the previous paragraph.

Then by special cases proved in the previous paragraph and in Lemma 26.30,

$$\begin{aligned}\int_\Omega\nabla\cdot Z\;dx &= \int_\Omega\nabla\cdot(\sum_i\varphi_i\;Z)\;dx = \sum_i\int_\Omega\nabla\cdot(\varphi_i Z)dx \\ &= \sum_i\int_{\partial\Omega}(\varphi_i Z)\cdot n\;d\sigma \\ &= \int_{\partial\Omega}\sum_i\varphi_i Z\cdot n\;d\sigma = \int_{\partial\Omega}Z\cdot n\;d\sigma.\end{aligned}$$

■

26.5.2 Extensions of the Divergence Theorem to Lipschitz domains

BRUCE: This should be done after the fact about Lip-functions being a.e. differentiable are proved.

The divergence theorem holds more generally for manifolds Ω with Lipschitz boundary. By this we mean, locally near a boundary point, Ω should be of the form

$$\Omega := \{(y,z) \in D \times \mathbb{R} \subset \mathbb{R}^n : z \geq g(y)\} = \{z \geq g\}$$

where $g : D \to \mathbb{R}$ is a Lipschitz function and D is the open unit ball in $\mathbb{R}^{n-1}$.

To prove this remark, first suppose that $Z \in C_c^1(\mathbb{R}^n, \mathbb{R}^n)$ such that $\mathrm{supp}(Z) \subset D \times \mathbb{R}$. Let $\delta_m(x) = m^n \rho(mx)$ where $\rho \in C_c^\infty(B(0,1),[0,\infty))$ such that $\int_{\mathbb{R}^n} \rho dm = 1$ and let $g_m := g * \delta_m$ defined on $D_{1-1/m}$ – the open ball of radius $1 - 1/m$ in $\mathbb{R}^{n-1}$ and let $\Omega_m := \{z \geq g_m\}$. For m large enough we will have $\mathrm{supp}(Z) \subset D_{1-1/m} \times \mathbb{R}$ and so by the divergence theorem we have already proved,

$$\int_{\Omega_m} \nabla \cdot Z dm = \int_{\partial \Omega_m} Z \cdot n d\sigma = \int_D Z(y, g_m(y)) \cdot (\nabla g_m(y), -1)\, dy.$$

Now

$$\left| 1_{z>g} - \lim_{m\to\infty} 1_{z>g_m} \right| \leq 1_{z=g(y)}$$

and by Fubini's theorem,

$$\int_{D\times\mathbb{R}} 1_{z=g(y)} dydz = \int_D dy \int_{\mathbb{R}} 1_{z=g(y)} dz = 0.$$

Hence by the dominated convergence theorem,

$$\lim_{m\to\infty} \int_{\Omega_m} \nabla \cdot Z dm = \lim_{m\to\infty} \int 1_{z>g_m} \nabla \cdot Z dm = \int \lim_{m\to\infty} 1_{z>g_m} \nabla \cdot Z dm$$
$$= \int 1_{z>g} \nabla \cdot Z dm = \int_\Omega \nabla \cdot Z dm.$$

Moreover we also have from results to be proved later in the course that $\nabla g(y)$ exists for a.e. y and is bounded by the Lipschitz constant K for g and

$$\nabla g_m = \nabla g * \delta_m \to \nabla g \text{ in } L^p_{loc} \text{ for any } 1 \leq p < \infty.$$

Therefore,

$$\lim_{m\to\infty} \int_D Z(y, g_m(y)) \cdot (\nabla g_m(y), -1)\, dy = \int_D Z(y, g(y)) \cdot (\nabla g(y), -1)\, dy$$
$$= \int_{\partial\Omega} Z \cdot n d\sigma$$

where $nd\sigma$ is the vector valued measure on $\partial\Omega$ determined in local coordinates by $(\nabla g_m(y), -1)\, dy$.

Finally if $Z \in C^1(\Omega^o) \cap C_c(\Omega)$ with $\int_\Omega |\nabla \cdot Z|\, dm < \infty$ with Ω as above. We can use the above result applied to the vector field $Z_\varepsilon(y,z) := Z(y, z+\varepsilon)$ which we may now view as an element of $C_c^1(\Omega)$. We then have

$$\int_\Omega \nabla \cdot Z(\cdot, \cdot + \varepsilon) dm = \int_D Z(y, g(y) + \varepsilon) \cdot (\nabla g(y), -1)\, dy$$
$$\to \int_D Z(y, g(y)) \cdot (\nabla g(y), -1)\, dy = \int_{\partial\Omega} Z \cdot n d\sigma. \quad (26.28)$$

And again by the dominated convergence theorem,

$$\lim_{\varepsilon\downarrow 0} \int_\Omega \nabla \cdot Z(\cdot, \cdot + \varepsilon) dm = \lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}^n} 1_\Omega(y,z) \nabla \cdot Z(y, z+\varepsilon) dydz$$
$$= \lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}^n} 1_\Omega(y, z-\varepsilon) \nabla \cdot Z(y,z) dydz$$
$$= \int_{\mathbb{R}^n} \lim_{\varepsilon\downarrow 0} 1_\Omega(y, z-\varepsilon) \nabla \cdot Z(y,z) dydz$$
$$= \int_{\mathbb{R}^n} 1_\Omega(y,z) \nabla \cdot Z(y,z) dydz = \int_\Omega \nabla \cdot Z dm \quad (26.29)$$

wherein we have used

$$\lim_{\varepsilon\downarrow 0} 1_\Omega(y, z-\varepsilon) = \lim_{\varepsilon\downarrow 0} 1_{z>g(y)+\varepsilon} = 1_{z>g(y)}.$$

Comparing Eqs. (26.28) and (26.29) finishes the proof of the extension.

26.6 Application to Holomorphic functions

Let $\Omega \subset \mathbb{C} \cong \mathbb{R}^2$ be a compact manifold with C^2 – boundary.

Definition 26.35. *Let $\Omega \subset \mathbb{C} \cong \mathbb{R}^2$ be a compact manifold with C^2 – boundary and $f \in C(\partial\Omega, \mathbb{C})$. The contour integral, $\int_{\partial\Omega} f(z)dz$, of f along $\partial\Omega$ is defined by*

$$\int_{\partial\Omega} f(z)dz := i \int_{\partial\Omega} f\ \mathbf{n}\ d\sigma$$

where $\mathbf{n} : \partial\Omega \to S^1 \subset \mathbb{C}$ is chosen so that $n := (\mathrm{Re}\,\mathbf{n}, \mathrm{Im}\,\mathbf{n})$ is the outward pointing normal, see Figure 26.12.

In order to carry out the integral in Definition 26.35 more effectively, suppose that $z = \gamma(t)$ with $a \leq t \leq b$ is a parametrization of a part of the boundary

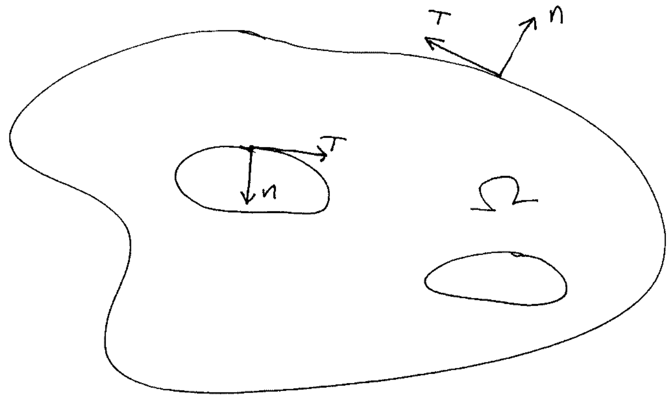


Fig. 26.12. The induced direction for countour integrals along boundaries of regions.

of Ω and γ is chosen so that $T := \dot{\gamma}(t)/|\dot{\gamma}(t)| = in(\gamma(t))$. That is to say T is gotten from n by a 90^o rotation in the counterclockwise direction. Combining this with $d\sigma = |\dot{\gamma}(t)|\,dt$ we see that

$$i\ n\ d\sigma = T\,|\dot{\gamma}(t)|\,dt = \dot{\gamma}(t)dt =: dz$$

so that

$$\int_\gamma f(z)dz = \int_a^b f(\gamma(t))\dot{\gamma}(t)dt.$$

Proposition 26.36. *Let $f \in C^1(\bar{\Omega}, \mathbb{C})$ and $\bar{\partial} := \frac{1}{2}(\partial_x + i\partial_y)$, then*

$$\int_{\partial\Omega} f(z)dz = 2i\int_\Omega \bar{\partial} f dm. \tag{26.30}$$

Now suppose $w \in \Omega$, then

$$f(w) = \frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(z)}{z-w}dz - \frac{1}{\pi}\int_\Omega \frac{\bar{\partial} f(z)}{z-w}dm(z). \tag{26.31}$$

Proof. By the divergence theorem,

$$\int_\Omega \bar{\partial} f dm = \frac{1}{2}\int_\Omega (\partial_x + i\partial_y) f dm = \frac{1}{2}\int_{\partial\Omega} f\,(n_1 + in_2)\,d\sigma$$
$$= \frac{1}{2}\int_{\partial\Omega} f n d\sigma = -\frac{i}{2}\int_{\partial\Omega} f(z)dz.$$

Given $\varepsilon > 0$ small, let $\Omega_\varepsilon := \Omega \setminus B(w,\varepsilon)$. Eq. (26.30) with $\Omega = \Omega_\varepsilon$ and f being replaced by $\frac{f(z)}{z-w}$ implies

$$\int_{\partial\Omega_\varepsilon}\frac{f(z)}{z-w}dz = 2i\int_{\Omega_\varepsilon}\frac{\bar{\partial} f}{z-w}dm \tag{26.32}$$

wherein we have used the product rule and the fact that $\bar{\partial}(z-w)^{-1} = 0$ to conclude

$$\bar{\partial}\left[\frac{f(z)}{z-w}\right] = \frac{\bar{\partial} f(z)}{z-w}.$$

Noting that $\partial\Omega_\varepsilon = \partial\Omega \cup \partial B(w,\varepsilon)$ and $\partial B(w,\varepsilon)$ may be parametrized by $z = w + \varepsilon e^{-i\theta}$ with $0 \le \theta \le 2\pi$, we have

$$\int_{\partial\Omega_\varepsilon}\frac{f(z)}{z-w}dz = \int_{\partial\Omega}\frac{f(z)}{z-w}dz + \int_0^{2\pi}\frac{f(w+\varepsilon e^{-i\theta})}{\varepsilon e^{-i\theta}}\left(-i\varepsilon\right)e^{-i\theta}d\theta$$
$$= \int_{\partial\Omega}\frac{f(z)}{z-w}dz - i\int_0^{2\pi} f(w+\varepsilon e^{-i\theta})d\theta$$

and hence

$$\int_{\partial\Omega}\frac{f(z)}{z-w}dz - i\int_0^{2\pi} f(w+\varepsilon e^{-i\theta})d\theta = 2i\int_{\Omega_\varepsilon}\frac{\bar{\partial} f(z)}{z-w}dm(z) \tag{26.33}$$

Since

$$\lim_{\varepsilon\downarrow 0}\int_0^{2\pi} f(w+\varepsilon e^{-i\theta})d\theta = 2\pi f(w)$$

and

$$\lim_{\varepsilon\downarrow 0}\int_{\Omega_\varepsilon}\frac{\bar{\partial} f}{z-w}dm = \int_\Omega \frac{\bar{\partial} f(z)}{z-w}dm(z).$$

we may pass to the limit in Eq. (26.33) to find

$$\int_{\partial\Omega}\frac{f(z)}{z-w}dz - 2\pi i f(w) = 2i\int_\Omega \frac{\bar{\partial} f(z)}{z-w}dm(z)$$

which is equivalent to Eq. (26.31). ∎

Remark 26.37. Eq. (26.31) implies $\bar{\partial}\frac{1}{z} = \pi\delta(z)$. Indeed if $f \in C_c^\infty\left(\mathbb{C} \cong \mathbb{R}^2\right)$, then by Eq. (26.31)

$$\langle\bar{\partial}\frac{1}{\pi z}, f\rangle := \langle\frac{1}{\pi z}, -\bar{\partial} f\rangle = -\frac{1}{\pi}\int_{\mathbb{C}}\frac{1}{z}\bar{\partial} f(z)dm(z) = f(0)$$

which is equivalent to $\bar{\partial}\frac{1}{z} = \pi\delta(z)$.

Exercise 26.3. Let Ω be as above and assume $f \in C^1(\bar{\Omega}, \mathbb{C})$ satisfies $g := \bar{\partial} f \in C^\infty(\Omega, \mathbb{C})$. Show $f \in C^\infty(\Omega, \mathbb{C})$. Hint, let $w_0 \in \Omega$ and $\varepsilon > 0$ be small and choose $\varphi \in C_c^\infty(B(z_0,\varepsilon))$ such that $\varphi = 1$ in a neighborhood of w_0 and let $\psi = 1-\varphi$. Then by Eq. (26.31),

$$f(w) = \frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(z)}{z-w}dz - \frac{1}{\pi}\int_\Omega \frac{g(z)}{z-w}\varphi(z)dm(z) - \frac{1}{\pi}\int_\Omega \frac{g(z)}{z-w}\psi(z)dm(z).$$

Now show each of the three terms above are smooth in w for w near w_0. To handle the middle term notice that

$$\int_\Omega \frac{g(z)}{z-w}\varphi(z)dm(z) = \int_{\mathbb{C}}\frac{g(z+w)}{z}\varphi(z+w)dm(z)$$

for w near w_0.

Definition 26.38. *A function $f \in C^1(\Omega,\mathbb{C})$ is said to be holomorphic if $\bar\partial f =$ 0.*

By Proposition 26.36, if $f \in C^1(\bar\Omega,\mathbb{C})$ and $\bar\partial f = 0$ on Ω, then **Cauchy's integral formula** holds for $w \in \Omega$, namely

$$f(w) = \frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(z)}{z-w}dz$$

and $f \in C^\infty(\Omega,\mathbb{C})$. For more details on Holomorphic functions, see the complex variable appendix.

26.7 Dirichlet Problems on D

BRUCE: This should be moved to the sections on Fourier Series.

Let $D := \{z \in \mathbb{C} : |z| < 1\}$ be the open unit disk in $\mathbb{C} \cong \mathbb{R}^2$, where we write $z = x+iy = re^{i\theta}$ in the usual way. Also let $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ and recall that Δ may be computed in polar coordinates by the formula,

$$\Delta u = r^{-1}\partial_r\left(r^{-1}\partial_r u\right) + \frac{1}{r^2}\partial_\theta^2 u. \tag{26.34}$$

Indeed if $v \in C_c^1(D)$, then

$$\begin{aligned}\int_D \Delta u v dm &= -\int_D \nabla u\cdot\nabla v dm = -\int 1_{0\le\theta\le 2\pi}1_{0\le r<1}\left(u_r v_r + \frac{1}{r^2}u_\theta v_\theta\right) r dr d\theta\\ &= \int 1_{0\le\theta\le 2\pi}1_{0\le r<1}\left((ru_r)_r v + \frac{1}{r}u_{\theta\theta}v\right) dr d\theta\\ &= \int 1_{0\le\theta\le 2\pi}1_{0\le r<1}\left(\frac{1}{r}(ru_r)_r + \frac{1}{r^2}u_{\theta\theta}\right) v r^2 dr d\theta\\ &= \int_D \left(\frac{1}{r}(ru_r)_r + \frac{1}{r^2}u_{\theta\theta}\right) v dm\end{aligned}$$

which shows Eq. (26.34) is valid. See Exercises 26.5 and 26.6 for more details.

Suppose that $u \in C(\bar D)\cap C^2(D)$ and $\Delta u(z) = 0$ for $z \in D$. Let $g = u|_{\partial D}$ and

$$A_k := \hat g(k) := \frac{1}{2\pi}\int_{-\pi}^{\pi} g(e^{ik\theta})e^{-ik\theta}d\theta.$$

(We are identifying $S^1 = \partial D := \left\{z \in \bar D : |z| = 1\right\}$ with $[-\pi,\pi]/(\pi \sim -\pi)$ by the map $\theta \in [-\pi,\pi] \to e^{i\theta} \in S^1$.) Let

$$\hat u(r,k) := \frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})e^{-ik\theta}d\theta \tag{26.35}$$

then:

1. $\tilde u(r,k)$ satisfies the ordinary differential equation
$$r^{-1}\partial_r\left(r\partial_r \hat u(r,k)\right) = \frac{1}{r^2}k^2\hat u(r,k) \text{ for } r \in (0,1).$$
2. Recall the general solution to
$$r\partial_r\left(r\partial_r y(r)\right) = k^2 y(r) \tag{26.36}$$
may be found by trying solutions of the form $y(r) = r^\alpha$ which then implies $\alpha^2 = k^2$ or $\alpha = \pm k$. From this one sees that $\tilde u(r,k)$ may be written as $\hat u(r,k) = A_k r^{|k|} + B_k r^{-|k|}$ for some constants A_k and B_k when $k \neq 0$. If $k = 0$, the solution to Eq. (26.36) is gotten by simple integration and the result is $\hat u(r,0) = A_0 + B_0\ln r$. Since $\hat u(r,k)$ is bounded near the origin for each k, it follows that $B_k = 0$ for all $k \in \mathbb{Z}$.
3. So we have shown
$$A_k r^{|k|} = \hat u(r,k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})e^{-ik\theta}d\theta$$
and letting $r \uparrow 1$ in this equation implies
$$A_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(e^{i\theta})e^{-ik\theta}d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta})e^{-ik\theta}d\theta.$$
Therefore,
$$u(re^{i\theta}) = \sum_{k\in\mathbb{Z}} A_k r^{|k|}e^{ik\theta} \tag{26.37}$$
for $r < 1$ or equivalently,
$$\begin{aligned}u(z) &= \sum_{k\in\mathbb{N}_0} A_k z^k + \sum_{k\in\mathbb{N}} A_{-k}\bar z^k = A_0 + \sum_{k\ge 1}A_k z^k + \sum_{k\ge 1}\overline{A_k}\bar z^k\\ &= \mathrm{Re}\left(A_0 + 2\sum_{k\ge 1}A_k z^k\right)\end{aligned}$$

In particular $\Delta u = 0$ implies $u(z)$ is the sum of a holomorphic and an anti-holomorphic functions and also that u is the real part of a holomorphic function $F(z) := A_0 + \frac{1}{2}\sum_{k\geq 1} A_k z^k$. The imaginary part $v(z) := \operatorname{Im} F(z)$ is harmonic as well and is given by

$$\begin{aligned} v(z) &= 2\operatorname{Im}\sum_{k\geq 1} A_k z^k = \frac{1}{i}\left(\sum_{k\geq 1} A_k z^k - \sum_{k\geq 1} \overline{A_k}\bar{z}^k\right) \\ &= \frac{1}{i}\left(\sum_{k\geq 1} A_k z^k - \sum_{k\geq 1} A_{-k}\bar{z}^k\right) \\ &= \frac{1}{i}\left(\sum_{k\geq 1} A_k r^k e^{ik\theta} - \sum_{k\geq 1} A_{-k} r^k e^{-ik\theta}\right) \\ &= \sum_{k\neq 0} \frac{1}{i}\operatorname{sgn}(k) A_k r^k e^{ik\theta} = -i\operatorname{sgn}(\frac{1}{i}\frac{d}{d\theta})u(z) \end{aligned}$$

wherein we are writing z as $re^{i\theta}$. Here $\operatorname{sgn}(\frac{1}{i}\frac{d}{d\theta})$ is the bounded self-adjoint operator on $L^2(S^1)$ which satisfies

$$\operatorname{sgn}(\frac{1}{i}\frac{d}{d\theta})e^{in\theta} = \operatorname{sgn}(n)e^{in\theta}$$

and

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

4. Inserting the formula for A_k into Eq. (26.37) gives

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{k\in\mathbb{Z}} r^{|k|}e^{ik(\theta-\alpha)}\right)u(e^{i\alpha})d\alpha \text{ for all } r < 1.$$

Now by simple geometric series considerations we find, setting $\delta = \theta - \alpha$, that

$$\begin{aligned} \sum_{k\in\mathbb{Z}} r^{|k|}e^{ik\delta} &= \sum_{k=0}^{\infty} r^k e^{ik\delta} + \sum_{k=0}^{\infty} r^k e^{-ik\delta} - 1 = 2\operatorname{Re}\sum_{k=0}^{\infty} r^k e^{ik\delta} - 1 \\ &= \operatorname{Re}\left[2\frac{1}{1-re^{i\delta}} - 1\right] = \operatorname{Re}\left[\frac{1+re^{i\delta}}{1-re^{i\delta}}\right] \\ &= \operatorname{Re}\left[\frac{\left(1+re^{i\delta}\right)\left(1-re^{-i\delta}\right)}{\left|1-re^{i\delta}\right|^2}\right] = \operatorname{Re}\left[\frac{1-r^2+2ir\sin\delta}{1-2r\cos\delta+r^2}\right] \qquad (26.38) \\ &= \frac{1-r^2}{\left|1-re^{i\delta}\right|^2} = \frac{1-r^2}{1-2r\cos\delta+r^2}. \end{aligned}$$

Putting this altogether we have shown

$$\begin{aligned} u(re^{i\theta}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta-\alpha)u(e^{i\alpha})d\alpha =: P_r * u(e^{i\theta}) \\ &= \frac{1}{2\pi}\operatorname{Re}\int_{-\pi}^{\pi}\frac{1+re^{i(\theta-\alpha)}}{1-re^{i(\theta-\alpha)}}u(e^{i\alpha})d\alpha \qquad (26.39) \end{aligned}$$

where

$$P_r(\delta) := \frac{1-r^2}{1-2r\cos\delta+r^2} \qquad (26.40)$$

is the so called Poisson kernel. The fact that $\frac{1}{2\pi}\operatorname{Re}\int_{-\pi}^{\pi} P_r(\theta)d\theta = 1$ follows from the fact that

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta)d\theta &= \operatorname{Re}\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k\in\mathbb{Z}} r^{|k|}e^{ik\theta}d\theta \\ &= \operatorname{Re}\frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\int_{-\pi}^{\pi} r^{|k|}e^{ik\theta}d\theta = 1. \end{aligned}$$

Writing $z = re^{i\theta}$, Eq. (26.39) may be rewritten as

$$u(z) = \frac{1}{2\pi}\operatorname{Re}\int_{-\pi}^{\pi}\frac{1+ze^{-i\alpha}}{1-ze^{-i\alpha}}u(e^{i\alpha})d\alpha$$

which shows $u = \operatorname{Re} F$ where

$$F(z) := \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1+ze^{-i\alpha}}{1-ze^{-i\alpha}}u(e^{i\alpha})d\alpha.$$

Moreover it follows from Eq. (26.38) that

$$\begin{aligned} \operatorname{Im} F(re^{i\theta}) &= \frac{1}{\pi}\operatorname{Im}\int_{-\pi}^{\pi}\frac{r\sin(\theta-\alpha)}{1-2r\cos(\theta-\alpha)+r^2}u(e^{i\alpha})d\alpha \\ &=: Q_r * u(e^{i\theta}) \end{aligned}$$

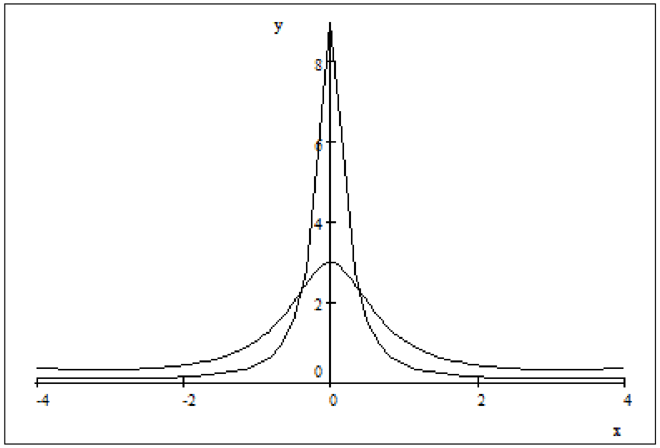


Fig. 26.13. Here is a plot of $p_r(x)$ for $r = .5$ and $r = .8$.

where

$$Q_r(\delta) := \frac{r\sin(\delta)}{1 - 2r\cos(\delta) + r^2}.$$

From these remarks it follows that v is the harmonic conjugate of u and $\tilde{P}_r = Q_r$. Summarizing these results gives

$$\tilde{f}(e^{i\theta}) = -i\operatorname{sgn}(\frac{1}{i}\frac{d}{d\theta})f(e^{i\theta}) = \lim_{r\uparrow 1}(Q_r * f)(e^{i\theta})$$

26.7.1 Appendix: More Proofs of Proposition 26.32

Exercise 26.4. $\det'(A)B = \det(A)\operatorname{tr}(A^{-1}B)$.

Proof. 2nd Proof of Proposition 26.32 by direct computation. Letting $A = \psi'$,

$$\begin{aligned}\frac{1}{\det A}\nabla\cdot(\det A\, Z^\psi) &= \frac{1}{\det A}\{Z^\psi\cdot\nabla\det A + \det A\ \nabla\cdot Z^\psi\}\\ &= \operatorname{tr}[A^{-1}\partial_{Z^\psi}A] + \nabla\cdot Z^\psi \qquad (26.41)\end{aligned}$$

and

$$\begin{aligned}\nabla\cdot Z^\psi &= \nabla\cdot(A^{-1}Z\circ\psi) = \partial_i(A^{-1}Z\circ\psi)\cdot e_i\\ &= e_i\cdot(-A^{-1}\partial_i A\,A^{-1}Z\circ\psi) + e_i\cdot A^{-1}(Z'\circ\psi)Ae_i\\ &= -e_i\cdot(A^{-1}\psi''\langle e_i, A^{-1}Z\circ\psi\rangle) + \operatorname{tr}(A^{-1}(Z'\circ\psi)A)\\ &= -e_i\cdot(A^{-1}\psi''\langle e_i, A^{-1}Z\circ\psi\rangle) + \operatorname{tr}(Z'\circ\psi)\\ &= -\operatorname{tr}(A^{-1}\psi''\langle Z^\psi, -\rangle) + (\nabla\cdot Z)\circ\psi\\ &= -\operatorname{tr}\left[A^{-1}\partial_{Z^\psi}A\right] + (\nabla\cdot Z)\circ\psi. \qquad (26.42)\end{aligned}$$

Combining Eqs. (26.41) and (26.42) gives the desired result:

$$\nabla\cdot(\det\psi'\, Z^\psi) = \det\psi'(\nabla\cdot Z)\circ\psi.$$

■

Lemma 26.39 (Flow interpretation of the divergence). *Let $Z \in C^1(\Omega, \mathbb{R}^n)$. Then*

$$\nabla\cdot Z = \frac{d}{dt}\Big|_0 \det(e^{tZ})'$$

and

$$\int_\Omega \nabla\cdot(f\ Z)dm = \frac{d}{dt}\Big|_0 \int_{e^{tZ}(\Omega)} f\ dm.$$

Proof. By Exercise 26.4 and the change of variables formula,

$$\frac{d}{dt}\Big|_0 \det(e^{tZ})' = \operatorname{tr}\left(\frac{d}{dt}\Big|_0 (e^{tZ})'\right) = \operatorname{tr}(Z') = \nabla\cdot Z$$

and

$$\begin{aligned}\frac{d}{dt}\Big|_0 \int_{e^{tZ}(\Omega)} f(x)dx &= \frac{d}{dt}\Big|_0 \int_\Omega f(e^{tZ}(y))\det(e^{tZ})'(y)dy\\ &= \int_\Omega \{\nabla f(y)\cdot Z(y) + f(y)\nabla\cdot Z(y)\}\ dy\\ &= \int_\Omega \nabla\cdot(f\ Z)\ dm.\end{aligned}$$

■

Proof. 3rd Proof of Proposition 26.32. Using Lemma 26.39 with $f = \det\psi'$ and $Z = Z^\psi$ and the change of variables formula,

$$\begin{aligned}
\int_{\Omega} \nabla \cdot (\det \psi' \, Z^{\psi}) dm &= \frac{d}{dt}\Big|_0 \int_{e^{tZ}(\Omega)} \det \psi' \; dm \\
&= \frac{d}{dt}\Big|_0 m(\psi \circ e^{tZ^{\psi}}(\Omega)) \\
&= \frac{d}{dt}\Big|_0 m(\psi \circ \psi^{-1} \circ e^{tZ} \circ \psi(\Omega)) \\
&= \frac{d}{dt}\Big|_0 m(e^{tZ}(\psi(\Omega))) \\
&= \frac{d}{dt}\Big|_0 \int_{e^{tZ}(\psi(\Omega))} 1 \; dm = \int_{\psi(\Omega)} \nabla \cdot Z dm \\
&= \int_{\Omega} (\nabla \cdot Z) \circ \psi \; \det \psi' \; dm.
\end{aligned}$$

Since this is true for all regions Ω, it follows that $\nabla \cdot (\det \psi' \, Z^{\psi}) = \det \psi' (\nabla \cdot Z^{\psi})$. ■

26.8 Exercises

See Exercises 12.5 as well

Exercise 26.5. Let $x = (x_1, \dots, x_n) = \psi(y_1, \dots, y_n) = \psi(y)$ be a C^2 – diffeomorphism, $\psi : V \to W$ where V and W are open subsets of $\mathbb{R}^n$. For $y \in V$ define

$$g_{ij}(y) = \frac{\partial \psi}{\partial y_i}(y) \cdot \frac{\partial \psi}{\partial y_j}(y)$$
$$g^{ij}(y) = (g_{ij}(y))^{-1}_{ij} \text{ and } \sqrt{g}(y) = \det(g_{ij}(y)).$$

Show

1. $g_{ij} = (\psi'^{tr}\psi')_{ij}$ and $\sqrt{g} = |\det \psi'|$. (So in the making the change of variables $x = \psi(y)$ we have $dx = \sqrt{g} dy$.)
2. Given functions $f, h \in C^1(W)$, let $f^{\psi} = f \circ \psi$ and $h^{\psi} = h \circ \psi$. Show
$$\nabla f(\psi) \cdot \nabla h(\psi) = g^{ij} \frac{\partial f^{\psi}}{\partial y_i} \frac{\partial h^{\psi}}{\partial y_j}.$$
3. For $f \in C^2(W)$, show
$$(\Delta f) \circ \psi = \frac{1}{\sqrt{g}} \frac{\partial}{\partial y_j} \left(\sqrt{g} g^{ij} \frac{\partial f^{\psi}}{\partial y_i} \right). \tag{26.43}$$

Hint: for $h \in C_c^2(W)$ compute we have
$$\int_W \Delta f(x) h(x) dx = -\int_W \nabla f(x) \cdot \nabla h(x) dx.$$
Now make the change of variables $x = \psi(y)$ in both of the above integrals and then do some more integration by parts to prove Eq. (26.43).

Notation 26.40 *We will usually abuse notation in the future and write Eq. (26.43) as*
$$\Delta f = \frac{1}{\sqrt{g}} \frac{\partial}{\partial y_j} \left(\sqrt{g} g^{ij} \frac{\partial f}{\partial y_i} \right).$$

Exercise 26.6. Let $\psi(\theta, \varphi_1, \dots, \varphi_{n-2}, r) = (x_1, \dots, x_n)$ where $(x_1, \dots, x_n)$ are as in Eq. (26.11). Show:

1. The vectors $\left\{ \frac{\partial \psi}{\partial \theta}, \frac{\partial \psi}{\partial \varphi_1}, \dots, \frac{\partial \psi}{\partial \varphi_{n-2}}, \frac{\partial \psi}{\partial r} \right\}$ form an orthogonal set and that
$$\left| \frac{\partial \psi}{\partial r} \right| = 1, \quad \left| \frac{\partial \psi}{\partial \varphi_{n-2}} \right| = r, \quad \left| \frac{\partial \psi}{\partial \theta} \right| = r \sin \varphi_{n-2} \dots \sin \varphi_1 \text{ and}$$
$$\left| \frac{\partial \psi}{\partial \varphi_j} \right| = r \sin \varphi_{n-2} \dots \sin \varphi_{j+1} \text{ for } j = 1, \dots, n-3.$$
2. Use item 1. to give another derivation of Eq. (26.13), i.e.
$$\sqrt{g} = |\det \psi'| = r^{n-1} \sin^{n-2} \varphi_{n-2} \dots \sin^2 \varphi_2 \sin \varphi_1$$
3. Use Eq. (26.43) to conclude
$$\Delta f = \frac{1}{r^{n-1}} \frac{\partial}{\partial r} \left(r^{n-1} \frac{\partial f}{\partial r} \right) + \frac{1}{r^2} \Delta_{S^{n-1}} f.$$
where
$$\Delta_{S^{n-1}} f := \sum_{j=1}^{n-2} \frac{1}{\sin^2 \varphi_{n-2} \dots \sin^2 \varphi_{j+1}} \frac{1}{\sin^j \varphi_j} \frac{\partial}{\partial \varphi_j} \left(\sin^j \varphi_j \frac{\partial f}{\partial \varphi_j} \right) + \frac{1}{\sin^2 \varphi_{n-2} \dots \sin^2 \varphi_1} \frac{\partial^2 f}{\partial \theta^2}$$
and
$$\frac{1}{\sin^2 \varphi_{n-2} \dots \sin^2 \varphi_{j+1}} := 1 \text{ if } j = n-2.$$
In particular if $f = F(r, \varphi_{n-2})$ we have

$$\Delta f = \frac{1}{r^{n-1}}\frac{\partial}{\partial r}\left(r^{n-1}\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{1}{\sin^{n-2}\varphi_{n-2}}\frac{\partial}{\partial \varphi_{n-2}}\left(\sin^{n-2}\varphi_{n-2}\frac{\partial f}{\partial \varphi_{n-2}}\right). \tag{26.44}$$

It is also worth noting that

$$\Delta_{S^{n-1}} f := \frac{1}{\sin^{n-2}\varphi_{n-2}}\frac{\partial}{\partial \varphi_{n-2}}\left(\sin^{n-2}\varphi_{n-2}\frac{\partial f}{\partial \varphi_{n-2}}\right) + \frac{1}{\sin^{n-2}\varphi_{n-2}}\Delta_{S^{n-2}} f.$$

Let us write $\psi := \varphi_{n-2}$ and suppose $f = r^\lambda w(\psi)$. According to Eq. (26.44),

$$\begin{aligned}
\Delta f &= \frac{1}{r^{n-1}}\frac{\partial}{\partial r}\left(r^{n-1}\frac{\partial\left(r^\lambda w(\psi)\right)}{\partial r}\right) + \frac{1}{r^2}\frac{1}{\sin^{n-2}\psi}\frac{\partial}{\partial \psi}\left(\sin^{n-2}\psi\frac{\partial\left(r^\lambda w(\psi)\right)}{\partial \psi}\right) \\
&= w(\psi)\frac{1}{r^{n-1}}\frac{\partial}{\partial r}\left(\lambda r^{n-1+\lambda-1}\right) + r^{\lambda-2}\frac{1}{\sin^{n-2}\psi}\frac{\partial}{\partial \psi}\left(\sin^{n-2}\psi\frac{\partial w}{\partial \psi}\right) \\
&= w(\psi)\lambda\left(n+\lambda-2\right)r^{\lambda-2} + r^{\lambda-2}\frac{1}{\sin^{n-2}\psi}\frac{\partial}{\partial \psi}\left(\sin^{n-2}\psi\frac{\partial w}{\partial \psi}\right) \\
&= r^{\lambda-2}\left[\lambda\left(n+\lambda-2\right)w(\psi) + \frac{1}{\sin^{n-2}\psi}\frac{\partial}{\partial \psi}\left(\sin^{n-2}\psi\frac{\partial w}{\partial \psi}\right)\right].
\end{aligned}$$

Write $w(\psi) = W(x)$ where $x = \cos\psi$, then $\frac{\partial w}{\partial \psi} = -W'(x)\sin\psi$ and hence

$$\begin{aligned}
\frac{1}{\sin^{n-2}\psi}\frac{\partial}{\partial \psi}\left(\sin^{n-2}\psi\frac{\partial w}{\partial \psi}\right) &= -\frac{1}{\sin^{n-2}\psi}\frac{\partial}{\partial \psi}\left(\sin^{n-1}\psi W'(x)\right) \\
&= \frac{-(n-1)\sin^{n-2}\psi\cos\psi W'(x)}{\sin^{n-2}\psi} - \frac{\sin^{n-1}\psi}{\sin^{n-2}\psi}\left\{-W''(x)\sin\psi\right\} \\
&= -(n-1)xW'(x) + (1-x^2)W''(x).
\end{aligned}$$

Hence we have shown, with $x = \cos\psi$ that

$$\Delta\left[r^\lambda W(x)\right] = r^{\lambda-2}\left[\lambda\left(n+\lambda-2\right)W(x) - (n-1)xW'(x) + (1-x^2)W''(x)\right].$$

27

Sard's Theorem

See p. 538 of Taylor and references. Also see Milnor's topology book. Add in the Brower's Fixed point theorem here as well. Also Spivak's calculus on manifolds.

Theorem 27.1. *Let $U \subset_o \mathbb{R}^m$, $f \in C^\infty(U, \mathbb{R}^d)$ and $C := \{x \in U : rank(f'(x)) < d\}$ be the set of critical points of f. Then the critical values, $f(C)$, is a Borel measurable subset of $\mathbb{R}^d$ of Lebesgue measure 0.*

Remark 27.2. This result clearly extends to manifolds.

For simplicity in the proof given below it will be convenient to use the norm, $|x| := \max_i |x_i|$. Recall that if $f \in C^1(U, \mathbb{R}^d)$ and $p \in U$, then

$$f(p+x) = f(p) + \int_0^1 f'(p+tx)xdt = f(p) + f'(p)x + \int_0^1 [f'(p+tx) - f'(p)]\, xdt$$

so that if

$$R(p,x) := f(p+x) - f(p) - f'(p)x = \int_0^1 [f'(p+tx) - f'(p)]\, xdt$$

we have

$$|R(p,x)| \le |x| \int_0^1 |f'(p+tx) - f'(p)|\, dt = |x|\, \varepsilon(p,x).$$

By uniform continuity, it follows for any compact subset $K \subset U$ that

$$\sup\{|\varepsilon(p,x)| : p \in K \text{ and } |x| \le \delta\} \to 0 \text{ as } \delta \downarrow 0.$$

Proof. (BRUCE: This proof needs to be gone through carefully. There are many misprints in the proof.) Notice that if $x \in U \setminus C$, then $f'(x) : \mathbb{R}^m \to \mathbb{R}^d$ is surjective, which is an open condition, so that $U \setminus C$ is an open subset of U. This shows C is relatively closed in U, i.e. there exists $\tilde{C} \sqsubset \mathbb{R}^m$ such that $C = \tilde{C} \cap U$. Let $K_n \subset U$ be compact subsets of U such that $K_n \uparrow U$, then $K_n \cap C \uparrow C$ and $K_n \cap C = K_n \cap \tilde{C}$ is compact for each n. Therefore, $f(K_n \cap C) \uparrow f(C)$ i.e. $f(C) = \cup_n f(K_n \cap C)$ is a countable union of compact sets and therefore is Borel measurable. Moreover, since $m(f(C)) = \lim_{n\to\infty} m(f(K_n \cap C))$, it suffices to show $m(f(K)) = 0$ for all compact subsets $K \subset C$.

Case 1. ($m \le d$) Let $K = [a, a+\gamma]$ be a cube contained in U and by scaling the domain we may assume $\gamma = (1,1,1,\dots,1)$. For $N \in \mathbb{N}$ and $j \in S_N := \{0, 1, \dots, N-1\}^d$ let $K_j = j/N + [a, a+\gamma/N]$ so that $K = \cup_{j\in S_N} K_j$ with $K_j^o \cap K_{j'}^o = \emptyset$ if $j \ne j'$. Let $\{Q_j : j = 1\dots, M\}$ be the collection of those $\{K_j : j \in S_N\}$ which intersect C. For each j, let $p_j \in Q_j \cap C$ and for $x \in Q_j - p_j$ we have

$$f(p_j + x) = f(p_j) + f'(p_j)x + R_j(x)$$

where $|R_j(x)| \le \varepsilon_j(N)/N$ and $\varepsilon(N) := \max_j \varepsilon_j(N) \to 0$ as $N \to \infty$. Now

$$\begin{aligned} m(f(Q_j)) &= m(f(p_j) + (f'(p_j) + R_j)(Q_j - p_j)) \\ &= m((f'(p_j) + R_j)(Q_j - p_j)) \\ &= m(O_j(f'(p_j) + R_j)(Q_j - p_j)) \end{aligned} \tag{27.1}$$

where $O_j \in SO(d)$ is chosen so that $O_j f'(p_j)\mathbb{R}^d \subset \mathbb{R}^{m-1} \times \{0\}$. Now $O_j f'(p_j)(Q_j - p_j)$ is contained in $\Gamma \times \{0\}$ where $\Gamma \subset \mathbb{R}^{m-1}$ is a cube centered at $0 \in \mathbb{R}^{m-1}$ with side length at most $2|f'(p_j)|/N \le 2M/N$ where $M = \max_{p\in K} |f'(p)|$. It now follows that $O_j(f'(p_j) + R_j)(Q_j - p_j)$ is contained the set of all points within $\varepsilon(N)/N$ of $\Gamma \times \{0\}$ and in particular

$$O_j(f'(p_j) + R_j)(Q_j - p_j) \subset (1 + \varepsilon(N)/N)\,\Gamma \times [\varepsilon(N)/N, \varepsilon(N)/N].$$

From this inclusion and Eq. (27.1) it follows that

$$\begin{aligned} m(f(Q_j)) &\le \left[2\frac{M}{N}(1 + \varepsilon(N)/N)\right]^{m-1} 2\varepsilon(N)/N \\ &= 2^m M^{m-1} [(1 + \varepsilon(N)/N)]^{m-1} \varepsilon(N) \frac{1}{N^m} \end{aligned}$$

and therefore,

$$\begin{aligned} m(f(C \cap K)) &\le \sum_j m(f(Q_j)) \le N^d 2^m M^{m-1} [(1 + \varepsilon(N)/N)]^{m-1} \varepsilon(N) \frac{1}{N^m} \\ &= 2^d M^{d-1} [(1 + \varepsilon(N)/N)]^{d-1} \varepsilon(N) \frac{1}{N^{m-d}} \to 0 \text{ as } N \to \infty \end{aligned}$$

since $m \ge d$. This proves the easy case since we may write U as a countable union of cubes K as above.

Remark. The case $(m < d)$ also follows from the case $m = d$ as follows. When $m < d$, $C = U$ and we must show $m(f(U)) = 0$. Letting $F : U \times \mathbb{R}^{d-m} \to \mathbb{R}^d$ be the map $F(x, y) = f(x)$. Then $F'(x, y)(v, w) = f'(x)v$, and hence $C_F := U \times \mathbb{R}^{d-m}$. So if the assertion holds for $m = d$ we have

$$m(f(U)) = m(F(U \times \mathbb{R}^{d-m})) = 0.$$

Case 2. $(m > d)$ This is the hard case and the case we will need in the co-area formula to be proved later. Here I will follow the proof in Milnor. Let

$$C_i := \{x \in U : \partial^\alpha f(x) = 0 \text{ when } |\alpha| \le i\}$$

so that $C \supset C_1 \supset C_2 \supset C_3 \supset \dots$. The proof is by induction on d and goes by the following steps:

1. $m(f(C \setminus C_1)) = 0$.
2. $m(f(C_i \setminus C_{i+1})) = 0$ for all $i \ge 1$.
3. $m(f(C_i)) = 0$ for all i sufficiently large.

Step 1. If $m = 1$, there is nothing to prove since $C = C_1$ so we may assume $m \ge 2$. Suppose that $x \in C \setminus C_1$, then $f'(p) \ne 0$ and so by reordering the components of x and $f(p)$ if necessary we may assume that $\partial_1 f_1(p) \ne 0$ where we are writing $\partial f(p)/\partial x_i$ as $\partial_i f(p)$. The map $h(x) := (f_1(x), x_2, \dots, x_d)$ has differential

$$h'(p) = \begin{bmatrix} \partial_1 f_1(p) & \partial_2 f_1(p) & \dots & \partial_d f_1(p) \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is not singular. So by the implicit function theorem, there exists there exists $V \in \tau_p$ such that $h : V \to h(V) \in \tau_{h(p)}$ is a diffeomorphism and in particular $\partial f_1(x)/\partial x_1 \ne 0$ for $x \in V$ and hence $V \subset U \setminus C_1$. Consider the map $g := f \circ h^{-1} : V' := h(V) \to \mathbb{R}^m$, which satisfies

$$(f_1(x), f_2(x), \dots, f_m(x)) = f(x) = g(h(x)) = g((f_1(x), x_2, \dots, x_d))$$

which implies $g(t, y) = (t, u(t, y))$ for $(t, y) \in V' := h(V) \in \tau_{h(p)}$, see Figure 27.1 below where $p = \bar{x}$ and $m = p$. Since

$$g'(t, y) = \begin{bmatrix} 1 & 0 \\ \partial_t u(t, y) & \partial_y u(t, y) \end{bmatrix}$$

it follows that (t, y) is a critical point of g iff $y \in C'_t$ – the set of critical points of $y \to u(t, y)$. Since h is a diffeomorphism we have $C' := h(C \cap V)$ are the critical points of g in V' and

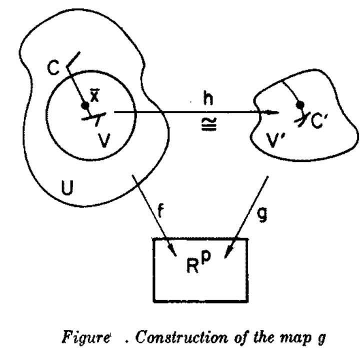


Fig. 27.1. Making a change of variable so as to apply induction.

$$f(C \cap V) = g(C') = \cup_t \left[\{t\} \times u_t(C'_t)\right].$$

By the induction hypothesis, $m_{m-1}(u_t(C'_t)) = 0$ for all t, and therefore by Fubini's theorem,

$$m(f(C \cap V)) = \int_{\mathbb{R}} m_{m-1}(u_t(C'_t)) 1_{V'_t \ne \emptyset} dt = 0.$$

Since $C \setminus C_1$ may be covered by a countable collection of open sets V as above, it follows that $m(f(C \setminus C_1)) = 0$. **Step 2.** Suppose that $p \in C_k \setminus C_{k+1}$, then there is an α such that $|\alpha| = k+1$ such that $\partial^\alpha f(p) = 0$ while $\partial^\beta f(p) = 0$ for all $|\beta| \le k$. Again by permuting coordinates we may assume that $\alpha_1 \ne 0$ and $\partial^\alpha f_1(p) \ne 0$. Let $w(x) := \partial^{\alpha - e_1} f_1(x)$, then $w(p) = 0$ while $\partial_1 w(p) \ne 0$. So again the implicit function theorem there exists $V \in \tau_p$ such that $h(x) := (w(x), x_2, \dots, x_d)$ maps $V \to V' := h(V) \in \tau_{h(p)}$ in a diffeomorphic way and in particular $\partial_1 w(x) \ne 0$ on V so that $V \subset U \setminus C_{k+1}$. As before, let $g := f \circ h^{-1}$ and notice that $C'_k := h(C_k \cap V) \subset \{0\} \times \mathbb{R}^{d-1}$ and

$$f(C_k \cap V) = g(C'_k) = \bar{g}(C'_k)$$

where $\bar{g} := g|_{(\{0\} \times \mathbb{R}^{d-1}) \cap V'}$. Clearly C'_k is contained in the critical points of $\bar{g}$, and therefore, by induction

$$0 = m(\bar{g}(C'_k)) = m(f(C_k \cap V)).$$

Since $C_k \setminus C_{k+1}$ is covered by a countable collection of such open sets, it follows that

$$m(f(C_k \setminus C_{k+1})) = 0 \text{ for all } k \ge 1.$$

Step 3. Suppose that Q is a closed cube with edge length δ contained in U and $k > d/m - 1$. We will show $m(f(Q \cap C_k)) = 0$ and since Q is arbitrary it will follows that $m(f(C_k)) = 0$ as desired. By Taylor's theorem with (integral) remainder, it follows for $x \in Q \cap C_k$ and h such that $x + h \in Q$ that

$$f(x+h) = f(x) + R(x,h)$$

where

$$|R(x,h)| \leq c \, \|h\|^{k+1}$$

where $c = c(Q,k)$. Now subdivide Q into r^d cubes of edge size δ/r and let Q' be one of the cubes in this subdivision such that $Q' \cap C_k \neq \emptyset$ and let $x \in Q' \cap C_k$. It then follows that $f(Q')$ is contained in a cube centered at $f(x) \in \mathbb{R}^m$ with side length at most $2c\,(\delta/r)^{k+1}$ and hence volume at most $(2c)^m\,(\delta/r)^{m(k+1)}$. Therefore, $f(Q \cap C_k)$ is contained in the union of at most r^d cubes of volume $(2c)^m\,(\delta/r)^{m(k+1)}$ and hence $m$each

$$m\left(f(Q \cap C_k)\right) \leq (2c)^m\,(\delta/r)^{m(k+1)}\,r^d = (2c)^m\,\delta^{m(k+1)} r^{d-m(k+1)} \to 0 \text{ as } r \uparrow \infty$$

provided that $d - m(k+1) < 0$, i.e. provided $k > d/m - 1$. ■

Part VI

More on L^p – Spaces

More Inequalities

28.1 Jensen's Inequality

Definition 28.1. *Given any function, $\varphi : (a, b) \to \mathbb{R}$, we say that φ is* ***convex*** *if for all $a < x_0 \le x_1 < b$ and $t \in [0, 1]$,*

$$\varphi(x_t) \le h_t := (1-t)\varphi(x_0) + t\varphi(x_1) \text{ for all } t \in [0,1], \tag{28.1}$$

where

$$x_t := x_0 + t(x_1 - x_0) = (1-t)x_0 + tx_1, \tag{28.2}$$

see Figure 28.1 below.

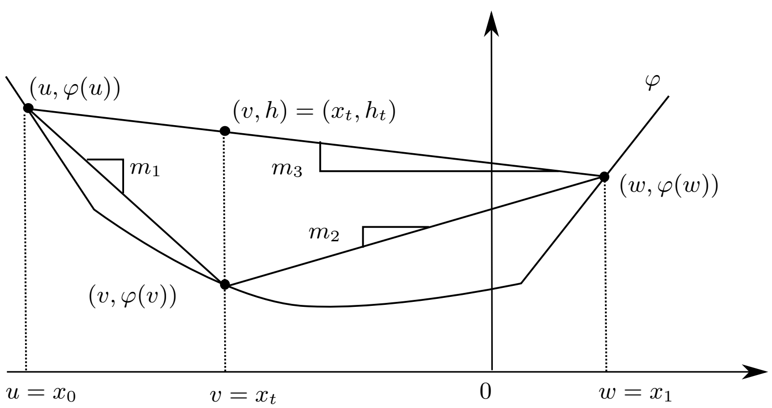


Fig. 28.1. A convex function with three cords. Notice the slope relationships; $m_1 \le m_3 \le m_2$.

Example 28.2. The functions $\exp(x)$ and $-\log(x)$ are convex and $|x|^p$ is convex iff $p \ge 1$ as follows from Corollary 28.5. below which, in part, states that any $\varphi \in C^2((a,b), \mathbb{R})$ such that $\varphi'' \ge 0$ is convex.

Given any function $\varphi : (a, b) \to \mathbb{R}$ and $x, y \in (a, b)$ with $x < y$, let[1]

$$F(x,y) = F_\varphi(x,y) := \frac{\varphi(y) - \varphi(x)}{y - x}. \tag{28.3}$$

In words, $F(x, y)$ is the slope of the line segment joining $(x, \varphi(x))$ to $(y, \varphi(y))$. The proof of the following elementary but useful lemma is left to the reader. The geometric meaning of this lemma is easily digested by referring to Figure 28.1.

Lemma 28.3. *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is a function and $a < u < v < w < b$ then each of the slope relations;*

$$F(u,v) \le F(u,w), \tag{28.4}$$
$$F(u,v) \le F(v,w), \text{ and} \tag{28.5}$$
$$F(u,w) \le F(v,w) \tag{28.6}$$

are equivalent to the inequality,

$$\varphi(v) \le \frac{w-v}{w-u}\varphi(u) + \frac{v-u}{w-u}\varphi(w). \tag{28.7}$$

Corollary 28.4. *Let $\varphi : (a, b) \to \mathbb{R}$ be a function then φ is convex iff any and hence all of the inequalities in Eqs. (28.4 – 28.7) hold for all $a < u < v < w < b$. We also have that φ is convex iff F is increasing in each of its arguments.*

Proof. The first assertion follows directly from Lemma 28.3. If φ is convex it follows from Eq. (28.4) that F is increasing in its second variable and from Eq. (28.6) that F is increasing in its first variable. Conversely if F is increasing in both of its variables and $a < u < v < w < b$ it follows (for example) that Eq. (28.4) holds and hence φ is convex. ∎

Corollary 28.5. *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is on once differentiable function. Then φ is convex iff $\varphi'(x)$ is non-decreasing in x. In particular if $\varphi \in C^2((a,b) \to \mathbb{R})$ then φ is convex iff $\varphi''(x) \ge 0$ for all $a < x < b$.*

Proof. If φ is convex and $a < u < v < b$ and $h > 0$ is such that $v + h < b$, then $F(u, u+h) \le F(v, v+h)$. Letting $h \downarrow 0$ shows $\varphi'(u) \le \varphi'(v)$, i.e. $.\varphi'$ is increasing. Conversely if φ' is increasing and $a < u < v < w < b$, then by the mean value theorem there exists $c \in (u, v)$ and $d \in (v, w)$ such that

[1] The same formula would define $F(x, y)$ for $x \neq y$. However, since $F(x, y) = F(y, x)$, we would gain no new information by this extension.

$$F(u,v) = \varphi'(c) \le \varphi'(d) = F(v,w)$$

which shows that Eq. (28.5) and hence φ is convex. ■

Theorem 28.6. *If $\varphi : (a,b) \to \mathbb{R}$ is a convex function, then for all $u \in (a,b)$ the limits*

$$\varphi'_+(u) = \frac{d}{du^+}\varphi(u) := \lim_{v \downarrow u} \frac{\varphi(v) - \varphi(u)}{v - u} \text{ and} \tag{28.8}$$

$$\varphi'_-(u) := \frac{d}{du^-}\varphi(u) = \lim_{v \uparrow u} \frac{\varphi(v) - \varphi(u)}{v - u} \tag{28.9}$$

exist and satisfy $-\infty < \varphi'_-(u) \le \varphi'_+(u) < \infty$. Furthermore $\varphi'_\pm$ are both increasing functions and further satisfy,

$$-\infty < \varphi'_-(u) \le \varphi'_+(u) \le \varphi'_-(v) < \infty \ \forall \ a < u < v < b. \tag{28.10}$$

Proof. Let $F = F_\varphi$ be as defined in Eq. (28.3). As F is increasing in each of its arguments the following monotone limits exist;

$$\varphi'_+(u) = F(u, u+) = \lim_{v \downarrow u} F(u,v) < \infty \text{ and}$$
$$\varphi'_-(u) = F(u-, u) := \lim_{v \uparrow u} F(v,u) > -\infty.$$

The monotonicity properties of F also implies

$$\varphi'_-(u) = F(u-, u) \le F(u, u+) = \varphi'_+(u)$$

and

$$\varphi'_+(u) = F(u, u+) \le F(v-, v) = \varphi'_-(v).$$

■

Theorem 28.7. *If $\varphi : (a,b) \to \mathbb{R}$ is a convex function then φ is continuous. In fact, if $a < \alpha < \beta < b$ and $K := \max\left\{\left|\varphi'_+(\alpha)\right|, \left|\varphi'_-(\beta)\right|\right\}$, then*

$$|\varphi(y) - \varphi(x)| \le K|y - x| \text{ for all } x, y \in [\alpha, \beta], \tag{28.11}$$

i.e. is φ is Lipschitz continuous on $[\alpha, \beta]$. Furthermore, for any $m \in \left[\varphi'_-(u), \varphi'_+(u)\right]$,

$$\varphi(x) \ge \varphi(u) + m(x - u) \text{ for all } a < x < b. \tag{28.12}$$

Proof. For $a < \alpha \le x < y \le \beta < b$, we have

$$\varphi'_+(\alpha) \le \varphi'_+(x) = F(x, x+) \le F(x,y) \le F(y-, y) = \varphi'_-(y) \le \varphi'_-(\beta). \tag{28.13}$$

From this inequality we may easily conclude that $|F(x,y)| \le K$ which is equivalent to Eq. (28.11).

Now let $m \in \left[\varphi'_-(u), \varphi'_+(u)\right]$. If $x \in (a,u)$ the monotonicity properties of F implies

$$\frac{\varphi(u) - \varphi(x)}{u - x} = F(x,u) \le F(u-, u) \le m$$

and solving this equality for $\varphi(x)$ (keep in mind $u - x < 0$) gives Eq. (28.12) for $a < x < u$. Similarly, for $x \in (u,b)$ we have

$$m \le F(u, u+) \le F(u,x) = \frac{\varphi(x) - \varphi(u)}{x - u}$$

which again gives Eq. (28.12). ■

Exercise 28.1. Suppose that φ is a C^1 – function on (a,b) such that $\varphi'(x)$ is non-decreasing. Give a direct proof of Eq. (28.12) with $m = \varphi'(u)$.

Theorem 28.8 (Jensen's Inequality). *Suppose that $(X, \mathcal{M}, \mu)$ is a probability space, i.e. μ is a positive measure and $\mu(X) = 1$. Also suppose that $f \in L^1(\mu)$, $f : X \to (a,b)$, and $\varphi : (a,b) \to \mathbb{R}$ is a convex function. Then*

$$\varphi\left(\int_X f d\mu\right) \le \int_X \varphi(f) d\mu$$

where if $\varphi \circ f \notin L^1(\mu)$, then $\varphi \circ f$ is integrable in the extended sense and $\int_X \varphi(f) d\mu = \infty$.

Proof. Let $u = \int_X f d\mu \in (a,b)$ and let $m \in \mathbb{R}$ be such that $\varphi(s) - \varphi(u) \ge m(s - u)$ for all $s \in (a,b)$. Then integrating the inequality, $\varphi(f) - \varphi(u) \ge m(f - u)$, implies that

$$0 \le \int_X \varphi(f) d\mu - \varphi(u) = \int_X \varphi(f) d\mu - \varphi(\int_X f d\mu).$$

Moreover, if $\varphi(f)$ is not integrable, then $\varphi(f) \ge \varphi(u) + m(f - u)$ which shows that negative part of $\varphi(f)$ is integrable. Therefore, $\int_X \varphi(f) d\mu = \infty$ in this case. ■

Example 28.9. The convex functions in Example 28.2 lead to the following inequalities,

$$\exp\left(\int_X f d\mu\right) \le \int_X e^f d\mu, \tag{28.14}$$

$$\int_X \log(|f|) d\mu \le \log\left(\int_X |f| d\mu\right)$$

and for $p \geq 1$,

$$\left|\int_X f d\mu\right|^p \leq \left(\int_X |f|\, d\mu\right)^p \leq \int_X |f|^p \, d\mu.$$

The last equation may also easily be derived using Hölder's inequality. As a special case of the first equation, we get another proof of Lemma 4.28. Indeed, more generally, suppose $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \ldots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \leq \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i} \tag{28.15}$$

where the inequality follows from Eq. (28.14) with $X = \{1, 2, \ldots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i} \delta_i$ and $f(i) := \ln s_i^{p_i}$. Of course Eq. (28.15) may be proved directly using the convexity of the exponential function.

Exercise 28.2. Use the inequality in Eq. (28.15) to give another proof of Corollary 16.3.

The next theorem may safely be skipped on your first reading.

Theorem 28.10 (Continuity properties of $\varphi_\pm$). *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is convex and $x, y \in (a, b)$ with $x < y$ and $\varphi'_\pm$ be the left and right derivatives of φ as in Theorem 28.7. Then;*

1. *The function φ'_+ is right continuous and φ'_- is left continuous.*
2. *The set of discontinuity points for φ'_+ and for φ'_- are the same as the set of points of non-differentiability of φ. Moreover this set is at most countable.*

Proof.

1. For $a < c < x < y < b$, we have $\varphi'_+(x) = F(x, x+) \leq F(x, y)$ and letting $x \downarrow c$ (using the continuity of F which follows from item 4.) we learn $\varphi'_+(c+) \leq F(c, y)$. We may now let $y \downarrow c$ to conclude $\varphi'_+(c+) \leq \varphi'_+(c)$. Since $\varphi'_+(c) \leq \varphi'_+(c+)$, it follows that $\varphi'_+(c) = \varphi'_+(c+)$ and hence that φ'_+ is right continuous.
Similarly, for $a < x < y < c < b$, we have $\varphi'_-(y) \geq F(x, y)$ and letting $y \uparrow c$ (using the continuity of F) we learn $\varphi'_-(c-) \geq F(x, c)$. Now let $x \uparrow c$ to conclude $\varphi'_-(c-) \geq \varphi'_-(c)$. Since $\varphi'_-(c) \geq \varphi'_-(c-)$, it follows that $\varphi'_-(c) = \varphi'_-(c-)$, i.e. φ'_- is left continuous.
2. Since $\varphi_\pm$ are increasing functions, they have at most countably many points of discontinuity. Letting $x \uparrow y$ in Eq. (28.10), using the left continuity of φ'_-, shows $\varphi'_-(y) = \varphi'_+(y-)$. Hence if φ'_- is continuous at y, $\varphi'_-(y) = \varphi'_-(y+) = \varphi'_+(y)$ and φ is differentiable at y. Conversely if φ is differentiable at y, then

$$\varphi'_+(y-) = \varphi'_-(y) = \varphi'(y) = \varphi'_+(y)$$

which shows φ'_+ is continuous at y. Thus we have shown that set of discontinuity points of φ'_+ is the same as the set of points of non-differentiability of φ. That the discontinuity set of φ'_- is the same as the non-differentiability set of φ is proved similarly.

■

28.2 Interpolation of L^p – spaces

The $L^p(\mu)$ – norm controls two types of behaviors of f, namely the "behavior at infinity" and the behavior of "local singularities." So in particular, if f blows up at a point $x_0 \in X$, then locally near x_0 it is harder for f to be in $L^p(\mu)$ as p increases. On the other hand a function $f \in L^p(\mu)$ is allowed to decay at "infinity" slower and slower as p increases. With these insights in mind, we should not in general expect $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$. However, there are two notable exceptions. (1) If $\mu(X) < \infty$, then there is no behavior at infinity to worry about and $L^q(\mu) \subset L^p(\mu)$ for all $q \geq p$ as is shown in Corollary 28.11 below. (2) If μ is counting measure, i.e. $\mu(A) = \#(A)$, then all functions in $L^p(\mu)$ for any p can not blow up on a set of positive measure, so there are no local singularities. In this case $L^p(\mu) \subset L^q(\mu)$ for all $q \geq p$, see Corollary 28.15 below.

Corollary 28.11. *If $\mu(X) < \infty$ and $0 < p < q \leq \infty$, then $L^q(\mu) \subset L^p(\mu)$, the inclusion map is bounded and in fact*

$$\|f\|_p \leq [\mu(X)]^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q.$$

Proof. Take $a \in [1, \infty]$ such that

$$\frac{1}{p} = \frac{1}{a} + \frac{1}{q}, \text{ i.e. } a = \frac{pq}{q - p}.$$

Then by Corollary 16.3,

$$\|f\|_p = \|f \cdot 1\|_p \leq \|f\|_q \cdot \|1\|_a = \mu(X)^{1/a} \|f\|_q = \mu(X)^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q.$$

The reader may easily check this final formula is correct even when $q = \infty$ provided we interpret $1/p - 1/\infty$ to be $1/p$. ■

Proposition 28.12. *Suppose that $0 < p_0 < p_1 \leq \infty$, $\lambda \in (0, 1)$ and $p_\lambda \in (p_0, p_1)$ be defined by*

$$\frac{1}{p_\lambda} = \frac{1 - \lambda}{p_0} + \frac{\lambda}{p_1} \tag{28.16}$$

with the interpretation that $\lambda/p_1 = 0$ *if* $p_1 = \infty$.[2] *Then* $L^{p_\lambda} \subset L^{p_0} + L^{p_1}$, *i.e. every function* $f \in L^{p_\lambda}$ *may be written as* $f = g + h$ *with* $g \in L^{p_0}$ *and* $h \in L^{p_1}$. *For* $1 \leq p_0 < p_1 \leq \infty$ *and* $f \in L^{p_0} + L^{p_1}$ *let*

$$\|f\| := \inf\left\{\|g\|_{p_0} + \|h\|_{p_1} : f = g + h\right\}.$$

Then $(L^{p_0} + L^{p_1}, \|\cdot\|)$ *is a Banach space and the inclusion map from* L^{p_λ} *to* $L^{p_0} + L^{p_1}$ *is bounded; in fact* $\|f\| \leq 2\|f\|_{p_\lambda}$ *for all* $f \in L^{p_\lambda}$.

Proof. Let $M > 0$, then the local singularities of f are contained in the set $E := \{|f| > M\}$ and the behavior of f at "infinity" is solely determined by f on E^c. Hence let $g = f1_E$ and $h = f1_{E^c}$ so that $f = g + h$. By our earlier discussion we expect that $g \in L^{p_0}$ and $h \in L^{p_1}$ and this is the case since,

$$\|g\|_{p_0}^{p_0} = \int |f|^{p_0} 1_{|f|>M} = M^{p_0} \int \left|\frac{f}{M}\right|^{p_0} 1_{|f|>M}$$
$$\leq M^{p_0} \int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|>M} \leq M^{p_0 - p_\lambda} \|f\|_{p_\lambda}^{p_\lambda} < \infty$$

and

$$\|h\|_{p_1}^{p_1} = \left\|f 1_{|f|\leq M}\right\|_{p_1}^{p_1} = \int |f|^{p_1} 1_{|f|\leq M} = M^{p_1} \int \left|\frac{f}{M}\right|^{p_1} 1_{|f|\leq M}$$
$$\leq M^{p_1} \int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|\leq M} \leq M^{p_1 - p_\lambda} \|f\|_{p_\lambda}^{p_\lambda} < \infty.$$

Moreover this shows

$$\|f\| \leq M^{1-p_\lambda/p_0} \|f\|_{p_\lambda}^{p_\lambda/p_0} + M^{1-p_\lambda/p_1} \|f\|_{p_\lambda}^{p_\lambda/p_1}.$$

Taking $M = \alpha \|f\|_{p_\lambda}$ with $\alpha > 0$ implies

$$\|f\| \leq \left(\alpha^{1-p_\lambda/p_0} + \alpha^{1-p_\lambda/p_1}\right) \|f\|_{p_\lambda}$$

and then taking $\alpha = 1$ shows $\|f\| \leq 2\|f\|_{p_\lambda}$. The proof that $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space is left as Exercise 28.7 to the reader. ■

Corollary 28.13 (Interpolation of L^p – norms). *Suppose that* $0 < p_0 < p_1 \leq \infty$, $\lambda \in (0,1)$ *and* $p_\lambda \in (p_0, p_1)$ *be defined as in Eq. (28.16), then* $L^{p_0} \cap L^{p_1} \subset L^{p_\lambda}$ *and*

$$\|f\|_{p_\lambda} \leq \|f\|_{p_0}^{1-\lambda} \|f\|_{p_1}^{\lambda}. \tag{28.17}$$

Further assume $1 \leq p_0 < p_\lambda < p_1 \leq \infty$, *and for* $f \in L^{p_0} \cap L^{p_1}$ *let*

$$\|f\| := \|f\|_{p_0} + \|f\|_{p_1}.$$

Then $(L^{p_0} \cap L^{p_1}, \|\cdot\|)$ *is a Banach space and the inclusion map of* $L^{p_0} \cap L^{p_1}$ *into* L^{p_λ} *is bounded, in fact*

$$\|f\|_{p_\lambda} \leq \max\left(\lambda^{-1}, (1-\lambda)^{-1}\right)\left(\|f\|_{p_0} + \|f\|_{p_1}\right). \tag{28.18}$$

The heuristic explanation of this corollary is that if $f \in L^{p_0} \cap L^{p_1}$, then f has local singularities no worse than an L^{p_1} function and behavior at infinity no worse than an L^{p_0} function. Hence $f \in L^{p_\lambda}$ for any p_λ between p_0 and p_1.

Proof. Let λ be determined as above, $a = p_0/(1-\lambda)$ and $b = p_1/\lambda$, then by Corollary 16.3,

$$\|f\|_{p_\lambda} = \left\||f|^\lambda |f|^{1-\lambda}\right\|_{p_\lambda} \leq \left\||f|^{1-\lambda}\right\|_a \left\||f|^\lambda\right\|_b = \|f\|_{p_0}^{1-\lambda} \|f\|_{p_1}^{\lambda}.$$

which proves Eq. (28.17).

It is easily checked that $\|\cdot\|$ is a norm on $L^{p_0} \cap L^{p_1}$. To show this space is complete, suppose that $\{f_n\} \subset L^{p_0} \cap L^{p_1}$ is a $\|\cdot\|$ – Cauchy sequence. Then $\{f_n\}$ is both L^{p_0} and L^{p_1} – Cauchy. Hence there exist $f \in L^{p_0}$ and $g \in L^{p_1}$ such that $\lim_{n\to\infty} \|f - f_n\|_{p_0} = 0$ and $\lim_{n\to\infty} \|g - f_n\|_{p_\lambda} = 0$. By Chebyshev's inequality (Lemma 16.18) $f_n \to f$ and $f_n \to g$ in measure and therefore by Theorem 16.20, $f = g$ a.e. It now is clear that $\lim_{n\to\infty} \|f - f_n\| = 0$. The estimate in Eq. (28.18) is left as Exercise 28.6 to the reader. ■

Exercise 28.3. Show that Eq. (28.17) may be alternatively stated by saying that $\varphi(p) := \ln \|f\|_p^p$ is a convex function of p.

Exercise 28.4. Give a second proof of that Eq. (28.17) may be alternatively stated by saying that $\varphi(p) := \ln \|f\|_p^p$ is a convex function of p by explicitly computing $\varphi''(p)$ when f is nice. Then pass to the limit to get the general case from these consideration.

Remark 28.14. Combining Proposition 28.12 and Corollary 28.13 gives

$$L^{p_0} \cap L^{p_1} \subset L^{p_\lambda} \subset L^{p_0} + L^{p_1}$$

for $0 < p_0 < p_1 \leq \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ as in Eq. (28.16).

Corollary 28.15. *Suppose now that* μ *is counting measure on* X. *Then* $L^p(\mu) \subset L^q(\mu)$ *for all* $0 < p < q \leq \infty$ *and* $\|f\|_q \leq \|f\|_p$.

[2] A little algebra shows that λ may be computed in terms of p_0, p_λ and p_1 by

$$\lambda = \frac{p_0}{p_\lambda} \cdot \frac{p_1 - p_\lambda}{p_1 - p_0}.$$

Proof. Suppose that $0 < p < q = \infty$, then

$$\|f\|_\infty^p = \sup\{|f(x)|^p : x \in X\} \le \sum_{x\in X} |f(x)|^p = \|f\|_p^p,$$

i.e. $\|f\|_\infty \le \|f\|_p$ for all $0 < p < \infty$. For $0 < p \le q \le \infty$, apply Corollary 28.13 with $p_0 = p$ and $p_1 = \infty$ and take $q = p_\lambda$ ($\frac{1}{q} = \frac{1-\lambda}{p}$ or $\lambda = 1 - p/q$) to find

$$\|f\|_q \le \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} \le \|f\|_p^{p/q} \|f\|_p^{1-p/q} = \|f\|_p.$$

■

28.2.1 Summary:

1. Since $\mu(|f| > \varepsilon) \le \varepsilon^{-p} \|f\|_p^p$, L^p – convergence implies L^0 – convergence.
2. L^0 – convergence implies almost everywhere convergence for some subsequence.
3. If $\mu(X) < \infty$ then almost everywhere convergence implies uniform convergence off certain sets of small measure and in particular we have L^0 – convergence.
4. If $\mu(X) < \infty$, then $L^q \subset L^p$ for all $p \le q$ and L^q – convergence implies L^p – convergence.
5. $L^{p_0} \cap L^{p_1} \subset L^q \subset L^{p_0} + L^{p_1}$ for any $q \in (p_0, p_1)$.
6. If $p \le q$, then $\ell^p \subset \ell^q$ and $\|f\|_q \le \|f\|_p$.

28.3 Exercises

Exercise 28.5. Let $f \in L^p \cap L^\infty$ for some $p < \infty$. Show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$. If we further assume $\mu(X) < \infty$, show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$ for all measurable functions $f : X \to \mathbb{C}$. In particular, $f \in L^\infty$ iff $\lim_{q\to\infty} \|f\|_q < \infty$. **Hints:** Use Corollary 28.13 on interpolation of L^p – norms to show $\limsup_{q\to\infty} \|f\|_q \le \|f\|_\infty$ and to show $\liminf_{q\to\infty} \|f\|_q \ge \|f\|_\infty$, let $M < \|f\|_\infty$ and make use of Chebyshev's inequality.

Exercise 28.6 (Part of Folland 6.3 on p. 186.). Prove Eq. (28.18) in Corollary 28.13. In detail suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0, p_1)$ be defined by

$$\frac{1}{p_\lambda} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1}$$

as in Eq. (28.16). Show

$$\|f\|_{p_\lambda} \le \max(\lambda, (1-\lambda)) \left(\|f\|_{p_0} + \|f\|_{p_1}\right).$$

Hint: Use the inequality

$$st \le \frac{s^a}{a} + \frac{t^b}{b},$$

where $a, b \ge 1$ with $a^{-1} + b^{-1} = 1$ are chosen appropriately,(see Lemma 4.28 for Eq. (28.15)) applied to the right side of the interpolation inequality;

$$\|f\|_{p_\lambda} \le \|f\|_{p_0}^{1-\lambda} \|f\|_{p_1}^{\lambda}. \tag{28.19}$$

Exercise 28.7. Complete the proof of Proposition 28.12 by showing $(L^p + L^r, \|\cdot\|)$ is a Banach space. **Hint:** you may find using Theorem 14.18 (on the sum – criteria for completeness) is helpful here.

Exercise 28.8 (Folland 6.5 on p. 186.). Suppose $0 < p < q \le \infty$. Then $L^p \nsubseteq L^q$ iff X contains sets of arbitrarily small positive measure. Also $L^q \nsubseteq L^p$ iff X contains sets of of arbitrarily large finite measure.

Exercise 28.9. Folland 6.27 on p. 196. **Hint**: Theorem 29.4.

29

The Dual of L^p

29.1 Converse of Hölder's Inequality

Exercise 29.1. Let $(X, \mathcal{B}, \mu)$ be a measure space and $g \in L^1(\mu)$. Show that $\left|\int_X g d\mu\right| = \|g\|_1$ iff there exists $z \in \mathbb{C}$ with $|z| = 1$ such that $g = |g| z$ a.e. (This may be equivalently stated as $\operatorname{sgn}(g(x)) := g(x) / |g(x)|$ is constant for μ -a.e. on the set where $g \neq 0$.) In particular for $f \in L^p(\mu)$ and $g \in L^q(\mu)$ with $q = p/(p-1)$, we have,

$$\left|\int_X fg d\mu\right| = \int_X |fg| \, d\mu \iff \operatorname{sgn}(f(x)) = \overline{\operatorname{sgn}(g(x))} \text{ for } \mu\text{–a.e. } x \in \{fg \neq 0\}.$$

Throughout this section we assume $(X, \mathcal{M}, \mu)$ is a σ – finite measure space, $q \in [1, \infty]$ and $p \in [1, \infty]$ are conjugate exponents, i.e. $p^{-1} + q^{-1} = 1$. For $g \in L^q$, let $\varphi_g \in (L^p)^*$ be given by

$$\varphi_g(f) = \int gf \; d\mu =: \langle g, f\rangle. \tag{29.1}$$

By Hölder's inequality

$$|\varphi_g(f)| \le \int |gf| d\mu \le \|g\|_q \|f\|_p \tag{29.2}$$

which implies that

$$\|\varphi_g\|_{(L^p)^*} := \sup\{|\varphi_g(f)| : \|f\|_p = 1\} \le \|g\|_q. \tag{29.3}$$

Proposition 29.1 (Converse of Hölder's Inequality). *Let $(X, \mathcal{M}, \mu)$ be a σ – finite measure space and $1 \le p \le \infty$ as above. For all $g \in L^q$,*

$$\|g\|_q = \|\varphi_g\|_{(L^p)^*} := \sup\left\{|\varphi_g(f)| : \|f\|_p = 1\right\} \tag{29.4}$$

and for any measurable function $g : X \to \mathbb{C}$,

$$\|g\|_q = \sup\left\{\int_X |g| \, f d\mu : \|f\|_p = 1 \text{ and } f \ge 0\right\}. \tag{29.5}$$

(In Theorem 29.6 below we will see that every element of $(L^p)^*$ is of the form φ_g for some $g \in L^q$.) Moreover, Eq. (29.4) holds for arbitrary measure spaces if $1 < p < \infty$ and $g \in L^q(\mu)$.

Proof. We begin by proving Eq. (29.4). Assume first that $q < \infty$ so $p > 1$. Then

$$|\varphi_g(f)| = \left|\int gf \; d\mu\right| \le \int |gf| \, d\mu \le \|g\|_q \|f\|_p$$

and equality occurs in the first inequality when $\operatorname{sgn}(gf)$ is constant a.e. while equality in the second occurs, by Theorem 16.1, when $|f|^p = c|g|^q$ for some constant $c > 0$. So let $f := \overline{\operatorname{sgn}(g)}|g|^{q/p}$ which for $p = \infty$ is to be interpreted as $f = \overline{\operatorname{sgn}(g)}$, i.e. $|g|^{q/\infty} \equiv 1$. When $p = \infty$,

$$|\varphi_g(f)| = \int_X g \, \overline{\operatorname{sgn}(g)} d\mu = \|g\|_{L^1(\mu)} = \|g\|_1 \|f\|_\infty$$

which shows that $\|\varphi_g\|_{(L^\infty)^*} \ge \|g\|_1$. If $p < \infty$, then

$$\|f\|_p^p = \int |f|^p = \int |g|^q = \|g\|_q^q$$

while

$$\varphi_g(f) = \int gf d\mu = \int |g||g|^{q/p} d\mu = \int |g|^q d\mu = \|g\|_q^q.$$

Hence

$$\frac{|\varphi_g(f)|}{\|f\|_p} = \frac{\|g\|_q^q}{\|g\|_q^{q/p}} = \|g\|_q^{q(1-\frac{1}{p})} = \|g\|_q.$$

This shows that $\|\varphi_g\| \ge \|g\|_q$ which combined with Eq. (29.3) implies Eq. (29.4).

The last case to consider is $p = 1$ and $q = \infty$. Let $M := \|g\|_\infty$ and choose $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ as $n \to \infty$ and $\mu(X_n) < \infty$ for all n. For any $\varepsilon > 0$, $\mu(|g| \ge M - \varepsilon) > 0$ and $X_n \cap \{|g| \ge M - \varepsilon\} \uparrow \{|g| \ge M - \varepsilon\}$. Therefore, $\mu(X_n \cap \{|g| \ge M - \varepsilon\}) > 0$ for n sufficiently large. Let

$$f = \overline{\operatorname{sgn}(g)} 1_{X_n \cap \{|g| \ge M - \varepsilon\}},$$

then

$$\|f\|_1 = \mu(X_n \cap \{|g| \ge M - \varepsilon\}) \in (0, \infty)$$

and

$$|\varphi_g(f)| = \int_{X_n\cap\{|g|\ge M-\varepsilon\}} \overline{\operatorname{sgn}(g)} g d\mu = \int_{X_n\cap\{|g|\ge M-\varepsilon\}} |g| d\mu$$
$$\ge (M-\varepsilon)\mu(X_n\cap\{|g|\ge M-\varepsilon\}) = (M-\varepsilon)\|f\|_1.$$

Since $\varepsilon > 0$ is arbitrary, it follows from this equation that $\|\varphi_g\|_{(L^1)^*} \ge M = \|g\|_\infty$.

Now for the proof of Eq. (29.5). The key new point is that we no longer are assuming that $g \in L^q$. Let $M(g)$ denote the right member in Eq. (29.5) and set $g_n := 1_{X_n\cap\{|g|\le n\}} g$. Then $|g_n| \uparrow |g|$ as $n \to \infty$ and it is clear that $M(g_n)$ is increasing in n. Therefore using Lemma 4.10 and the monotone convergence theorem,

$$\begin{aligned}
\lim_{n\to\infty} M(g_n) &= \sup_n M(g_n) = \sup_n \sup\left\{\int_X |g_n|\, f d\mu : \|f\|_p = 1 \text{ and } f \ge 0\right\} \\
&= \sup\left\{\sup_n \int_X |g_n|\, f d\mu : \|f\|_p = 1 \text{ and } f \ge 0\right\} \\
&= \sup\left\{\lim_{n\to\infty} \int_X |g_n|\, f d\mu : \|f\|_p = 1 \text{ and } f \ge 0\right\} \\
&= \sup\left\{\int_X |g|\, f d\mu : \|f\|_p = 1 \text{ and } f \ge 0\right\} = M(g).
\end{aligned}$$

Since $g_n \in L^q$ for all n and $M(g_n) = \|\varphi_{g_n}\|_{(L^p)^*}$ (as you should verify), it follows from Eq. (29.4) that $M(g_n) = \|g_n\|_q$. When $q < \infty$ (by the monotone convergence theorem) and when $q = \infty$ (directly from the definitions) one learns that $\lim_{n\to\infty} \|g_n\|_q = \|g\|_q$. Combining this fact with $\lim_{n\to\infty} M(g_n) = M(g)$ just proved shows $M(g) = \|g\|_q$. ■

As an application we can derive a sweeping generalization of Minkowski's inequality. (See Reed and Simon, Vol II. Appendix IX.4 for a more thorough discussion of complex interpolation theory.)

Theorem 29.2 (Minkowski's Inequality for Integrals). *Let $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ be σ – finite measure spaces and $1 \le p \le \infty$. If f is a $\mathcal{M}\otimes\mathcal{N}$ measurable function, then $y \to \|f(\cdot,y)\|_{L^p(\mu)}$ is measurable and*

1. *if f is a positive $\mathcal{M}\otimes\mathcal{N}$ measurable function, then*
$$\left\|\int_Y f(\cdot,y) d\nu(y)\right\|_{L^p(\mu)} \le \int_Y \|f(\cdot,y)\|_{L^p(\mu)} d\nu(y). \tag{29.6}$$
2. *If $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M}\otimes\mathcal{N}$ measurable function and $\int_Y \|f(\cdot,y)\|_{L^p(\mu)} d\nu(y) < \infty$ then*
   a) for μ – a.e. x, $f(x,\cdot) \in L^1(\nu)$,
   b) the μ –a.e. defined function, $x \to \int_Y f(x,y)d\nu(y)$, is in $L^p(\mu)$ and
   c) the bound in Eq. (29.6) holds.

Proof. For $p \in [1,\infty]$, let $F_p(y) := \|f(\cdot,y)\|_{L^p(\mu)}$. If $p \in [1,\infty)$

$$F_p(y) = \|f(\cdot,y)\|_{L^p(\mu)} = \left(\int_X |f(x,y)|^p \, d\mu(x)\right)^{1/p}$$

is a measurable function on Y by Fubini's theorem. To see that F_∞ is measurable, let $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\mu(X_n) < \infty$ for all n. Then by Exercise 28.5,

$$F_\infty(y) = \lim_{n\to\infty}\lim_{p\to\infty} \|f(\cdot,y)1_{X_n}\|_{L^p(\mu)}$$

which shows that F_∞ is $(Y,\mathcal{N})$ – measurable as well. This shows that integral on the right side of Eq. (29.6) is well defined.

Now suppose that $f \ge 0$, $q = p/(p-1)$and $g \in L^q(\mu)$ such that $g \ge 0$ and $\|g\|_{L^q(\mu)} = 1$. Then by Tonelli's theorem and Hölder's inequality,

$$\begin{aligned}
\int_X \left[\int_Y f(x,y)d\nu(y)\right] g(x) d\mu(x) &= \int_Y d\nu(y) \int_X d\mu(x) f(x,y) g(x) \\
&\le \|g\|_{L^q(\mu)} \int_Y \|f(\cdot,y)\|_{L^p(\mu)} d\nu(y) \\
&= \int_Y \|f(\cdot,y)\|_{L^p(\mu)} d\nu(y).
\end{aligned}$$

Therefore by the converse to Hölder's inequality (Proposition 29.1),

$$\begin{aligned}
&\left\|\int_Y f(\cdot,y)d\nu(y)\right\|_{L^p(\mu)} \\
&\quad = \sup\left\{\int_X \left[\int_Y f(x,y)d\nu(y)\right] g(x)d\mu(x) : \|g\|_{L^q(\mu)} = 1 \text{ and } g \ge 0\right\} \\
&\quad \le \int_Y \|f(\cdot,y)\|_{L^p(\mu)} d\nu(y)
\end{aligned}$$

proving Eq. (29.6) in this case.

Now let $f : X \times Y \to \mathbb{C}$ be as in item 2) of the theorem. Applying the first part of the theorem to $|f|$ shows

$$\int_Y |f(x,y)|\, d\nu(y) < \infty \text{ for } \mu\text{– a.e. } x,$$

i.e. $f(x,\cdot) \in L^1(\nu)$ for the μ –a.e. x. Since $\left|\int_Y f(x,y)d\nu(y)\right| \le \int_Y |f(x,y)|\, d\nu(y)$ it follows by item 1) that

$$\left\| \int_Y f(\cdot,y)d\nu(y) \right\|_{L^p(\mu)} \le \left\| \int_Y |f(\cdot,y)|\, d\nu(y) \right\|_{L^p(\mu)} \le \int_Y \|f(\cdot,y)\|_{L^p(\mu)}\, d\nu(y).$$

Hence the function, $x \in X \to \int_Y f(x,y)d\nu(y)$, is in $L^p(\mu)$ and the bound in Eq. (29.6) holds. ■

Example 29.3. Suppose $p \in [1,\infty]$, $u \in L^1(\mathbb{R}^n, m)$ and $v \in L^p(\mathbb{R}^n, m)$, then $u * v(x) := \int_{\mathbb{R}^n} u(y)\, v(x-y)\, dy$ exists for almost every x, $u * v \in L^p$ and

$$\|u * v\|_p \le \|u\|_1 \|v\|_p .$$

Take $f(x,y) := u(y)\, v(x-y)$ and observe that

$$\int_{\mathbb{R}^n} \|f(\cdot,y)\|_p\, dy = \int_{\mathbb{R}^n} |u(y)|\, \|v\|_p\, dy = \|u\|_1 \cdot \|v\|_p < \infty.$$

The result now follows from Theorem 29.2.

Here is another application of Minkowski's inequality for integrals. In this theorem we will be using the convention that $x^{-1/\infty} := 1$.

Theorem 29.4 (Theorem 6.20 in Folland). *Suppose that $k : (0,\infty) \times (0,\infty) \to \mathbb{C}$ is a measurable function such that k is homogenous of degree -1, i.e. $k(\lambda x, \lambda y) = \lambda^{-1} k(x,y)$ for all $\lambda > 0$. If, for some $p \in [1,\infty]$,*

$$C_p := \int_0^\infty |k(x,1)|\, x^{-1/p} dx < \infty$$

then for $f \in L^p((0,\infty), m)$, $k(x,\cdot)f(\cdot) \in L^1((0,\infty), m)$ for m – a.e. x. Moreover, the m – a.e. defined function

$$(Kf)(x) = \int_0^\infty k(x,y)f(y)dy \tag{29.7}$$

is in $L^p((0,\infty), m)$ and

$$\|Kf\|_{L^p((0,\infty),m)} \le C_p \|f\|_{L^p((0,\infty),m)}.$$

Proof. By the homogeneity of k, $k(x,y) = x^{-1}k(1, \frac{y}{x})$. Using this relation and making the change of variables, $y = zx$, gives

$$\int_0^\infty |k(x,y)f(y)|\, dy = \int_0^\infty x^{-1} \left| k(1, \frac{y}{x}) f(y) \right| dy$$
$$= \int_0^\infty x^{-1} |k(1,z) f(xz)|\, x dz = \int_0^\infty |k(1,z) f(xz)|\, dz.$$

Since

$$\|f(\cdot\, z)\|^p_{L^p((0,\infty),m)} = \int_0^\infty |f(yz)|^p dy = \int_0^\infty |f(x)|^p \ \frac{dx}{z},$$
$$\|f(\cdot\, z)\|_{L^p((0,\infty),m)} = z^{-1/p} \|f\|_{L^p((0,\infty),m)}.$$

Using Minkowski's inequality for integrals then shows

$$\left\| \int_0^\infty |k(\cdot,y)f(y)|\, dy \right\|_{L^p((0,\infty),m)} \le \int_0^\infty |k(1,z)|\, \|f(\cdot z)\|_{L^p((0,\infty),m)}\, dz$$
$$= \|f\|_{L^p((0,\infty),m)} \int_0^\infty |k(1,z)|\, z^{-1/p} dz$$
$$= C_p \|f\|_{L^p((0,\infty),m)} < \infty.$$

This shows that Kf in Eq. (29.7) is well defined from m – a.e. x. The proof is finished by observing

$$\|Kf\|_{L^p((0,\infty),m)} \le \left\| \int_0^\infty |k(\cdot,y)f(y)|\, dy \right\|_{L^p((0,\infty),m)} \le C_p \|f\|_{L^p((0,\infty),m)}$$

for all $f \in L^p((0,\infty), m)$. ■

The following theorem is a strengthening of Proposition 29.1. It may be skipped on the first reading.

Theorem 29.5 (Converse of Hölder's Inequality II). *Assume that $(X, \mathcal{M}, \mu)$ is a σ – finite measure space, $q, p \in [1,\infty]$ are conjugate exponents and let $\mathbb{S}_f$ denote the set of simple functions φ on X such that $\mu(\varphi \neq 0) < \infty$. Let $g : X \to \mathbb{C}$ be a measurable function such that $\varphi g \in L^1(\mu)$ for all $\varphi \in \mathbb{S}_f$,*[1] *and define*

$$M_q(g) := \sup\left\{ \left| \int_X \varphi g d\mu \right| : \varphi \in \mathbb{S}_f \text{ with } \|\varphi\|_p = 1 \right\}. \tag{29.8}$$

If $M_q(g) < \infty$ then $g \in L^q(\mu)$ and $M_q(g) = \|g\|_q$.

Proof. Let $X_n \in \mathcal{M}$ be sets such that $\mu(X_n) < \infty$ and $X_n \uparrow X$ as $n \uparrow \infty$. Suppose that $q = 1$ and hence $p = \infty$. Choose simple functions φ_n on X such that $|\varphi_n| \le 1$ and $\overline{\operatorname{sgn}(g)} = \lim_{n\to\infty} \varphi_n$ in the pointwise sense. Then $1_{X_m}\varphi_n \in \mathbb{S}_f$ and therefore

$$\left| \int_X 1_{X_m} \varphi_n g d\mu \right| \le M_q(g)$$

for all m, n. By assumption $1_{X_m} g \in L^1(\mu)$ and therefore by the dominated convergence theorem we may let $n \to \infty$ in this equation to find

[1] This is equivalent to requiring $1_A g \in L^1(\mu)$ for all $A \in \mathcal{M}$ such that $\mu(A) < \infty$.

$$\int_X 1_{X_m} |g| \, d\mu \le M_q(g)$$

for all m. The monotone convergence theorem then implies that

$$\int_X |g| \, d\mu = \lim_{m\to\infty} \int_X 1_{X_m} |g| \, d\mu \le M_q(g)$$

showing $g \in L^1(\mu)$ and $\|g\|_1 \le M_q(g)$. Since Holder's inequality implies that $M_q(g) \le \|g\|_1$, we have proved the theorem in case $q = 1$. For $q > 1$, we will begin by assuming that $g \in L^q(\mu)$. Since $p \in [1, \infty)$ we know that $\mathbb{S}_f$ is a dense subspace of $L^p(\mu)$ and therefore, using φ_g is continuous on $L^p(\mu)$,

$$M_q(g) = \sup\left\{\left|\int_X \varphi g d\mu\right| : \varphi \in L^p(\mu) \text{ with } \|\varphi\|_p = 1\right\} = \|g\|_q$$

where the last equality follows by Proposition 29.1. So it remains to show that if $\varphi g \in L^1$ for all $\varphi \in \mathbb{S}_f$ and $M_q(g) < \infty$ then $g \in L^q(\mu)$. For $n \in \mathbb{N}$, let $g_n := 1_{X_n} 1_{|g|\le n} g$. Then $g_n \in L^q(\mu)$, in fact $\|g_n\|_q \le n\mu(X_n)^{1/q} < \infty$. So by the previous paragraph, $\|g_n\|_q = M_q(g_n)$ and hence

$$\begin{aligned}\|g_n\|_q &= \sup\left\{\left|\int_X \varphi 1_{X_n} 1_{|g|\le n} g d\mu\right| : \varphi \in L^p(\mu) \text{ with } \|\varphi\|_p = 1\right\} \\ &\le M_q(g) \left\|\varphi 1_{X_n} 1_{|g|\le n}\right\|_p \le M_q(g) \cdot 1 = M_q(g)\end{aligned}$$

wherein the second to last inequality we have made use of the definition of $M_q(g)$ and the fact that $\varphi 1_{X_n} 1_{|g|\le n} \in \mathbb{S}_f$. If $q \in (1, \infty)$, an application of the monotone convergence theorem (or Fatou's Lemma) along with the continuity of the norm, $\|\cdot\|_p$, implies

$$\|g\|_q = \lim_{n\to\infty} \|g_n\|_q \le M_q(g) < \infty.$$

If $q = \infty$, then $\|g_n\|_\infty \le M_q(g) < \infty$ for all n implies $|g_n| \le M_q(g)$ a.e. which then implies that $|g| \le M_q(g)$ a.e. since $|g| = \lim_{n\to\infty} |g_n|$. That is $g \in L^\infty(\mu)$ and $\|g\|_\infty \le M_\infty(g)$. ■

Theorem 29.6 (Dual of L^p – spaces). *Let $(X, \mathcal{M}, \mu)$ be a measure space and suppose that $p, q \in (1, \infty)$ are conjugate exponents. Then the map $g \in L^q \to \varphi_g \in (L^p)^*$ (where $\varphi_g = \langle \cdot, g\rangle_\mu$ was defined in Eq. 29.1) is an isometric isomorphism of Banach spaces. We summarize this by writing $(L^p)^* = L^q$ for all $1 < p < \infty$. Moreover, if we further assume that $(X, \mathcal{M}, \mu)$ is a σ – finite measure space, then the above results hold for $p = 1$ $(q = \infty)$ as well. (The result is in general false for $p = \infty$ as can be seen from Theorem 21.20 and Lemma 21.21 below.)*

Proof. The only results of this theorem which are not covered in Proposition 29.1 is the surjectivity of the map $g \in L^q \to \varphi_g \in (L^p)^*$. When $p = 2$, this surjectivity is a direct consequence of the Riesz Theorem 18.17.

Case 1. We will begin the proof under the extra assumption that $\mu(X) < \infty$ in which cased bounded functions are in $L^p(\mu)$ for all p. So let $\varphi \in (L^p)^*$. We need to find $g \in L^q(\mu)$ such that $\varphi = \varphi_g$. When $p \in [1, 2]$, $L^2(\mu) \subset L^p(\mu)$ so that we may restrict φ to $L^2(\mu)$ and again the result follows fairly easily from the Riesz Theorem, see Exercise 24.4 below. To handle general $p \in [1, \infty)$, define $\nu(A) := \varphi(1_A)$. If $A = \coprod_{n=1}^\infty A_n$ with $A_n \in \mathcal{M}$, then

$$\|1_A - \sum_{n=1}^N 1_{A_n}\|_{L^p} = \|1_{\cup_{n=N+1}^\infty A_n}\|_{L^p} = \left[\mu(\cup_{n=N+1}^\infty A_n)\right]^{\frac{1}{p}} \to 0 \text{ as } N \to \infty.$$

Therefore

$$\nu(A) = \varphi(1_A) = \sum_1^\infty \varphi(1_{A_n}) = \sum_1^\infty \nu(A_n)$$

showing ν is a complex measure.[2] For $A \in \mathcal{M}$, let $|\nu|(A)$ be the "total variation" of A defined by

$$|\nu|(A) := \sup\{|\varphi(f 1_A)| : |f| \le 1\} \tag{29.9}$$

and notice that

$$|\nu(A)| \le |\nu|(A) \le \|\varphi\|_{(L^p)^*} \mu(A)^{1/p} \text{ for all } A \in \mathcal{M}. \tag{29.10}$$

You are asked to show in Exercise 24.5 that $|\nu|$ is a measure on $(X, \mathcal{M})$. (This can also be deduced from Lemma 24.28 and Proposition 24.32 below.) By Eq. (29.10) $|\nu| \ll \mu$, by Theorem 24.6 $d\nu = hd|\nu|$ for some $|h| \le 1$ and by Theorem 24.13 $d|\nu| = \rho d\mu$ for some $\rho \in L^1(\mu)$. Hence, letting $g = \rho h \in L^1(\mu)$, $d\nu = gd\mu$ or equivalently

$$\varphi(1_A) = \int_X g 1_A d\mu \;\forall\; A \in \mathcal{M}. \tag{29.11}$$

By linearity this equation implies

$$\varphi(f) = \int_X g f d\mu \tag{29.12}$$

for all simple functions f on X. Replacing f by $1_{\{|g|\le M\}} f$ in Eq. (29.12) shows

$$\varphi(f 1_{\{|g|\le M\}}) = \int_X 1_{\{|g|\le M\}} g f d\mu$$

holds for all simple functions f and then by continuity for all $f \in L^p(\mu)$. By the converse to Holder's inequality, (Proposition 29.1) we learn that

[2] It is at this point that the proof breaks down when $p = \infty$.

$$\left\|1_{\{|g|\le M\}}g\right\|_q = \sup_{\|f\|_p=1}\left|\varphi(f1_{\{|g|\le M\}})\right|$$
$$\le \sup_{\|f\|_p=1}\|\varphi\|_{(L^p)^*}\left\|f1_{\{|g|\le M\}}\right\|_p \le \|\varphi\|_{(L^p)^*}.$$

Using the monotone convergence theorem we may let $M\to\infty$ in the previous equation to learn $\|g\|_q \le \|\varphi\|_{(L^p)^*}$ .With this result, Eq. (29.12) extends by continuity to hold for all $f\in L^p(\mu)$ and hence we have shown that $\varphi=\varphi_g$.

Case 2. Now suppose that μ is σ – finite and $X_n\in\mathcal{M}$ are sets such that $\mu(X_n)<\infty$ and $X_n\uparrow X$ as $n\to\infty$. We will identify $f\in L^p(X_n,\mu)$ with $f1_{X_n}\in L^p(X,\mu)$ and this way we may consider $L^p(X_n,\mu)$ as a subspace of $L^p(X,\mu)$ for all n and $p\in[1,\infty]$. By Case 1. there exists $g_n\in L^q(X_n,\mu)$ such that

$$\varphi(f)=\int_{X_n}g_nfd\mu \text{ for all } f\in L^p(X_n,\mu)$$

and

$$\|g_n\|_q=\sup\left\{|\varphi(f)| : f\in L^p(X_n,\mu) \text{ and } \|f\|_{L^p(X_n,\mu)}=1\right\}\le\|\varphi\|_{[L^p(\mu)]^*}.$$

It is easy to see that $g_n=g_m$ a.e. on $X_n\cap X_m$ for all m,n so that $g:=\lim_{n\to\infty}g_n$ exists μ – a.e. By the above inequality and Fatou's lemma, $\|g\|_q\le\|\varphi\|_{[L^p(\mu)]^*}<\infty$ and since $\varphi(f)=\int_{X_n}gfd\mu$ for all $f\in L^p(X_n,\mu)$ and n and $\cup_{n=1}^\infty L^p(X_n,\mu)$ is dense in $L^p(X,\mu)$ it follows by continuity that $\varphi(f)=\int_X gfd\mu$ for all $f\in L^p(X,\mu)$, i.e. $\varphi=\varphi_g$.

Case 3. Now suppose that $(X,\mathcal{M},\mu)$ is a general measure space and $1<p<\infty$. Given $E\in\mathcal{M}$ we will identify $f\in L^p(E,\mu)$ with its $f1_E$ (i.e. its extension to 0 on $X\setminus E$) inside of $L^p(X,\mu)$. If μ is σ – finite on $E\in\mathcal{M}$, then by Case 2. there exists $g_E\in L^q(E,\mu)$ such that

$$\varphi(f)=\int_E fg_Ed\mu \text{ for all } f\in L^p(E,\mu).$$

Moreover we have

$$\|g_E\|_{L^q(\mu)}=\left\|\varphi|_{L^p(E,\mu)}\right\|_{[L^p(\mu)]^*}\le\|\varphi\|_{[L^p(\mu)]^*}.$$

Let us observe that if μ is σ – finite on F and $E\subset F$ then $g_F=g_E$ a.e. on E and $\|g_E\|_{L^q(\mu)}\le\|g_F\|_{L^q(\mu)}$. So if we let

$$M:=\sup\left\{\|g_E\|_{L^q(\mu)} : \mu \text{ is } \sigma-\text{finite on } E\right\}$$

we can find $E_n\in\mathcal{M}$ on which μ is σ – finite so that $M=\lim_{n\to\infty}\|g_{E_n}\|_{L^q(\mu)}$. Moreover by the previous comments if we now let $E:=\cup_{n=1}^\infty E_n$ on which μ is σ – finite, then $\|g_E\|_{L^q(\mu)}\ge\|g_{E_n}\|_{L^q(\mu)}$ for all n and therefore $M=\|g_E\|_{L^q(\mu)}$.

If $f\in L^p(\mu)$, the set $F=\{|f|>0\}$ is σ – finite and therefore $E\cup F$ is σ – finite. Since $g_E=g_{E\cup F}$ a.e. on E (using $q<\infty$ as $p>1$)

$$M^q\ge\|g_{E\cup F}\|^q_{L^q(\mu)}=\|g_E\|^q_{L^q(\mu)}+\left\|g_{E\cup F}1_{[E\cup F]\setminus E}\right\|^q_{L^q(\mu)}$$
$$=M^q+\left\|g_{E\cup F}1_{[E\cup F]\setminus E}\right\|^q_{L^q(\mu)}$$

from which it follows that $g_{EUF}=0$ a.e. on $[E\cup F]\setminus E$. Therefore $g_{E\cup F}=g_E$ a.e. and so

$$\varphi(f)=\int_{E\cup F}g_{E\cup F}fd\mu=\int_E g_Efd\mu=\int_X g_Efd\mu.$$

As $f\in L^p(\mu)$ was arbitrary, this shows that $\varphi=\varphi_{g_E}$ to complete the proof. ■

30

Uniform Integrability

This section will address the question as to what extra conditions are needed in order that an L^0 – convergent sequence is L^p – convergent.

Notation 30.1 *For $f \in L^1(\mu)$ and $E \in \mathcal{M}$, let*

$$\mu(f : E) := \int_E f d\mu.$$

and more generally if $A, B \in \mathcal{M}$ let

$$\mu(f : A, B) := \int_{A\cap B} f d\mu.$$

Lemma 30.2. *Suppose $g \in L^1(\mu)$, then for any $\varepsilon > 0$ there exist a $\delta > 0$ such that $\mu(|g| : E) < \varepsilon$ whenever $\mu(E) < \delta$.*

Proof. If the Lemma is false, there would exist $\varepsilon > 0$ and sets E_n such that $\mu(E_n) \to 0$ while $\mu(|g| : E_n) \geq \varepsilon$ for all n. Since $|1_{E_n} g| \leq |g| \in L^1$ and for any $\delta \in (0,1)$, $\mu(1_{E_n}|g| > \delta) \leq \mu(E_n) \to 0$ as $n \to \infty$, the dominated convergence theorem of Corollary 16.21 implies $\lim_{n\to\infty} \mu(|g| : E_n) = 0$. This contradicts $\mu(|g| : E_n) \geq \varepsilon$ for all n and the proof is complete. ■

Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of measurable functions which converge in $L^1(\mu)$ to a function f. Then for $E \in \mathcal{M}$ and $n \in \mathbb{N}$,

$$|\mu(f_n : E)| \leq |\mu(f - f_n : E)| + |\mu(f : E)| \leq \|f - f_n\|_1 + |\mu(f : E)|.$$

Let $\varepsilon_N := \sup_{n>N} \|f - f_n\|_1$, then $\varepsilon_N \downarrow 0$ as $N \uparrow \infty$ and

$$\sup_n |\mu(f_n : E)| \leq \sup_{n\leq N} |\mu(f_n : E)| \vee (\varepsilon_N + |\mu(f : E)|) \leq \varepsilon_N + \mu(g_N : E), \quad (30.1)$$

where $g_N = |f| + \sum_{n=1}^N |f_n| \in L^1$. From Lemma 30.2 and Eq. (30.1) one easily concludes,

$$\forall\ \varepsilon > 0\ \exists\ \delta > 0\ \ni\ \sup_n |\mu(f_n : E)| < \varepsilon \text{ when } \mu(E) < \delta. \quad (30.2)$$

Definition 30.3. *Functions $\{f_n\}_{n=1}^\infty \subset L^1(\mu)$ satisfying Eq. (30.2) are said to be* uniformly integrable.

Remark 30.4. Let $\{f_n\}$ be real functions satisfying Eq. (30.2), E be a set where $\mu(E) < \delta$ and $E_n = E \cap \{f_n \geq 0\}$. Then $\mu(E_n) < \delta$ so that $\mu(f_n^+ : E) = \mu(f_n : E_n) < \varepsilon$ and similarly $\mu(f_n^- : E) < \varepsilon$. Therefore if Eq. (30.2) holds then

$$\sup_n \mu(|f_n| : E) < 2\varepsilon \text{ when } \mu(E) < \delta. \quad (30.3)$$

Similar arguments work for the complex case by looking at the real and imaginary parts of f_n. Therefore $\{f_n\}_{n=1}^\infty \subset L^1(\mu)$ is uniformly integrable iff

$$\forall\ \varepsilon > 0\ \exists\ \delta > 0\ \ni\ \sup_n \mu(|f_n| : E) < \varepsilon \text{ when } \mu(E) < \delta. \quad (30.4)$$

Lemma 30.5. *Assume that $\mu(X) < \infty$, then $\{f_n\}$ is uniformly bounded in $L^1(\mu)$ (i.e. $K = \sup_n \|f_n\|_1 < \infty$) and $\{f_n\}$ is uniformly integrable iff*

$$\lim_{M\to\infty} \sup_n \mu(|f_n| : |f_n| \geq M) = 0. \quad (30.5)$$

Proof. Since $\{f_n\}$ is uniformly bounded in $L^1(\mu)$, $\mu(|f_n| \geq M) \leq K/M$. So if (30.4) holds and $\varepsilon > 0$ is given, we may choose M sufficiently large so that $\mu(|f_n| \geq M) < \delta(\varepsilon)$ for all n and therefore,

$$\sup_n \mu(|f_n| : |f_n| \geq M) \leq \varepsilon.$$

Since ε is arbitrary, we concluded that Eq. (30.5) must hold. Conversely, suppose that Eq. (30.5) holds, then automatically $K = \sup_n \mu(|f_n|) < \infty$ because

$$\begin{aligned}\mu(|f_n|) &= \mu(|f_n| : |f_n| \geq M) + \mu(|f_n| : |f_n| < M)\\ &\leq \sup_n \mu(|f_n| : |f_n| \geq M) + M\mu(X) < \infty.\end{aligned}$$

Moreover,

$$\begin{aligned}\mu(|f_n| : E) &= \mu(|f_n| : |f_n| \geq M, E) + \mu(|f_n| : |f_n| < M, E)\\ &\leq \sup_n \mu(|f_n| : |f_n| \geq M) + M\mu(E).\end{aligned}$$

So given $\varepsilon > 0$ choose M so large that $\sup_n \mu(|f_n| : |f_n| \geq M) < \varepsilon/2$ and then take $\delta = \varepsilon/(2M)$. ■

Lemma 30.6 (Saks' Lemma [4, Lemma 7 on p. 308]). *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space such that P has no atoms. (Recall that $A \in \mathcal{B}$ is an atom if $P(A) > 0$ and for any $B \subset A$ with $B \in \mathcal{B}$ we have either $P(B) = 0$ or $P(B) = P(A)$.) Then for every $\delta > 0$ there exists a partition $\{E_\ell\}_{\ell=1}^n$ of Ω with $\mu(E_\ell) < \delta$. (For related results along this line also see [3, 5, 8, 11] to name a few.)*

Proof. For any $A \in \mathcal{B}$ let

$$\beta(A) := \sup\{P(B) : B \subset A \text{ and } P(B) \le \delta\}.$$

We begin by showing if $\mu(A) > 0$ then $\beta(A) > 0$. As there are no atoms there exists $A_1 \subset A$ such that $0 < P(A_1) < P(A)$. Similarly there exists $A_2 \subset A \setminus A_1$ such that $0 < P(A_2) < P(A \setminus A_1)$ and continuing inductively we find $\{A_n\}_{n=1}^\infty$ disjoint subsets of A such that $A_n \subset A \setminus (A_1 \cup \dots \cup A_{n-1})$ and

$$0 < P(A_n) < P(A \setminus (A_1 \cup \dots \cup A_{n-1})).$$

As $\sum_{n=1}^\infty A_n \subset A$ we must have $\sum_{n=1}^\infty P(A_n) \le P(A) < \infty$ and therefore $\lim_{n\to\infty} P(A_n) = 0$. Thus for sufficiently large n we have $0 < P(A_n) \le \delta$ and therefore $\beta(A) \ge P(A_n) > 0$.

Now to construct the desired partition. Choose $A_1 \subset \Omega$ such that $\delta \ge P(A_1) \ge \frac{1}{2}\beta(\Omega)$. If $P(\Omega \setminus A_1) > 0$ we may then choose $A_2 \subset \Omega \setminus A_1$ such that $\delta \ge P(A_2) \ge \frac{1}{2}\beta(\Omega \setminus [A_1 \cup A_2])$. We may continue on this way inductively to find disjoint subsets $\{A_k\}_{k=1}^n$ of Ω

$$\delta \ge P(A_k) \ge \frac{1}{2}\beta(\Omega \setminus [A_1 \cup \dots \cup A_{k-1}])$$

with either $P(\Omega \setminus [A_1 \cup \dots \cup A_{n-1}]) > 0$. If it happens that $P(\Omega \setminus [A_1 \cup \dots \cup A_n]) = 0$ it is easy to see we are done. So we may assume that process can be carried on indefinitely. We then let $F := \Omega \setminus \cup_{k=1}^\infty A_k$ and observe that

$$\beta(F) \le \beta(\Omega \setminus [A_1 \cup \dots \cup A_{n-1}]) \le 2P(A_n) \to 0 \text{ as } n \to \infty$$

as

$$\sum_{n=1}^\infty P(A_n) \le P(\Omega) < \infty.$$

But by the first paragraph this implies that $P(F) = 0$. Hence there exists $n < \infty$ such that $P\left(\Omega \setminus \cup_{k=1}^{n-1} A_k\right) \le \delta$. We may then define $E_k = A_k$ for $1 \le k \le n-1$ and $E_n = \Omega \setminus \cup_{k=1}^{n-1} A_k$ in order to construct the desired partition. ■

Corollary 30.7. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space such that P has no atoms. Then for any $\alpha \in (0,1)$ there exists $A \in \mathcal{B}$ with $P(A) = \alpha$.*

Proof. We may assume the $\alpha \in (0, 1/2)$. By dividing Ω into a partition $\{E_l\}_{l=1}^N$ with $P(E_l) \le \alpha/2$ we may let $A_1 := \cup_{l=1}^k E_l$ with k chosen so that $P(A_1) \le \alpha$ but

$$\alpha < P(A_1 \cup E_{k+1}) \le \frac{3}{2}\alpha.$$

Notice that $\alpha/2 \le P(A_1) \le \alpha$. Apply this procedure to $\Omega \setminus A_1$ in order to find $A_2 \supset A_1$ such that $\alpha/4 \le P(A_2) \le \alpha$. Continue this way inductively to find $A_n \uparrow A$ such that $P(A_n) \uparrow \alpha = P(A)$. (BRUCE: clean this proof up.) ■

Remark 30.8. It is not in general true that if $\{f_n\} \subset L^1(\mu)$ is uniformly integrable then $\sup_n \mu(|f_n|) < \infty$. For example take $X = \{*\}$ and $\mu(\{*\}) = 1$. Let $f_n(*) = n$. Since for $\delta < 1$ a set $E \subset X$ such that $\mu(E) < \delta$ is in fact the empty set, we see that Eq. (30.3) holds in this example. However, for finite measure spaces with out "atoms", for every $\delta > 0$ we may find a finite partition of X by sets $\{E_\ell\}_{\ell=1}^k$ with $\mu(E_\ell) < \delta$. Then if Eq. (30.3) holds with $2\varepsilon = 1$, then

$$\mu(|f_n|) = \sum_{\ell=1}^k \mu(|f_n| : E_\ell) \le k$$

showing that $\mu(|f_n|) \le k$ for all n.

The following Lemmas gives a concrete necessary and sufficient conditions for verifying a sequence of functions is uniformly bounded and uniformly integrable.

Lemma 30.9. *Suppose that $\mu(X) < \infty$, and $\Lambda \subset L^0(X)$ is a collection of functions.*

1. *If there exists a non decreasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{x\to\infty} \varphi(x)/x = \infty$ and*
$$K := \sup_{f\in\Lambda} \mu(\varphi(|f|)) < \infty \tag{30.6}$$
then
$$\lim_{M\to\infty} \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\ge M}\right) = 0. \tag{30.7}$$
2. *Conversely if Eq. (30.7) holds, there exists a non-decreasing continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(0) = 0$, $\lim_{x\to\infty} \varphi(x)/x = \infty$ and Eq. (30.6) is valid.*

Proof. 1. Let φ be as in item 1. above and set $\varepsilon_M := \sup_{x\ge M} \frac{x}{\varphi(x)} \to 0$ as $M \to \infty$ by assumption. Then for $f \in \Lambda$

$$\mu(|f|:|f|\geq M)=\mu(\frac{|f|}{\varphi\left(|f|\right)}\varphi\left(|f|\right):|f|\geq M)\leq\varepsilon_{M}\mu(\varphi\left(|f|\right):|f|\geq M)$$
$$\leq\varepsilon_{M}\mu(\varphi\left(|f|\right))\leq K\varepsilon_{M}$$

and hence

$$\lim_{M\to\infty}\sup_{f\in\Lambda}\mu\left(|f|\,1_{|f|\geq M}\right)\leq\lim_{M\to\infty}K\varepsilon_{M}=0.$$

2. By assumption, $\varepsilon_{M}:=\sup_{f\in\Lambda}\mu\left(|f|\,1_{|f|\geq M}\right)\to 0$ as $M\to\infty$. Therefore we may choose $M_{n}\uparrow\infty$ such that

$$\sum_{n=0}^{\infty}(n+1)\,\varepsilon_{M_{n}}<\infty$$

where by convention $M_{0}:=0$. Now define φ so that $\varphi(0)=0$ and

$$\varphi'(x)=\sum_{n=0}^{\infty}(n+1)\,1_{(M_{n},M_{n+1}]}(x),$$

i.e.

$$\varphi(x)=\int_{0}^{x}\varphi'(y)dy=\sum_{n=0}^{\infty}(n+1)\left(x\wedge M_{n+1}-x\wedge M_{n}\right).$$

By construction φ is continuous, $\varphi(0)=0$, $\varphi'(x)$ is increasing (so φ is convex) and $\varphi'(x)\geq(n+1)$ for $x\geq M_{n}$. In particular

$$\frac{\varphi(x)}{x}\geq\frac{\varphi(M_{n})+(n+1)x}{x}\geq n+1\text{ for }x\geq M_{n}$$

from which we conclude $\lim_{x\to\infty}\varphi(x)/x=\infty$. We also have $\varphi'(x)\leq(n+1)$ on $[0,M_{n+1}]$ and therefore

$$\varphi(x)\leq(n+1)x\text{ for }x\leq M_{n+1}.$$

So for $f\in\Lambda$,

$$\begin{aligned}\mu\left(\varphi(|f|)\right)&=\sum_{n=0}^{\infty}\mu\left(\varphi(|f|)1_{(M_{n},M_{n+1}]}(|f|)\right)\\&\leq\sum_{n=0}^{\infty}(n+1)\,\mu\left(|f|\,1_{(M_{n},M_{n+1}]}(|f|)\right)\\&\leq\sum_{n=0}^{\infty}(n+1)\,\mu\left(|f|\,1_{|f|\geq M_{n}}\right)\leq\sum_{n=0}^{\infty}(n+1)\,\varepsilon_{M_{n}}\end{aligned}$$

and hence

$$\sup_{f\in\Lambda}\mu\left(\varphi(|f|)\right)\leq\sum_{n=0}^{\infty}(n+1)\,\varepsilon_{M_{n}}<\infty.$$

∎

Theorem 30.10 (Vitali Convergence Theorem). *(Folland 6.15) Suppose that $1\leq p<\infty$. A sequence $\{f_{n}\}\subset L^{p}$ is Cauchy iff*

1. *$\{f_{n}\}$ is L^{0} – Cauchy,*
2. *$\{|f_{n}|^{p}\}$ – is uniformly integrable.*
3. *For all $\varepsilon>0$, there exists a set $E\in\mathcal{M}$ such that $\mu(E)<\infty$ and $\int_{E^{c}}|f_{n}|^{p}\,d\mu<\varepsilon$ for all n. (This condition is vacuous when $\mu(X)<\infty$.)*

Proof. ($\Longrightarrow$) Suppose $\{f_{n}\}\subset L^{p}$ is Cauchy. Then (1) $\{f_{n}\}$ is L^{0} – Cauchy by Lemma 16.18. (2) By completeness of L^{p}, there exists $f\in L^{p}$ such that $\|f_{n}-f\|_{p}\to 0$ as $n\to\infty$. By the mean value theorem,

$$||f|^{p}-|f_{n}|^{p}|\leq p(\max(|f|\,,|f_{n}|))^{p-1}\,||f|-|f_{n}||\leq p(|f|+|f_{n}|)^{p-1}\,||f|-|f_{n}||$$

and therefore by Hölder's inequality,

$$\begin{aligned}\int||f|^{p}-|f_{n}|^{p}|\,d\mu&\leq p\int(|f|+|f_{n}|)^{p-1}\,||f|-|f_{n}||\,d\mu\leq p\int(|f|+|f_{n}|)^{p-1}|f-f_{n}|d\mu\\&\leq p\|f-f_{n}\|_{p}\|(|f|+|f_{n}|)^{p-1}\|_{q}=p\|\,|f|+|f_{n}|\|_{p}^{p/q}\|f-f_{n}\|_{p}\\&\leq p(\|f\|_{p}+\|f_{n}\|_{p})^{p/q}\|f-f_{n}\|_{p}\end{aligned}$$

where $q:=p/(p-1)$. This shows that $\int||f|^{p}-|f_{n}|^{p}|\,d\mu\to 0$ as $n\to\infty$.[1] By the remarks prior to Definition 30.3, $\{|f_{n}|^{p}\}$ is uniformly integrable. To verify (3), for $M>0$ and $n\in\mathbb{N}$ let $E_{M}=\{|f|\geq M\}$ and $E_{M}(n)=\{|f_{n}|\geq M\}$. Then $\mu(E_{M})\leq\frac{1}{M^{p}}\|f\|_{p}^{p}<\infty$ and by the dominated convergence theorem,

$$\int_{E_{M}^{c}}|f|^{p}\,d\mu=\int|f|^{p}\,1_{|f|<M}d\mu\to 0\text{ as }M\to 0.$$

Moreover,

$$\left\|f_{n}1_{E_{M}^{c}}\right\|_{p}\leq\left\|f1_{E_{M}^{c}}\right\|_{p}+\left\|(f_{n}-f)1_{E_{M}^{c}}\right\|_{p}\leq\left\|f1_{E_{M}^{c}}\right\|_{p}+\|f_{n}-f\|_{p}\,.\quad(30.8)$$

So given $\varepsilon>0$, choose N sufficiently large such that for all $n\geq N$, $\|f-f_{n}\|_{p}^{p}<\varepsilon$. Then choose M sufficiently small such that $\int_{E_{M}^{c}}|f|^{p}\;d\mu<\varepsilon$ and $\int_{E_{M}^{c}(n)}|f|^{p}\;d\mu<\varepsilon$ for all $n=1,2,\dots,N-1$. Letting $E:=E_{M}\cup E_{M}(1)\cup\dots\cup E_{M}(N-1)$, we have

$$\mu(E)<\infty,\;\int_{E^{c}}|f_{n}|^{p}\;d\mu<\varepsilon\text{ for }n\leq N-1$$

[1] Here is an alternative proof. Let $h_{n}\equiv||f_{n}|^{p}-|f|^{p}|\leq|f_{n}|^{p}+|f|^{p}=:g_{n}\in L^{1}$ and $g\equiv 2|f|^{p}$. Then $g_{n}\overset{\mu}{\to}g$, $h_{n}\overset{\mu}{\to}0$ and $\int g_{n}\to\int g$. Therefore by the dominated convergence theorem in Corollary 16.21, $\lim_{n\to\infty}\int h_{n}\,d\mu=0$.

and by Eq. (30.8)

$$\int_{E^c} |f_n|^p \, d\mu < (\varepsilon^{1/p} + \varepsilon^{1/p})^p \le 2^p \varepsilon \text{ for } n \ge N.$$

Therefore we have found $E \in \mathcal{M}$ such that $\mu(E) < \infty$ and

$$\sup_n \int_{E^c} |f_n|^p \, d\mu \le 2^p \varepsilon$$

which verifies (3) since $\varepsilon > 0$ was arbitrary.

($\Longleftarrow$) Now suppose$\{f_n\} \subset L^p$ satisfies conditions (1) - (3). Let $\varepsilon > 0$, E be as in (3) and

$$A_{mn} := \{x \in E | f_m(x) - f_n(x)| \ge \varepsilon\}.$$

Then

$$\|(f_n - f_m)\ 1_{E^c}\|_p \le \|f_n 1_{E^c}\|_p + \|f_m\, 1_{E^c}\|_p < 2\varepsilon^{1/p}$$

and

$$\begin{aligned} \|f_n - f_m\|_p &= \|(f_n - f_m)1_{E^c}\|_p + \|(f_n - f_m)1_{E \setminus A_{mn}}\|_p \\ &\quad + \|(f_n - f_m)1_{A_{mn}}\|_p \\ &\le \|(f_n - f_m)1_{E \setminus A_{mn}}\|_p + \|(f_n - f_m)1_{A_{mn}}\|_p + 2\varepsilon^{1/p}. \end{aligned} \tag{30.9}$$

Using properties (1) and (3) and $1_{E \cap \{|f_m - f_n| < \varepsilon\}} |f_m - f_n|^p \le \varepsilon^p 1_E \in L^1$, the dominated convergence theorem in Corollary 16.21 implies

$$\|(f_n - f_m)\ 1_{E \setminus A_{mn}}\|_p^p = \int 1_{E \cap \{|f_m - f_n| < \varepsilon\}}\ |f_m - f_n|^p \underset{m,n\to\infty}{\longrightarrow} 0.$$

which combined with Eq. (30.9) implies

$$\limsup_{m,n\to\infty} \|f_n - f_m\|_p \le \limsup_{m,n\to\infty} \|(f_n - f_m)1_{A_{mn}}\|_p + 2\varepsilon^{1/p}.$$

Finally

$$\|(f_n - f_m)1_{A_{mn}}\|_p \le \|f_n 1_{A_{mn}}\|_p + \|f_m\ 1_{A_{mn}}\|_p \le 2\delta(\varepsilon)$$

where

$$\delta(\varepsilon) := \sup_n\ \sup\{\ \|f_n\ 1_E\|_p : E \in \mathcal{M}\ \ni \mu(E) \le \varepsilon\}$$

By property (2), $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Therefore

$$\limsup_{m,n\to\infty} \|f_n - f_m\|_p \le 2\varepsilon^{1/p} + 0 + 2\delta(\varepsilon) \to 0 \text{ as } \varepsilon \downarrow 0$$

and therefore $\{f_n\}$ is L^p-Cauchy. ∎

Here is another version of Vitali's Convergence Theorem.

Theorem 30.11 (Vitali Convergence Theorem). *(This is problem 9 on p. 133 in Rudin.) Assume that $\mu(X) < \infty$, $\{f_n\}$ is uniformly integrable, $f_n \to f$ a.e. and $|f| < \infty$ a.e., then $f \in L^1(\mu)$ and $f_n \to f$ in $L^1(\mu)$.*

Proof. Let $\varepsilon > 0$ be given and choose $\delta > 0$ as in the Eq. (30.3). Now use Egoroff's Theorem 16.22 to choose a set E^c where $\{f_n\}$ converges uniformly on E^c and $\mu(E) < \delta$. By uniform convergence on E^c, there is an integer $N < \infty$ such that $|f_n - f_m| \le 1$ on E^c for all $m, n \ge N$. Letting $m \to \infty$, we learn that

$$|f_N - f| \le 1 \text{ on } E^c.$$

Therefore $|f| \le |f_N| + 1$ on E^c and hence

$$\begin{aligned} \mu(|f|) &= \mu(|f| : E^c) + \mu(|f| : E) \\ &\le \mu(|f_N|) + \mu(X) + \mu(|f| : E). \end{aligned}$$

Now by Fatou's lemma,

$$\mu(|f| : E) \le \liminf_{n\to\infty} \mu(|f_n| : E) \le 2\varepsilon < \infty$$

by Eq. (30.3). This shows that $f \in L^1$. Finally

$$\begin{aligned} \mu(|f - f_n|) &= \mu(|f - f_n| : E^c) + \mu(|f - f_n| : E) \\ &\le \mu(|f - f_n| : E^c) + \mu(|f| + |f_n| : E) \\ &\le \mu(|f - f_n| : E^c) + 4\varepsilon \end{aligned}$$

and so by the Dominated convergence theorem we learn that

$$\limsup_{n\to\infty} \mu(|f - f_n|) \le 4\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary this completes the proof. ∎

Theorem 30.12 (Vitali again). *Suppose that $f_n \to f$ in μ measure and Eq. (30.5) holds, then $f_n \to f$ in L^1.*

Proof. This could of course be proved using 30.11 after passing to subsequences to get $\{f_n\}$ to converge a.s. However I wish to give another proof. First off, by Fatou's lemma, $f \in L^1(\mu)$. Now let

$$\varphi_K(x) = x 1_{|x| \le K} + K 1_{|x| > K}.$$

then $\varphi_K(f_n) \overset{\mu}{\to} \varphi_K(f)$ because $|\varphi_K(f) - \varphi_K(f_n)| \le |f - f_n|$ and since

$$|f - f_n| \le |f - \varphi_K(f)| + |\varphi_K(f) - \varphi_K(f_n)| + |\varphi_K(f_n) - f_n|$$

we have that

$$\begin{aligned}\mu|f-f_n| &\le \mu|f-\varphi_K(f)| + \mu|\varphi_K(f)-\varphi_K(f_n)| + \mu|\varphi_K(f_n)-f_n| \\ &= \mu(|f| : |f| \ge K) + \mu|\varphi_K(f)-\varphi_K(f_n)| + \mu(|f_n| : |f_n| \ge K).\end{aligned}$$

Therefore by the dominated convergence theorem

$$\limsup_{n\to\infty} \mu|f-f_n| \le \mu(|f| : |f| \ge K) + \limsup_{n\to\infty} \mu(|f_n| : |f_n| \ge K).$$

This last expression goes to zero as $K \to \infty$ by uniform integrability. ■

30.1 Exercises

Exercise 30.1. Show $\ell^\infty(\mathbb{N})$ is not separable. Hint: find an uncountable set $\Lambda \subset \ell^\infty(\mathbb{N})$ as in the hypothesis of Exercise 13.6.

Exercise 30.2. Let $(X, \mathcal{B}, \mu)$ be a σ – finite measure space. Suppose that $1 \le p < \infty$ and to each $\varphi \in L^\infty(\mu)$ show $M_\varphi \in L(L^p(\mu))$ be defined by $M_\varphi f = \varphi f$ for all $f \in L^p(\mu)$ – so M_φ is multiplication by φ. Show $\|M_\varphi\|_{op} = \|\varphi\|_\infty$, i.e. $L^\infty(\mu) \ni \varphi \to M_\varphi \in L(L^p(\mu))$ is an isometry.

Exercise 30.3. Let m be Lebesgue measure on $([0,\infty), \mathcal{B}, m)$.

1. Show $L^\infty(([0,\infty), \mathcal{B}, m))$ is not separable. Hint:you might produce an isometry from $\ell^\infty(\mathbb{N})$ into $L^\infty([0,\infty), m)$ and then use Exercise 30.1. find an uncountable set $\Lambda \subset L^\infty(\mathbb{R}, m)$ as in the hypothesis of Exercise 13.6.
2. Use this result along with Exercise 30.2 in order to show $L(L^p(([0,\infty), \mathcal{B}, m)))$ is not separable for all $1 \le p < \infty$.

Definition 30.13. *The* ***essential range*** *of* f, essran(f), *consists of those* $\lambda \in \mathbb{C}$ *such that* $\mu(|f-\lambda| < \varepsilon) > 0$ *for all* $\varepsilon > 0$.

Definition 30.14. *Let* (X, τ) *be a topological space and* ν *be a measure on* $\mathcal{B}_X = \sigma(\tau)$. *The* ***support*** *of* ν, supp(ν), *consists of those* $x \in X$ *such that* $\nu(V) > 0$ *for all open neighborhoods,* V, *of* x.

Exercise 30.4. Let (X, τ) be a second countable topological space and ν be a measure on $\mathcal{B}_X$ – the Borel σ – algebra on X. Show

1. supp(ν) is a closed set. (This is actually true on all topological spaces.)
2. $\nu(X \setminus \mathrm{supp}(\nu)) = 0$ and use this to conclude that $W := X \setminus \mathrm{supp}(\nu)$ is the largest open set in X such that $\nu(W) = 0$. **Hint:** let $\mathcal{U} \subset \tau$ be a countable base for the topology τ. Show that W may be written as a union of elements from $V \in \mathcal{V}$ with the property that $\mu(V) = 0$.

Exercise 30.5. Prove the following facts about essran(f).

1. Let $\nu = f_*\mu := \mu \circ f^{-1}$ – a Borel measure on $\mathbb{C}$. Show essran$(f) = \mathrm{supp}(\nu)$.
2. essran(f) is a closed set and $f(x) \in \mathrm{essran}(f)$ for almost every x, i.e. $\mu(f \notin \mathrm{essran}(f)) = 0$.
3. If $F \subset \mathbb{C}$ is a closed set such that $f(x) \in F$ for almost every x then essran$(f) \subset F$. So essran(f) is the smallest closed set F such that $f(x) \in F$ for almost every x.
4. $\|f\|_\infty = \sup\{|\lambda| : \lambda \in \mathrm{essran}(f)\}$.

Exercise 30.6. By making the change of variables, $u = \ln x$, prove the following facts:

$$\begin{aligned}\int_0^{1/2} x^{-a}|\ln x|^b\,dx < \infty &\iff a < 1 \text{ or } a = 1 \text{ and } b < -1 \\ \int_2^\infty x^{-a}|\ln x|^b\,dx < \infty &\iff a > 1 \text{ or } a = 1 \text{ and } b < -1 \\ \int_0^1 x^{-a}|\ln x|^b\,dx < \infty &\iff a < 1 \text{ and } b > -1 \\ \int_1^\infty x^{-a}|\ln x|^b\,dx < \infty &\iff a > 1 \text{ and } b > -1.\end{aligned}$$

Suppose $0 < p_0 < p_1 \le \infty$ and m is Lebesgue measure on $(0,\infty)$. Use the above results to manufacture a function f on $(0,\infty)$ such that $f \in L^p((0,\infty), m)$ iff (a) $p \in (p_0, p_1)$, (b) $p \in [p_0, p_1]$ and (c) $p = p_0$.

Exercise 30.7. Folland 6.9 on p. 186.

Exercise 30.8. Folland 6.10 on p. 186. Use the strong form of Theorem **??**.

Exercise 30.9. Folland 2.32 on p. 63. Suppose that $\mu(X) < \infty$. If f and g are complex-valued measurable functions on X, define

$$\rho(f,g) := \int_X \frac{|f-g|}{1+|f-g|}\,d\mu.$$

Show ρ is a metric on the space of measurable functions provided we agree to identify functions that are equal a.e.. Also show that $\rho(f_n, f) \to 0$ as $n \to \infty$ iff $f_n \overset{\mu}{\to} f$ as $n \to \infty$.

Exercise 30.10. Folland 2.38 on p. 63.

Exercise 30.11. Suppose A is an index set, $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are two collections of random variables. If $\{g_\alpha\}_{\alpha\in A}$ is uniformly integrable and $|f_\alpha| \le |g_\alpha|$ for all $\alpha \in A$, show $\{f_\alpha\}_{\alpha\in A}$ is uniformly integrable as well.

31

Approximation Theorems and Convolutions

31.1 Density Theorems

BRUCE: there is a lot of redundancy of this Section with Section 16.2. In this section, $(X, \mathcal{M}, \mu)$ will be a measure space $\mathcal{A}$ will be a subalgebra of $\mathcal{M}$.

Notation 31.1 *Suppose $(X, \mathcal{M}, \mu)$ is a measure space and $\mathcal{A} \subset \mathcal{M}$ is a subalgebra of $\mathcal{M}$. Let $\mathbb{S}(\mathcal{A})$ denote those simple functions $\varphi : X \to \mathbb{C}$ such that $\varphi^{-1}(\{\lambda\}) \in \mathcal{A}$ for all $\lambda \in \mathbb{C}$ and let $\mathbb{S}_f(\mathcal{A}, \mu)$ denote those $\varphi \in \mathbb{S}(\mathcal{A})$ such that $\mu(\varphi \neq 0) < \infty$.*

Remark 31.2. For $\varphi \in \mathbb{S}_f(\mathcal{A}, \mu)$ and $p \in [1, \infty)$, $|\varphi|^p = \sum_{z \neq 0} |z|^p 1_{\{\varphi = z\}}$ and hence

$$\int |\varphi|^p \, d\mu = \sum_{z \neq 0} |z|^p \mu(\varphi = z) < \infty \tag{31.1}$$

so that $\mathbb{S}_f(\mathcal{A}, \mu) \subset L^p(\mu)$. Conversely if $\varphi \in \mathbb{S}(\mathcal{A}) \cap L^p(\mu)$, then from Eq. (31.1) it follows that $\mu(\varphi = z) < \infty$ for all $z \neq 0$ and therefore $\mu(\varphi \neq 0) < \infty$. Hence we have shown, for any $1 \leq p < \infty$,

$$\mathbb{S}_f(\mathcal{A}, \mu) = \mathbb{S}(\mathcal{A}) \cap L^p(\mu).$$

Lemma 31.3 (Simple Functions are Dense). *The simple functions, $\mathbb{S}_f(\mathcal{M}, \mu)$, form a dense subspace of $L^p(\mu)$ for all $1 \leq p < \infty$.*

Proof. Let $\{\varphi_n\}_{n=1}^\infty$ be the simple functions in the approximation Theorem **??**. Since $|\varphi_n| \leq |f|$ for all n, $\varphi_n \in \mathbb{S}_f(\mathcal{M}, \mu)$ and

$$|f - \varphi_n|^p \leq (|f| + |\varphi_n|)^p \leq 2^p |f|^p \in L^1(\mu).$$

Therefore, by the dominated convergence theorem,

$$\lim_{n\to\infty} \int |f - \varphi_n|^p d\mu = \int \lim_{n\to\infty} |f - \varphi_n|^p d\mu = 0.$$

■

The goal of this section is to find a number of other dense subspaces of $L^p(\mu)$ for $p \in [1, \infty)$. The next theorem is the key result of this section.

Theorem 31.4 (Density Theorem). *Let $p \in [1, \infty)$, $(X, \mathcal{M}, \mu)$ be a measure space and M be an algebra of bounded $\mathbb{F}$ – valued ($\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) measurable functions such that*

1. *$M \subset L^p(\mu, \mathbb{F})$ and $\sigma(M) = \mathcal{M}$.*
2. *There exists $\psi_k \in M$ such that $\psi_k \to 1$ boundedly.*
3. *If $\mathbb{F} = \mathbb{C}$ we further assume that M is closed under complex conjugation.*

Then to every function $f \in L^p(\mu, \mathbb{F})$, there exist $\varphi_n \in M$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(\mu)} = 0$, i.e. M is dense in $L^p(\mu, \mathbb{F})$.

Proof. Fix $k \in \mathbb{N}$ for the moment and let $\mathcal{H}$ denote those bounded $\mathcal{M}$ – measurable functions, $f : X \to \mathbb{F}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset M$ such that $\lim_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} = 0$. A routine check shows $\mathcal{H}$ is a subspace of $\ell^\infty(\mathcal{M}, \mathbb{F})$ such that $1 \in \mathcal{H}$, $M \subset \mathcal{H}$ and $\mathcal{H}$ is closed under complex conjugation if $\mathbb{F} = \mathbb{C}$. Moreover, $\mathcal{H}$ is closed under bounded convergence. To see this suppose $f_n \in \mathcal{H}$ and $f_n \to f$ boundedly. Then, by the dominated convergence theorem, $\lim_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} = 0$.[1] (Take the dominating function to be $g = [2C |\psi_k|]^p$ where C is a constant bounding all of the $\{|f_n|\}_{n=1}^\infty$.) We may now choose $\varphi_n \in M$ such that $\|\varphi_n - \psi_k f_n\|_{L^p(\mu)} \leq \frac{1}{n}$ then

$$\begin{aligned} \limsup_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} &\leq \limsup_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} \\ &\quad + \limsup_{n\to\infty} \|\psi_k f_n - \varphi_n\|_{L^p(\mu)} = 0 \end{aligned} \tag{31.2}$$

which implies $f \in \mathcal{H}$. An application of Dynkin's Multiplicative System Theorem 11.28 if $\mathbb{F} = \mathbb{R}$ or Theorem 11.29 if $\mathbb{F} = \mathbb{C}$ now shows $\mathcal{H}$ contains all bounded measurable functions on X.

Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f| \leq k\}} f - f\right\|_{L^p(\mu)} = 0$. (Take the dominating function to be $g = [2C |f|]^p$ where C is a bound on all of the $|\psi_k|$.) Using this and what we have just proved, there exists $\varphi_k \in M$ such that

$$\left\|\psi_k 1_{\{|f| \leq k\}} f - \varphi_k\right\|_{L^p(\mu)} \leq \frac{1}{k}.$$

[1] It is at this point that the proof would break down if $p = \infty$.

The same line of reasoning used in Eq. (31.2) now implies $\lim_{k\to\infty}\|f-\varphi_k\|_{L^p(\mu)}=0$. ■

Definition 31.5. *Let (X,τ) be a topological space and μ be a measure on $\mathcal{B}_X=\sigma(\tau)$. A* ***locally integrable*** *function is a Borel measurable function $f:X\to\mathbb{C}$ such that $\int_K |f|\,d\mu<\infty$ for all compact subsets $K\subset X$. We will write $L^1_{loc}(\mu)$ for the space of locally integrable functions. More generally we say $f\in L^p_{loc}(\mu)$ iff $\|1_K f\|_{L^p(\mu)}<\infty$ for all compact subsets $K\subset X$.*

Definition 31.6. *Let (X,τ) be a topological space. A* ***K-finite measure*** *on X is Borel measure μ such that $\mu(K)<\infty$ for all compact subsets $K\subset X$.*

Lebesgue measure on $\mathbb{R}$ is an example of a K-finite measure while counting measure on $\mathbb{R}$ is not a K-finite measure.

Example 31.7. Suppose that μ is a K-finite measure on $\mathcal{B}_{\mathbb{R}^d}$. An application of Theorem 31.4 shows $C_c(\mathbb{R},\mathbb{C})$ is dense in $L^p(\mathbb{R}^d,\mathcal{B}_{\mathbb{R}^d},\mu;\mathbb{C})$. To apply Theorem 31.4, let $M:=C_c(\mathbb{R}^d,\mathbb{C})$ and $\psi_k(x):=\psi(x/k)$ where $\psi\in C_c(\mathbb{R}^d,\mathbb{C})$ with $\psi(x)=1$ in a neighborhood of 0. The proof is completed by showing $\sigma(M)=\sigma(C_c(\mathbb{R}^d,\mathbb{C}))=\mathcal{B}_{\mathbb{R}^d}$, which follows directly from Lemma 11.34.

We may also give a more down to earth proof as follows. Let $x_0\in\mathbb{R}^d$, $R>0$, $A:=B(x_0,R)^c$ and $f_n(x):=d_A^{1/n}(x)$. Then $f_n\in M$ and $f_n\to 1_{B(x_0,R)}$ as $n\to\infty$ which shows $1_{B(x_0,R)}$ is $\sigma(M)$-measurable, i.e. $B(x_0,R)\in\sigma(M)$. Since $x_0\in\mathbb{R}^d$ and $R>0$ were arbitrary, $\sigma(M)=\mathcal{B}_{\mathbb{R}^d}$.

More generally we have the following result.

Theorem 31.8. *Let (X,τ) be a second countable locally compact Hausdorff space and $\mu:\mathcal{B}_X\to[0,\infty]$ be a K-finite measure. Then $C_c(X)$ (the space of continuous functions with compact support) is dense in $L^p(\mu)$ for all $p\in[1,\infty)$. (See also Proposition 38.17 below.)*

Proof. Let $M:=C_c(X)$ and use Item 3. of Lemma 11.34 to find functions $\psi_k\in M$ such that $\psi_k\to 1$ to boundedly as $k\to\infty$. The result now follows from an application of Theorem 31.4 along with the aid of item 4. of Lemma 11.34. ■

Exercise 31.1. Show that $BC(\mathbb{R},\mathbb{C})$ is not dense in $L^\infty(\mathbb{R},\mathcal{B}_\mathbb{R},m;\mathbb{C})$. Hence the hypothesis that $p<\infty$ in Theorem 31.4 can not be removed.

Corollary 31.9. *Suppose $X\subset\mathbb{R}^n$ is an open set, $\mathcal{B}_X$ is the Borel σ – algebra on X and μ be a K-finite measure on $(X,\mathcal{B}_X)$. Then $C_c(X)$ is dense in $L^p(\mu)$ for all $p\in[1,\infty)$.*

Corollary 31.10. *Suppose that X is a compact subset of $\mathbb{R}^n$ and μ is a finite measure on $(X,\mathcal{B}_X)$, then polynomials are dense in $L^p(X,\mu)$ for all $1\le p<\infty$.*

Proof. Consider X to be a metric space with usual metric induced from $\mathbb{R}^n$. Then X is a locally compact separable metric space and therefore $C_c(X,\mathbb{C})=C(X,\mathbb{C})$ is dense in $L^p(\mu)$ for all $p\in[1,\infty)$. Since, by the dominated convergence theorem, uniform convergence implies $L^p(\mu)$ – convergence, it follows from the Weierstrass approximation theorem (see Theorem 32.39 and Corollary 32.41 or Theorem 37.31 and Corollary 37.32) that polynomials are also dense in $L^p(\mu)$. ■

Lemma 31.11. *Let (X,τ) be a second countable locally compact Hausdorff space and $\mu:\mathcal{B}_X\to[0,\infty]$ be a K-finite measure on X. If $h\in L^1_{loc}(\mu)$ is a function such that*

$$\int_X fh\,d\mu=0 \text{ for all } f\in C_c(X) \tag{31.3}$$

then $h(x)=0$ for μ – a.e. x. (See also Corollary 38.20 below.)

Proof. Let $d\nu(x)=|h(x)|\,dx$, then ν is a K-finite measure on X and hence $C_c(X)$ is dense in $L^1(\nu)$ by Theorem 31.8. Notice that

$$\int_X f\cdot\operatorname{sgn}(h)d\nu=\int_X fh\,d\mu=0 \text{ for all } f\in C_c(X). \tag{31.4}$$

Let $\{K_k\}_{k=1}^\infty$ be a sequence of compact sets such that $K_k\uparrow X$ as in Lemma 36.5. Then $1_{K_k}\overline{\operatorname{sgn}(h)}\in L^1(\nu)$ and therefore there exists $f_m\in C_c(X)$ such that $f_m\to 1_{K_k}\overline{\operatorname{sgn}(h)}$ in $L^1(\nu)$. So by Eq. (31.4),

$$\nu(K_k)=\int_X 1_{K_k}d\nu=\lim_{m\to\infty}\int_X f_m\operatorname{sgn}(h)d\nu=0.$$

Since $K_k\uparrow X$ as $k\to\infty$, $0=\nu(X)=\int_X|h|\,d\mu$, i.e. $h(x)=0$ for μ – a.e. x. ■

As an application of Lemma 31.11 and Example 37.34, we will show that the Laplace transform is injective.

Theorem 31.12 (Injectivity of the Laplace Transform). *For $f\in L^1([0,\infty),dx)$, the Laplace transform of f is defined by*

$$\mathcal{L}f(\lambda):=\int_0^\infty e^{-\lambda x}f(x)dx \text{ for all } \lambda>0.$$

If $\mathcal{L}f(\lambda):=0$ then $f(x)=0$ for m -a.e. x.

Proof. Suppose that $f\in L^1([0,\infty),dx)$ such that $\mathcal{L}f(\lambda)\equiv 0$. Let $g\in C_0([0,\infty),\mathbb{R})$ and $\varepsilon>0$ be given. By Example 37.34 we may choose $\{a_\lambda\}_{\lambda>0}$ such that $\#(\{\lambda>0:a_\lambda\neq 0\})<\infty$ and

$$|g(x)-\sum_{\lambda>0}a_\lambda e^{-\lambda x}|<\varepsilon \text{ for all } x\ge 0.$$

Then

$$\left|\int_0^\infty g(x)f(x)dx\right| = \left|\int_0^\infty \left(g(x) - \sum_{\lambda>0} a_\lambda e^{-\lambda x}\right) f(x)dx\right|$$
$$\le \int_0^\infty \left|g(x) - \sum_{\lambda>0} a_\lambda e^{-\lambda x}\right| |f(x)|\, dx \le \varepsilon \|f\|_1.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\int_0^\infty g(x)f(x)dx = 0$ for all $g \in C_0([0,\infty),\mathbb{R})$. The proof is finished by an application of Lemma 31.11. ■

Here is another variant of Theorem 31.8.

Theorem 31.13. *Let (X,d) be a metric space, τ_d be the topology on X generated by d and $\mathcal{B}_X = \sigma(\tau_d)$ be the Borel σ – algebra. Suppose $\mu : \mathcal{B}_X \to [0,\infty]$ is a measure which is σ – finite on τ_d and let $BC_f(X)$ denote the bounded continuous functions on X such that $\mu(f \neq 0) < \infty$. Then $BC_f(X)$ is a dense subspace of $L^p(\mu)$ for any $p \in [1,\infty)$.*

Proof. Let $X_k \in \tau_d$ be open sets such that $X_k \uparrow X$ and $\mu(X_k) < \infty$ and let

$$\psi_k(x) = \min(1, k \cdot d_{X_k^c}(x)) = \varphi_k(d_{X_k^c}(x)),$$

see Figure 31.1 below. It is easily verified that $M := BC_f(X)$ is an algebra,

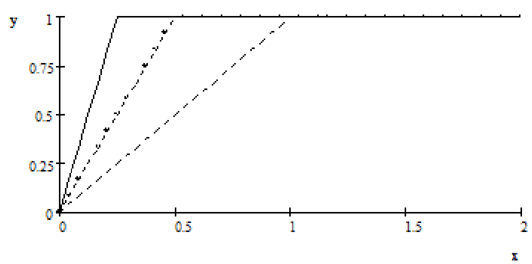


Fig. 31.1. The plot of ϕ_n for $n = 1$, 2, and 4. Notice that $\phi_n \to 1_{(0,\infty)}$.

$\psi_k \in M$ for all k and $\psi_k \to 1$ boundedly as $k \to \infty$. Given $V \in \tau$ and $k, n \in \mathbb{N}$,let

$$f_{k,n}(x) := \min(1, n \cdot d_{(V\cap X_k)^c}(x)).$$

Then $\{f_{k,n} \neq 0\} = V \cap X_k$ so $f_{k,n} \in BC_f(X)$. Moreover

$$\lim_{k\to\infty}\lim_{n\to\infty} f_{k,n} = \lim_{k\to\infty} 1_{V\cap X_k} = 1_V$$

which shows $V \in \sigma(M)$ and hence $\sigma(M) = \mathcal{B}_X$. The proof is now completed by an application of Theorem 31.4. ■

Exercise 31.2. (BRUCE: Should drop this exercise.) Suppose that (X,d) is a metric space, μ is a measure on $\mathcal{B}_X := \sigma(\tau_d)$ which is finite on bounded measurable subsets of X. Show $BC_b(X,\mathbb{R})$, defined in Eq. (**??**), is dense in $L^p(\mu)$. **Hints:** let ψ_k be as defined in Eq. (**??**) which incidentally may be used to show $\sigma(BC_b(X,\mathbb{R})) = \sigma(BC(X,\mathbb{R}))$. Then use the argument in the proof of Corollary 11.32 to show $\sigma(BC(X,\mathbb{R})) = \mathcal{B}_X$.

Theorem 31.14. *Suppose $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{M}$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and μ is σ – finite on $\mathcal{A}$. Then $\mathbb{S}_f(\mathcal{A},\mu)$ is dense in $L^p(\mu)$. (See also Remark* **??** *below.)*

Proof. Let $M := \mathbb{S}_f(\mathcal{A},\mu)$. By assumption there exists $X_k \in \mathcal{A}$ such that $\mu(X_k) < \infty$ and $X_k \uparrow X$ as $k \to \infty$. If $A \in \mathcal{A}$, then $X_k \cap A \in \mathcal{A}$ and $\mu(X_k \cap A) < \infty$ so that $1_{X_k\cap A} \in M$. Therefore $1_A = \lim_{k\to\infty} 1_{X_k\cap A}$ is $\sigma(M)$ – measurable for every $A \in \mathcal{A}$. So we have shown that $\mathcal{A} \subset \sigma(M) \subset \mathcal{M}$ and therefore $\mathcal{M} = \sigma(\mathcal{A}) \subset \sigma(M) \subset \mathcal{M}$, i.e. $\sigma(M) = \mathcal{M}$. The theorem now follows from Theorem 31.4 after observing $\psi_k := 1_{X_k} \in M$ and $\psi_k \to 1$ boundedly. ■

Theorem 31.15 (Separability of L^p – Spaces). *Suppose, $p \in [1,\infty)$, $\mathcal{A} \subset \mathcal{M}$ is a countable algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and μ is σ – finite on $\mathcal{A}$. Then $L^p(\mu)$ is separable and*

$$\mathbb{D} = \{\sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q},\ A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty\}$$

is a countable dense subset.

Proof. It is left to reader to check $\mathbb{D}$ is dense in $\mathbb{S}_f(\mathcal{A},\mu)$ relative to the $L^p(\mu)$ – norm. The proof is then complete since $\mathbb{S}_f(\mathcal{A},\mu)$ is a dense subspace of $L^p(\mu)$ by Theorem 31.14. ■

Example 31.16. The collection of functions of the form $\varphi = \sum_{k=1}^n c_k 1_{(a_k,b_k]}$ with $a_k, b_k \in \mathbb{Q}$ and $a_k < b_k$ are dense in $L^p(\mathbb{R},\mathcal{B}_\mathbb{R},m;\mathbb{C})$ and $L^p(\mathbb{R},\mathcal{B}_\mathbb{R},m;\mathbb{C})$ is separable for any $p \in [1,\infty)$. To prove this simply apply Theorem 31.14 with $\mathcal{A}$ being the algebra on $\mathbb{R}$ generated by the half open intervals $(a,b] \cap \mathbb{R}$ with $a < b$ and $a, b \in \mathbb{Q}\cup\{\pm\infty\}$, i.e. $\mathcal{A}$ consists of sets of the form $\coprod_{k=1}^n (a_k,b_k] \cap \mathbb{R}$, where $a_k, b_k \in \mathbb{Q}\cup\{\pm\infty\}$.

Exercise 31.3. Show $L^\infty([0,1],\mathcal{B}_\mathbb{R},m;\mathbb{C})$ is not separable. **Hint:** Suppose Γ is a dense subset of $L^\infty([0,1],\mathcal{B}_\mathbb{R},m;\mathbb{C})$ and for $\lambda \in (0,1)$, let $f_\lambda(x) := 1_{[0,\lambda]}(x)$. For each $\lambda \in (0,1)$, choose $g_\lambda \in \Gamma$ such that $\|f_\lambda - g_\lambda\|_\infty < 1/2$ and then show the map $\lambda \in (0,1) \to g_\lambda \in \Gamma$ is injective. Use this to conclude that Γ must be uncountable.

Corollary 31.17 (Riemann Lebesgue Lemma). *Suppose that $f \in L^1(\mathbb{R},m)$, then*

$$\lim_{\lambda\to\pm\infty}\int_\mathbb{R} f(x)e^{i\lambda x}dm(x) = 0.$$

Proof. By Example 31.16, given $\varepsilon > 0$ there exists $\varphi = \sum_{k=1}^n c_k 1_{(a_k,b_k]}$ with $a_k, b_k \in \mathbb{R}$ such that

$$\int_{\mathbb{R}} |f - \varphi| dm < \varepsilon.$$

Notice that

$$\begin{aligned}\int_{\mathbb{R}} \varphi(x) e^{i\lambda x} dm(x) &= \int_{\mathbb{R}} \sum_{k=1}^n c_k 1_{(a_k,b_k]}(x) e^{i\lambda x} dm(x) \\ &= \sum_{k=1}^n c_k \int_{a_k}^{b_k} e^{i\lambda x} dm(x) = \sum_{k=1}^n c_k \lambda^{-1} e^{i\lambda x}|_{a_k}^{b_k} \\ &= \lambda^{-1} \sum_{k=1}^n c_k \left(e^{i\lambda b_k} - e^{i\lambda a_k}\right) \to 0 \text{ as } |\lambda| \to \infty.\end{aligned}$$

Combining these two equations with

$$\begin{aligned}\left|\int_{\mathbb{R}} f(x) e^{i\lambda x} dm(x)\right| &\le \left|\int_{\mathbb{R}} (f(x) - \varphi(x)) e^{i\lambda x} dm(x)\right| + \left|\int_{\mathbb{R}} \varphi(x) e^{i\lambda x} dm(x)\right| \\ &\le \int_{\mathbb{R}} |f - \varphi| dm + \left|\int_{\mathbb{R}} \varphi(x) e^{i\lambda x} dm(x)\right| \\ &\le \varepsilon + \left|\int_{\mathbb{R}} \varphi(x) e^{i\lambda x} dm(x)\right|\end{aligned}$$

we learn that

$$\limsup_{|\lambda|\to\infty} \left|\int_{\mathbb{R}} f(x) e^{i\lambda x} dm(x)\right| \le \varepsilon + \limsup_{|\lambda|\to\infty} \left|\int_{\mathbb{R}} \varphi(x) e^{i\lambda x} dm(x)\right| = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof of the Riemann Lebesgue lemma. ∎

Corollary 31.18. *Suppose $\mathcal{A} \subset \mathcal{M}$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{M}$ and μ is σ – finite on $\mathcal{A}$. Then for every $B \in \mathcal{M}$ such that $\mu(B) < \infty$ and $\varepsilon > 0$ there exists $D \in \mathcal{A}$ such that $\mu(B \triangle D) < \varepsilon$. (See also Remark* **??** *below.)*

Proof. By Theorem 31.14, there exists a collection, $\{A_i\}_{i=1}^n$, of pairwise disjoint subsets of $\mathcal{A}$ and $\lambda_i \in \mathbb{R}$ such that $\int_X |1_B - f| d\mu < \varepsilon$ where $f = \sum_{i=1}^n \lambda_i 1_{A_i}$. Let $A_0 := X \setminus \cup_{i=1}^n A_i \in \mathcal{A}$ then

$$\begin{aligned}\int_X |1_B - f| d\mu &= \sum_{i=0}^n \int_{A_i} |1_B - f| d\mu \\ &= \mu(A_0 \cap B) + \sum_{i=1}^n \left[\int_{A_i \cap B} |1_B - \lambda_i| d\mu + \int_{A_i \setminus B} |1_B - \lambda_i| d\mu\right] \\ &= \mu(A_0 \cap B) + \sum_{i=1}^n [|1 - \lambda_i| \mu(B \cap A_i) + |\lambda_i| \mu(A_i \setminus B)] \qquad (31.5) \\ &\ge \mu(A_0 \cap B) + \sum_{i=1}^n \min\{\mu(B \cap A_i), \mu(A_i \setminus B)\} \qquad (31.6)\end{aligned}$$

where the last equality is a consequence of the fact that $1 \le |\lambda_i| + |1 - \lambda_i|$. Let

$$\alpha_i = \begin{cases} 0 \text{ if } \mu(B \cap A_i) < \mu(A_i \setminus B) \\ 1 \text{ if } \mu(B \cap A_i) \ge \mu(A_i \setminus B) \end{cases}$$

and $g = \sum_{i=1}^n \alpha_i 1_{A_i} = 1_D$ where

$$D := \cup\{A_i : i > 0 \ \&\ \alpha_i = 1\} \in \mathcal{A}.$$

Equation (31.5) with λ_i replaced by α_i and f by g implies

$$\int_X |1_B - 1_D| d\mu = \mu(A_0 \cap B) + \sum_{i=1}^n \min\{\mu(B \cap A_i), \mu(A_i \setminus B)\}.$$

The latter expression, by Eq. (31.6), is bounded by $\int_X |1_B - f| d\mu < \varepsilon$ and therefore,

$$\mu(B \triangle D) = \int_X |1_B - 1_D| d\mu < \varepsilon.$$

∎

Remark 31.19. We have to assume that $\mu(B) < \infty$ as the following example shows. Let $X = \mathbb{R}$, $\mathcal{M} = \mathcal{B}$, $\mu = m$, $\mathcal{A}$ be the algebra generated by half open intervals of the form $(a, b]$, and $B = \cup_{n=1}^\infty (2n, 2n+1]$. It is easily checked that for every $D \in \mathcal{A}$, that $m(B \Delta D) = \infty$.

31.2 Convolution and Young's Inequalities

Throughout this section we will be solely concerned with d – dimensional Lebesgue measure, m, and we will simply write L^p for $L^p(\mathbb{R}^d, m)$.

Definition 31.20 (Convolution). *Let $f, g : \mathbb{R}^d \to \mathbb{C}$ be measurable functions. We define*

$$f * g(x) = \int_{\mathbb{R}^d} f(x-y)g(y)dy \tag{31.7}$$

whenever the integral is defined, i.e. either $f(x-\cdot)\, g(\cdot) \in L^1(\mathbb{R}^d, m)$ or $f(x-\cdot)\, g(\cdot) \geq 0$. Notice that the condition that $f(x-\cdot)\, g(\cdot) \in L^1(\mathbb{R}^d, m)$ is equivalent to writing $|f| * |g|(x) < \infty$. By convention, if the integral in Eq. (31.7) is not defined, let $f * g(x) := 0$.

Notation 31.21 *Given a multi-index $\alpha \in \mathbb{Z}_+^d$, let $|\alpha| = \alpha_1 + \cdots + \alpha_d$,*

$$x^\alpha := \prod_{j=1}^d x_j^{\alpha_j}, \text{ and } \partial_x^\alpha = \left(\frac{\partial}{\partial x}\right)^\alpha := \prod_{j=1}^d \left(\frac{\partial}{\partial x_j}\right)^{\alpha_j}.$$

For $z \in \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{C}$, let $\tau_z f : \mathbb{R}^d \to \mathbb{C}$ be defined by $\tau_z f(x) = f(x-z)$.

Remark 31.22 (The Significance of Convolution).

1. Suppose that $f, g \in L^1(m)$ are positive functions and let μ be the measure on $\left(\mathbb{R}^d\right)^2$ defined by
$$d\mu(x,y) := f(x)\, g(y)\, dm(x)\, dm(y).$$
Then if $h : \mathbb{R} \to [0, \infty]$ is a measurable function we have
$$\begin{aligned}\int_{(\mathbb{R}^d)^2} h(x+y)\, d\mu(x,y) &= \int_{(\mathbb{R}^d)^2} h(x+y)\, f(x)\, g(y)\, dm(x)\, dm(y) \\ &= \int_{(\mathbb{R}^d)^2} h(x)\, f(x-y)\, g(y)\, dm(x)\, dm(y) \\ &= \int_{\mathbb{R}^d} h(x)\, f * g(x)\, dm(x).\end{aligned}$$
In other words, this shows the measure $(f * g)\, m$ is the same as $S_*\mu$ where $S(x,y) := x + y$. In probability lingo, the distribution of a sum of two "independent" (i.e. product measure) random variables is the the convolution of the individual distributions.
2. Suppose that $L = \sum_{|\alpha| \leq k} a_\alpha \partial^\alpha$ is a constant coefficient differential operator and suppose that we can solve (uniquely) the equation $Lu = g$ in the form
$$u(x) = Kg(x) := \int_{\mathbb{R}^d} k(x,y)g(y)dy$$
where $k(x,y)$ is an "integral kernel." (This is a natural sort of assumption since, in view of the fundamental theorem of calculus, integration is the inverse operation to differentiation.) Since $\tau_z L = L\tau_z$ for all $z \in \mathbb{R}^d$, (this is another way to characterize constant coefficient differential operators) and $L^{-1} = K$ we should have $\tau_z K = K\tau_z$. Writing out this equation then says
$$\begin{aligned}\int_{\mathbb{R}^d} k(x-z,y)g(y)dy &= (Kg)(x-z) = \tau_z Kg(x) = (K\tau_z g)(x) \\ &= \int_{\mathbb{R}^d} k(x,y)g(y-z)dy = \int_{\mathbb{R}^d} k(x,y+z)g(y)dy.\end{aligned}$$
Since g is arbitrary we conclude that $k(x-z,y) = k(x,y+z)$. Taking $y = 0$ then gives
$$k(x,z) = k(x-z,0) =: \rho(x-z).$$
We thus find that $Kg = \rho * g$. Hence we expect the convolution operation to appear naturally when solving constant coefficient partial differential equations. More about this point later.

Proposition 31.23. *Suppose $p \in [1, \infty]$, $f \in L^1$ and $g \in L^p$, then $f * g(x)$ exists for almost every x, $f * g \in L^p$ and*

$$\|f * g\|_p \leq \|f\|_1 \|g\|_p.$$

Proof. This follows directly from Minkowski's inequality for integrals, Theorem 29.2, and was explained in Example 29.3. ∎

Example 31.24. Suppose that Ω is a bounded Borel subset of $\mathbb{R}^n$, $\alpha < n$, then the operator $K : L^2(\Omega, m) \to L^2(\Omega, m)$ defined by

$$Kf(x) := \int_\Omega \frac{1}{|x-y|^\alpha} f(y)dy$$

is compact.

Proof. For $\varepsilon \geq 0$, let

$$K_\varepsilon f(x) := \int_\Omega \frac{1}{|x-y|^\alpha + \varepsilon} f(y)dy = [g_\varepsilon * (1_\Omega f)](x)$$

where $g_\varepsilon(x) = \frac{1}{|x|^\alpha + \varepsilon} 1_C(x)$ with $C \subset \mathbb{R}^n$ a sufficiently large ball such that $\Omega - \Omega \subset C$. Since $\alpha < n$, it follows that

$$g_\varepsilon \leq g_0 = |\cdot|^{-\alpha} 1_C \in L^1(\mathbb{R}^n, m).$$

Hence it follows by Proposition 31.23 that

$$\begin{aligned}\|(K - K_\varepsilon) f\|_{L^2(\Omega)} &\leq \|(g_0 - g_\varepsilon) * (1_\Omega f)\|_{L^2(\mathbb{R}^n)} \\ &\leq \|(g_0 - g_\varepsilon)\|_{L^1(\mathbb{R}^n)} \|1_\Omega f\|_{L^2(\mathbb{R}^n)} \\ &= \|(g_0 - g_\varepsilon)\|_{L^1(\mathbb{R}^n)} \|f\|_{L^2(\Omega)}\end{aligned}$$

which implies

$$\|K - K_\varepsilon\|_{B(L^2(\Omega))} \le \|g_0 - g_\varepsilon\|_{L^1(\mathbb{R}^n)} = \int_C \left| \frac{1}{|x|^\alpha + \varepsilon} - \frac{1}{|x|^\alpha} \right| dx \to 0 \text{ as } \varepsilon \downarrow 0 \tag{31.8}$$

by the dominated convergence theorem. For any $\varepsilon > 0$,

$$\int_{\Omega\times\Omega} \left[\frac{1}{|x-y|^\alpha + \varepsilon} \right]^2 dxdy < \infty,$$

and hence K_ε is Hilbert Schmidt and hence compact. By Eq. (31.8), $K_\varepsilon \to K$ as $\varepsilon \downarrow 0$ and hence it follows that K is compact as well. ■

Proposition 31.25. *Suppose that $p \in [1,\infty)$, then $\tau_z : L^p \to L^p$ is an isometric isomorphism and for $f \in L^p$, $z \in \mathbb{R}^d \to \tau_z f \in L^p$ is continuous.*

Proof. The assertion that $\tau_z : L^p \to L^p$ is an isometric isomorphism follows from translation invariance of Lebesgue measure and the fact that $\tau_{-z} \circ \tau_z = id$. For the continuity assertion, observe that

$$\|\tau_z f - \tau_y f\|_p = \|\tau_{-y}(\tau_z f - \tau_y f)\|_p = \|\tau_{z-y} f - f\|_p$$

from which it follows that it is enough to show $\tau_z f \to f$ in L^p as $z \to 0 \in \mathbb{R}^d$. When $f \in C_c(\mathbb{R}^d)$, $\tau_z f \to f$ uniformly and since the $K := \cup_{|z|\le 1} \mathrm{supp}(\tau_z f)$ is compact, it follows by the dominated convergence theorem that $\tau_z f \to f$ in L^p as $z \to 0 \in \mathbb{R}^d$. For general $g \in L^p$ and $f \in C_c(\mathbb{R}^d)$,

$$\begin{aligned}\|\tau_z g - g\|_p &\le \|\tau_z g - \tau_z f\|_p + \|\tau_z f - f\|_p + \|f - g\|_p \\ &= \|\tau_z f - f\|_p + 2\|f - g\|_p\end{aligned}$$

and thus

$$\limsup_{z\to 0} \|\tau_z g - g\|_p \le \limsup_{z\to 0} \|\tau_z f - f\|_p + 2\|f - g\|_p = 2\|f - g\|_p.$$

Because $C_c(\mathbb{R}^d)$ is dense in L^p, the term $\|f - g\|_p$ may be made as small as we please. ■

Exercise 31.4. Let $p \in [1,\infty]$ and $\|\tau_z - I\|_{L(L^p(m))}$ be the operator norm $\tau_z - I$. Show $\|\tau_z - I\|_{L(L^p(m))} = 2$ for all $z \in \mathbb{R}^d \setminus \{0\}$ and conclude from this that $z \in \mathbb{R}^d \to \tau_z \in L(L^p(m))$ is **not** continuous.

Hints: 1) Show $\|\tau_z - I\|_{L(L^p(m))} = \left\|\tau_{|z|e_1} - I\right\|_{L(L^p(m))}$. 2) Let $z = te_1$ with $t > 0$ and look for $f \in L^p(m)$ such that $\tau_z f$ is approximately equal to $-f$. (In fact, if $p = \infty$, you can find $f \in L^\infty(m)$ such that $\tau_z f = -f$.) (BRUCE: add on a problem somewhere showing that $\sigma(\tau_z) = S^1 \subset \mathbb{C}$. This is very simple to prove if $p = 2$ by using the Fourier transform.)

Definition 31.26. *Suppose that (X,τ) is a topological space and μ is a measure on $\mathcal{B}_X = \sigma(\tau)$. For a measurable function $f : X \to \mathbb{C}$ we define the essential support of f by*

$$\mathrm{supp}_\mu(f) = \{x \in X : \mu(\{y \in V : f(y) \ne 0\}) > 0\ \forall \text{ neighborhoods } V \text{ of } x\}. \tag{31.9}$$

Equivalently, $x \notin \mathrm{supp}_\mu(f)$ iff there exists an open neighborhood V of x such that $1_V f = 0$ a.e.

It is not hard to show that if $\mathrm{supp}(\mu) = X$ (see Definition 30.14) and $f \in C(X)$ then $\mathrm{supp}_\mu(f) = \mathrm{supp}(f) := \overline{\{f \ne 0\}}$, see Exercise 31.7.

Lemma 31.27. *Suppose (X,τ) is second countable and $f : X \to \mathbb{C}$ is a measurable function and μ is a measure on $\mathcal{B}_X$. Then $X := U \setminus \mathrm{supp}_\mu(f)$ may be described as the largest open set W such that $f1_W(x) = 0$ for μ – a.e. x. Equivalently put, $C := \mathrm{supp}_\mu(f)$ is the smallest closed subset of X such that $f = f1_C$ a.e.*

Proof. To verify that the two descriptions of $\mathrm{supp}_\mu(f)$ are equivalent, suppose $\mathrm{supp}_\mu(f)$ is defined as in Eq. (31.9) and $W := X \setminus \mathrm{supp}_\mu(f)$. Then

$$\begin{aligned}W &= \{x \in X : \exists\ \tau \ni V \ni x \text{ such that } \mu(\{y \in V : f(y) \ne 0\}) = 0\} \\ &= \cup\{V \subset_o X : \mu(f1_V \ne 0) = 0\} \\ &= \cup\{V \subset_o X : f1_V = 0 \text{ for } \mu - \text{a.e.}\}.\end{aligned}$$

So to finish the argument it suffices to show $\mu(f1_W \ne 0) = 0$. To to this let $\mathcal{U}$ be a countable base for τ and set

$$\mathcal{U}_f := \{V \in \mathcal{U} : f1_V = 0 \text{ a.e.}\}.$$

Then it is easily seen that $W = \cup\mathcal{U}_f$ and since $\mathcal{U}_f$ is countable

$$\mu(f1_W \ne 0) \le \sum_{V\in\mathcal{U}_f} \mu(f1_V \ne 0) = 0.$$

■

Lemma 31.28. *Suppose $f, g, h : \mathbb{R}^d \to \mathbb{C}$ are measurable functions and assume that x is a point in $\mathbb{R}^d$ such that $|f| * |g|(x) < \infty$ and $|f| * (|g| * |h|)(x) < \infty$, then*

1. *$f * g(x) = g * f(x)$*
2. *$f * (g * h)(x) = (f * g) * h(x)$*
3. *If $z \in \mathbb{R}^d$ and $\tau_z(|f| * |g|)(x) = |f| * |g|(x - z) < \infty$, then*

$$\tau_z(f * g)(x) = \tau_z f * g(x) = f * \tau_z g(x)$$

4. *If $x \notin \operatorname{supp}_m(f) + \operatorname{supp}_m(g)$ then $f * g(x) = 0$ and in particular,*

$$\operatorname{supp}_m(f * g) \subset \overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}$$

where in defining $\operatorname{supp}_m(f*g)$ we will use the convention that "$f*g(x) \neq 0$" when $|f| * |g|(x) = \infty$.

Proof. For item 1.,

$$|f| * |g|(x) = \int_{\mathbb{R}^d} |f|(x-y)\, |g|(y) dy = \int_{\mathbb{R}^d} |f|(y)\, |g|(y-x) dy = |g| * |f|(x)$$

where in the second equality we made use of the fact that Lebesgue measure invariant under the transformation $y \to x - y$. Similar computations prove all of the remaining assertions of the first three items of the lemma. Item 4. Since $f * g(x) = \tilde{f} * \tilde{g}(x)$ if $f = \tilde{f}$ and $g = \tilde{g}$ a.e. we may, by replacing f by $f 1_{\operatorname{supp}_m(f)}$ and g by $g 1_{\operatorname{supp}_m(g)}$ if necessary, assume that $\{f \neq 0\} \subset \operatorname{supp}_m(f)$ and $\{g \neq 0\} \subset \operatorname{supp}_m(g)$. So if $x \notin (\operatorname{supp}_m(f) + \operatorname{supp}_m(g))$ then $x \notin (\{f \neq 0\} + \{g \neq 0\})$ and for all $y \in \mathbb{R}^d$, either $x - y \notin \{f \neq 0\}$ or $y \notin \{g \neq 0\}$. That is to say either $x - y \in \{f = 0\}$ or $y \in \{g = 0\}$ and hence $f(x-y)g(y) = 0$ for all y and therefore $f * g(x) = 0$. This shows that $f * g = 0$ on $\mathbb{R}^d \setminus \left(\overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}\right)$ and therefore

$$\mathbb{R}^d \setminus \left(\overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}\right) \subset \mathbb{R}^d \setminus \operatorname{supp}_m(f * g),$$

i.e. $\operatorname{supp}_m(f * g) \subset \overline{\operatorname{supp}_m(f) + \operatorname{supp}_m(g)}$. ∎

Remark 31.29. Let A, B be closed sets of $\mathbb{R}^d$, it is not necessarily true that $A + B$ is still closed. For example, take

$$A = \{(x, y) : x > 0 \text{ and } y \geq 1/x\} \text{ and } B = \{(x, y) : x < 0 \text{ and } y \geq 1/|x|\},$$

then every point of $A + B$ has a positive y - component and hence is not zero. On the other hand, for $x > 0$ we have $(x, 1/x) + (-x, 1/x) = (0, 2/x) \in A + B$ for all x and hence $0 \in \overline{A + B}$ showing $A + B$ is not closed. Nevertheless if one of the sets A or B is compact, then $A + B$ is closed again. Indeed, if A is compact and $x_n = a_n + b_n \in A + B$ and $x_n \to x \in \mathbb{R}^d$, then by passing to a subsequence if necessary we may assume $\lim_{n\to\infty} a_n = a \in A$ exists. In this case

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} (x_n - a_n) = x - a \in B$$

exists as well, showing $x = a + b \in A + B$.

Proposition 31.30. *Suppose that $p, q \in [1, \infty]$ and p and q are conjugate exponents, $f \in L^p$ and $g \in L^q$, then $f * g \in BC(\mathbb{R}^d)$, $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$ and if $p, q \in (1, \infty)$ then $f * g \in C_0(\mathbb{R}^d)$.*

Proof. The existence of $f * g(x)$ and the estimate $|f * g|(x) \leq \|f\|_p \|g\|_q$ for all $x \in \mathbb{R}^d$ is a simple consequence of Holders inequality and the translation invariance of Lebesgue measure. In particular this shows $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$. By relabeling p and q if necessary we may assume that $p \in [1, \infty)$. Since

$$\begin{aligned} \|\tau_z (f * g) - f * g\|_u &= \|\tau_z f * g - f * g\|_u \\ &\leq \|\tau_z f - f\|_p \|g\|_q \to 0 \text{ as } z \to 0 \end{aligned}$$

it follows that $f * g$ is uniformly continuous. Finally if $p, q \in (1, \infty)$, we learn from Lemma 31.28 and what we have just proved that $f_m * g_m \in C_c(\mathbb{R}^d)$ where $f_m = f 1_{|f| \leq m}$ and $g_m = g 1_{|g| \leq m}$. Moreover,

$$\begin{aligned} \|f * g - f_m * g_m\|_\infty &\leq \|f * g - f_m * g\|_\infty + \|f_m * g - f_m * g_m\|_\infty \\ &\leq \|f - f_m\|_p \|g\|_q + \|f_m\|_p \|g - g_m\|_q \\ &\leq \|f - f_m\|_p \|g\|_q + \|f\|_p \|g - g_m\|_q \to 0 \text{ as } m \to \infty \end{aligned}$$

showing, with the aid of Proposition 37.23, $f * g \in C_0(\mathbb{R}^d)$. ∎

Theorem 31.31 (Young's Inequality). *Let $p, q, r \in [1, \infty]$ satisfy*

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}. \tag{31.10}$$

If $f \in L^p$ and $g \in L^q$ then $|f| * |g|(x) < \infty$ for m – a.e. x and

$$\|f * g\|_r \leq \|f\|_p \|g\|_q. \tag{31.11}$$

In particular L^1 is closed under convolution. (The space $(L^1, *)$ is an example of a "Banach algebra" without unit.)

Remark 31.32. Before going to the formal proof, let us first understand Eq. (31.10) by the following scaling argument. For $\lambda > 0$, let $f_\lambda(x) := f(\lambda x)$, then after a few simple change of variables we find

$$\|f_\lambda\|_p = \lambda^{-d/p} \|f\| \text{ and } (f * g)_\lambda = \lambda^d f_\lambda * g_\lambda.$$

Therefore if Eq. (31.11) holds for some $p, q, r \in [1, \infty]$, we would also have

$$\|f * g\|_r = \lambda^{d/r} \|(f * g)_\lambda\|_r \leq \lambda^{d/r} \lambda^d \|f_\lambda\|_p \|g_\lambda\|_q = \lambda^{(d + d/r - d/p - d/q)} \|f\|_p \|g\|_q$$

for all $\lambda > 0$. This is only possible if Eq. (31.10) holds.

Proof. By the usual sorts of arguments, we may assume f and g are positive functions. Let $\alpha, \beta \in [0, 1]$ and $p_1, p_2 \in (0, \infty]$ satisfy $p_1^{-1} + p_2^{-1} + r^{-1} = 1$. Then by Hölder's inequality, Corollary 16.3,

$$\begin{aligned} f*g(x) &= \int_{\mathbb{R}^d} \left[f(x-y)^{(1-\alpha)}g(y)^{(1-\beta)}\right] f(x-y)^{\alpha} g(y)^{\beta} dy \\ &\leq \left(\int_{\mathbb{R}^d} f(x-y)^{(1-\alpha)r} g(y)^{(1-\beta)r} dy\right)^{1/r} \left(\int_{\mathbb{R}^d} f(x-y)^{\alpha p_1} dy\right)^{1/p_1} \times \\ &\qquad \times \left(\int_{\mathbb{R}^d} g(y)^{\beta p_2} dy\right)^{1/p_2} \\ &= \left(\int_{\mathbb{R}^d} f(x-y)^{(1-\alpha)r} g(y)^{(1-\beta)r} dy\right)^{1/r} \|f\|_{\alpha p_1}^{\alpha} \|g\|_{\beta p_2}^{\beta}. \end{aligned}$$

Taking the r^{th} power of this equation and integrating on x gives

$$\begin{aligned} \|f*g\|_r^r &\leq \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-y)^{(1-\alpha)r} g(y)^{(1-\beta)r} dy\right) dx \cdot \|f\|_{\alpha p_1}^{\alpha} \|g\|_{\beta p_2}^{\beta} \\ &= \|f\|_{(1-\alpha)r}^{(1-\alpha)r} \|g\|_{(1-\beta)r}^{(1-\beta)r} \|f\|_{\alpha p_1}^{\alpha r} \|g\|_{\beta p_2}^{\beta r}. \end{aligned} \tag{31.12}$$

Let us now suppose, $(1-\alpha)r = \alpha p_1$ and $(1-\beta)r = \beta p_2$, in which case Eq. (31.12) becomes,

$$\|f*g\|_r^r \leq \|f\|_{\alpha p_1}^r \|g\|_{\beta p_2}^r$$

which is Eq. (31.11) with

$$p := (1-\alpha)r = \alpha p_1 \text{ and } q := (1-\beta)r = \beta p_2. \tag{31.13}$$

So to finish the proof, it suffices to show p and q are arbitrary indices in $[1,\infty]$ satisfying $p^{-1}+q^{-1} = 1+r^{-1}$. If α, β, p_1, p_2 satisfy the relations above, then

$$\alpha = \frac{r}{r+p_1} \text{ and } \beta = \frac{r}{r+p_2}$$

and

$$\begin{aligned} \frac{1}{p} + \frac{1}{q} &= \frac{1}{\alpha p_1} + \frac{1}{\alpha p_2} = \frac{1}{p_1}\frac{r+p_1}{r} + \frac{1}{p_2}\frac{r+p_2}{r} \\ &= \frac{1}{p_1} + \frac{1}{p_2} + \frac{2}{r} = 1 + \frac{1}{r}. \end{aligned}$$

Conversely, if p,q,r satisfy Eq. (31.10), then let α and β satisfy $p = (1-\alpha)r$ and $q = (1-\beta)r$, i.e.

$$\alpha := \frac{r-p}{r} = 1 - \frac{p}{r} \leq 1 \text{ and } \beta = \frac{r-q}{r} = 1 - \frac{q}{r} \leq 1.$$

Using Eq. (31.10) we may also express α and β as

$$\alpha = p(1-\frac{1}{q}) \geq 0 \text{ and } \beta = q(1-\frac{1}{p}) \geq 0$$

and in particular we have shown $\alpha, \beta \in [0,1]$. If we now define $p_1 := p/\alpha \in (0,\infty]$ and $p_2 := q/\beta \in (0,\infty]$, then

$$\begin{aligned} \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{r} &= \beta\frac{1}{q} + \alpha\frac{1}{p} + \frac{1}{r} \\ &= (1-\frac{1}{q}) + (1-\frac{1}{p}) + \frac{1}{r} \\ &= 2 - \left(1+\frac{1}{r}\right) + \frac{1}{r} = 1 \end{aligned}$$

as desired. ∎

Theorem 31.33 (Approximate δ − functions). *Let $p \in [1,\infty]$, $\varphi \in L^1(\mathbb{R}^d)$, $a := \int_{\mathbb{R}^d} \varphi(x)dx$, and for $t > 0$ let $\varphi_t(x) = t^{-d}\varphi(x/t)$. Then*

1. *If $f \in L^p$ with $p < \infty$ then $\varphi_t * f \to af$ in L^p as $t \downarrow 0$.*
2. *If $f \in BC(\mathbb{R}^d)$ and f is uniformly continuous then $\|\varphi_t * f - af\|_\infty \to 0$ as $t \downarrow 0$.*
3. *If $f \in L^\infty$ and f is continuous on $U \subset_o \mathbb{R}^d$ then $\varphi_t * f \to af$ uniformly on compact subsets of U as $t \downarrow 0$.*

(See Proposition 31.34 below and for a statement about almost everywhere convergence.)

Proof. Making the change of variables $y = tz$ implies

$$\varphi_t * f(x) = \int_{\mathbb{R}^d} f(x-y)\varphi_t(y)dy = \int_{\mathbb{R}^d} f(x-tz)\varphi(z)dz$$

so that

$$\begin{aligned} \varphi_t * f(x) - af(x) &= \int_{\mathbb{R}^d} \left[f(x-tz) - f(x)\right]\varphi(z)dz \\ &= \int_{\mathbb{R}^d} \left[\tau_{tz}f(x) - f(x)\right]\varphi(z)dz. \end{aligned} \tag{31.14}$$

Hence by Minkowski's inequality for integrals (Theorem 29.2), Proposition 31.25 and the dominated convergence theorem,

$$\|\varphi_t * f - af\|_p \leq \int_{\mathbb{R}^d} \|\tau_{tz}f - f\|_p \, |\varphi(z)|\, dz \to 0 \text{ as } t \downarrow 0.$$

Item 2. is proved similarly. Indeed, form Eq. (31.14)

$$\|\varphi_t * f - af\|_\infty \leq \int_{\mathbb{R}^d} \|\tau_{tz}f - f\|_\infty \, |\varphi(z)|\, dz$$

which again tends to zero by the dominated convergence theorem because $\lim_{t\downarrow 0}\|\tau_{tz}f - f\|_\infty = 0$ uniformly in z by the uniform continuity of f.

Item 3. Let $B_R = B(0,R)$ be a large ball in $\mathbb{R}^d$ and $K \sqsubset\sqsubset U$, then

$$\begin{aligned}
&\sup_{x\in K}|\varphi_t * f(x) - af(x)| \\
&\quad\le \left|\int_{B_R}[f(x-tz)-f(x)]\,\varphi(z)dz\right| + \left|\int_{B_R^c}[f(x-tz)-f(x)]\,\varphi(z)dz\right| \\
&\quad\le \int_{B_R}|\varphi(z)|\,dz\cdot\sup_{x\in K, z\in B_R}|f(x-tz)-f(x)| + 2\|f\|_\infty\int_{B_R^c}|\varphi(z)|\,dz \\
&\quad\le \|\varphi\|_1\cdot\sup_{x\in K, z\in B_R}|f(x-tz)-f(x)| + 2\|f\|_\infty\int_{|z|>R}|\varphi(z)|\,dz
\end{aligned}$$

so that using the uniform continuity of f on compact subsets of U,

$$\limsup_{t\downarrow 0}\sup_{x\in K}|\varphi_t * f(x) - af(x)| \le 2\|f\|_\infty\int_{|z|>R}|\varphi(z)|\,dz \to 0 \text{ as } R\to\infty.$$

■

The next two results give a version of Theorem 31.33 where the convergence holds almost everywhere. For $f\in L^1_{loc}(\mathbb{R}^n)$ let

$$\mathcal{L}(f) := \left\{x\in\mathbb{R}^n : \lim_{r\downarrow 0}\frac{1}{|B(x,r)|}\int_{B(x,r)}|f(y)-f(x)|dy = 0\right\}$$

be the **Lebesgue set** of f. We will see below in Theorem 25.13 that $m(\mathbb{R}^n\setminus\mathcal{L}(f)) = 0$.

Proposition 31.34 (Theorem 31.33 continued). *Let $p\in[1,\infty)$, $\rho>0$ and $\varphi\in L^\infty(\mathbb{R}^d)$ such that $0\le\varphi\le C1_{B(0,\rho)}$ for some $C<\infty$ and $\int_{\mathbb{R}^d}\varphi(x)\,dx = 1$. If $f\in L^1_{loc}(m)$, and $x\in\mathcal{L}(f)$, then*

$$\lim_{t\downarrow 0}(\varphi_t * f)(x) = f(x),$$

where $\varphi_t(x) := t^{-d}\varphi(x/t)$. In particular, $\varphi_t * f\to f$ a.e. as $t\downarrow 0$.

Proof. Notice that $0\le\varphi_t\le Ct^{-d}1_{B(0,\rho t)}$ and therefore for $x\in\mathcal{L}(f)$ we have,

$$\begin{aligned}
|\varphi_t * f(x) - f(x)| &= \left|\int_{\mathbb{R}^d}[f(x-y)-f(x)]\,\varphi_t(y)\,dy\right| \\
&\le \int_{\mathbb{R}^d}|f(x-y)-f(x)|\,\varphi_t(y)\,dy \\
&\le Ct^{-d}\int_{B(0,\rho t)}|f(x-y)-f(x)|\,dy \\
&= C(\rho,d)\frac{1}{|B(0,\rho t)|}\int_{B(0,\rho t)}|f(x-y)-f(x)|\,dy \to 0 \text{ as } t\downarrow 0.
\end{aligned}$$

■

Theorem 31.35 (Theorem 8.15 of Folland). *More general version, assume that $|\varphi(x)|\le C(1+|x|)^{-(d+\varepsilon)}$ and $\int_{\mathbb{R}^d}\varphi(x)\,dx = a$. Then for all $x\in\mathcal{L}(f)$,*

$$\lim_{t\downarrow 0}(\varphi_t * f)(x) = af(x)$$

and in fact,

$$L(x) := \limsup_{t\downarrow 0}\int|f(x-y)-f(x)|\,|\varphi_t(y)|\,dy = 0.$$

Proof. Throughout this proof $f\in L^1(\mathbb{R}^d)$ and $x\in\mathcal{L}(f)$ be fixed and for $b>0$ let

$$\delta(b) := \frac{1}{b^d}\int_{|y|\le b}|f(x-y)-f(x)|\,dy.$$

From the definition if $\mathcal{L}(f)$ we know that $\lim_{b\downarrow 0}\delta(b) = 0$. The remainder of the proof will be broken into a number of steps.

1. For any $\eta>0$,

$$L(x) = \limsup_{t\downarrow 0}\int_{|y|\le\eta}|f(x-y)-f(x)|\,|\varphi_t(y)|\,dy$$

which is seen as follows;

$$\begin{aligned}
&\int_{|y|>\eta}|f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \\
&\quad\le \int_{|y|>\eta}|f(x-y)|\,|\varphi_t(y)|\,dy + |f(x)|\int_{|y|>\eta}|\varphi_t(y)|\,dy \\
&\quad\le Ct^{-n}\int_{|y|>\eta}|f(x-y)|\left(\frac{1}{1+|y|/t}\right)^{n+\varepsilon}dy + |f(x)|\int_{|z|>\eta/t}|\varphi(z)|\,dy \\
&\quad\le \frac{Ct^{\varepsilon}}{(t+\eta)^{n+\varepsilon}}\|f\|_1 + |f(x)|\int_{|z|>\eta/t}|\varphi(z)|\,dy \to 0 \text{ as } t\downarrow 0.
\end{aligned}$$

2. For any $\rho > 0$,

$$\int_{|y|\le\rho} |f(x-y)-f(x)|\,|\varphi_t(y)|\,dy = t^{-d}\int_{|y|\le\rho} |f(x-y)-f(x)|\,|\varphi(y/t)|\,dy$$
$$\le Ct^{-d}\delta(\rho)\cdot\rho^d = C\delta(\rho)\cdot\left(\frac{\rho}{t}\right)^d.$$

In particular $\rho \le kt$ for some k, then

$$\int_{|y|\le\rho} |f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \le Ck^d\delta(kt) \to 0 \text{ as } t\downarrow 0\cdot$$

3. Given items 1. and 2., in order to finish the proof we must estimate the integral over the annular region $\left\{y\in\mathbb{R}^d : kt \le |y| \le \eta\right\}$. In order to control this integral we are going to have to divide this annular region up into a number of concentric annular regions which we will do shortly. For the moment, let $0 < a < b < \infty$ be given, then

$$\begin{aligned}
\int_{a<|y|\le b} &|f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \\
&\le Ct^{-d}\int_{a<|y|\le b} |f(x-y)-f(x)|\left(1+\left|\frac{y}{t}\right|\right)^{-(d+\varepsilon)} dy \\
&\le Ct^{-d}\int_{a<|y|\le b} |f(x-y)-f(x)|\left(1+\frac{a}{t}\right)^{-(d+\varepsilon)} dy \\
&\le Ct^{-d}\delta(b)\,b^d\left(1+\frac{a}{t}\right)^{-(d+\varepsilon)} \\
&= Ct^{-(d+\varepsilon)}\delta(b)\,b^{(d+\varepsilon)}\left(1+\frac{a}{t}\right)^{-(d+\varepsilon)} t^\varepsilon b^{-\varepsilon} \\
&= C\delta(b)\left(\frac{t}{b}\right)^\varepsilon \frac{1}{\left(t+\frac{a}{b}\right)^{d+\varepsilon}}.
\end{aligned}$$

Taking $a = b/2$ in this expression shows,

$$\begin{aligned}
\int_{\frac{b}{2}<|y|\le b} &|f(x-y)-f(x)|\cdot|\varphi_t(y)|\,dy \\
&\le C\delta(b)\left(\frac{t}{b}\right)^\varepsilon \frac{1}{\left(t+\frac{1}{2}\right)^{d+\varepsilon}} \\
&= C\delta(b)\left(\frac{t}{b}\right)^\varepsilon \frac{1}{(2t+1)^{d+\varepsilon}}.
\end{aligned}$$

Taking $b = 2^{-k}\eta$ and summing the result on $0 \le k \le K-1$ shows

$$\begin{aligned}
\sum_{k=0}^{K-1}\int_{2^{-(k+1)}\eta<|y|\le 2^{-k}\eta} &|f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \\
&\le C\sum_{k=0}^{K-1}\delta\left(2^{-k}\eta\right)\left(\frac{t}{2^{-k}\eta}\right)^\varepsilon \frac{1}{(2t+1)^{d+\varepsilon}} \\
&= \frac{C\delta(\eta)}{(2t+1)^{d+\varepsilon}}\left(\frac{t}{\eta}\right)^\varepsilon \sum_{k=0}^{K-1} 2^{\varepsilon k} \\
&= \frac{C\delta(\eta)}{(2t+1)^{d+\varepsilon}}\left(\frac{t}{\eta}\right)^\varepsilon \frac{2^{\varepsilon K}-1}{2^\varepsilon-1}.
\end{aligned}$$

We now choose K so that $2^K\frac{t}{\eta}\sim 1$ (i.e. $2^{-K}\eta\sim t$) and we have shown,

$$\begin{aligned}
\int_{2^{-K}\eta<|y|\le\eta} &|f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \\
&= \sum_{k=0}^{K-1}\int_{2^{-(k+1)}\eta<|y|\le 2^{-k}\eta} |f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \le C\delta(\eta).
\end{aligned}$$

4. Combining item **2.** with $\rho = 2^{-K}\eta\sim t$ with item **3.** shows

$$\int_{|y|\le\eta} |f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \le C\delta(\eta).$$

Combining this result with item **1.** implies,

$$\begin{aligned}
L(x) &= \limsup_{t\downarrow 0}\int_{|y|\le\eta} |f(x-y)-f(x)|\,|\varphi_t(y)|\,dy \\
&\le C\delta(\eta) \to 0 \text{ as } \eta\downarrow 0.
\end{aligned}$$

■

Exercise 31.5. Let

$$f(t) = \begin{cases} e^{-1/t} & \text{if } t > 0 \\ 0 & \text{if } t \le 0. \end{cases}$$

Show $f \in C^\infty(\mathbb{R},[0,1])$.

Lemma 31.36. *There exists $\varphi \in C_c^\infty(\mathbb{R}^d,[0,\infty))$ such that $\varphi(0) > 0$, $\operatorname{supp}(\varphi)\subset \bar{B}(0,1)$ and $\int_{\mathbb{R}^d}\varphi(x)dx = 1$.*

Proof. Define $h(t) = f(1-t)f(t+1)$ where f is as in Exercise 31.5. Then $h \in C_c^\infty(\mathbb{R},[0,1])$, $\operatorname{supp}(h)\subset[-1,1]$ and $h(0) = e^{-2} > 0$. Define $c = \int_{\mathbb{R}^d} h(|x|^2)dx$. Then $\varphi(x) = c^{-1}h(|x|^2)$ is the desired function. ■

The reader asked to prove the following proposition in Exercise 31.9 below.

Proposition 31.37. *Suppose that $f \in L^1_{loc}(\mathbb{R}^d, m)$ and $\varphi \in C^1_c(\mathbb{R}^d)$, then $f * \varphi \in C^1(\mathbb{R}^d)$ and $\partial_i(f * \varphi) = f * \partial_i \varphi$. Moreover if $\varphi \in C^\infty_c(\mathbb{R}^d)$ then $f * \varphi \in C^\infty(\mathbb{R}^d)$.*

Corollary 31.38 (C^∞ − Uryshon's Lemma). *Given $K \sqsubset\sqsubset U \subset_o \mathbb{R}^d$, there exists $f \in C^\infty_c(\mathbb{R}^d, [0,1])$ such that $\operatorname{supp}(f) \subset U$ and $f = 1$ on K.*

Proof. Let φ be as in Lemma 31.36, $\varphi_t(x) = t^{-d}\varphi(x/t)$ be as in Theorem 31.33, d be the standard metric on $\mathbb{R}^d$ and $\varepsilon = d(K, U^c)$. Since K is compact and U^c is closed, $\varepsilon > 0$. Let $V_\delta = \left\{x \in \mathbb{R}^d : d(x,K) < \delta\right\}$ and $f = \varphi_{\varepsilon/3} * 1_{V_{\varepsilon/3}}$, then

$$\operatorname{supp}(f) \subset \overline{\operatorname{supp}(\varphi_{\varepsilon/3}) + V_{\varepsilon/3}} \subset \bar{V}_{2\varepsilon/3} \subset U.$$

Since $\bar{V}_{2\varepsilon/3}$ is closed and bounded, $f \in C^\infty_c(U)$ and for $x \in K$,

$$f(x) = \int_{\mathbb{R}^d} 1_{d(y,K)<\varepsilon/3} \cdot \varphi_{\varepsilon/3}(x-y)dy = \int_{\mathbb{R}^d} \varphi_{\varepsilon/3}(x-y)dy = 1.$$

The proof will be finished after the reader (easily) verifies $0 \le f \le 1$. ■

Here is an application of this corollary whose proof is left to the reader, Exercise 31.10.

Lemma 31.39 (Integration by Parts). *Suppose f and g are measurable functions on $\mathbb{R}^d$ such that $t \to f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_d)$ and $t \to g(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_d)$ are continuously differentiable functions on $\mathbb{R}$ for each fixed $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Moreover assume $f \cdot g$, $\frac{\partial f}{\partial x_i} \cdot g$ and $f \cdot \frac{\partial g}{\partial x_i}$ are in $L^1(\mathbb{R}^d, m)$. Then*

$$\int_{\mathbb{R}^d} \frac{\partial f}{\partial x_i} \cdot g dm = -\int_{\mathbb{R}^d} f \cdot \frac{\partial g}{\partial x_i} dm.$$

With this result we may give another proof of the Riemann Lebesgue Lemma.

Lemma 31.40 (Riemann Lebesgue Lemma). *For $f \in L^1(\mathbb{R}^d, m)$ let*

$$\hat{f}(\xi) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\xi\cdot x} dm(x)$$

be the Fourier transform of f. Then $\hat{f} \in C_0(\mathbb{R}^d)$ and $\left\|\hat{f}\right\|_\infty \le (2\pi)^{-d/2} \|f\|_1$. (The choice of the normalization factor, $(2\pi)^{-d/2}$, in $\hat{f}$ is for later convenience.)

Proof. The fact that $\hat{f}$ is continuous is a simple application of the dominated convergence theorem. Moreover,

$$\left|\hat{f}(\xi)\right| \le \int_{\mathbb{R}^d} |f(x)|\, dm(x) \le (2\pi)^{-d/2} \|f\|_1$$

so it only remains to see that $\hat{f}(\xi) \to 0$ as $|\xi| \to \infty$. First suppose that $f \in C^\infty_c(\mathbb{R}^d)$ and let $\Delta = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}$ be the Laplacian on $\mathbb{R}^d$. Notice that $\frac{\partial}{\partial x_j} e^{-i\xi\cdot x} = -i\xi_j e^{-i\xi\cdot x}$ and $\Delta e^{-i\xi\cdot x} = -|\xi|^2 e^{-i\xi\cdot x}$. Using Lemma 31.39 repeatedly,

$$\begin{aligned}\int_{\mathbb{R}^d} \Delta^k f(x) e^{-i\xi\cdot x} dm(x) &= \int_{\mathbb{R}^d} f(x) \Delta_x^k e^{-i\xi\cdot x} dm(x) = -|\xi|^{2k} \int_{\mathbb{R}^d} f(x) e^{-i\xi\cdot x} dm(x) \\ &= -(2\pi)^{d/2} |\xi|^{2k} \hat{f}(\xi)\end{aligned}$$

for any $k \in \mathbb{N}$. Hence

$$(2\pi)^{d/2} \left|\hat{f}(\xi)\right| \le |\xi|^{-2k} \left\|\Delta^k f\right\|_1 \to 0$$

as $|\xi| \to \infty$ and $\hat{f} \in C_0(\mathbb{R}^d)$. Suppose that $f \in L^1(m)$ and $f_k \in C^\infty_c(\mathbb{R}^d)$ is a sequence such that $\lim_{k\to\infty} \|f - f_k\|_1 = 0$, then $\lim_{k\to\infty} \left\|\hat{f} - \hat{f}_k\right\|_\infty = 0$. Hence $\hat{f} \in C_0(\mathbb{R}^d)$ by an application of Proposition 37.23. ■

Corollary 31.41. *Let $X \subset \mathbb{R}^d$ be an open set and μ be a K-finite measure on $\mathcal{B}_X$.*

1. *Then $C^\infty_c(X)$ is dense in $L^p(\mu)$ for all $1 \le p < \infty$.*
2. *If $h \in L^1_{loc}(\mu)$ satisfies*

$$\int_X f h d\mu = 0 \text{ for all } f \in C^\infty_c(X) \tag{31.15}$$

then $h(x) = 0$ for μ – a.e. x.

Proof. Let $f \in C_c(X)$, φ be as in Lemma 31.36, φ_t be as in Theorem 31.33 and set $\psi_t := \varphi_t * (f 1_X)$. Then by Proposition 31.37 $\psi_t \in C^\infty(X)$ and by Lemma 31.28 there exists a compact set $K \subset X$ such that $\operatorname{supp}(\psi_t) \subset K$ for all t sufficiently small. By Theorem 31.33, $\psi_t \to f$ uniformly on X as $t \downarrow 0$

1. The dominated convergence theorem (with dominating function being $\|f\|_\infty 1_K$), shows $\psi_t \to f$ in $L^p(\mu)$ as $t \downarrow 0$. This proves Item 1., since Theorem 31.8 guarantees that $C_c(X)$ is dense in $L^p(\mu)$.
2. Keeping the same notation as above, the dominated convergence theorem (with dominating function being $\|f\|_\infty |h| 1_K$) implies

$$0 = \lim_{t\downarrow 0} \int_X \psi_t h d\mu = \int_X \lim_{t\downarrow 0} \psi_t h d\mu = \int_X f h d\mu.$$

The proof is now finished by an application of Lemma 31.11.

■

31.2.1 Smooth Partitions of Unity

We have the following smooth variants of Proposition 37.16, Theorem 37.18 and Corollary 37.20. The proofs of these results are the same as their continuous counterparts. One simply uses the smooth version of Urysohn's Lemma of Corollary 31.38 in place of Lemma 37.8.

Proposition 31.42 (Smooth Partitions of Unity for Compacts). *Suppose that X is an open subset of $\mathbb{R}^d$, $K \subset X$ is a compact set and $\mathcal{U} = \{U_j\}_{j=1}^n$ is an open cover of K. Then there exists a smooth (i.e. $h_j \in C^\infty(X,[0,1])$) partition of unity $\{h_j\}_{j=1}^n$ of K such that $h_j \prec U_j$ for all $j = 1, 2, \dots, n$.*

Theorem 31.43 (Locally Compact Partitions of Unity). *Suppose that X is an open subset of $\mathbb{R}^d$ and $\mathcal{U}$ is an open cover of X. Then there exists a smooth partition of unity of $\{h_i\}_{i=1}^N$ ($N = \infty$ is allowed here) subordinate to the cover $\mathcal{U}$ such that $\operatorname{supp}(h_i)$ is compact for all i.*

Corollary 31.44. *Suppose that X is an open subset of $\mathbb{R}^d$ and $\mathcal{U} = \{U_\alpha\}_{\alpha \in A} \subset \tau$ is an open cover of X. Then there exists a smooth partition of unity of $\{h_\alpha\}_{\alpha\in A}$ subordinate to the cover $\mathcal{U}$ such that $\operatorname{supp}(h_\alpha) \subset U_\alpha$ for all $\alpha \in A$. Moreover if $\bar{U}_\alpha$ is compact for each $\alpha \in A$ we may choose h_α so that $h_\alpha \prec U_\alpha$.*

31.3 Classical Weierstrass Approximation Theorem

Lemma 31.45 (More approximate $\delta -$ sequences). *Suppose that $\{q_n\}_{n=1}^\infty$ is a sequence non-negative continuous real valued functions on $\mathbb{R}$ with compact support that satisfy*

$$\int_{\mathbb{R}} q_n(x)\ dx = 1 \text{ and} \tag{31.16}$$

$$\lim_{n\to\infty} \int_{|x|\geq\varepsilon} q_n(x)dx = 0 \text{ for all } \varepsilon > 0. \tag{31.17}$$

If $f \in BC(\mathbb{R}, Z)$ where Z is a Banach space, then

$$q_n * f(x) := \int_{\mathbb{R}} q_n(y) f(x-y)dy$$

converges to f uniformly on compact subsets of $\mathbb{R}$.

Proof. Let $x \in \mathbb{R}$, then because of Eq. (31.16),

$$\begin{aligned}\|q_n * f(x) - f(x)\| &= \left\|\int_{\mathbb{R}} q_n(y)\left(f(x-y)-f(x)\right)dy\right\| \\ &\leq \int_{\mathbb{R}} q_n(y) \|f(x-y)-f(x)\|\,dy.\end{aligned}$$

Let $M = \sup\{\|f(x)\| : x \in \mathbb{R}\}$. Then for any $\varepsilon > 0$, using Eq. (31.16),

$$\begin{aligned}\|q_n * f(x) - f(x)\| &\leq \int_{|y|\leq\varepsilon} q_n(y)\|f(x-y)-f(x)\|\,dy \\ &\quad + \int_{|y|>\varepsilon} q_n(y)\|f(x-y)-f(x)\|\,dy \\ &\leq \sup_{|w|\leq\varepsilon} \|f(x+w)-f(x)\| + 2M\int_{|y|>\varepsilon} q_n(y)dy.\end{aligned}$$

So if K is a compact subset of $\mathbb{R}$ (for example a large interval) we have

$$\begin{aligned}&\sup_{(x)\in K} \|q_n * f(x) - f(x)\| \\ &\quad\leq \sup_{|w|\leq\varepsilon,\ x\in K} \|f(x+w)-f(x)\| + 2M\int_{\|y\|>\varepsilon} q_n(y)dy\end{aligned}$$

and hence by Eq. (31.17),

$$\begin{aligned}&\lim_{n\to\infty}\sup\ \sup_{x\in K} \|q_n * f(x) - f(x)\| \\ &\quad\leq \sup_{|w|\leq\varepsilon,\ x\in K} \|f(x+w)-f(x)\|.\end{aligned}$$

This finishes the proof since the right member of this equation tends to 0 as $\varepsilon \downarrow 0$ by uniform continuity of f on compact subsets of $\mathbb{R}$. ■

Let $q_n : \mathbb{R} \to [0,\infty)$ be defined by

$$q_n(x) := \frac{1}{c_n}(1-x^2)^n 1_{|x|\leq 1} \text{ where } c_n := \int_{-1}^1 (1-x^2)^n dx. \tag{31.18}$$

Figure 31.2 displays the key features of the functions q_n.

Lemma 31.46. *The sequence $\{q_n\}_{n=1}^\infty$ is an approximate δ – sequence, i.e. they satisfy Eqs. (31.16) and (31.17).*

Proof. By construction, $q_n \in C_c(\mathbb{R}, [0,\infty))$ for each n and Eq. 31.16 holds. Since

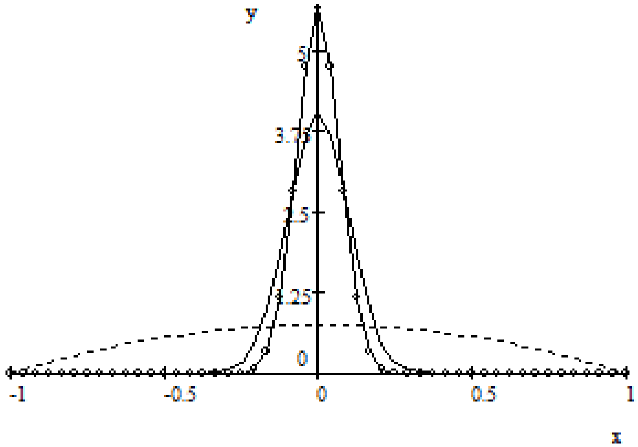


Fig. 31.2. A plot of q_1, q_{50}, and q_{100}. The most peaked curve is q_{100} and the least is q_1. The total area under each of these curves is one.

$$\begin{aligned}\int_{|x|\geq\varepsilon} q_n(x)dx &= \frac{2\int_\varepsilon^1 (1-x^2)^n dx}{2\int_0^\varepsilon (1-x^2)^n dx + 2\int_\varepsilon^1 (1-x^2)^n dx} \\ &\leq \frac{\int_\varepsilon^1 \frac{x}{\varepsilon}(1-x^2)^n dx}{\int_0^\varepsilon \frac{x}{\varepsilon}(1-x^2)^n dx} = \frac{(1-x^2)^{n+1}|_\varepsilon^1}{(1-x^2)^{n+1}|_0^\varepsilon} \\ &= \frac{(1-\varepsilon^2)^{n+1}}{1-(1-\varepsilon^2)^{n+1}} \to 0 \text{ as } n\to\infty,\end{aligned}$$

the proof is complete. ∎

Theorem 31.47 (Classical Weierstrass Approximation Theorem). *Suppose $-\infty < a < b < \infty$, $J = [a,b]$ and $f \in C(J,Z)$. Then there exists polynomials p_n on $\mathbb{R}$ with values in Z such that $p_n \to f$ uniformly on J.*

Proof. By replacing f by F where

$$F(t) := f(a + t(b-a)) - [f(a) + t(f(b) - f(a))] \text{ for } t \in [0,1],$$

it suffices to assume $a = 0$, $b = 1$ and $f(0) = f(1) = 0$. Furthermore we may now extend f to a continuous function on all $\mathbb{R}$ by setting $f \equiv 0$ on $\mathbb{R} \setminus [0,1]$.

With q_n defined as in Eq. (31.18), let $f_n(x) := (q_n * f)(x)$ and recall from Lemma 31.45 that $f_n(x) \to f(x)$ as $n \to \infty$ with the convergence being uniform in $x \in [0,1]$. This completes the proof since f_n is equal to a polynomial function on $[0,1]$. Indeed, there are polynomials, $a_k(y)$, such that

$$(1-(x-y)^2)^n = \sum_{k=0}^{2n} a_k(y)\, x^k,$$

and therefore, for $x \in [0,1]$,

$$\begin{aligned}f_n(x) &= \int_{\mathbb{R}} q_n(x-y) f(y) dy \\ &= \frac{1}{c_n}\int_{[0,1]} f(y)\left[(1-(x-y)^2)^n 1_{|x-y|\leq 1}\right] dy \\ &= \frac{1}{c_n}\int_{[0,1]} f(y)(1-(x-y)^2)^n dy \\ &= \frac{1}{c_n}\int_{[0,1]} f(y)\sum_{k=0}^{2n} a_k(y)\, x^k dy = \sum_{k=0}^{2n} A_k x^k\end{aligned}$$

where

$$A_k = \frac{1}{c_n}\int_{[0,1]} f(y)\, a_k(y)\, dy \in Z.$$

∎

31.4 Exercises

Exercise 31.6. Let (X,τ) be a topological space, μ a measure on $\mathcal{B}_X = \sigma(\tau)$ and $f : X \to \mathbb{C}$ be a measurable function. Letting ν be the measure, $d\nu = |f|\, d\mu$, show $\operatorname{supp}(\nu) = \operatorname{supp}_\mu(f)$, where $\operatorname{supp}(\nu)$ is defined in Definition 30.14).

Exercise 31.7. Let (X,τ) be a topological space, μ a measure on $\mathcal{B}_X = \sigma(\tau)$ such that $\operatorname{supp}(\mu) = X$ (see Definition 30.14). Show $\operatorname{supp}_\mu(f) = \operatorname{supp}(f) = \overline{\{f \neq 0\}}$ for all $f \in C(X)$.

Exercise 31.8. Prove the following strong version of item 3. of Proposition 35.52, namely to every pair of points, x_0, x_1, in a connected open subset V of $\mathbb{R}^d$ there exists $\sigma \in C^\infty(\mathbb{R}, V)$ such that $\sigma(0) = x_0$ and $\sigma(1) = x_1$. **Hint**: First choose a continuous path $\gamma : [0,1] \to V$ such that $\gamma(t) = x_0$ for t near 0 and $\gamma(t) = x_1$ for t near 1 and then use a convolution argument to smooth γ.

Exercise 31.9. Prove Proposition 31.37 by appealing to Corollary 10.30.

Exercise 31.10 (Integration by Parts). Suppose that $(x,y) \in \mathbb{R}\times\mathbb{R}^{d-1} \to f(x,y) \in \mathbb{C}$ and $(x,y) \in \mathbb{R}\times\mathbb{R}^{d-1} \to g(x,y) \in \mathbb{C}$ are measurable functions such that for each fixed $y \in \mathbb{R}^d$, $x \to f(x,y)$ and $x \to g(x,y)$ are continuously differentiable. Also assume $f\cdot g$, $\partial_x f \cdot g$ and $f \cdot \partial_x g$ are integrable relative to Lebesgue measure on $\mathbb{R}\times\mathbb{R}^{d-1}$, where $\partial_x f(x,y) := \frac{d}{dt} f(x+t,y)|_{t=0}$. Show

$$\int_{\mathbb{R}\times\mathbb{R}^{d-1}} \partial_x f(x,y)\cdot g(x,y)dxdy = -\int_{\mathbb{R}\times\mathbb{R}^{d-1}} f(x,y)\cdot \partial_x g(x,y)dxdy. \quad (31.19)$$

(Note: this result and Fubini's theorem proves Lemma 31.39.)

Hints: Let $\psi \in C_c^\infty(\mathbb{R})$ be a function which is 1 in a neighborhood of $0\in\mathbb{R}$ and set $\psi_\varepsilon(x) = \psi(\varepsilon x)$. First verify Eq. (31.19) with $f(x,y)$ replaced by $\psi_\varepsilon(x)f(x,y)$ by doing the $x-$integral first. Then use the dominated convergence theorem to prove Eq. (31.19) by passing to the limit, $\varepsilon\downarrow 0$.

Exercise 31.11 (Folland 8.4 on p. 239). If $f \in L^\infty(\mathbb{R}^n, m)$ and $\|\tau_y f - f\|_\infty \to 0$ as $|y|\to\infty$, then f agrees a.e. with a uniformly continuous function. (See hints in the book.)

Definition 31.48 (Strong Differentiability). *Let $1\le p\le\infty$, $v\in\mathbb{R}^d$, and $u\in L^p(\mathbb{R}^d)$, then $\partial_v u$ is said to exist* ***strongly*** *in L^p if the $\frac{d}{dt}|_0\tau_{-tv}u$ exists in L^p. To be more precise, there should exist a $g\in L^p$ such that*

$$0 = \lim_{t\to 0}\left\|\frac{\tau_{-tv}u - u}{t} - g\right\|_p.$$

Exercise 31.12. Suppose that $1\le p<\infty$, $\partial_v u = g$ exists strongly in L^p as in Definition 31.48, then $\frac{d}{dt}\tau_{-tv}u = \tau_{-tv}g$ for all $t\in\mathbb{R}$. Note that we already know that $\mathbb{R}\ni t\to\tau_{-tv}g\in L^p$ is continuous.

Remark 31.49. It is not hard to develop Riemann style integrals of continuous functions, $f:\mathbb{R}\to X$ where X is a Banach space. This integrals satisfy;

1. The fundamental theorem of calculus holds; i.e.
$$f(t) - f(s) = \int_s^t \dot f(\tau)\,d\tau \;\forall\; f\in C^1(\mathbb{R},X)$$
and if $f\in C(\mathbb{R},X)$ then
$$\frac{d}{dt}\int_a^t f(s)\,ds = f(t)$$
where the derivative is take as an $X-$ valued function.
2. the triangle inequality holds, i.e.
$$\left\|\int_a^b f(t)\,dt\right\|_X \le \int_a^b \|f(t)\|_X\,dt.$$

Exercise 31.13. 1. In particular in the context of Exercise **??** we have
$$\tau_{-tv}u - u = \int_0^t (\tau_{-sv}g)\,ds$$
whenever $\partial_v u = g$ exists strongly in L^p where $1\le p<\infty$.

Exercise 31.14. Suppose that $1\le p<\infty$, $v\in\mathbb{R}^d$, $f\in L^p(\mathbb{R}^d)\cap C^1(\mathbb{R}^d)$ such that $\partial_v f\in L^p(\mathbb{R}^d)$. Then $L^p - \frac{d}{dt}|_0\tau_{-tv}f$ exists and equals $\partial_v f$.

Exercise 31.15. Suppose $1\le p<\infty$, $v\in\mathbb{R}^d$, and $\{u_n\}_{n=1}^\infty\subset L^p(\mathbb{R}^d)$ are such that $\partial_v u_n$ exists strongly for all n and $u_n\to u$ and $\partial_v u_n\to g$ in $L^p(\mathbb{R}^d)$. Show $\partial_v u$ exists strongly and $\partial_v u = g$. **Hint:** Make use of Remark 31.49.

Definition 31.50 (Weak Differentiability). *Let $v\in\mathbb{R}^d$ and $u\in L^p(\mathbb{R}^d)$, then $\partial_v u$ is said to* ***exist weakly*** *in $L^p(\mathbb{R}^d)$ if there exists a function $g\in L^p(\mathbb{R}^d)$ such that*

$$\langle u, \partial_v\varphi\rangle = -\langle g,\varphi\rangle \text{ for all } \varphi\in C_c^\infty(\mathbb{R}^d) \quad (31.20)$$

where

$$\langle u, v\rangle := \int_{\mathbb{R}^d} u(x)\,v(x)\,dx.$$

More generally if $p(\xi) = \sum_{|\alpha|\le N} a_\alpha \xi^\alpha$ is a polynomial in $\xi\in\mathbb{R}^d$ and $p(\partial) := \sum_{|\alpha|\le N} a_\alpha\partial^\alpha$, then we say $p(\partial)u$ exists weakly in $L^p(\mathbb{R}^d)$ if there exists a function $g\in L^p(\mathbb{R}^d)$ such that

$$\langle u, p(-\partial)\varphi\rangle = \langle g,\varphi\rangle \text{ for all } \varphi\in C_c^\infty(\mathbb{R}^d). \quad (31.21)$$

Exercise 31.16. Suppose $1\le p<\infty$, $v\in\mathbb{R}^d$, and $u,g\in L^p(\mathbb{R}^d)$ such that $\partial_v u = g$ exists strongly in L^p. Show $\partial_v u$ exists weakly as is still equal to g. **Hint:** observe that

$$\langle u, \tau_y v\rangle = \int_{\mathbb{R}^d} u(x)\,v(x-y)\,dx = \int_{\mathbb{R}^d} u(x+y)\,v(x)\,dx = \langle \tau_y u, v\rangle.$$

Exercise 31.17. Suppose $1\le p<\infty$, $v\in\mathbb{R}^d$, and $u,g\in L^p(\mathbb{R}^d)$ such that $\partial_v u = g$ exists weakly in L^p. Show for all $\varphi\in C_c^\infty(\mathbb{R}^d)$ that $\partial_v(u*\varphi) = g*\varphi$ **strongly** in L^p.

Exercise 31.18. Suppose $1\le p<\infty$, $v\in\mathbb{R}^d$, and $u,g\in L^p(\mathbb{R}^d)$ such that $\partial_v u = g$ exists weakly in L^p. Show $\partial_v u$ exists strongly in L^p and the strong derivative is g. **Hint,** let $\varphi\in C_c^\infty(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d}\varphi(x)\,dx = 1$ and set $\varphi_n(x) := n^d\varphi(nx)$ so that $\{\varphi_n\}_{n=1}^\infty$ is an approximate $\delta-$ sequence. Then consider $u_n := u*\varphi_n$.

Remark 31.51. Because of the above results, if $1\le p<\infty$, $v\in\mathbb{R}^d$, $u\in L^p(\mathbb{R}^d)$ we will simply say $\partial_v u = g$ in L^p to mean $\partial_v u$ exists strongly or equivalently weakly in L^p.

Exercise 31.19. Suppose $1\le p<\infty$, $v\in\mathbb{R}^d$, and $u,g\in L^p(\mathbb{R}^d)$ such that there exists $\{t_n\}_{n=1}^\infty\subset\mathbb{R}\backslash\{0\}$ such that

$$\lim_{n\to\infty}\left\langle \frac{\tau_{-t_n v}u-u}{t_n},\varphi\right\rangle = \langle g,\varphi\rangle \text{ for all } \varphi\in C_c^\infty\left(\mathbb{R}^d\right),$$

then $\partial_v u$ exists in L^p and is equal to g. [**Hint:** show $\partial_v u = g$ in the weak sense.]

Remark 31.52 (Bounded difference quotients implies differentiability). Combining these exercises with the Banach - Alaoglu's Theorem (see Proposition 41.16), the fact that $L^p\left(\mathbb{R}^d\right)\cong\left[L^{p/(p-1)}\left(\mathbb{R}^d\right)\right]^*$ for $1\le p<\infty$, and the fact that $L^q\left(\mathbb{R}^d\right)$ is separable for $1\le q<\infty$, one can show; if $1<p<\infty$, $u\in L^p\left(\mathbb{R}^d\right)$, and there exists $\{t_n\}_{n=1}^\infty\subset\mathbb{R}\setminus\{0\}$ such that

$$\sup_n\left\|\frac{\tau_{-t_n v}u-u}{t_n}\right\|_p<\infty,$$

then $\partial_v u$ exists in $L^p\left(\mathbb{R}^d\right)$!

Exercise 31.20. Suppose $1\le p<\infty$, $v\in\mathbb{R}^d$, $p(\xi)$ is a polynomial function on $\mathbb{R}^d$, and $u,g\in L^p(\mathbb{R}^d)$ such that $p(\partial)u=g$ exists weakly in L^p. Show there exists $u_n\in C_c^\infty(\mathbb{R}^n)$ such that $u_n\to u$ and $p(\partial)u_n\to g$ in L^p. **Hint,** let $\varphi\in C_c^\infty\left(\mathbb{R}^d\right)$ such that $\int_{\mathbb{R}^d}\varphi(x)\,dx=1$ and set $\varphi_n(x):=n^d\varphi(nx)$ so that $\{\varphi_n\}_{n=1}^\infty$ is an approximate δ – sequence. Then consider $u_n:=\psi_n\cdot u*\varphi_n$ where $\psi_n(x)=\psi(x/n)$ and $\psi\in C_c^\infty\left(\mathbb{R}^d\right)$ such that $\psi=1$ near $0\in\mathbb{R}^d$.

Here is a summary of what the previous exercises prove.

Theorem 31.53 (Weak and Strong Differentiability). *Suppose* $p\in[1,\infty)$, $u\in L^p(\mathbb{R}^d)$ *and* $v\in\mathbb{R}^d\setminus\{0\}$ *and let*

$$\partial_v^h u:=\frac{u(\cdot+hv)-u(\cdot)}{h} \text{ for all } h\ne 0.$$

Then the following are equivalent:

1. *There exists* $g\in L^p(\mathbb{R}^d)$ *and* $\{h_n\}_{n=1}^\infty\subset\mathbb{R}\setminus\{0\}$ *such that* $\lim_{n\to\infty}h_n=0$ *and*
$$\lim_{n\to\infty}\langle\partial_v^{h_n}u,\varphi\rangle=\langle g,\varphi\rangle \text{ for all } \varphi\in C_c^\infty(\mathbb{R}^d).$$
2. $\partial_v^{(w)}u$ *exists and is equal to* $g\in L^p(\mathbb{R}^d)$, *i.e.* $\langle u,\partial_v\varphi\rangle=-\langle g,\varphi\rangle$ *for all* $\varphi\in C_c^\infty(\mathbb{R}^d)$.
3. *There exists* $g\in L^p(\mathbb{R}^d)$ *and* $u_n\in C_c^\infty(\mathbb{R}^d)$ *such that* $u_n\overset{L^p}{\to}u$ *and* $\partial_v u_n\overset{L^p}{\to}g$ *as* $n\to\infty$.
4. $\partial_v^{(s)}u$ *exists and is is equal to* $g\in L^p(\mathbb{R}^d)$, *i.e.* $\partial_v^h u\to g$ *in* L^p *as* $h\to 0$.

Moreover if $p\in(1,\infty)$ *any one of the equivalent conditions 1. – 4. above are implied by the following condition.*

1′. There exists $\{h_n\}_{n=1}^\infty\subset\mathbb{R}\setminus\{0\}$ *such that* $\lim_{n\to\infty}h_n=0$ *and* $\sup_n\left\|\partial_v^{h_n}u\right\|_p<\infty$.

Exercise 31.21. Suppose $v\in\mathbb{R}^d$, $p,q,r\in[1,\infty]$ satisfy $1+\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, $f\in L^p\left(\mathbb{R}^d\right)$, and $g\in L^q\left(\mathbb{R}^d\right)$ such that $\partial_v f$ exists strongly in L^q. Show $\partial_v(f*g)$ exists strongly in L^r and $f*\partial_v g=f*\partial_v g$.

31.5 Consider putting Chapter 40 here.

Part VII

Calculus and Ordinary Differential Equations in Banach Spaces

The Riemann Integral

In this Chapter, the Riemann integral for Banach space valued functions is defined and developed. Our exposition will be brief, since the Lebesgue integral and the Bochner Lebesgue integral will subsume the content of this chapter. In Definition 35.54 below, we will give a general notion of a compact subset of a "topological" space. However, by Corollary 35.62 below, when we are working with subsets of $\mathbb{R}^d$ this definition is equivalent to the following definition.

Definition 32.1. *A subset $A \subset \mathbb{R}^d$ is said to be* ***compact*** *if A is closed and bounded.*

Theorem 32.2. *Suppose that $K \subset \mathbb{R}^d$ is a compact set and $f \in C(K, X)$. Then*

1. *Every sequence $\{u_n\}_{n=1}^\infty \subset K$ has a convergent subsequence.*
2. *The function f is uniformly continuous on K, namely for every $\varepsilon > 0$ there exists a $\delta > 0$ only depending on ε such that $\|f(u) - f(v)\| < \varepsilon$ whenever $u, v \in K$ and $|u - v| < \delta$ where $|\cdot|$ is the standard Euclidean norm on $\mathbb{R}^d$.*

Proof.

1. (This is a special case of Theorem 35.60 and Corollary 35.62 below.) Since K is bounded, $K \subset [-R, R]^d$ for some sufficiently large d. Let t_n be the first component of u_n so that $t_n \in [-R, R]$ for all n. Let $J_1 = [0, R]$ if $t_n \in J_1$ for infinitely many n otherwise let $J_1 = [-R, 0]$. Similarly split J_1 in half and let $J_2 \subset J_1$ be one of the halves such that $t_n \in J_2$ for infinitely many n. Continue this way inductively to find a nested sequence of intervals $J_1 \supset J_2 \supset J_3 \supset J_4 \supset \dots$ such that the length of J_k is $2^{-(k-1)}R$ and for each k, $t_n \in J_k$ for infinitely many n. We may now choose a subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$ such that $\tau_k := t_{n_k} \in J_k$ for all k. The sequence $\{\tau_k\}_{k=1}^\infty$ is Cauchy and hence convergent. Thus by replacing $\{u_n\}_{n=1}^\infty$ by a subsequence if necessary we may assume the first component of $\{u_n\}_{n=1}^\infty$ is convergent. Repeating this argument for the second, then the third and all the way through the d^{th} – components of $\{u_n\}_{n=1}^\infty$, we may, by passing to further subsequences, assume all of the components of u_n are convergent. But this implies $\lim u_n = u$ exists and since K is closed, $u \in K$.
2. (This is a special case of Exercise 35.20 below.) If f were not uniformly continuous on K, there would exists an $\varepsilon > 0$ and sequences $\{u_n\}_{n=1}^\infty$ and $\{v_n\}_{n=1}^\infty$ in K such that
$$\|f(u_n) - f(v_n)\| \geq \varepsilon \text{ while } \lim_{n\to\infty} |u_n - v_n| = 0.$$
By passing to subsequences if necessary we may assume that $\lim_{n\to\infty} u_n$ and $\lim_{n\to\infty} v_n$ exists. Since $\lim_{n\to\infty} |u_n - v_n| = 0$, we must have
$$\lim_{n\to\infty} u_n = u = \lim_{n\to\infty} v_n$$
for some $u \in K$. Since f is continuous, vector addition is continuous and the norm is continuous, we may now conclude that
$$\varepsilon \leq \lim_{n\to\infty} \|f(u_n) - f(v_n)\| = \|f(u) - f(u)\| = 0$$
which is a contradiction.

■

For the remainder of the chapter, let $[a, b]$ be a fixed compact interval and X be a Banach space. The collection $\mathcal{S} = \mathcal{S}([a, b], X)$ of **step functions,** $f : [a, b] \to X$, consists of those functions f which may be written in the form

$$f(t) = x_0 1_{[a,t_1]}(t) + \sum_{i=1}^{n-1} x_i 1_{(t_i,t_{i+1}]}(t), \tag{32.1}$$

where $\pi := \{a = t_0 < t_1 < \cdots < t_n = b\}$ is a partition of $[a, b]$ and $x_i \in X$. For f as in Eq. (32.1), let

$$I(f) := \sum_{i=0}^{n-1} (t_{i+1} - t_i) x_i \in X. \tag{32.2}$$

Exercise 32.1. Show that $I(f)$ is well defined, independent of how f is represented as a step function. (**Hint:** show that adding a point to a partition π of $[a, b]$ does not change the right side of Eq. (32.2).) Also verify that $I : \mathcal{S} \to X$ is a linear operator.

Notation 32.3 *Let $\bar{\mathcal{S}}$ denote the closure of $\mathcal{S}$ inside the Banach space, $\ell^\infty([a, b], X)$ as defined in Remark 14.7.*

The following simple "Bounded Linear Transformation" theorem will often be used in the sequel to define linear transformations.

Theorem 32.4 (B. L. T. Theorem). *Suppose that Z is a normed space, X is a Banach space, and $\mathcal{S} \subset Z$ is a dense linear subspace of Z. If $T : \mathcal{S} \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \le C \|z\|$ for all $z \in \mathcal{S}$), then T has a unique extension to an element $\bar{T} \in L(Z, X)$ and this extension still satisfies*

$$\left\|\bar{T}z\right\| \le C \|z\| \text{ for all } z \in \bar{\mathcal{S}}.$$

Exercise 32.2. Prove Theorem 32.4.

Proposition 32.5 (Riemann Integral). *The linear function $I : \mathcal{S} \to X$ extends uniquely to a continuous linear operator $\bar{I}$ from $\bar{\mathcal{S}}$ to X and this operator satisfies,*

$$\|\bar{I}(f)\| \le (b-a)\|f\|_\infty \text{ for all } f \in \bar{\mathcal{S}}. \tag{32.3}$$

Furthermore, $C([a,b], X) \subset \bar{\mathcal{S}} \subset \ell^\infty([a,b], X)$ and for $f \in C([a,b], X)$, $\bar{I}(f)$ may be computed as

$$\bar{I}(f) = \lim_{|\pi| \to 0} \sum_{i=0}^{n-1} f(c_i^\pi)(t_{i+1} - t_i) \tag{32.4}$$

where $\pi := \{a = t_0 < t_1 < \cdots < t_n = b\}$ denotes a partition of $[a,b]$, $|\pi| = \max\{|t_{i+1} - t_i| : i = 0, \ldots, n-1\}$ is the mesh size of π and c_i^π may be chosen arbitrarily inside $[t_i, t_{i+1}]$. See Figure 32.1.

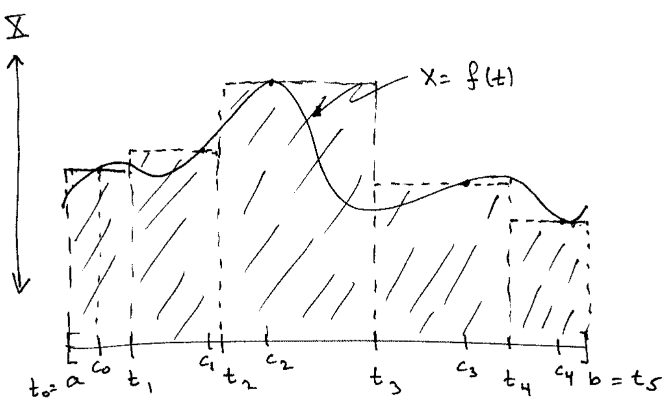


Fig. 32.1. The usual picture associated to the Riemann integral.

Proof. Taking the norm of Eq. (32.2) and using the triangle inequality shows,

$$\|I(f)\| \le \sum_{i=0}^{n-1} (t_{i+1} - t_i)\|x_i\| \le \sum_{i=0}^{n-1} (t_{i+1} - t_i)\|f\|_\infty \le (b-a)\|f\|_\infty. \tag{32.5}$$

The existence of $\bar{I}$ satisfying Eq. (32.3) is a consequence of Theorem 32.4. Given $f \in C([a,b], X)$, $\pi := \{a = t_0 < t_1 < \cdots < t_n = b\}$ a partition of $[a,b]$, and $c_i^\pi \in [t_i, t_{i+1}]$ for $i = 0, 1, 2 \ldots, n-1$, let $f_\pi \in \mathcal{S}$ be defined by

$$f_\pi(t) := f(c_0)_0 1_{[t_0, t_1]}(t) + \sum_{i=1}^{n-1} f(c_i^\pi) 1_{(t_i, t_{i+1}]}(t).$$

Then by the uniform continuity of f on $[a,b]$ (Theorem 32.2), $\lim_{|\pi| \to 0} \|f - f_\pi\|_\infty = 0$ and therefore $f \in \bar{\mathcal{S}}$. Moreover,

$$I(f) = \lim_{|\pi| \to 0} I(f_\pi) = \lim_{|\pi| \to 0} \sum_{i=0}^{n-1} f(c_i^\pi)(t_{i+1} - t_i)$$

which proves Eq. (32.4). ∎

If $f_n \in \mathcal{S}$ and $f \in \bar{\mathcal{S}}$ such that $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$, then for $a \le \alpha < \beta \le b$, then $1_{(\alpha,\beta]} f_n \in \mathcal{S}$ and $\lim_{n\to\infty} \left\|1_{(\alpha,\beta]} f - 1_{(\alpha,\beta]} f_n\right\|_\infty = 0$. This shows $1_{(\alpha,\beta]} f \in \bar{\mathcal{S}}$ whenever $f \in \bar{\mathcal{S}}$.

Notation 32.6 *For $f \in \bar{\mathcal{S}}$ and $a \le \alpha \le \beta \le b$ we will denote $\bar{I}(1_{(\alpha,\beta]} f)$ by $\int_\alpha^\beta f(t)\,dt$ or $\int_{(\alpha,\beta]} f(t) dt$. Also following the usual convention, if $a \le \beta \le \alpha \le b$, we will let*

$$\int_\alpha^\beta f(t)\,dt = -\bar{I}(1_{(\beta,\alpha]} f) = -\int_\beta^\alpha f(t)\,dt.$$

The next Lemma, whose proof is left to the reader, contains some of the many familiar properties of the Riemann integral.

Lemma 32.7. *For $f \in \bar{\mathcal{S}}([a,b], X)$ and $\alpha, \beta, \gamma \in [a,b]$, the Riemann integral satisfies:*

1. $\left\|\int_\alpha^\beta f(t)\,dt\right\|_X \le (\beta - \alpha) \sup\{\|f(t)\| : \alpha \le t \le \beta\}$.
2. $\int_\alpha^\gamma f(t)\,dt = \int_\alpha^\beta f(t)\,dt + \int_\beta^\gamma f(t)\,dt$.
3. *The function $G(t) := \int_a^t f(\tau) d\tau$ is continuous on $[a,b]$.*
4. *If Y is another Banach space and $T \in L(X,Y)$, then $Tf \in \bar{\mathcal{S}}([a,b], Y)$ and*

$$T\left(\int_\alpha^\beta f(t)\ dt\right) = \int_\alpha^\beta Tf(t)\ dt.$$

5. *The function* $t \to \|f(t)\|_X$ *is in* $\bar{\mathcal{S}}([a,b],\mathbb{R})$ *and*

$$\left\|\int_a^b f(t)\ dt\right\|_X \le \int_a^b \|f(t)\|_X\ dt.$$

6. *If* $f, g \in \bar{\mathcal{S}}([a,b],\mathbb{R})$ *and* $f \le g$, *then*

$$\int_a^b f(t)\ dt \le \int_a^b g(t)\ dt.$$

Exercise 32.3. Prove Lemma 32.7.

Remark 32.8 (BRUCE: todo?). Perhaps the Riemann Stieljtes integral, Lemma **??**, should be done here. Maybe this should be done in the more genral context of Banach valued functions in preparation of T. Lyon's rough path analysis. The point would be to let X_t take values in a Banach space and assume that X_t had finite variation. Then define $\mu_X(t) := \sup_\pi \sum_l \left\|X_{t\wedge t_l} - X_{t\wedge t_{l-1}}\right\|$. Then we could define

$$\int_0^T Z_t dX_t := \lim_{|\pi|\to 0} \sum Z_{t_{l-1}}\left(X_{t\wedge t_l} - X_{t\wedge t_{l-1}}\right)$$

for continuous operator valued paths, Z_t. This integral would then satisfy the estimates,

$$\left\|\int_0^T Z_t dX_t\right\| \le \int_0^T \|Z_t\|\, d\mu_X(t) \le \sup_{0\le t\le T}\|Z_t\|\,\mu_X(T).$$

32.1 The Fundamental Theorem of Calculus

Our next goal is to show that our Riemann integral interacts well with differentiation, namely the fundamental theorem of calculus holds. Before doing this we will need a couple of basic definitions and results of differential calculus, more details and the next few results below will be done in greater detail in Chapter 34.

Definition 32.9. *Let* $(a,b) \subset \mathbb{R}$. *A function* $f : (a,b) \to X$ *is differentiable at* $t \in (a,b)$ *iff*

$$L := \lim_{h\to 0}\left(h^{-1}\left[f(t+h) - f(t)\right]\right) = \lim_{h\to 0} \text{“}\frac{f(t+h)-f(t)}{h}\text{”}$$

exists in X. *The limit* L, *if it exists, will be denoted by* $\dot{f}(t)$ *or* $\frac{df}{dt}(t)$. *We also say that* $f \in C^1((a,b),X)$ *if* f *is differentiable at all points* $t \in (a,b)$ *and* $\dot{f} \in C((a,b),X)$.

As for the case of real valued functions, the derivative operator $\frac{d}{dt}$ is easily seen to be linear. The next two results have proofs very similar to their real valued function analogues.

Lemma 32.10 (Product Rules). *Suppose that* $t \to U(t) \in L(X)$, $t \to V(t) \in L(X)$ *and* $t \to x(t) \in X$ *are differentiable at* $t = t_0$, *then*

1. $\frac{d}{dt}|_{t_0}[U(t)x(t)] \in X$ *exists and*
$$\frac{d}{dt}|_{t_0}[U(t)x(t)] = \left[\dot{U}(t_0)x(t_0) + U(t_0)\dot{x}(t_0)\right]$$
and
2. $\frac{d}{dt}|_{t_0}[U(t)V(t)] \in L(X)$ *exists and*
$$\frac{d}{dt}|_{t_0}[U(t)V(t)] = \left[\dot{U}(t_0)V(t_0) + U(t_0)\dot{V}(t_0)\right].$$
3. *If* $U(t_0)$ *is invertible, then* $t \to U(t)^{-1}$ *is differentiable at* $t = t_0$ *and*
$$\frac{d}{dt}|_{t_0}U(t)^{-1} = -U(t_0)^{-1}\dot{U}(t_0)U(t_0)^{-1}. \tag{32.6}$$

Proof. The reader is asked to supply the proof of the first two items in Exercise 32.9. Before proving item 3., let us assume that $U(t)^{-1}$ is differentiable, then using the product rule we would learn

$$0 = \frac{d}{dt}|_{t_0}I = \frac{d}{dt}|_{t_0}\left[U(t)^{-1}U(t)\right] = \left[\frac{d}{dt}|_{t_0}U(t)^{-1}\right]U(t_0) + U(t_0)^{-1}\dot{U}(t_0).$$

Solving this equation for $\frac{d}{dt}|_{t_0}U(t)^{-1}$ gives the formula in Eq. (32.6). The problem with this argument is that we have not yet shown $t \to U(t)^{-1}$ is invertible at t_0. Here is the formal proof. Since $U(t)$ is differentiable at t_0, $U(t) \to U(t_0)$ as $t \to t_0$ and by Corollary 14.27, $U(t_0+h)$ is invertible for h near 0 and

$$U(t_0+h)^{-1} \to U(t_0)^{-1} \text{ as } h \to 0.$$

Therefore, using Lemma 14.16, we may let $h \to 0$ in the identity,

$$\frac{U(t_0+h)^{-1} - U(t_0)^{-1}}{h} = U(t_0+h)^{-1}\left(\frac{U(t_0) - U(t_0+h)}{h}\right)U(t_0)^{-1},$$

to learn

$$\lim_{h\to 0}\frac{U(t_0+h)^{-1} - U(t_0)^{-1}}{h} = -U(t_0)^{-1}\dot{U}(t_0)U(t_0)^{-1}.$$

■

Proposition 32.11 (Chain Rule). *Suppose $s \to x(s) \in X$ is differentiable at $s = s_0$ and $t \to T(t) \in \mathbb{R}$ is differentiable at $t = t_0$ and $T(t_0) = s_0$, then $t \to x(T(t))$ is differentiable at t_0 and*

$$\frac{d}{dt}|_{t_0} x(T(t)) = x'(T(t_0)) T'(t_0).$$

The proof of the chain rule is essentially the same as the real valued function case, see Exercise 32.10.

Proposition 32.12. *Suppose that $f : [a,b] \to X$ is a continuous function such that $\dot{f}(t)$ exists and is equal to zero for $t \in (a,b)$. Then f is constant.*

Proof. Let $\varepsilon > 0$ and $\alpha \in (a,b)$ be given. (We will later let $\varepsilon \downarrow 0$.) By the definition of the derivative, for all $\tau \in (a,b)$ there exists $\delta_\tau > 0$ such that

$$\|f(t) - f(\tau)\| = \left\|f(t) - f(\tau) - \dot{f}(\tau)(t-\tau)\right\| \le \varepsilon |t-\tau| \text{ if } |t-\tau| < \delta_\tau. \tag{32.7}$$

Let

$$A = \{t \in [\alpha, b] : \|f(t) - f(\alpha)\| \le \varepsilon(t-\alpha)\} \tag{32.8}$$

and t_0 be the least upper bound for A. We will now use a standard argument which is sometimes referred to as **continuous induction** to show $t_0 = b$. Eq. (32.7) with $\tau = \alpha$ shows $t_0 > \alpha$ and a simple continuity argument shows $t_0 \in A$, i.e.

$$\|f(t_0) - f(\alpha)\| \le \varepsilon(t_0 - \alpha). \tag{32.9}$$

For the sake of contradiction, suppose that $t_0 < b$. By Eqs. (32.7) and (32.9),

$$\begin{aligned}\|f(t) - f(\alpha)\| &\le \|f(t) - f(t_0)\| + \|f(t_0) - f(\alpha)\| \\ &\le \varepsilon(t_0 - \alpha) + \varepsilon(t - t_0) = \varepsilon(t-\alpha)\end{aligned}$$

for $0 \le t - t_0 < \delta_{t_0}$ which violates the definition of t_0 being an upper bound. Thus we have shown $b \in A$ and hence

$$\|f(b) - f(\alpha)\| \le \varepsilon(b - \alpha).$$

Since $\varepsilon > 0$ was arbitrary we may let $\varepsilon \downarrow 0$ in the last equation to conclude $f(b) = f(\alpha)$. Since $\alpha \in (a,b)$ was arbitrary it follows that $f(b) = f(\alpha)$ for all $\alpha \in (a,b]$ and then by continuity for all $\alpha \in [a,b]$, i.e. f is constant. ∎

Remark 32.13. The usual real variable proof of Proposition 32.12 makes use of Rolle's theorem which in turn uses the extreme value theorem. This latter theorem is not available to vector valued functions. However with the aid of the Hahn Banach Theorem 21.7 (or Corollary 21.8) below and Lemma 32.7, it is possible to reduce the proof of Proposition 32.12 and the proof of the Fundamental Theorem of Calculus 32.14 to the real valued case, see Exercise 21.6.

Theorem 32.14 (Fundamental Theorem of Calculus). *Suppose that $f \in C([a,b], X)$, Then*

1. *$\frac{d}{dt}\int_a^t f(\tau)\, d\tau = f(t)$ for all $t \in (a,b)$.*
2. *Now assume that $F \in C([a,b], X)$, F is continuously differentiable on (a,b) (i.e. $\dot{F}(t)$ exists and is continuous for $t \in (a,b)$) and $\dot{F}$ extends to a continuous function on $[a,b]$ which is still denoted by $\dot{F}$. Then*

$$\int_a^b \dot{F}(t)\, dt = F(b) - F(a).$$

Proof. Let $h > 0$ be a small number and consider

$$\begin{aligned}\left\|\int_a^{t+h} f(\tau)d\tau - \int_a^t f(\tau)d\tau - f(t)h\right\| &= \left\|\int_t^{t+h} (f(\tau) - f(t))\, d\tau\right\| \\ &\le \int_t^{t+h} \|(f(\tau) - f(t))\|\, d\tau \le h\varepsilon(h),\end{aligned}$$

where $\varepsilon(h) := \max_{\tau \in [t, t+h]} \|(f(\tau) - f(t))\|$. Combining this with a similar computation when $h < 0$ shows, for all $h \in \mathbb{R}$ sufficiently small, that

$$\left\|\int_a^{t+h} f(\tau)d\tau - \int_a^t f(\tau)d\tau - f(t)h\right\| \le |h|\varepsilon(h),$$

where now $\varepsilon(h) := \max_{\tau \in [t-|h|, t+|h|]} \|(f(\tau) - f(t))\|$. By continuity of f at t, $\varepsilon(h) \to 0$ and hence $\frac{d}{dt}\int_a^t f(\tau)\, d\tau$ exists and is equal to $f(t)$. For the second item, set $G(t) := \int_a^t \dot{F}(\tau)\, d\tau - F(t)$. Then G is continuous by Lemma 32.7 and $\dot{G}(t) = 0$ for all $t \in (a,b)$ by item 1. An application of Proposition 32.12 shows G is a constant and in particular $G(b) = G(a)$, i.e. $\int_a^b \dot{F}(\tau)\, d\tau - F(b) = -F(a)$. ∎

Corollary 32.15 (Mean Value Inequality). *Suppose that $f : [a,b] \to X$ is a continuous function such that $\dot{f}(t)$ exists for $t \in (a,b)$ and $\dot{f}$ extends to a continuous function on $[a,b]$. Then*

$$\|f(b) - f(a)\| \le \int_a^b \|\dot{f}(t)\| dt \le (b-a) \cdot \left\|\dot{f}\right\|_\infty. \tag{32.10}$$

Proof. By the fundamental theorem of calculus, $f(b) - f(a) = \int_a^b \dot{f}(t)dt$ and then by Lemma 32.7,

$$\begin{aligned}\|f(b) - f(a)\| &= \left\|\int_a^b \dot{f}(t)dt\right\| \le \int_a^b \|\dot{f}(t)\| dt \\ &\le \int_a^b \left\|\dot{f}\right\|_\infty dt = (b-a) \cdot \left\|\dot{f}\right\|_\infty.\end{aligned}$$

■

Corollary 32.16 (Change of Variable Formula). *Suppose that $f \in C([a,b],X)$ and $T : [c,d] \to (a,b)$ is a continuous function such that $T(s)$ is continuously differentiable for $s \in (c,d)$ and $T'(s)$ extends to a continuous function on $[c,d]$. Then*

$$\int_c^d f(T(s))\,T'(s)\,ds = \int_{T(c)}^{T(d)} f(t)\,dt.$$

Proof. For $t \in (a,b)$ define $F(t) := \int_{T(c)}^t f(\tau)\,d\tau$. Then $F \in C^1((a,b),X)$ and by the fundamental theorem of calculus and the chain rule,

$$\frac{d}{ds}F(T(s)) = F'(T(s))\,T'(s) = f(T(s))\,T'(s).$$

Integrating this equation on $s \in [c,d]$ and using the chain rule again gives

$$\int_c^d f(T(s))\,T'(s)\,ds = F(T(d)) - F(T(c)) = \int_{T(c)}^{T(d)} f(t)\,dt.$$

■

32.2 Integral Operators as Examples of Bounded Operators

In the examples to follow, all integrals are the standard Riemann integrals and we will make use of the following notation.

Notation 32.17 *Given an open set $U \subset \mathbb{R}^d$, let $C_c(U)$ denote the collection of real valued continuous functions f on U such that*

$$\operatorname{supp}(f) := \overline{\{x \in U : f(x) \neq 0\}}$$

is a compact subset of U.

Example 32.18. Suppose that $K : [0,1] \times [0,1] \to \mathbb{C}$ is a continuous function. For $f \in C([0,1])$, let

$$Tf(x) = \int_0^1 K(x,y)f(y)dy.$$

Since

$$\begin{aligned}|Tf(x) - Tf(z)| &\le \int_0^1 |K(x,y) - K(z,y)|\,|f(y)|\,dy \\ &\le \|f\|_\infty \max_y |K(x,y) - K(z,y)| \qquad (32.11)\end{aligned}$$

and the latter expression tends to 0 as $x \to z$ by uniform continuity of K. Therefore $Tf \in C([0,1])$ and by the linearity of the Riemann integral, $T : C([0,1]) \to C([0,1])$ is a linear map. Moreover,

$$|Tf(x)| \le \int_0^1 |K(x,y)|\,|f(y)|\,dy \le \int_0^1 |K(x,y)|\,dy \cdot \|f\|_\infty \le A\,\|f\|_\infty$$

where

$$A := \sup_{x\in[0,1]} \int_0^1 |K(x,y)|\,dy < \infty. \qquad (32.12)$$

This shows $\|T\| \le A < \infty$ and therefore T is bounded. We may in fact show $\|T\| = A$. To do this let $x_0 \in [0,1]$ be such that

$$\sup_{x\in[0,1]} \int_0^1 |K(x,y)|\,dy = \int_0^1 |K(x_0,y)|\,dy.$$

Such an x_0 can be found since, using a similar argument to that in Eq. (32.11), $x \to \int_0^1 |K(x,y)|\,dy$ is continuous. Given $\varepsilon > 0$, let

$$f_\varepsilon(y) := \frac{\overline{K(x_0,y)}}{\sqrt{\varepsilon + |K(x_0,y)|^2}}$$

and notice that $\lim_{\varepsilon\downarrow 0} \|f_\varepsilon\|_\infty = 1$ and

$$\|Tf_\varepsilon\|_\infty \ge |Tf_\varepsilon(x_0)| = Tf_\varepsilon(x_0) = \int_0^1 \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}}dy.$$

Therefore,

$$\begin{aligned}\|T\| &\ge \lim_{\varepsilon\downarrow 0} \frac{1}{\|f_\varepsilon\|_\infty} \int_0^1 \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}}dy \\ &= \lim_{\varepsilon\downarrow 0} \int_0^1 \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}}dy = A\end{aligned}$$

since

$$\begin{aligned}0 &\le |K(x_0,y)| - \frac{|K(x_0,y)|^2}{\sqrt{\varepsilon + |K(x_0,y)|^2}} \\ &= \frac{|K(x_0,y)|}{\sqrt{\varepsilon + |K(x_0,y)|^2}}\left[\sqrt{\varepsilon + |K(x_0,y)|^2} - |K(x_0,y)|\right] \\ &\le \sqrt{\varepsilon + |K(x_0,y)|^2} - |K(x_0,y)|\end{aligned}$$

and the latter expression tends to zero uniformly in y as $\varepsilon \downarrow 0$.

We may also consider other norms on $C([0,1])$. Let (for now) $L^1([0,1])$ denote $C([0,1])$ with the norm

$$\|f\|_1 = \int_0^1 |f(x)|\, dx,$$

then $T : L^1([0,1], dm) \to C([0,1])$ is bounded as well. Indeed, let $M = \sup\{|K(x,y)| : x, y \in [0,1]\}$, then

$$|(Tf)(x)| \le \int_0^1 |K(x,y)f(y)|\, dy \le M \|f\|_1$$

which shows $\|Tf\|_\infty \le M \|f\|_1$ and hence,

$$\|T\|_{L^1 \to C} \le \max\{|K(x,y)| : x, y \in [0,1]\} < \infty.$$

We can in fact show that $\|T\| = M$ as follows. Let $(x_0, y_0) \in [0,1]^2$ satisfying $|K(x_0,y_0)| = M$. Then given $\varepsilon > 0$, there exists a neighborhood $U = I \times J$ of (x_0, y_0) such that $|K(x,y) - K(x_0,y_0)| < \varepsilon$ for all $(x,y) \in U$. Let $f \in C_c(I, [0,\infty))$ such that $\int_0^1 f(x)dx = 1$. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha K(x_0,y_0) = M$, then

$$\begin{aligned} |(T\alpha f)(x_0)| &= \left|\int_0^1 K(x_0,y)\alpha f(y)dy\right| = \left|\int_I K(x_0,y)\alpha f(y)dy\right| \\ &\ge \operatorname{Re}\int_I \alpha K(x_0,y) f(y)dy \\ &\ge \int_I (M-\varepsilon) f(y)dy = (M-\varepsilon)\|\alpha f\|_{L^1} \end{aligned}$$

and hence

$$\|T\alpha f\|_C \ge (M-\varepsilon)\|\alpha f\|_{L^1}$$

showing that $\|T\| \ge M - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we learn that $\|T\| \ge M$ and hence $\|T\| = M$.

One may also view T as a map from $T : C([0,1]) \to L^1([0,1])$ in which case one may show

$$\|T\|_{L^1 \to C} \le \int_0^1 \max_y |K(x,y)|\, dx < \infty.$$

32.3 Linear Ordinary Differential Equations

Let X be a Banach space, $J = (a,b) \subset \mathbb{R}$ be an open interval with $0 \in J$, $h \in C(J,X)$ and $A \in C(J, L(X))$. In this section we are going to consider the ordinary differential equation,

$$\dot{y}(t) = A(t)y(t) + h(t) \text{ and } y(0) = x \in X, \tag{32.13}$$

where y is an unknown function in $C^1(J,X)$. This equation may be written in its equivalent (as the reader should verify) integral form, namely we are looking for $y \in C(J,X)$ such that

$$y(t) = x + \int_0^t h(\tau)\, d\tau + \int_0^t A(\tau)y(\tau)d\tau. \tag{32.14}$$

In what follows, we will abuse notation and use $\|\cdot\|$ to denote the operator norm on $L(X)$ associated to the norm, $\|\cdot\|$, on X and let $\|\varphi\|_\infty := \max_{t\in J}\|\varphi(t)\|$ for $\varphi \in BC(J,X)$ or $BC(J, L(X))$.

Notation 32.19 *For $t \in \mathbb{R}$ and $n \in \mathbb{N}$, let*

$$\Delta_n(t) = \begin{cases} \{(\tau_1,\dots,\tau_n) \in \mathbb{R}^n : 0 \le \tau_1 \le \dots \le \tau_n \le t\} & \text{if } t \ge 0 \\ \{(\tau_1,\dots,\tau_n) \in \mathbb{R}^n : t \le \tau_n \le \dots \le \tau_1 \le 0\} & \text{if } t \le 0 \end{cases}$$

and also write $d\tau = d\tau_1 \dots d\tau_n$ and

$$\int_{\Delta_n(t)} f(\tau_1,\dots,\tau_n)d\tau := (-1)^{n\cdot 1_{t<0}} \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \dots \int_0^{\tau_2} d\tau_1 f(\tau_1,\dots,\tau_n)$$

where

$$1_{t<0} = \begin{cases} 1 & \text{if } t < 0 \\ 0 & \text{if } t \ge 0 \end{cases}.$$

Lemma 32.20. *Suppose that $\psi \in C(\mathbb{R},\mathbb{R})$, then*

$$(-1)^{n\cdot 1_{t<0}} \int_{\Delta_n(t)} \psi(\tau_1)\dots\psi(\tau_n)d\tau = \frac{1}{n!}\left(\int_0^t \psi(\tau)d\tau\right)^n. \tag{32.15}$$

Proof. Let $\Psi(t) := \int_0^t \psi(\tau)d\tau$. The proof will go by induction on n. The case $n = 1$ is easily verified since

$$(-1)^{1\cdot 1_{t<0}} \int_{\Delta_1(t)} \psi(\tau_1)d\tau_1 = \int_0^t \psi(\tau)d\tau = \Psi(t).$$

Now assume the truth of Eq. (32.15) for $n-1$ for some $n \ge 2$, then

$$\begin{aligned} (-1)^{n\cdot 1_{t<0}} &\int_{\Delta_n(t)} \psi(\tau_1)\dots\psi(\tau_n)d\tau \\ &= \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \dots \int_0^{\tau_2} d\tau_1 \psi(\tau_1)\dots\psi(\tau_n) \\ &= \int_0^t d\tau_n \frac{\Psi^{n-1}(\tau_n)}{(n-1)!}\psi(\tau_n) = \int_0^t d\tau_n \frac{\Psi^{n-1}(\tau_n)}{(n-1)!}\dot{\Psi}(\tau_n) \\ &= \int_0^{\Psi(t)} \frac{u^{n-1}}{(n-1)!}du = \frac{\Psi^n(t)}{n!}, \end{aligned}$$

wherein we made the change of variables, $u = \Psi(\tau_n)$, in the second to last equality. ■

Remark 32.21. Eq. (32.15) is equivalent to

$$\int_{\Delta_n(t)} \psi(\tau_1)\dots\psi(\tau_n)d\tau = \frac{1}{n!}\left(\int_{\Delta_1(t)} \psi(\tau)d\tau\right)^n$$

and another way to understand this equality is to view $\int_{\Delta_n(t)} \psi(\tau_1)\dots\psi(\tau_n)d\tau$ as a multiple integral (see Chapter **??** below) rather than an iterated integral. Indeed, taking $t>0$ for simplicity and letting S_n be the permutation group on $\{1,2,\dots,n\}$ we have

$$[0,t]^n = \cup_{\sigma\in S_n}\{(\tau_1,\dots,\tau_n)\in\mathbb{R}^n : 0\le\tau_{\sigma 1}\le\dots\le\tau_{\sigma n}\le t\}$$

with the union being "essentially" disjoint. Therefore, making a change of variables and using the fact that $\psi(\tau_1)\dots\psi(\tau_n)$ is invariant under permutations, we find

$$\begin{aligned}\left(\int_0^t \psi(\tau)d\tau\right)^n &= \int_{[0,t]^n}\psi(\tau_1)\dots\psi(\tau_n)d\tau\\ &= \sum_{\sigma\in S_n}\int_{\{(\tau_1,\dots,\tau_n)\in\mathbb{R}^n:0\le\tau_{\sigma 1}\le\dots\le\tau_{\sigma n}\le t\}}\psi(\tau_1)\dots\psi(\tau_n)d\tau\\ &= \sum_{\sigma\in S_n}\int_{\{(s_1,\dots,s_n)\in\mathbb{R}^n:0\le s_1\le\dots\le s_n\le t\}}\psi(s_{\sigma^{-1}1})\dots\psi(s_{\sigma^{-1}n})d\mathbf{s}\\ &= \sum_{\sigma\in S_n}\int_{\{(s_1,\dots,s_n)\in\mathbb{R}^n:0\le s_1\le\dots\le s_n\le t\}}\psi(s_1)\dots\psi(s_n)d\mathbf{s}\\ &= n!\int_{\Delta_n(t)}\psi(\tau_1)\dots\psi(\tau_n)d\tau.\end{aligned}$$

Theorem 32.22. *Let $\varphi\in BC(J,X)$, then the integral equation*

$$y(t) = \varphi(t) + \int_0^t A(\tau)y(\tau)d\tau \tag{32.16}$$

has a unique solution given by

$$y(t) = \varphi(t) + \sum_{n=1}^\infty (-1)^{n\cdot 1_{t<0}}\int_{\Delta_n(t)} A(\tau_n)\dots A(\tau_1)\varphi(\tau_1)d\tau \tag{32.17}$$

and this solution satisfies the bound

$$\|y\|_\infty \le \|\varphi\|_\infty e^{\int_J \|A(\tau)\|d\tau}.$$

Proof. Define $\Lambda : BC(J,X)\to BC(J,X)$ by

$$(\Lambda y)(t) = \int_0^t A(\tau)y(\tau)d\tau.$$

Then y solves Eq. (32.14) iff $y=\varphi+\Lambda y$ or equivalently iff $(I-\Lambda)y=\varphi$. An induction argument shows

$$\begin{aligned}(\Lambda^n\varphi)(t) &= \int_0^t d\tau_n A(\tau_n)(\Lambda^{n-1}\varphi)(\tau_n)\\ &= \int_0^t d\tau_n\int_0^{\tau_n} d\tau_{n-1}A(\tau_n)A(\tau_{n-1})(\Lambda^{n-2}\varphi)(\tau_{n-1})\\ &\vdots\\ &= \int_0^t d\tau_n\int_0^{\tau_n} d\tau_{n-1}\dots\int_0^{\tau_2} d\tau_1 A(\tau_n)\dots A(\tau_1)\varphi(\tau_1)\\ &= (-1)^{n\cdot 1_{t<0}}\int_{\Delta_n(t)} A(\tau_n)\dots A(\tau_1)\varphi(\tau_1)d\tau.\end{aligned}$$

Taking norms of this equation and using the triangle inequality along with Lemma 32.20 gives,

$$\begin{aligned}\|(\Lambda^n\varphi)(t)\| &\le \|\varphi\|_\infty\cdot\int_{\Delta_n(t)}\|A(\tau_n)\|\dots\|A(\tau_1)\|d\tau\\ &\le \|\varphi\|_\infty\cdot\frac{1}{n!}\left(\int_{\Delta_1(t)}\|A(\tau)\|d\tau\right)^n\\ &\le \|\varphi\|_\infty\cdot\frac{1}{n!}\left(\int_J\|A(\tau)\|d\tau\right)^n.\end{aligned}$$

Therefore,

$$\|\Lambda^n\|_{op}\le\frac{1}{n!}\left(\int_J\|A(\tau)\|d\tau\right)^n \tag{32.18}$$

and

$$\sum_{n=0}^\infty\|\Lambda^n\|_{op}\le e^{\int_J\|A(\tau)\|d\tau}<\infty$$

where $\|\cdot\|_{op}$ denotes the operator norm on $L(BC(J,X))$. An application of Proposition 14.26 now shows $(I-\Lambda)^{-1}=\sum_{n=0}^\infty\Lambda^n$ exists and

$$\left\|(I-\Lambda)^{-1}\right\|_{op}\le e^{\int_J\|A(\tau)\|d\tau}.$$

It is now only a matter of working through the notation to see that these assertions prove the theorem. ■

Corollary 32.23. *Suppose $h \in C(J,X)$ and $x \in X$, then there exists a unique solution, $y \in C^1(J,X)$, to the linear ordinary differential Eq. (32.13).*

Proof. Let

$$\varphi(t) = x + \int_0^t h(\tau)\,d\tau.$$

By applying Theorem 32.22 with J replaced by any open interval J_0 such that $0 \in J_0$ and $\bar{J}_0$ is a compact subinterval[1] of J, there exists a unique solution y_{J_0} to Eq. (32.13) which is valid for $t \in J_0$. By uniqueness of solutions, if J_1 is a subinterval of J such that $J_0 \subset J_1$ and $\bar{J}_1$ is a compact subinterval of J, we have $y_{J_1} = y_{J_0}$ on J_0. Because of this observation, we may construct a solution y to Eq. (32.13) which is defined on the full interval J by setting $y(t) = y_{J_0}(t)$ for any J_0 as above which also contains $t \in J$. ∎

Corollary 32.24. *Suppose that $A \in L(X)$ is independent of time, then the solution to*

$$\dot{y}(t) = Ay(t) \text{ with } y(0) = x$$

is given by $y(t) = e^{tA}x$ where

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n. \tag{32.19}$$

Moreover,

$$e^{(t+s)A} = e^{tA}e^{sA} \text{ for all } s,t \in \mathbb{R}. \tag{32.20}$$

Proof. The first assertion is a simple consequence of Eq. 32.17 and Lemma 32.20 with $\psi = 1$. The assertion in Eq. (32.20) may be proved by explicit computation but the following proof is more instructive. Given $x \in X$, let $y(t) := e^{(t+s)A}x$. By the chain rule,

$$\begin{aligned}\frac{d}{dt}y(t) &= \frac{d}{d\tau}|_{\tau=t+s} e^{\tau A}x = Ae^{\tau A}x|_{\tau=t+s} \\ &= Ae^{(t+s)A}x = Ay(t) \text{ with } y(0) = e^{sA}x.\end{aligned}$$

The unique solution to this equation is given by

$$y(t) = e^{tA}x(0) = e^{tA}e^{sA}x.$$

This completes the proof since, by definition, $y(t) = e^{(t+s)A}x$. ∎

We also have the following converse to this corollary whose proof is outlined in Exercise 32.20 below.

Theorem 32.25. *Suppose that $T_t \in L(X)$ for $t \geq 0$ satisfies*

1. *(Semi-group property.) $T_0 = Id_X$ and $T_tT_s = T_{t+s}$ for all $s,t \geq 0$.*
2. *(Norm Continuity) $t \to T_t$ is continuous at 0, i.e. $\|T_t - I\|_{L(X)} \to 0$ as $t \downarrow 0$.*

Then there exists $A \in L(X)$ such that $T_t = e^{tA}$ where e^{tA} is defined in Eq. (32.19).

[1] We do this so that $\phi|_{J_0}$ will be bounded.

32.4 Operator Logarithms

Our goal in this section is to find an explicit local inverse to the exponential function, $A \to e^A$ for A near zero. The existence of such an inverse can be deduced from the inverse function theorem although we will not need this fact here. We begin with the real variable fact that

$$\ln(1+x) = \int_0^1 \frac{d}{ds}\ln(1+sx)\,ds = \int_0^1 x(1+sx)^{-1}\,ds.$$

Definition 32.26. *When $A \in L(X)$ satisfies $1 + sA$ is invertible for $0 \leq s \leq 1$ we define*

$$\ln(1+A) = \int_0^1 A(1+sA)^{-1}\,ds. \tag{32.21}$$

The invertibility of $1 + sA$ for $0 \leq s \leq 1$ is satisfied if;

1. A is nilpotent, i.e. $A^N = 0$ for some $N \in \mathbb{N}$ or more generally if
2. $\sum_{n=0}^{\infty} \|A^n\| < \infty$ (for example assume that $\|A\| < 1$), of
3. if X is a Hilbert space and $A^* = A$ with $A \geq 0$.

In the first two cases

$$(1+sA)^{-1} = \sum_{n=0}^{\infty} (-s)^n A^n.$$

Proposition 32.27. *If $1 + sA$ is invertible for $0 \leq s \leq 1$, then*

$$\partial_B \ln(1+A) = \int_0^1 (1+sA)^{-1} B (1+sA)^{-1}\,ds. \tag{32.22}$$

If $0 = [A,B] := AB - BA$, Eq. (32.22) reduces to

$$\partial_B \ln(1+A) = B(1+A)^{-1}. \tag{32.23}$$

Proof. Differentiating Eq. (32.21) shows

$$\begin{aligned}\partial_B \ln(1+A) &= \int_0^1 \left[B(1+sA)^{-1} - A(1+sA)^{-1} sB(1+sA)^{-1}\right] ds \\ &= \int_0^1 \left[B - sA(1+sA)^{-1} B\right](1+sA)^{-1} ds.\end{aligned}$$

Combining this last equality with

$$sA(1+sA)^{-1} = (1+sA-1)(1+sA)^{-1} = 1-(1+sA)^{-1}$$

gives Eq. (32.22). In case $[A,B]=0$,

$$\begin{aligned}(1+sA)^{-1} B(1+sA)^{-1} &= B(1+sA)^{-2} \\ &= B\frac{d}{ds}\left[-A^{-1}(1+sA)^{-1}\right]\end{aligned}$$

and so by the fundamental theorem of calculus

$$\begin{aligned}\partial_B \ln(1+A) &= B\int_0^1 (1+sA)^{-2} ds = B\left[-A^{-1}(1+sA)^{-1}\right]_{s=0}^{s=1} \\ &= B\left[A^{-1} - A^{-1}(1+A)^{-1}\right] = BA^{-1}\left[1-(1+A)^{-1}\right] \\ &= B\left[A^{-1}(1+A) - A^{-1}\right](1+A)^{-1} = B(1+A)^{-1}.\end{aligned}$$

■

Corollary 32.28. *Suppose that $t \to A(t) \in L(X)$ is a C^1 – function $1+sA(t)$ is invertible for $0 \le s \le 1$ for all $t \in J = (a,b) \subset \mathbb{R}$. If $g(t) := 1 + A(t)$ and $t \in J$, then*

$$\frac{d}{dt}\ln(g(t)) = \int_0^1 (1-s+sg(t))^{-1}\dot{g}(t)(1-s+sg(t))^{-1} ds. \tag{32.24}$$

Moreover if $[A(t), A(\tau)] = 0$ for all $t,\tau \in J$ then,

$$\frac{d}{dt}\ln(g(t)) = \dot{A}(t)(1+A(t))^{-1}. \tag{32.25}$$

Proof. Differentiating past the integral and then using Eq. (32.22) gives

$$\begin{aligned}\frac{d}{dt}\ln(g(t)) &= \int_0^1 (1+sA(t))^{-1}\dot{A}(t)(1+sA(t))^{-1} ds \\ &= \int_0^1 (1+s(g(t)-1))^{-1}\dot{g}(t)(1+s(g(t)-1))^{-1} ds \\ &= \int_0^1 (1-s+sg(t))^{-1}\dot{g}(t)(1-s+sg(t))^{-1} ds.\end{aligned}$$

For the second assertion we may use Eq. (32.23) instead Eq. (32.22) in order to immediately arrive at Eq. (32.25). ■

Theorem 32.29. *If $A \in L(X)$ satisfies, $1+sA$ is invertible for $0 \le s \le 1$, then*

$$e^{\ln(I+A)} = I + A. \tag{32.26}$$

If $C \in L(X)$ satisfies $\sum_{n=1}^\infty \frac{1}{n!}\|C^n\|^n < 1$ (for example assume $\|C\| < \ln 2$, i.e. $e^{\|C\|} < 2$), then

$$\ln e^C = C. \tag{32.27}$$

This equation also holds of C is nilpotent or if X is a Hilbert space and $C = C^*$ with $C \ge 0$.

Proof. For $0 \le t \le 1$ let

$$C(t) = \ln(I+tA) = t\int_0^1 A(1+stA)^{-1} ds.$$

Since $[C(t), C(\tau)] = 0$ for all $\tau, t \in [0,1]$, if we let $g(t) := e^{C(t)}$, then

$$\dot{g}(t) = \frac{d}{dt}e^{C(t)} = \dot{C}(t)e^{C(t)} = A(1+tA)^{-1} g(t) \text{ with } g(0) = I.$$

Noting that $g(t) = 1 + tA$ solves this ordinary differential equation, it follows by uniqueness of solutions to ODE's that $e^{C(t)} = g(t) = 1 + tA$. Evaluating this equation at $t = 1$ implies Eq. (32.26).

Now let $C \in L(X)$ as in the statement of the theorem and for $t \in \mathbb{R}$ set

$$A(t) := e^{tC} - 1 = \sum_{n=1}^\infty \frac{t^n}{n!}C^n.$$

Therefore,

$$1 + sA(t) = 1 + s\sum_{n=1}^\infty \frac{t^n}{n!}C^n$$

with

$$\left\| s\sum_{n=1}^\infty \frac{t^n}{n!}C^n \right\| \le s\sum_{n=1}^\infty \frac{t^n}{n!}\|C^n\|^n < 1 \text{ for } 0 \le s,t \le 1.$$

Because of this observation, $\ln(e^{tC}) := \ln(1+A(t))$ is well defined and because $[A(t), A(\tau)] = 0$ for all τ and t we may use Eq. (32.25) to learn,

$$\frac{d}{dt}\ln(e^{tC}) := \dot{A}(t)(1+A(t))^{-1} = Ce^{tC}e^{-tC} = C \text{ with } \ln(e^{0C}) = 0.$$

The unique solution to this simple ODE is $\ln(e^{tC}) = tC$ and evaluating this at $t = 1$ gives Eq. (32.27). ■

32.5 Classical Weierstrass Approximation Theorem

Definition 32.30 (Support). *Let $f : X \to Z$ be a function from a metric space (X, ρ) to a vector space Z. The support of f is the closed subset,* $\operatorname{supp}(f)$, *of X defined by*

$$\operatorname{supp}(f) := \overline{\{x \in X : f(x) \neq 0\}}.$$

Example 32.31. For example if $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = \sin(x) 1_{[0,4\pi]}(x) \in \mathbb{R}$, then

$$\{f \neq 0\} = (0, 4\pi) \setminus \{\pi, 2\pi, 3\pi\}$$

and therefore $\operatorname{supp}(f) = [0, 4\pi]$.

For the remainder of this section, Z will be used to denote a Banach space.

Definition 32.32 (Convolution). *For $f, g \in C(\mathbb{R})$ with either f or g having compact support, we define the **convolution** of f and g by*

$$f * g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy = \int_{\mathbb{R}} f(y)g(x-y)dy.$$

We will also use this definition when one of the functions, either f or g, takes values in a Banach space Z.

Lemma 32.33 (Approximate δ – sequences). *Suppose that $\{q_n\}_{n=1}^{\infty}$ is a sequence non-negative continuous real valued functions on $\mathbb{R}$ with compact support that satisfy*

$$\int_{\mathbb{R}} q_n(x)\, dx = 1 \text{ and} \tag{32.28}$$

$$\lim_{n\to\infty} \int_{|x|\geq\varepsilon} q_n(x)dx = 0 \text{ for all } \varepsilon > 0. \tag{32.29}$$

If $f \in BC(\mathbb{R}, Z)$, then

$$q_n * f\,(x) := \int_{\mathbb{R}} q_n(y) f(x-y)dy$$

converges to f uniformly on compact subsets of $\mathbb{R}$.

Proof. Let $x \in \mathbb{R}$, then because of Eq. (32.28),

$$\begin{aligned}\|q_n * f(x) - f(x)\| &= \left\| \int_{\mathbb{R}} q_n(y) \left(f(x-y) - f(x)\right) dy \right\| \\ &\leq \int_{\mathbb{R}} q_n(y) \|f(x-y) - f(x)\|\, dy.\end{aligned}$$

Let $M = \sup\{\|f(x)\| : x \in \mathbb{R}\}$. Then for any $\varepsilon > 0$, using Eq. (32.28),

$$\begin{aligned}\|q_n * f(x) - f(x)\| &\leq \int_{|y|\leq\varepsilon} q_n(y) \|f(x-y) - f(x)\|\, dy \\ &\quad + \int_{|y|>\varepsilon} q_n(y) \|f(x-y) - f(x)\|\, dy \\ &\leq \sup_{|w|\leq\varepsilon} \|f(x+w) - f(x)\| + 2M \int_{|y|>\varepsilon} q_n(y)dy.\end{aligned}$$

So if K is a compact subset of $\mathbb{R}$ (for example a large interval) we have

$$\begin{aligned}&\sup_{(x)\in K} \|q_n * f(x) - f(x)\| \\ &\quad\leq \sup_{|w|\leq\varepsilon,\ x\in K} \|f(x+w) - f(x)\| + 2M \int_{\|y\|>\varepsilon} q_n(y)dy\end{aligned}$$

and hence by Eq. (32.29),

$$\begin{aligned}&\limsup_{n\to\infty} \sup_{x\in K} \|q_n * f(x) - f(x)\| \\ &\quad\leq \sup_{|w|\leq\varepsilon,\ x\in K} \|f(x+w) - f(x)\|.\end{aligned}$$

This finishes the proof since the right member of this equation tends to 0 as $\varepsilon \downarrow 0$ by uniform continuity of f on compact subsets of $\mathbb{R}$. ■

Let $q_n : \mathbb{R} \to [0, \infty)$ be defined by

$$q_n(x) := \frac{1}{c_n}(1-x^2)^n 1_{|x|\leq 1} \text{ where } c_n := \int_{-1}^{1} (1-x^2)^n dx. \tag{32.30}$$

Figure 32.2 displays the key features of the functions q_n.

Lemma 32.34. *The sequence $\{q_n\}_{n=1}^{\infty}$ is an approximate δ – sequence, i.e. they satisfy Eqs. (32.28) and (32.29).*

Proof. By construction, $q_n \in C_c(\mathbb{R}, [0, \infty))$ for each n and Eq. 32.28 holds. Since

$$\begin{aligned}\int_{|x|\geq\varepsilon} q_n(x)dx &= \frac{2\int_\varepsilon^1 (1-x^2)^n dx}{2\int_0^\varepsilon (1-x^2)^n dx + 2\int_\varepsilon^1 (1-x^2)^n dx} \\ &\leq \frac{\int_\varepsilon^1 \frac{x}{\varepsilon}(1-x^2)^n dx}{\int_0^\varepsilon \frac{x}{\varepsilon}(1-x^2)^n dx} = \frac{(1-x^2)^{n+1}|_\varepsilon^1}{(1-x^2)^{n+1}|_0^\varepsilon} \\ &= \frac{(1-\varepsilon^2)^{n+1}}{1-(1-\varepsilon^2)^{n+1}} \to 0 \text{ as } n \to \infty,\end{aligned}$$

the proof is complete. ■

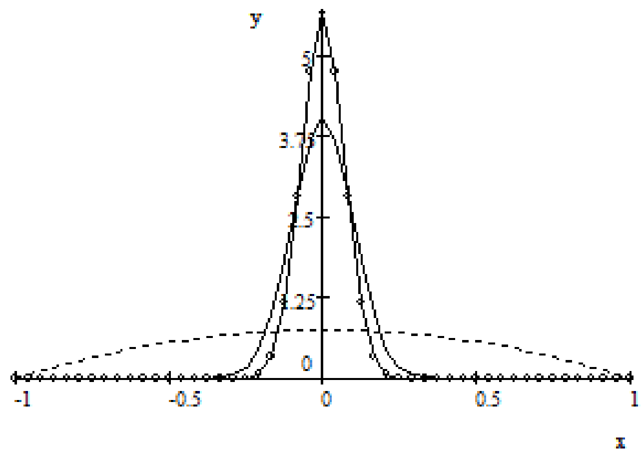


Fig. 32.2. A plot of q_1, q_{50}, and q_{100}. The most peaked curve is q_{100} and the least is q_1. The total area under each of these curves is one.

Notation 32.35 *Let $\mathbb{Z}_+ := \mathbb{N} \cup \{0\}$ and for $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{Z}_+^d$ let $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$ and $|\alpha| = \sum_{i=1}^d \alpha_i$. A polynomial on $\mathbb{R}^d$ with values in Z is a function $p : \mathbb{R}^d \to Z$ of the form*

$$p(x) = \sum_{\alpha:|\alpha|\le N} p_\alpha x^\alpha \text{ with } p_\alpha \in Z \text{ and } N \in \mathbb{Z}_+.$$

If $p_\alpha \neq 0$ for some α such that $|\alpha| = N$, then we define $\deg(p) := N$ to be the degree of p. If Z is a complex Banach space, the function p has a natural extension to $z \in \mathbb{C}^d$, namely $p(z) = \sum_{\alpha:|\alpha|\le N} p_\alpha z^\alpha$ where $z^\alpha = \prod_{i=1}^d z_i^{\alpha_i}$.

Given a compact subset $K \subset \mathbb{R}^d$ and $f \in C(K, \mathbb{C})$[2], we are going to show, in the Weierstrass approximation Theorem 32.39 below, that f may be uniformly approximated by polynomial functions on K. The next theorem addresses this question when K is a compact subinterval of $\mathbb{R}$.

Theorem 32.36 (Weierstrass Approximation Theorem). *Suppose $-\infty < a < b < \infty$, $J = [a, b]$ and $f \in C(J, Z)$. Then there exists polynomials p_n on $\mathbb{R}$ such that $p_n \to f$ uniformly on J.*

[2] Note that f is automatically bounded because if not there would exist $u_n \in K$ such that $\lim_{n\to\infty} |f(u_n)| = \infty$. Using Theorem 32.2 we may, by passing to a subsequence if necessary, assume $u_n \to u \in K$ as $n \to \infty$. Now the continuity of f would then imply

$$\infty = \lim_{n\to\infty} |f(u_n)| = |f(u)|$$

which is absurd since f takes values in $\mathbb{C}$.

Proof. By replacing f by F where

$$F(t) := f(a + t(b-a)) - [f(a) + t(f(b) - f(a))] \text{ for } t \in [0,1],$$

it suffices to assume $a = 0$, $b = 1$ and $f(0) = f(1) = 0$. Furthermore we may now extend f to a continuous function on all $\mathbb{R}$ by setting $f \equiv 0$ on $\mathbb{R} \setminus [0,1]$.

With q_n defined as in Eq. (32.30), let $f_n(x) := (q_n * f)(x)$ and recall from Lemma 32.33 that $f_n(x) \to f(x)$ as $n \to \infty$ with the convergence being uniform in $x \in [0,1]$. This completes the proof since f_n is equal to a polynomial function on $[0,1]$. Indeed, there are polynomials, $a_k(y)$, such that

$$(1-(x-y)^2)^n = \sum_{k=0}^{2n} a_k(y)\, x^k,$$

and therefore, for $x \in [0,1]$,

$$\begin{aligned} f_n(x) &= \int_{\mathbb{R}} q_n(x-y) f(y) dy \\ &= \frac{1}{c_n} \int_{[0,1]} f(y) \left[(1-(x-y)^2)^n 1_{|x-y|\le 1}\right] dy \\ &= \frac{1}{c_n} \int_{[0,1]} f(y)(1-(x-y)^2)^n dy \\ &= \frac{1}{c_n} \int_{[0,1]} f(y) \sum_{k=0}^{2n} a_k(y)\, x^k dy = \sum_{k=0}^{2n} A_k x^k \end{aligned}$$

where

$$A_k = \frac{1}{c_n} \int_{[0,1]} f(y)\, a_k(y)\, dy.$$

■

Lemma 32.37. *Suppose $J = [a, b]$ is a compact subinterval of $\mathbb{R}$ and K is a compact subset of $\mathbb{R}^{d-1}$, then the linear mapping $R : C(J \times K, Z) \to C(J, C(K, Z))$ defined by $(Rf)(t) = f(t, \cdot) \in C(K, Z)$ for $t \in J$ is an isometric isomorphism of Banach spaces.*

Proof. By uniform continuity of f on $J \times K$ (see Theorem 32.2),

$$\|(Rf)(t) - (Rf)(s)\|_{C(K,Z)} = \max_{y\in K} \|f(t,y) - f(s,y)\|_Z \to 0 \text{ as } s \to t$$

which shows that Rf is indeed in $C(J, C(K, Z))$. Moreover,

$$\begin{aligned} \|Rf\|_{C(J\to C(K,Z))} &= \max_{t\in J} \|(Rf)(t)\|_{C(K,Z)} \\ &= \max_{t\in J} \max_{y\in K} \|f(t,y)\|_Z = \|f\|_{C(J\times K,Z)}, \end{aligned}$$

showing R is isometric and therefore injective.

To see that R is surjective, let $F \in C(J, C(K,Z))$ and define $f(t,y) := F(t)(y)$. Since

$$\begin{aligned}\|f(t,y) - f(s,y')\|_Z &\le \|f(t,y) - f(s,y)\|_Z + \|f(s,y) - f(s,y')\|_Z \\ &\le \|F(t) - F(s)\|_{C(K,Z)} + \|F(s)(y) - F(s)(y')\|_Z ,\end{aligned}$$

it follows by the continuity of $t \to F(t)$ and $y \to F(s)(y)$ that

$$\|f(t,y) - f(s,y')\|_Z \to 0 \text{ as } (t,y) \to (s,y').$$

This shows $f \in C(J \times K, Z)$ and thus completes the proof because $Rf = F$ by construction. ■

Corollary 32.38 (Weierstrass Approximation Theorem). *Let $d \in \mathbb{N}$, $J_i = [a_i, b_i]$ be compact subintervals of $\mathbb{R}$ for $i = 1, 2, \ldots, d$, $J := J_1 \times \cdots \times J_d$ and $f \in C(J,Z)$. Then there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on J.*

Proof. The proof will be by induction on d with the case $d = 1$ being the content of Theorem 32.36. Now suppose that $d > 1$ and the theorem holds with d replaced by $d-1$. Let $K := J_2 \times \cdots \times J_d$, $Z_0 = C(K,Z)$, $R : C(J_1 \times K, Z) \to C(J_1, Z_0)$ be as in Lemma 32.37 and $F := Rf$. By Theorem 32.36, for any $\varepsilon > 0$ there exists a polynomial function

$$p(t) = \sum_{k=0}^{n} c_k t^k$$

with $c_k \in Z_0 = C(K,Z)$ such that $\|F - p\|_{C(J_1, Z_0)} < \varepsilon$. By the induction hypothesis, there exists polynomial functions $q_k : K \to Z$ such that

$$\|c_k - q_k\|_{Z_0} < \frac{\varepsilon}{n(|a| + |b|)^k}.$$

It is now easily verified (you check) that the polynomial function,

$$\rho(x) := \sum_{k=0}^{n} x_1^k q_k(x_2, \ldots, x_d) \text{ for } x \in J$$

satisfies $\|f - \rho\|_{C(J,Z)} < 2\varepsilon$ and this completes the induction argument and hence the proof. ■

The reader is referred to Chapter **??** for two more alternative proofs of this corollary.

Theorem 32.39 (Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{R}^d$ is a compact subset and $f \in C(K, \mathbb{C})$. Then there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K.*

Proof. Choose $\lambda > 0$ and $b \in \mathbb{R}^d$ such that

$$K_0 := \lambda K - b := \{\lambda x - b : x \in K\} \subset B_d$$

where $B_d := (0,1)^d$. The function $F(y) := f\left(\lambda^{-1}(y + b)\right)$ for $y \in K_0$ is in $C(K_0, \mathbb{C})$ and if $\hat{p}_n(y)$ are polynomials on $\mathbb{R}^d$ such that $\hat{p}_n \to F$ uniformly on K_0 then $p_n(x) := \hat{p}_n(\lambda x - b)$ are polynomials on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K. Hence we may now assume that K is a compact subset of B_d. Let $g \in C(K \cup B_d^c)$ be defined by

$$g(x) = \begin{cases} f(x) & \text{if } x \in K \\ 0 & \text{if } x \in B_d^c \end{cases}$$

and then use the Tietze extension Theorem 14.5 (applied to the real and imaginary parts of F) to find a continuous function $F \in C(\mathbb{R}^d, \mathbb{C})$ such that $F = g|_{K \cup B_d^c}$. If p_n are polynomials on $\mathbb{R}^d$ such that $p_n \to F$ uniformly on $[0,1]^d$ then p_n also converges to f uniformly on K. Hence, by replacing f by F, we may now assume that $f \in C(\mathbb{R}^d, \mathbb{C})$, $K = \bar{B}_d = [0,1]^d$, and $f \equiv 0$ on B_d^c. The result now follows by an application of Corollary 32.38 with $Z = \mathbb{C}$. ■

Remark 32.40. The mapping $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \to z = x + iy \in \mathbb{C}^d$ is an isomorphism of vector spaces. Letting $\bar{z} = x - iy$ as usual, we have $x = \frac{z + \bar{z}}{2}$ and $y = \frac{z - \bar{z}}{2i}$. Therefore under this identification any polynomial $p(x,y)$ on $\mathbb{R}^d \times \mathbb{R}^d$ may be written as a polynomial q in $(z, \bar{z})$, namely

$$q(z, \bar{z}) = p\left(\frac{z + \bar{z}}{2}, \frac{z - \bar{z}}{2i}\right).$$

Conversely a polynomial q in $(z, \bar{z})$ may be thought of as a polynomial p in (x,y), namely $p(x,y) = q(x + iy, x - iy)$.

Corollary 32.41 (Complex Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{C}^d$ is a compact set and $f \in C(K, \mathbb{C})$. Then there exists polynomials $p_n(z, \bar{z})$ for $z \in \mathbb{C}^d$ such that $\sup_{z \in K} |p_n(z, \bar{z}) - f(z)| \to 0$ as $n \to \infty$.*

Proof. This is an immediate consequence of Theorem 32.39 and Remark 32.40. ■

Example 32.42. Let $K = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}$ be the set of polynomials in $(z, \bar{z})$ restricted to S^1. Then $\mathcal{A}$ is dense in $C(S^1)$.[3] Since $\bar{z} = z^{-1}$ on S^1,

[3] Note that it is easy to extend $f \in C(S^1)$ to a function $F \in C(\mathbb{C})$ by setting $F(z) = |z| f(\frac{z}{|z|})$ for $z \neq 0$ and $F(0) = 0$. So this special case does not require the Tietze extension theorem.

we have shown polynomials in z and z^{-1} are dense in $C(S^1)$. This example generalizes in an obvious way to $K = \left(S^1\right)^d \subset \mathbb{C}^d$.

Exercise 32.4. Suppose $-\infty < a < b < \infty$ and $f \in C([a,b],\mathbb{C})$ satisfies

$$\int_a^b f(t)\, t^n dt = 0 \text{ for } n = 0,1,2\dots.$$

Show $f \equiv 0$.

Exercise 32.5. Suppose $f \in C(\mathbb{R},\mathbb{C})$ is a 2π – periodic function (i.e. $f(x+2\pi) = f(x)$ for all $x \in \mathbb{R}$) and

$$\int_0^{2\pi} f(x)\, e^{inx} dx = 0 \text{ for all } n \in \mathbb{Z},$$

show again that $f \equiv 0$. **Hint**: Use Example 32.42 to show that any 2π – periodic continuous function g on $\mathbb{R}$ is the uniform limit of trigonometric polynomials of the form

$$p(x) = \sum_{k=-n}^{n} p_k e^{ikx} \text{ with } p_k \in \mathbb{C} \text{ for all } k.$$

32.6 Iterated Integrals

Theorem 32.43 (Baby Fubini Theorem). *Let $a_i, b_i \in \mathbb{R}$ with $a_i \neq b_i$ for $i = 1, 2\dots, n$, $f(t_1, t_2, \dots, t_n) \in Z$ be a continuous function of $(t_1, t_2, \dots, t_n)$ where t_i between a_i and b_i for each i and for any given permutation, σ, of $\{1, 2\dots, n\}$ let*

$$I_\sigma(f) := \int_{a_{\sigma_1}}^{b_{\sigma_1}} dt_{\sigma_1} \dots \int_{a_{\sigma_n}}^{b_{\sigma_n}} dt_{\sigma_n} f(t_1, t_2, \dots, t_n). \tag{32.31}$$

Then $I_\sigma(f)$ is well defined and independent of σ, i.e. the order of iterated integrals is irrelevant under these hypothesis.

Proof. Let $J_i := [\min(a_i, b_i), \max(a_i, b_i)]$, $J := J_1 \times \dots \times J_n$ and $|J_i| := \max(a_i, b_i) - \min(a_i, b_i)$. Using the uniform continuity of f (Theorem 32.2) and the continuity of the Riemann integral, it is easy to prove (compare with the proof of Lemma 32.37) that the map

$$(t_1, \dots, \hat{t}_{\sigma n}, \dots, t_n) \in (J_1 \times \dots \times \hat{J}_{\sigma n} \times \dots \times J_n) \to \int_{a_{\sigma n}}^{b_{\sigma n}} dt_{\sigma n} f(t_1, t_2, \dots, t_n)$$

is continuous, where the hat is used to denote a missing element from a list. From this remark, it follows that each of the integrals in Eq. (32.31) are well defined and hence so is $I_\sigma(f)$. Moreover by an induction argument using Lemma 32.37 and the boundedness of the Riemann integral, we have the estimate,

$$\|I_\sigma(f)\|_Z \leq \left(\prod_{i=1}^n |J_i|\right) \|f\|_{C(J,Z)}. \tag{32.32}$$

Now suppose τ is another permutation. Because of Eq. (32.32), I_σ and I_τ are bounded operators on $C(J,Z)$ and so to shows $I_\sigma = I_\tau$ is suffices to shows there are equal on the dense set of polynomial functions (see Corollary 32.38) in $C(J,Z)$. Moreover by linearity, it suffices to show $I_\sigma(f) = I_\tau(f)$ when f has the form

$$f(t_1, t_2, \dots, t_n) = t_1^{k_1} \dots t_n^{k_n} z$$

for some $k_i \in \mathbb{N}_0$ and $z \in Z$. However for this function, explicit computations show

$$I_\sigma(f) = I_\tau(f) = \left(\prod_{i=1}^n \frac{b_i^{k_i+1} - a_i^{k_i+1}}{k_i + 1}\right) \cdot z.$$

■

Proposition 32.44 (Equality of Mixed Partial Derivatives). *Let $Q = (a,b) \times (c,d)$ be an open rectangle in $\mathbb{R}^2$ and $f \in C(Q,Z)$. Assume that $\frac{\partial}{\partial t} f(s,t)$, $\frac{\partial}{\partial s} f(s,t)$ and $\frac{\partial}{\partial t}\frac{\partial}{\partial s} f(s,t)$ exists and are continuous for $(s,t) \in Q$, then $\frac{\partial}{\partial s}\frac{\partial}{\partial t} f(s,t)$ exists for $(s,t) \in Q$ and*

$$\frac{\partial}{\partial s}\frac{\partial}{\partial t} f(s,t) = \frac{\partial}{\partial t}\frac{\partial}{\partial s} f(s,t) \text{ for } (s,t) \in Q. \tag{32.33}$$

Proof. Fix $(s_0, t_0) \in Q$. By two applications of Theorem 32.14,

$$\begin{aligned} f(s,t) &= f(s_0,t) + \int_{s_0}^s \frac{\partial}{\partial \sigma} f(\sigma,t) d\sigma \\ &= f(s_0,t) + \int_{s_0}^s \frac{\partial}{\partial \sigma} f(\sigma,t_0) d\sigma + \int_{s_0}^s d\sigma \int_{t_0}^t d\tau \frac{\partial}{\partial \tau}\frac{\partial}{\partial \sigma} f(\sigma,\tau) \end{aligned} \tag{32.34}$$

and then by Fubini's Theorem 32.43 we learn

$$f(s,t) = f(s_0,t) + \int_{s_0}^s \frac{\partial}{\partial \sigma} f(\sigma,t_0) d\sigma + \int_{t_0}^t d\tau \int_{s_0}^s d\sigma \frac{\partial}{\partial \tau}\frac{\partial}{\partial \sigma} f(\sigma,\tau).$$

Differentiating this equation in t and then in s (again using two more applications of Theorem 32.14) shows Eq. (32.33) holds. ■

32.7 Exercises

Throughout these problems, $(X, \|\cdot\|)$ is a Banach space.

Exercise 32.6. Show $f = (f_1, \dots, f_n) \in \bar{\mathcal{S}}([a,b], \mathbb{R}^n)$ iff $f_i \in \bar{\mathcal{S}}([a,b], \mathbb{R})$ for $i = 1, 2, \dots, n$ and

$$\int_a^b f(t)dt = \left(\int_a^b f_1(t)dt, \dots, \int_a^b f_n(t)dt\right).$$

Here $\mathbb{R}^n$ is to be equipped with the usual Euclidean norm. **Hint:** Use Lemma 32.7 to prove the forward implication.

Exercise 32.7. Give another proof of Proposition 32.44 which does not use Fubini's Theorem 32.43 as follows.

1. By a simple translation argument we may assume $(0,0) \in Q$ and we are trying to prove Eq. (32.33) holds at $(s,t) = (0,0)$.
2. Let $h(s,t) := \frac{\partial}{\partial t}\frac{\partial}{\partial s} f(s,t)$ and

   $$G(s,t) := \int_0^s d\sigma \int_0^t d\tau h(\sigma, \tau)$$

   so that Eq. (32.34) states

   $$f(s,t) = f(0,t) + \int_0^s \frac{\partial}{\partial \sigma} f(\sigma, t_0) d\sigma + G(s,t)$$

   and differentiating this equation at $t = 0$ shows

   $$\frac{\partial}{\partial t} f(s,0) = \frac{\partial}{\partial t} f(0,0) + \frac{\partial}{\partial t} G(s,0). \tag{32.35}$$

   Now show using the definition of the derivative that

   $$\frac{\partial}{\partial t} G(s,0) = \int_0^s d\sigma h(\sigma, 0). \tag{32.36}$$

   Hint: Consider

   $$G(s,t) - t\int_0^s d\sigma h(\sigma,0) = \int_0^s d\sigma \int_0^t d\tau \left[h(\sigma,\tau) - h(\sigma,0)\right].$$

3. Now differentiate Eq. (32.35) in s using Theorem 32.14 to finish the proof.

Exercise 32.8. Give another proof of Eq. (32.31) in Theorem 32.43 based on Proposition 32.44. To do this let $t_0 \in (c,d)$ and $s_0 \in (a,b)$ and define

$$G(s,t) := \int_{t_0}^t d\tau \int_{s_0}^s d\sigma f(\sigma, \tau)$$

Show G satisfies the hypothesis of Proposition 32.44 which combined with two applications of the fundamental theorem of calculus implies

$$\frac{\partial}{\partial t}\frac{\partial}{\partial s} G(s,t) = \frac{\partial}{\partial s}\frac{\partial}{\partial t} G(s,t) = f(s,t).$$

Use two more applications of the fundamental theorem of calculus along with the observation that $G = 0$ if $t = t_0$ or $s = s_0$ to conclude

$$G(s,t) = \int_{s_0}^s d\sigma \int_{t_0}^t d\tau \frac{\partial}{\partial \tau}\frac{\partial}{\partial \sigma} G(\sigma,\tau) = \int_{s_0}^s d\sigma \int_{t_0}^t d\tau \frac{\partial}{\partial \tau} f(\sigma, \tau). \tag{32.37}$$

Finally let $s = b$ and $t = d$ in Eq. (32.37) and then let $s_0 \downarrow a$ and $t_0 \downarrow c$ to prove Eq. (32.31).

Exercise 32.9 (Product Rule). Prove items 1. and 2. of Lemma 32.10. This can be modeled on the standard proof for real valued functions.

Exercise 32.10 (Chain Rule). Prove the chain rule in Proposition 32.11. Again this may be modeled on the standard proof for real valued functions.

Exercise 32.11. To each $A \in L(X)$, we may define $L_A, R_A : L(X) \to L(X)$ by

$$L_A B = AB \text{ and } R_A B = BA \text{ for all } B \in L(X).$$

Show $L_A, R_A \in L(L(X))$ and that

$$\|L_A\|_{L(L(X))} = \|A\|_{L(X)} = \|R_A\|_{L(L(X))}.$$

Exercise 32.12. Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $U, V : \mathbb{R} \to L(X)$ are the unique solution to the linear differential equations

$$\dot{V}(t) = A(t)V(t) \text{ with } V(0) = I \tag{32.38}$$

and

$$\dot{U}(t) = -U(t)A(t) \text{ with } U(0) = I. \tag{32.39}$$

Prove that $V(t)$ is invertible and that $V^{-1}(t) = U(t)$[4], where by abuse of notation I am writing $V^{-1}(t)$ for $[V(t)]^{-1}$. **Hints:** 1) show $\frac{d}{dt}[U(t)V(t)] = 0$ (which is sufficient if $\dim(X) < \infty$) and 2) show $y(t) := V(t)U(t)$ solves a linear differential ordinary differential equation that has $y \equiv I$ as an obvious solution. (The results of Exercise 32.11 may be useful here.) Then use the uniqueness of solutions to linear ODEs.

[4] The fact that $U(t)$ must be defined as in Eq. (32.39) follows from Lemma 32.10.

Exercise 32.13. Suppose that $(X, \|\cdot\|)$ is a Banach space, $J = (a, b)$ with $-\infty \le a < b \le \infty$ and $f_n : \mathbb{R} \to X$ are continuously differentiable functions such that there exists a summable sequence $\{a_n\}_{n=1}^{\infty}$ satisfying

$$\|f_n(t)\| + \left\|\dot{f}_n(t)\right\| \le a_n \text{ for all } t \in J \text{ and } n \in \mathbb{N}.$$

Show:

1. $\sup\left\{\left\|\frac{f_n(t+h)-f_n(t)}{h}\right\| : (t,h) \in J \times \mathbb{R} \ni t+h \in J \text{ and } h \neq 0\right\} \le a_n.$
2. The function $F : \mathbb{R} \to X$ defined by
$$F(t) := \sum_{n=1}^{\infty} f_n(t) \text{ for all } t \in J$$
is differentiable and for $t \in J$,
$$\dot{F}(t) = \sum_{n=1}^{\infty} \dot{f}_n(t).$$

Exercise 32.14. Suppose that $A \in L(X)$. Show directly that:

1. e^{tA} define in Eq. (32.19) is convergent in $L(X)$ when equipped with the operator norm.
2. e^{tA} is differentiable in t and that $\frac{d}{dt}e^{tA} = Ae^{tA}$.

Exercise 32.15. Suppose that $A \in L(X)$ and $v \in X$ is an eigenvector of A with eigenvalue λ, i.e. that $Av = \lambda v$. Show $e^{tA}v = e^{t\lambda}v$. Also show that if $X = \mathbb{R}^n$ and A is a diagonalizable $n \times n$ matrix with

$$A = SDS^{-1} \text{ with } D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

then $e^{tA} = Se^{tD}S^{-1}$ where $e^{tD} = \operatorname{diag}(e^{t\lambda_1}, \dots, e^{t\lambda_n})$. Here $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ denotes the diagonal matrix Λ such that $\Lambda_{ii} = \lambda_i$ for $i = 1, 2, \dots, n$.

Exercise 32.16. Suppose that $A, B \in L(X)$ and $[A, B] := AB - BA = 0$. Show that $e^{(A+B)} = e^A e^B$.

Exercise 32.17. Suppose $A \in C(\mathbb{R}, L(X))$ satisfies $[A(t), A(s)] = 0$ for all $s, t \in \mathbb{R}$. Show

$$y(t) := e^{\left(\int_0^t A(\tau)d\tau\right)}x$$

is the unique solution to $\dot{y}(t) = A(t)y(t)$ with $y(0) = x$.

Exercise 32.18. Compute e^{tA} when

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and use the result to prove the formula

$$\cos(s+t) = \cos s \cos t - \sin s \sin t.$$

Hint: Sum the series and use $e^{tA}e^{sA} = e^{(t+s)A}$.

Exercise 32.19. Compute e^{tA} when

$$A = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}$$

with $a, b, c \in \mathbb{R}$. Use your result to compute $e^{t(\lambda I + A)}$ where $\lambda \in \mathbb{R}$ and I is the 3×3 identity matrix. **Hint:** Sum the series.

Exercise 32.20. Prove Theorem 32.25 using the following outline.

1. Using the right continuity at 0 and the semi-group property for T_t, show there are constants M and C such that $\|T_t\|_{L(X)} \le MC^t$ for all $t > 0$.
2. Show $t \in [0, \infty) \to T_t \in L(X)$ is continuous.
3. For $\varepsilon > 0$, let $S_\varepsilon := \frac{1}{\varepsilon}\int_0^\varepsilon T_\tau d\tau \in L(X)$. Show $S_\varepsilon \to I$ as $\varepsilon \downarrow 0$ and conclude from this that S_ε is invertible when $\varepsilon > 0$ is sufficiently small. For the remainder of the proof fix such a small $\varepsilon > 0$.
4. Show
$$T_t S_\varepsilon = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} T_\tau d\tau$$
and conclude from this that
$$\lim_{t\downarrow 0}\left(\frac{T_t - I}{t}\right)S_\varepsilon = \frac{1}{\varepsilon}\left(T_\varepsilon - Id_X\right).$$
5. Using the fact that S_ε is invertible, conclude $A = \lim_{t\downarrow 0} t^{-1}(T_t - I)$ exists in $L(X)$ and that
$$A = \frac{1}{\varepsilon}\left(T_\varepsilon - I\right)S_\varepsilon^{-1}.$$
6. Now show, using the semigroup property and step 5., that $\frac{d}{dt}T_t = AT_t$ for all $t > 0$.
7. Using step 6., show $\frac{d}{dt}e^{-tA}T_t = 0$ for all $t > 0$ and therefore $e^{-tA}T_t = e^{-0A}T_0 = I$.

Exercise 32.21 (Duhamel' s Principle I). Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $V : \mathbb{R} \to L(X)$ is the unique solution to the linear differential equation in Eq. (32.38). Let $x \in X$ and $h \in C(\mathbb{R}, X)$ be given. Show that the unique solution to the differential equation:

$$\dot{y}(t) = A(t)y(t) + h(t) \text{ with } y(0) = x \tag{32.40}$$

is given by

$$y(t) = V(t)x + V(t) \int_0^t V(\tau)^{-1} h(\tau)\, d\tau. \tag{32.41}$$

Hint: compute $\frac{d}{dt}[V^{-1}(t)y(t)]$ (see Exercise 32.12) when y solves Eq. (32.40).

Exercise 32.22 (Duhamel' s Principle II). Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $V : \mathbb{R} \to L(X)$ is the unique solution to the linear differential equation in Eq. (32.38). Let $W_0 \in L(X)$ and $H \in C(\mathbb{R}, L(X))$ be given. Show that the unique solution to the differential equation:

$$\dot{W}(t) = A(t)W(t) + H(t) \text{ with } W(0) = W_0 \tag{32.42}$$

is given by

$$W(t) = V(t)W_0 + V(t) \int_0^t V(\tau)^{-1} H(\tau)\, d\tau. \tag{32.43}$$

33

Ordinary Differential Equations in a Banach Space

Let X be a Banach space, $U \subset_o X$, $J = (a, b) \ni 0$ and $Z \in C(J \times U, X)$. The function Z is to be interpreted as a time dependent vector-field on $U \subset X$. In this section we will consider the ordinary differential equation (ODE for short)

$$\dot{y}(t) = Z(t, y(t)) \text{ with } y(0) = x \in U. \tag{33.1}$$

The reader should check that any solution $y \in C^1(J, U)$ to Eq. (33.1) gives a solution $y \in C(J, U)$ to the integral equation:

$$y(t) = x + \int_0^t Z(\tau, y(\tau))d\tau \tag{33.2}$$

and conversely if $y \in C(J, U)$ solves Eq. (33.2) then $y \in C^1(J, U)$ and y solves Eq. (33.1).

Remark 33.1. For notational simplicity we have assumed that the initial condition for the ODE in Eq. (33.1) is taken at $t = 0$. There is no loss in generality in doing this since if $\tilde{y}$ solves

$$\frac{d\tilde{y}}{dt}(t) = \tilde{Z}(t, \tilde{y}(t)) \text{ with } \tilde{y}(t_0) = x \in U$$

iff $y(t) := \tilde{y}(t + t_0)$ solves Eq. (33.1) with $Z(t, x) = \tilde{Z}(t + t_0, x)$.

33.1 Examples

Let $X = \mathbb{R}$, $Z(x) = x^n$ with $n \in \mathbb{N}$ and consider the ordinary differential equation

$$\dot{y}(t) = Z(y(t)) = y^n(t) \text{ with } y(0) = x \in \mathbb{R}. \tag{33.3}$$

If y solves Eq. (33.3) with $x \neq 0$, then $y(t)$ is not zero for t near 0. Therefore up to the first time y possibly hits 0, we must have

$$t = \int_0^t \frac{\dot{y}(\tau)}{y(\tau)^n} d\tau = \int_{y(0)}^{y(t)} u^{-n} du = \begin{cases} \frac{[y(t)]^{1-n} - x^{1-n}}{1-n} & \text{if } n > 1 \\ \ln\left|\frac{y(t)}{x}\right| & \text{if } n = 1 \end{cases}$$

and solving these equations for $y(t)$ implies

$$y(t) = y(t, x) = \begin{cases} \frac{x}{\sqrt[n-1]{1-(n-1)tx^{n-1}}} & \text{if } n > 1 \\ e^t x & \text{if } n = 1. \end{cases} \tag{33.4}$$

The reader should verify by direct calculation that $y(t, x)$ defined above does indeed solve Eq. (33.3). The above argument shows that these are the only possible solutions to the Equations in (33.3).

Notice that when $n = 1$, the solution exists for all time while for $n > 1$, we must require

$$1 - (n-1)tx^{n-1} > 0$$

or equivalently that

$$t < \frac{1}{(1-n)x^{n-1}} \text{ if } x^{n-1} > 0 \text{ and}$$
$$t > -\frac{1}{(1-n)|x|^{n-1}} \text{ if } x^{n-1} < 0.$$

Moreover for $n > 1$, $y(t, x)$ blows up as t approaches $(n-1)^{-1} x^{1-n}$. The reader should also observe that, at least for s and t close to 0,

$$y(t, y(s, x)) = y(t + s, x) \tag{33.5}$$

for each of the solutions above. Indeed, if $n = 1$ Eq. (33.5) is equivalent to the well know identity, $e^t e^s = e^{t+s}$ and for $n > 1$,

$$\begin{aligned}
y(t, y(s,x)) &= \frac{y(s,x)}{\sqrt[n-1]{1-(n-1)ty(s,x)^{n-1}}} \\
&= \frac{\frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}}}}{\sqrt[n-1]{1-(n-1)t\left[\frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}}}\right]^{n-1}}} \\
&= \frac{\frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}}}}{\sqrt[n-1]{1-(n-1)t\frac{x^{n-1}}{1-(n-1)sx^{n-1}}}} \\
&= \frac{x}{\sqrt[n-1]{1-(n-1)sx^{n-1}-(n-1)tx^{n-1}}} \\
&= \frac{x}{\sqrt[n-1]{1-(n-1)(s+t)x^{n-1}}} = y(t+s,x).
\end{aligned}$$

Now suppose $Z(x) = |x|^\alpha$ with $0 < \alpha < 1$ and we now consider the ordinary differential equation

$$\dot{y}(t) = Z(y(t)) = |y(t)|^\alpha \text{ with } y(0) = x \in \mathbb{R}. \tag{33.6}$$

Working as above we find, if $x \neq 0$ that

$$t = \int_0^t \frac{\dot{y}(\tau)}{|y(t)|^\alpha} d\tau = \int_{y(0)}^{y(t)} |u|^{-\alpha} du = \frac{[y(t)]^{1-\alpha} - x^{1-\alpha}}{1-\alpha},$$

where $u^{1-\alpha} := |u|^{1-\alpha} \operatorname{sgn}(u)$. Since $\operatorname{sgn}(y(t)) = \operatorname{sgn}(x)$ the previous equation implies

$$\begin{aligned}
\operatorname{sgn}(x)(1-\alpha)t &= \operatorname{sgn}(x)\left[\operatorname{sgn}(y(t))\,|y(t)|^{1-\alpha} - \operatorname{sgn}(x)\,|x|^{1-\alpha}\right] \\
&= |y(t)|^{1-\alpha} - |x|^{1-\alpha}
\end{aligned}$$

and therefore,

$$y(t,x) = \operatorname{sgn}(x)\left(|x|^{1-\alpha} + \operatorname{sgn}(x)(1-\alpha)t\right)^{\frac{1}{1-\alpha}} \tag{33.7}$$

is uniquely determined by this formula until the first time t where $|x|^{1-\alpha} + \operatorname{sgn}(x)(1-\alpha)t = 0$.

As before $y(t) = 0$ is a solution to Eq. (33.6) when $x = 0$, however it is far from being the unique solution. For example letting $x \downarrow 0$ in Eq. (33.7) gives a function

$$y(t, 0+) = ((1-\alpha)t)^{\frac{1}{1-\alpha}}$$

which solves Eq. (33.6) for $t > 0$. Moreover if we define

$$y(t) := \begin{cases} ((1-\alpha)t)^{\frac{1}{1-\alpha}} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases},$$

(for example if $\alpha = 1/2$ then $y(t) = \frac{1}{4}t^2 1_{t \geq 0}$) then the reader may easily check y also solve Eq. (33.6). Furthermore, $y_a(t) := y(t-a)$ also solves Eq. (33.6) for all $a \geq 0$, see Figure 33.1 below.

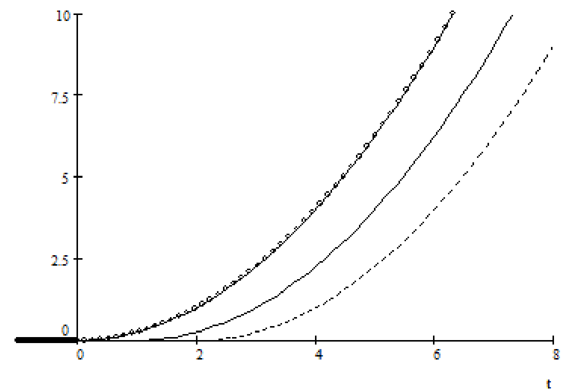


Fig. 33.1. Three different solutions to the ODE $\dot{y}(t) = |y(t)|^{1/2}$ with $y(0) = 0$.

With these examples in mind, let us now go to the general theory. The case of linear ODE's has already been studied in Section 32.3 above.

33.2 Uniqueness Theorem and Continuous Dependence on Initial Data

Lemma 33.2 (Gronwall's Lemma). *Suppose that f, ε, and k are non-negative locally integrable functions of $t \in [0, \infty)$ such that*

$$f(t) \leq \varepsilon(t) + \left|\int_0^t k(\tau) f(\tau) d\tau\right|. \tag{33.8}$$

Then

$$f(t) \leq \varepsilon(t) + \left|\int_0^t k(\tau)\varepsilon(\tau) e^{\left|\int_\tau^t k(s)ds\right|} d\tau\right|, \tag{33.9}$$

and in particular if ε and k are constants we find that

$$f(t) \leq \varepsilon e^{k|t|}. \tag{33.10}$$

Proof. I will only prove the case $t \geq 0$. The case $t \leq 0$ can be derived by applying the $t \geq 0$ to $\tilde{f}(t) = f(-t)$, $\tilde{k}(t) = k(-t)$ and $\varepsilon(t) = \varepsilon(-t)$.

Set $F(t) = \int_0^t k(\tau)f(\tau)d\tau$. Then by (33.8) and the Lebesgue version of the fundamental theorem of calculus,

$$\dot{F} = kf \leq k\varepsilon + kF \text{ a.e.}$$

Hence,

$$\frac{d}{dt}\left(e^{-\int_0^t k(s)ds}F(t)\right) \overset{\text{a.e.}}{=} e^{-\int_0^t k(s)ds}(\dot{F}(t) - k(t)F(t)) \overset{\text{a.e.}}{\leq} k(t)\varepsilon(t)e^{-\int_0^t k(s)ds}.$$

Integrating this last inequality from 0 to t and then solving for F yields:

$$F(t) \leq e^{\int_0^t k(s)ds} \cdot \int_0^t d\tau k(\tau)\varepsilon(\tau)e^{-\int_0^\tau k(s)ds} = \int_0^t d\tau k(\tau)\varepsilon(\tau)e^{\int_\tau^t k(s)ds}.$$

But by the definition of F and Eq. (33.8) we have,

$$f(t) \leq \varepsilon(t) + F(t) \leq \int_0^t d\tau k(\tau)\varepsilon(\tau)e^{\int_\tau^t k(s)ds}$$

which is Eq. (33.9). Equation (33.10) follows from (33.9) by a simple integration. ■

Corollary 33.3 (Continuous Dependence on Initial Data). *Let $U \subset_o X$, $0 \in (a,b)$ and $Z : (a,b) \times U \to X$ be a continuous function which is K–Lipschitz function on U, i.e. $\|Z(t,x) - Z(t,x')\| \leq K\|x - x'\|$ for all x and x' in U. Suppose $y_1, y_2 : (a,b) \to U$ solve*

$$\frac{dy_i(t)}{dt} = Z(t, y_i(t)) \quad \text{with } y_i(0) = x_i \quad \text{for } i = 1,2. \tag{33.11}$$

Then

$$\|y_2(t) - y_1(t)\| \leq \|x_2 - x_1\|e^{K|t|} \text{ for } t \in (a,b) \tag{33.12}$$

and in particular, there is at most one solution to Eq. (33.1) under the above Lipschitz assumption on Z.

Proof. Let $f(t) := \|y_2(t) - y_1(t)\|$. Then by the fundamental theorem of calculus,

$$\begin{aligned} f(t) &= \|y_2(0) - y_1(0) + \int_0^t (\dot{y}_2(\tau) - \dot{y}_1(\tau))\,d\tau\| \\ &\leq f(0) + \left|\int_0^t \|Z(\tau, y_2(\tau)) - Z(\tau, y_1(\tau))\|\,d\tau\right| \\ &= \|x_2 - x_1\| + K\left|\int_0^t f(\tau)\,d\tau\right|. \end{aligned}$$

Therefore by Gronwall's inequality we have,

$$\|y_2(t) - y_1(t)\| = f(t) \leq \|x_2 - x_1\|e^{K|t|}.$$

■

33.3 Local Existence (Non-Linear ODE)

Lemma 33.4. *Suppose that $K(t)$ is a locally integrable function of $t \in [0,\infty)$ and $\{\varepsilon_n(t)\}_{n=0}^\infty$ is a sequence of non-negative continuous functions such that*

$$\varepsilon_{n+1}(t) \leq \int_0^t K(\tau)\varepsilon_n(\tau)\,d\tau \text{ for all } n \geq 0 \tag{33.13}$$

and $\varepsilon_0(t) \leq \delta < \infty$ for all $t \in [0,\infty)$. Then

$$\varepsilon_n(t) \leq \frac{\delta}{n!}\left[\int_0^t K(s)\,d\tau\right]^n. \tag{33.14}$$

Proof. The proof is by induction. Notice that

$$\varepsilon_1(t) \leq \int_0^t K(\tau)\varepsilon_0(\tau)\,d\tau \leq \delta\int_0^t K(\tau)\,d\tau$$

as desired. If Eq. (33.14) holds for level n, then

$$\begin{aligned} \varepsilon_{n+1}(t) &\leq \int_0^t K(\tau)\varepsilon_n(\tau)\,d\tau \leq \frac{\delta}{n!}\int_0^t K(\tau)\left[\int_0^\tau K(s)\,d\tau\right]^n d\tau \\ &= \frac{\delta}{n!}\int_0^t \frac{1}{n+1}\frac{d}{d\tau}\left[\int_0^\tau K(s)\,d\tau\right]^{n+1} d\tau \\ &= \frac{\delta}{(n+1)!}\left[\int_0^t K(s)\,d\tau\right]^{n+1} \end{aligned}$$

which is Eq. (33.14) at level $n+1$. ■

We now show that Eq. (33.1) has a unique solution when Z satisfies the **Lipschitz condition** in Eq. (33.16). See Exercise 36.17 below for another existence theorem.

Theorem 33.5 (Local Existence). *Let $T > 0$, $J = (-T,T)$, $x_0 \in X$, $r > 0$ and*

$$C(x_0, r) := \{x \in X : \|x - x_0\| \leq r\}$$

be the closed r – ball centered at $x_0 \in X$. Assume

$$M = \sup\{\|Z(t,x)\| : (t,x) \in J \times C(x_0,r)\} < \infty \tag{33.15}$$

and there exists $K < \infty$ such that

$$\|Z(t,x) - Z(t,y)\| \le K\|x-y\| \text{ for all } x,y \in C(x_0,r) \text{ and } t \in J. \tag{33.16}$$

Let $T_0 < \min\{r/M, T\}$ and $J_0 := (-T_0, T_0)$, then for each $x \in B(x_0, r - MT_0)$ there exists a unique solution $y(t) = y(t,x)$ to Eq. (33.2) in $C(J_0, C(x_0,r))$. Moreover $y(t,x)$ is jointly continuous in (t,x), $y(t,x)$ is differentiable in t, $\dot{y}(t,x)$ is jointly continuous for all $(t,x) \in J_0 \times B(x_0, r-MT_0)$ and satisfies Eq. (33.1).

Proof. The uniqueness assertion has already been proved in Corollary 33.3. To prove existence, let $C_r := C(x_0,r)$, $Y := C(J_0, C(x_0,r))$ and

$$S_x(y)(t) := x + \int_0^t Z(\tau, y(\tau))d\tau. \tag{33.17}$$

With this notation, Eq. (33.2) becomes $y = S_x(y)$, i.e. we are looking for a fixed point of S_x. If $y \in Y$, then

$$\begin{aligned}\|S_x(y)(t) - x_0\| &\le \|x - x_0\| + \left|\int_0^t \|Z(\tau, y(\tau))\| d\tau\right| \le \|x-x_0\| + M|t| \\ &\le \|x-x_0\| + MT_0 \le r - MT_0 + MT_0 = r,\end{aligned}$$

showing $S_x(Y) \subset Y$ for all $x \in B(x_0, r - MT_0)$. Moreover if $y, z \in Y$,

$$\begin{aligned}\|S_x(y)(t) - S_x(z)(t)\| &= \left\|\int_0^t [Z(\tau,y(\tau)) - Z(\tau,z(\tau))]\, d\tau\right\| \\ &\le \left|\int_0^t \|Z(\tau,y(\tau)) - Z(\tau,z(\tau))\| d\tau\right| \\ &\le K\left|\int_0^t \|y(\tau) - z(\tau)\| d\tau\right|. \end{aligned} \tag{33.18}$$

Let $y_0(t,x) = x$ and $y_n(\cdot,x) \in Y$ defined inductively by

$$y_n(\cdot,x) := S_x(y_{n-1}(\cdot,x)) = x + \int_0^t Z(\tau, y_{n-1}(\tau,x))d\tau. \tag{33.19}$$

Using the estimate in Eq. (33.18) repeatedly we find

$$\begin{aligned}&\| y_{n+1}(t) - y_n(t) \| \\ &\quad\le K\left|\int_0^t \|y_n(\tau) - y_{n-1}(\tau)\| d\tau\right| \\ &\quad\le K^2\left|\int_0^t dt_1 \left|\int_0^{t_1} dt_2 \|y_{n-1}(t_2) - y_{n-2}(t_2)\|\right|\right| \\ &\quad\vdots \\ &\quad\le K^n\left|\int_0^t dt_1 \left|\int_0^{t_1} dt_2 \dots \left|\int_0^{t_{n-1}} dt_n \|y_1(t_n) - y_0(t_n)\|\right| \dots\right|\right| \\ &\quad\le K^n \|y_1(\cdot,x) - y_0(\cdot,x)\|_\infty \int_{\Delta_n(t)} d\tau \\ &\quad= \frac{K^n|t|^n}{n!}\|y_1(\cdot,x) - y_0(\cdot,x)\|_\infty\end{aligned} \tag{33.20}$$

wherein we have also made use of Lemma 32.20 (or see Lemma 33.4 BRUCE perhaps one lemma should be deleted.) Combining this estimate with

$$\|y_1(t,x) - y_0(t,x)\| = \left\|\int_0^t Z(\tau,x)d\tau\right\| \le \left|\int_0^t \|Z(\tau,x)\| d\tau\right| \le M_0,$$

where

$$M_0 = \max\left\{\int_0^{T_0} \|Z(\tau,x)\| d\tau, \int_{-T_0}^0 \|Z(\tau,x)\| d\tau\right\} \le MT_0,$$

shows

$$\|y_{n+1}(t,x) - y_n(t,x)\| \le M_0\frac{K^n|t|^n}{n!} \le M_0\frac{K^nT_0^n}{n!}$$

and this implies

$$\begin{aligned}&\sum_{n=0}^\infty \sup\{\|y_{n+1}(\cdot,x) - y_n(\cdot,x)\|_{\infty,J_0} : t \in J_0\} \\ &\qquad\le \sum_{n=0}^\infty M_0\frac{K^nT_0^n}{n!} = M_0e^{KT_0} < \infty\end{aligned}$$

where

$$\|y_{n+1}(\cdot,x) - y_n(\cdot,x)\|_{\infty,J_0} := \sup\{\|y_{n+1}(t,x) - y_n(t,x)\| : t \in J_0\}.$$

So $y(t,x) := \lim_{n\to\infty} y_n(t,x)$ exists uniformly for $t \in J$ and using Eq. (33.16) we also have

$$\sup\{ \|Z(t,y(t)) - Z(t,y_{n-1}(t))\| : t \in J_0\}$$
$$\leq K \|y(\cdot,x) - y_{n-1}(\cdot,x)\|_{\infty,J_0} \to 0 \text{ as } n \to \infty.$$

Now passing to the limit in Eq. (33.19) shows y solves Eq. (33.2). From this equation it follows that $y(t,x)$ is differentiable in t and y satisfies Eq. (33.1). The continuity of $y(t,x)$ follows from Corollary 33.3 and mean value inequality (Corollary 32.15):

$$\begin{aligned}
\|y(t,x) - y(t',x')\| &\leq \|y(t,x) - y(t,x')\| + \|y(t,x') - y(t',x')\| \\
&= \|y(t,x) - y(t,x')\| + \left\|\int_{t'}^{t} Z(\tau, y(\tau,x'))d\tau\right\| \\
&\leq \|y(t,x) - y(t,x')\| + \left|\int_{t'}^{t} \|Z(\tau, y(\tau,x'))\| \, d\tau\right| \\
&\leq \|x - x'\|e^{KT} + \left|\int_{t'}^{t} \|Z(\tau, y(\tau,x'))\| \, d\tau\right| \qquad (33.21) \\
&\leq \|x - x'\|e^{KT} + M\,|t - t'|\,.
\end{aligned}$$

The continuity of $\dot{y}(t,x)$ is now a consequence Eq. (33.1) and the continuity of y and Z. ■

Corollary 33.6. *Let $J = (a,b) \ni 0$ and suppose $Z \in C(J \times X, X)$ satisfies*

$$\|Z(t,x) - Z(t,y)\| \leq K\,\|x - y\| \text{ for all } x,y \in X \text{ and } t \in J. \qquad (33.22)$$

Then for all $x \in X$, there is a unique solution $y(t,x)$ (for $t \in J$) to Eq. (33.1). Moreover $y(t,x)$ and $\dot{y}(t,x)$ are jointly continuous in (t,x).

Proof. Let $J_0 = (a_0,b_0) \ni 0$ be a precompact subinterval of J and $Y := BC\,(J_0,X)$. By compactness, $M := \sup_{t\in\bar{J}_0} \|Z(t,0)\| < \infty$ which combined with Eq. (33.22) implies

$$\sup_{t\in\bar{J}_0} \|Z(t,x)\| \leq M + K\,\|x\| \text{ for all } x \in X.$$

Using this estimate and Lemma 32.7 one easily shows $S_x(Y) \subset Y$ for all $x \in X$. The proof of Theorem 33.5 now goes through without any further change. ■

33.4 Global Properties

Definition 33.7 (Local Lipschitz Functions). *Let $U \subset_o X$, J be an open interval and $Z \in C(J \times U, X)$. The function Z is said to be locally Lipschitz in x if for all $x \in U$ and all compact intervals $I \subset J$ there exists $K = K(x,I) < \infty$ and $\varepsilon = \varepsilon(x,I) > 0$ such that $B(x,\varepsilon(x,I)) \subset U$ and*

$$\|Z(t,x_1) - Z(t,x_0)\| \leq K(x,I)\|x_1 - x_0\| \;\forall\; x_0,x_1 \in B(x,\varepsilon(x,I)) \;\&\; t \in I. \qquad (33.23)$$

For the rest of this section, we will assume J is an open interval containing 0, U is an open subset of X and $Z \in C(J \times U, X)$ is a locally Lipschitz function.

Lemma 33.8. *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in X and E be a compact subset of U and I be a compact subset of J. Then there exists $\varepsilon > 0$ such that $Z(t,x)$ is bounded for $(t,x) \in I \times E_\varepsilon$ and and $Z(t,x)$ is K – Lipschitz on E_ε for all $t \in I$, where*

$$E_\varepsilon := \{x \in U : \operatorname{dist}(x,E) < \varepsilon\}\,.$$

Proof. Let $\varepsilon(x,I)$ and $K(x,I)$ be as in Definition 33.7 above. Since E is compact, there exists a finite subset $\Lambda \subset E$ such that $E \subset V := \cup_{x\in\Lambda} B(x,\varepsilon(x,I)/2)$. If $y \in V$, there exists $x \in \Lambda$ such that $\|y - x\| < \varepsilon(x,I)/2$ and therefore

$$\begin{aligned}
\|Z(t,y)\| &\leq \|Z(t,x)\| + K(x,I)\,\|y - x\| \leq \|Z(t,x)\| + K(x,I)\varepsilon(x,I)/2 \\
&\leq \sup_{x\in\Lambda, t\in I} \{\|Z(t,x)\| + K(x,I)\varepsilon(x,I)/2\} =: M < \infty.
\end{aligned}$$

This shows Z is bounded on $I \times V$. Let

$$\varepsilon := d(E,V^c) \leq \frac{1}{2} \min_{x\in\Lambda} \varepsilon(x,I)$$

and notice that $\varepsilon > 0$ since E is compact, V^c is closed and $E \cap V^c = \emptyset$. If $y,z \in E_\varepsilon$ and $\|y - z\| < \varepsilon$, then as before there exists $x \in \Lambda$ such that $\|y - x\| < \varepsilon(x,I)/2$. Therefore

$$\|z - x\| \leq \|z - y\| + \|y - x\| < \varepsilon + \varepsilon(x,I)/2 \leq \varepsilon(x,I)$$

and since $y,z \in B(x,\varepsilon(x,I))$, it follows that

$$\|Z(t,y) - Z(t,z)\| \leq K(x,I)\|y - z\| \leq K_0\|y - z\|$$

where $K_0 := \max_{x\in\Lambda} K(x,I) < \infty$. On the other hand if $y,z \in E_\varepsilon$ and $\|y - z\| \geq \varepsilon$, then

$$\|Z(t,y) - Z(t,z)\| \leq 2M \leq \frac{2M}{\varepsilon}\,\|y - z\|\,.$$

Thus if we let $K := \max\{2M/\varepsilon, K_0\}$, we have shown

$$\|Z(t,y) - Z(t,z)\| \leq K\|y - z\| \text{ for all } y,z \in E_\varepsilon \text{ and } t \in I.$$

■

Proposition 33.9 (Maximal Solutions). *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x and let $x \in U$ be fixed. Then there is an interval $J_x = (a(x), b(x))$ with $a \in [-\infty, 0)$ and $b \in (0, \infty]$ and a C^1–function $y : J \to U$ with the following properties:*

1. *y solves ODE in Eq. (33.1).*
2. *If $\tilde{y} : \tilde{J} = (\tilde{a}, \tilde{b}) \to U$ is another solution of Eq. (33.1) (we assume that $0 \in \tilde{J}$) then $\tilde{J} \subset J$ and $\tilde{y} = y|_{\tilde{J}}$.*

The function $y : J \to U$ is called the maximal solution to Eq. (33.1).

Proof. Suppose that $y_i : J_i = (a_i, b_i) \to U$, $i = 1, 2$, are two solutions to Eq. (33.1). We will start by showing that $y_1 = y_2$ on $J_1 \cap J_2$. To do this[1] let $J_0 = (a_0, b_0)$ be chosen so that $0 \in \bar{J}_0 \subset J_1 \cap J_2$, and let $E := y_1(\bar{J}_0) \cup y_2(\bar{J}_0)$ – a compact subset of X. Choose $\varepsilon > 0$ as in Lemma 33.8 so that Z is Lipschitz on E_ε. Then $y_1|_{J_0}, y_2|_{J_0} : J_0 \to E_\varepsilon$ both solve Eq. (33.1) and therefore are equal by Corollary 33.3. Since $J_0 = (a_0, b_0)$ was chosen arbitrarily so that $[a_0, b_0] \subset J_1 \cap J_2$, we may conclude that $y_1 = y_2$ on $J_1 \cap J_2$. Let $(y_\alpha, J_\alpha = (a_\alpha, b_\alpha))_{\alpha \in A}$ denote the possible solutions to (33.1) such that $0 \in J_\alpha$. Define $J_x = \cup J_\alpha$ and set $y = y_\alpha$ on J_α. We have just checked that y is well defined and the reader may easily check that this function $y : J_x \to U$ satisfies all the conclusions of the theorem. ∎

Notation 33.10 *For each $x \in U$, let $J_x = (a(x), b(x))$ be the maximal interval on which Eq. (33.1) may be solved, see Proposition 33.9. Set $\mathcal{D}(Z) := \cup_{x \in U} (J_x \times \{x\}) \subset J \times U$ and let $\varphi : \mathcal{D}(Z) \to U$ be defined by $\varphi(t, x) = y(t)$ where y is the maximal solution to Eq. (33.1). (So for each $x \in U$, $\varphi(\cdot, x)$ is the maximal solution to Eq. (33.1).)*

Proposition 33.11. *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x and $y : J_x = (a(x), b(x)) \to U$ be the maximal solution to Eq. (33.1). If $b(x) < b$, then either $\limsup_{t \uparrow b(x)} \|Z(t, y(t))\| = \infty$ or $y(b(x)-) := \lim_{t \uparrow b(x)} y(t)$ exists and $y(b(x)-) \notin U$. Similarly, if $a > a(x)$, then either $\limsup_{t \downarrow a(x)} \|y(t)\| = \infty$ or $y(a(x)+) := \lim_{t \downarrow a(x)} y(t)$ exists and $y(a(x)+) \notin U$.*

[1] Here is an alternate proof of the uniqueness. Let

$$T \equiv \sup\{t \in [0, \min\{b_1, b_2\}) : y_1 = y_2 \text{ on } [0, t]\}.$$

(T is the first positive time after which y_1 and y_2 disagree.)

Suppose, for sake of contradiction, that $T < \min\{b_1, b_2\}$. Notice that $y_1(T) = y_2(T) =: x'$. Applying the local uniqueness theorem to $y_1(\cdot - T)$ and $y_2(\cdot - T)$ thought as function from $(-\delta, \delta) \to B(x', \epsilon(x'))$ for some δ sufficiently small, we learn that $y_1(\cdot - T) = y_2(\cdot - T)$ on $(-\delta, \delta)$. But this shows that $y_1 = y_2$ on $[0, T + \delta)$ which contradicts the definition of T. Hence we must have the $T = \min\{b_1, b_2\}$, i.e. $y_1 = y_2$ on $J_1 \cap J_2 \cap [0, \infty)$. A similar argument shows that $y_1 = y_2$ on $J_1 \cap J_2 \cap (-\infty, 0]$ as well.

Proof. Suppose that $b < b(x)$ and $M := \limsup_{t \uparrow b(x)} \|Z(t, y(t))\| < \infty$. Then there is a $b_0 \in (0, b(x))$ such that $\|Z(t, y(t))\| \le 2M$ for all $t \in (b_0, b(x))$. Thus, by the usual fundamental theorem of calculus argument,

$$\|y(t) - y(t')\| \le \left| \int_t^{t'} \|Z(t, y(\tau))\| \, d\tau \right| \le 2M |t - t'|$$

for all $t, t' \in (b_0, b(x))$. From this it is easy to conclude that $y(b(x)-) = \lim_{t \uparrow b(x)} y(t)$ exists. If $y(b(x)-) \in U$, by the local existence Theorem 33.5, there exists $\delta > 0$ and $w \in C^1((b(x) - \delta, b(x) + \delta), U)$ such that

$$\dot{w}(t) = Z(t, w(t)) \text{ and } w(b(x)) = y(b(x)-).$$

Now define $\tilde{y} : (a, b(x) + \delta) \to U$ by

$$\tilde{y}(t) = \begin{cases} y(t) & \text{if } t \in J_x \\ w(t) & \text{if } t \in [b(x), b(x) + \delta) \end{cases}.$$

The reader may now easily show $\tilde{y}$ solves the integral Eq. (33.2) and hence also solves Eq. 33.1 for $t \in (a(x), b(x) + \delta)$.[2] But this violates the maximality of y and hence we must have that $y(b(x)-) \notin U$. The assertions for t near $a(x)$ are proved similarly. ∎

Example 33.12. Let $X = \mathbb{R}^2$, $J = \mathbb{R}$, $U = \{(x, y) \in \mathbb{R}^2 : 0 < r < 1\}$ where $r^2 = x^2 + y^2$ and

$$Z(x, y) = \frac{1}{r}(x, y) + \frac{1}{1 - r^2}(-y, x).$$

Then the unique solution $(x(t), y(t))$ to

$$\frac{d}{dt}(x(t), y(t)) = Z(x(t), y(t)) \text{ with } (x(0), y(0)) = (\frac{1}{2}, 0)$$

is given by

$$(x(t), y(t)) = \left(t + \frac{1}{2}\right) \left(\cos\left(\frac{1}{1/2 - t}\right), \sin\left(\frac{1}{1/2 - t}\right)\right)$$

for $t \in J_{(1/2, 0)} = (-1/2, 1/2)$. Notice that $\|Z(x(t), y(t))\| \to \infty$ as $t \uparrow 1/2$ and $\operatorname{dist}((x(t), y(t)), U^c) \to 0$ as $t \uparrow 1/2$.

[2] See the argument in Proposition 33.14 for a slightly different method of extending y which avoids the use of the integral equation (33.2).

Example 33.13. (Not worked out completely.) Let $X = U = \ell^2$, $\psi \in C^\infty(\mathbb{R}^2)$ be a smooth function such that $\psi = 1$ in a neighborhood of the line segment joining $(1,0)$ to $(0,1)$ and being supported within the $1/10$ – neighborhood of this segment. Choose $a_n \uparrow \infty$ and $b_n \uparrow \infty$ and define

$$Z(x) = \sum_{n=1}^{\infty} a_n \psi(b_n(x_n, x_{n+1}))(e_{n+1} - e_n). \tag{33.24}$$

For any $x \in \ell^2$, only a finite number of terms are non-zero in the above sum in a neighborhood of x. Therefor $Z : \ell^2 \to \ell^2$ is a smooth and hence locally Lipschitz vector field. Let $(y(t), J = (a,b))$ denote the maximal solution to

$$\dot{y}(t) = Z(y(t)) \text{ with } y(0) = e_1.$$

Then if the a_n and b_n are chosen appropriately, then $b < \infty$ and there will exist $t_n \uparrow b$ such that $y(t_n)$ is approximately e_n for all n. So again $y(t_n)$ does not have a limit yet $\sup_{t\in[0,b)} \|y(t)\| < \infty$. The idea is that Z is constructed to "blow" the particle from e_1 to e_2 to e_3 to e_4 etc. etc. with the time it takes to travel from e_n to e_{n+1} being on order $1/2^n$. The vector field in Eq. (33.24) is a first approximation at such a vector field, it may have to be adjusted a little more to provide an honest example. In this example, we are having problems because $y(t)$ is "going off in dimensions."

Here is another version of Proposition 33.11 which is more useful when $\dim(X) < \infty$.

Proposition 33.14. *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x and $y : J_x = (a(x), b(x)) \to U$ be the maximal solution to Eq. (33.1).*

1. *If $b(x) < b$, then for every compact subset $K \subset U$ there exists $T_K < b(x)$ such that $y(t) \notin K$ for all $t \in [T_K, b(x))$.*
2. *When $\dim(X) < \infty$, we may write this condition as: if $b(x) < b$, then either*

$$\limsup_{t\uparrow b(x)} \|y(t)\| = \infty \text{ or } \liminf_{t\uparrow b(x)} dist(y(t), U^c) = 0.$$

Proof. 1) Suppose that $b(x) < b$ and, for sake of contradiction, there exists a compact set $K \subset U$ and $t_n \uparrow b(x)$ such that $y(t_n) \in K$ for all n. Since K is compact, by passing to a subsequence if necessary, we may assume $y_\infty := \lim_{n\to\infty} y(t_n)$ exists in $K \subset U$. By the local existence Theorem 33.5, there exists $T_0 > 0$ and $\delta > 0$ such that for each $x' \in B(y_\infty, \delta)$ there exists a unique solution $w(\cdot, x') \in C^1((-T_0, T_0), U)$ solving

$$w(t, x') = Z(t, w(t, x')) \text{ and } w(0, x') = x'.$$

Now choose n sufficiently large so that $t_n \in (b(x) - T_0/2, b(x))$ and $y(t_n) \in B(y_\infty, \delta)$. Define $\tilde{y} : (a(x), b(x) + T_0/2) \to U$ by

$$\tilde{y}(t) = \begin{cases} y(t) & \text{if } t \in J_x \\ w(t - t_n, y(t_n)) & \text{if } t \in (t_n - T_0, b(x) + T_0/2). \end{cases}$$

wherein we have used $(t_n - T_0, b(x) + T_0/2) \subset (t_n - T_0, t_n + T_0)$. By uniqueness of solutions to ODE's $\tilde{y}$ is well defined, $\tilde{y} \in C^1((a(x), b(x) + T_0/2), X)$ and $\tilde{y}$ solves the ODE in Eq. 33.1. But this violates the maximality of y.

2) For each $n \in \mathbb{N}$ let

$$K_n := \{x \in U : \|x\| \le n \text{ and } \operatorname{dist}(x, U^c) \ge 1/n\}.$$

Then $K_n \uparrow U$ and each K_n is a closed bounded set and hence compact if $\dim(X) < \infty$. Therefore if $b(x) < b$, by item 1., there exists $T_n \in [0, b(x))$ such that $y(t) \notin K_n$ for all $t \in [T_n, b(x))$ or equivalently $\|y(t)\| > n$ or $\operatorname{dist}(y(t), U^c) < 1/n$ for all $t \in [T_n, b(x))$. ∎

Remark 33.15 (This remark is still rather rough.). In general it is **not** true that the functions a and b are continuous. For example, let U be the region in $\mathbb{R}^2$ described in polar coordinates by $r > 0$ and $0 < \theta < 3\pi/2$ and $Z(x,y) = (0,-1)$ as in Figure 33.2 below. Then $b(x,y) = y$ for all $x \ge 0$ and $y > 0$ while $b(x,y) = \infty$ for all $x < 0$ and $y \in \mathbb{R}$ which shows b is discontinuous. On the other hand notice that

$$\{b > t\} = \{x < 0\} \cup \{(x,y) : x \ge 0, y > t\}$$

is an open set for all $t > 0$. An example of a vector field for which $b(x)$ is discontinuous is given in the top left hand corner of Figure 33.2. The map $\psi(r(\cos\theta, \sin\theta)) := \left(\ln r, \tan\left(\frac{2}{3}\theta - \frac{\pi}{2}\right)\right)$, would allow the reader to find an example on $\mathbb{R}^2$ if so desired. Some calculations shows that Z transferred to $\mathbb{R}^2$ by the map ψ is given by the new vector

$$\tilde{Z}(x,y) = -e^{-x}\left(\sin\left(\frac{3\pi}{8} + \frac{3}{4}\tan^{-1}(y)\right), \cos\left(\frac{3\pi}{8} + \frac{3}{4}\tan^{-1}(y)\right)\right).$$

(Bruce: Check this!)

Theorem 33.16 (Global Continuity). *Let $Z \in C(J \times U, X)$ be a locally Lipschitz function in x. Then $\mathcal{D}(Z)$ is an open subset of $J \times U$ and the functions $\varphi : \mathcal{D}(Z) \to U$ and $\varphi : \mathcal{D}(Z) \to U$ are continuous. More precisely, for all $x_0 \in U$ and all open intervals J_0 such that $0 \in J_0 \sqsubset\sqsubset J_{x_0}$ there exists $\delta = \delta(x_0, J_0, Z) > 0$ and $C = C(x_0, J_0, Z) < \infty$ such that for all $x \in B(x_0, \delta)$, $J_0 \subset J_x$ and*

$$\|\varphi(\cdot, x) - \varphi(\cdot, x_0)\|_{BC(J_0, U)} \le C \|x - x_0\|. \tag{33.25}$$

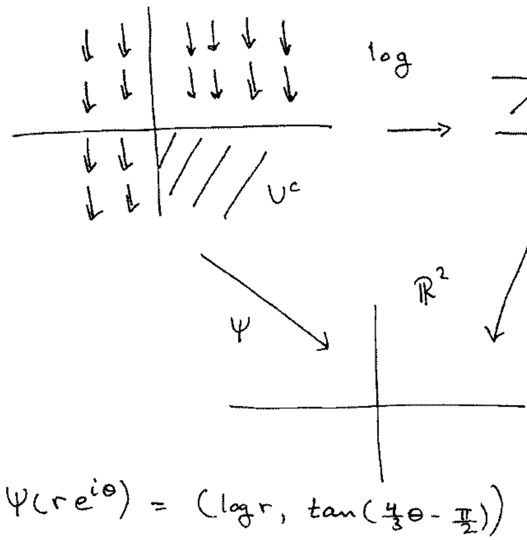


Fig. 33.2. Manufacturing vector fields where $b(x)$ is discontinuous.

Proof. Let $|J_0| = b_0 - a_0$, $I = \bar{J}_0$ and $E := y(\bar{J}_0)$ – a compact subset of U and let $\varepsilon > 0$ and $K < \infty$ be given as in Lemma 33.8, i.e. K is the Lipschitz constant for Z on E_ε. Also recall the notation: $\Delta_1(t) = [0, t]$ if $t > 0$ and $\Delta_1(t) = [t, 0]$ if $t < 0$. Suppose that $x \in E_\varepsilon$, then by Corollary 33.3,

$$\|\varphi(t, x) - \varphi(t, x_0)\| \le \|x - x_0\| e^{K|t|} \le \|x - x_0\| e^{K|J_0|} \tag{33.26}$$

for all $t \in J_0 \cap J_x$ such that such that $\varphi(\Delta_1(t), x) \subset E_\varepsilon$. Letting $\delta := \varepsilon e^{-K|J_0|}/2$, and assuming $x \in B(x_0, \delta)$, the previous equation implies

$$\|\varphi(t, x) - \varphi(t, x_0)\| \le \varepsilon/2 < \varepsilon \;\forall\; t \in J_0 \cap J_x \;\ni\; \varphi(\Delta_1(t), x) \subset E_\varepsilon.$$

This estimate further shows that $\varphi(t, x)$ remains bounded and strictly away from the boundary of U for all such t. Therefore, it follows from Proposition 33.9 and "continuous induction[3]" that $J_0 \subset J_x$ and Eq. (33.26) is valid for all $t \in J_0$. This proves Eq. (33.25) with $C := e^{K|J_0|}$. Suppose that $(t_0, x_0) \in \mathcal{D}(Z)$ and let $0 \in J_0 \sqsubset\sqsubset J_{x_0}$ such that $t_0 \in J_0$ and δ be as above. Then we have just shown $J_0 \times B(x_0, \delta) \subset \mathcal{D}(Z)$ which proves $\mathcal{D}(Z)$ is open. Furthermore, since the evaluation map

$$(t_0, y) \in J_0 \times BC(J_0, U) \xrightarrow{e} y(t_0) \in X$$

is continuous (as the reader should check) it follows that $\varphi = e \circ (x \to \varphi(\cdot, x)) : J_0 \times B(x_0, \delta) \to U$ is also continuous; being the composition of continuous maps. The continuity of $\varphi(t_0, x)$ is a consequence of the continuity of φ and the differential equation 33.1 Alternatively using Eq. (33.2),

$$\begin{aligned} \|\varphi(t_0, x) - \varphi(t, x_0)\| &\le \|\varphi(t_0, x) - \varphi(t_0, x_0)\| + \|\varphi(t_0, x_0) - \varphi(t, x_0)\| \\ &\le C\|x - x_0\| + \left| \int_t^{t_0} \|Z(\tau, \varphi(\tau, x_0))\| \, d\tau \right| \\ &\le C\|x - x_0\| + M|t_0 - t| \end{aligned}$$

where C is the constant in Eq. (33.25) and $M = \sup_{\tau \in J_0} \|Z(\tau, \varphi(\tau, x_0))\| < \infty$. This clearly shows φ is continuous. ■

[3] See the argument in the proof of Proposition 32.12.

33.5 Semi-Group Properties of time independent flows

To end this chapter we investigate the semi-group property of the flow associated to the vector-field Z. It will be convenient to introduce the following suggestive notation. For $(t, x) \in \mathcal{D}(Z)$, set $e^{tZ}(x) = \varphi(t, x)$. So the path $t \to e^{tZ}(x)$ is the maximal solution to

$$\frac{d}{dt} e^{tZ}(x) = Z(e^{tZ}(x)) \quad \text{with } e^{0Z}(x) = x.$$

This exponential notation will be justified shortly. It is convenient to have the following conventions.

Notation 33.17 *We write $f : X \to X$ to mean a function defined on some open subset $D(f) \subset X$. The open set $D(f)$ will be called the domain of f. Given two functions $f : X \to X$ and $g : X \to X$ with domains $D(f)$ and $D(g)$ respectively, we define the composite function $f \circ g : X \to X$ to be the function with domain*

$$D(f \circ g) = \{x \in X : x \in D(g) \;\text{ and } g(x) \in D(f)\} = g^{-1}(D(f))$$

given by the rule $f \circ g(x) = f(g(x))$ for all $x \in D(f \circ g)$. We now write $f = g$ iff $D(f) = D(g)$ and $f(x) = g(x)$ for all $x \in D(f) = D(g)$. We will also write $f \subset g$ iff $D(f) \subset D(g)$ and $g|_{D(f)} = f$.

Theorem 33.18. *For fixed $t \in \mathbb{R}$ we consider e^{tZ} as a function from X to X with domain $D(e^{tZ}) = \{x \in U : (t, x) \in \mathcal{D}(Z)\}$, where $D(\varphi) = \mathcal{D}(Z) \subset \mathbb{R} \times U$, $\mathcal{D}(Z)$ and φ are defined in Notation 33.10. Conclusions:*

1. *If $t, s \in \mathbb{R}$ and $t \cdot s \ge 0$, then $e^{tZ} \circ e^{sZ} = e^{(t+s)Z}$.*
2. *If $t \in \mathbb{R}$, then $e^{tZ} \circ e^{-tZ} = Id_{D(e^{-tZ})}$.*

3. *For arbitrary* $t, s \in \mathbb{R}$, $e^{tZ} \circ e^{sZ} \subset e^{(t+s)Z}$.

Proof. Item 1. For simplicity assume that $t, s \geq 0$. The case $t, s \leq 0$ is left to the reader. Suppose that $x \in D(e^{tZ} \circ e^{sZ})$. Then by assumption $x \in D(e^{sZ})$ and $e^{sZ}(x) \in D(e^{tZ})$. Define the path $y(\tau)$ via:

$$y(\tau) = \begin{cases} e^{\tau Z}(x) & \text{if } 0 \leq \tau \leq s \\ e^{(\tau - s)Z}(x) & \text{if } s \leq \tau \leq t+s \end{cases}.$$

It is easy to check that y solves $\dot{y}(\tau) = Z(y(\tau))$ with $y(0) = x$. But since, $e^{\tau Z}(x)$ is the maximal solution we must have that $x \in D(e^{(t+s)Z})$ and $y(t+s) = e^{(t+s)Z}(x)$. That is $e^{(t+s)Z}(x) = e^{tZ} \circ e^{sZ}(x)$. Hence we have shown that $e^{tZ} \circ e^{sZ} \subset e^{(t+s)Z}$. To finish the proof of item 1. it suffices to show that $D(e^{(t+s)Z}) \subset D(e^{tZ} \circ e^{sZ})$. Take $x \in D(e^{(t+s)Z})$, then clearly $x \in D(e^{sZ})$. Set $y(\tau) = e^{(\tau+s)Z}(x)$ defined for $0 \leq \tau \leq t$. Then y solves

$$\dot{y}(\tau) = Z(y(\tau)) \text{ with } y(0) = e^{sZ}(x).$$

But since $\tau \to e^{\tau Z}(e^{sZ}(x))$ is the maximal solution to the above initial valued problem we must have that $y(\tau) = e^{\tau Z}(e^{sZ}(x))$, and in particular at $\tau = t$, $e^{(t+s)Z}(x) = e^{tZ}(e^{sZ}(x))$. This shows that $x \in D(e^{tZ} \circ e^{sZ})$ and in fact $e^{(t+s)Z} \subset e^{tZ} \circ e^{sZ}$.

Item 2. Let $x \in D(e^{-tZ})$ – again assume for simplicity that $t \geq 0$. Set $y(\tau) = e^{(\tau - t)Z}(x)$ defined for $0 \leq \tau \leq t$. Notice that $y(0) = e^{-tZ}(x)$ and $\dot{y}(\tau) = Z(y(\tau))$. This shows that $y(\tau) = e^{\tau Z}(e^{-tZ}(x))$ and in particular that $x \in D(e^{tZ} \circ e^{-tZ})$ and $e^{tZ} \circ e^{-tZ}(x) = x$. This proves item 2.

Item 3. I will only consider the case that $s < 0$ and $t + s \geq 0$, the other cases are handled similarly. Write u for $t + s$, so that $t = -s + u$. We know that $e^{tZ} = e^{uZ} \circ e^{-sZ}$ by item 1. Therefore

$$e^{tZ} \circ e^{sZ} = (e^{uZ} \circ e^{-sZ}) \circ e^{sZ}.$$

Notice in general, one has $(f \circ g) \circ h = f \circ (g \circ h)$ (you prove). Hence, the above displayed equation and item 2. imply that

$$e^{tZ} \circ e^{sZ} = e^{uZ} \circ (e^{-sZ} \circ e^{sZ}) = e^{(t+s)Z} \circ I_{D(e^{sZ})} \subset e^{(t+s)Z}.$$

■

The following result is trivial but conceptually illuminating partial converse to Theorem 33.18.

Proposition 33.19 (Flows and Complete Vector Fields). *Suppose* $U \subset_o X$, $\varphi \in C(\mathbb{R} \times U, U)$ *and* $\varphi_t(x) = \varphi(t, x)$. *Suppose* φ *satisfies:*

1. $\varphi_0 = I_U$,
2. $\varphi_t \circ \varphi_s = \varphi_{t+s}$ *for all* $t, s \in \mathbb{R}$, *and*
3. $Z(x) := \dot{\varphi}(0, x)$ *exists for all* $x \in U$ *and* $Z \in C(U, X)$ *is locally Lipschitz.*

Then $\varphi_t = e^{tZ}$.

Proof. Let $x \in U$ and $y(t) := \varphi_t(x)$. Then using Item 2.,

$$\dot{y}(t) = \frac{d}{ds}|_0 y(t+s) = \frac{d}{ds}|_0 \varphi_{(t+s)}(x) = \frac{d}{ds}|_0 \varphi_s \circ \varphi_t(x) = Z(y(t)).$$

Since $y(0) = x$ by Item 1. and Z is locally Lipschitz by Item 3., we know by uniqueness of solutions to ODE's (Corollary 33.3) that $\varphi_t(x) = y(t) = e^{tZ}(x)$. ■

33.6 Exercises

Exercise 33.1. Find a vector field Z such that $e^{(t+s)Z}$ is not contained in $e^{tZ} \circ e^{sZ}$.

Definition 33.20. *A locally Lipschitz function* $Z : U \subset_o X \to X$ *is said to be a complete vector field if* $\mathcal{D}(Z) = \mathbb{R} \times U$. *That is for any* $x \in U$, $t \to e^{tZ}(x)$ *is defined for all* $t \in \mathbb{R}$.

Exercise 33.2. Suppose that $Z : X \to X$ is a locally Lipschitz function. Assume there is a constant $C > 0$ such that

$$\|Z(x)\| \leq C(1 + \|x\|) \text{ for all } x \in X.$$

Then Z is complete. **Hint:** use Gronwall's Lemma 33.2 and Proposition 33.11.

Exercise 33.3. Suppose y is a solution to $\dot{y}(t) = |y(t)|^{1/2}$ with $y(0) = 0$. Show there exists $a, b \in [0, \infty]$ such that

$$y(t) = \begin{cases} \frac{1}{4}(t-b)^2 & \text{if } \quad t \geq b \\ 0 & \text{if } -a < t < b \\ -\frac{1}{4}(t+a)^2 & \text{if } \quad t \leq -a. \end{cases}$$

Exercise 33.4. Using the fact that the solutions to Eq. (33.3) are never 0 if $x \neq 0$, show that $y(t) = 0$ is the only solution to Eq. (33.3) with $y(0) = 0$.

Exercise 33.5 (Higher Order ODE). Let X be a Banach space, , $\mathcal{U} \subset_o X^n$ and $f \in C(J \times \mathcal{U}, X)$ be a Locally Lipschitz function in $\mathbf{x} = (x_1, \ldots, x_n)$. Show the n^{th} ordinary differential equation,

$$y^{(n)}(t) = f(t, y(t), \dot{y}(t), \ldots, y^{(n-1)}(t)) \text{ with } y^{(k)}(0) = y_0^k \text{ for } k < n \quad (33.27)$$

where $(y_0^0, \dots, y_0^{n-1})$ is given in $\mathcal{U}$, has a unique solution for small $t \in J$. **Hint**: let $\mathbf{y}(t) = \left(y(t), \dot{y}(t), \dots, y^{(n-1)}(t)\right)$ and rewrite Eq. (33.27) as a first order ODE of the form

$$\dot{\mathbf{y}}(t) = Z(t, \mathbf{y}(t)) \text{ with } \mathbf{y}(0) = (y_0^0, \dots, y_0^{n-1}).$$

Exercise 33.6. Use the results of Exercises 32.19 and 33.5 to solve

$$\ddot{y}(t) - 2\dot{y}(t) + y(t) = 0 \text{ with } y(0) = a \text{ and } \dot{y}(0) = b.$$

Hint: The 2×2 matrix associated to this system, A, has only one eigenvalue 1 and may be written as $A = I + B$ where $B^2 = 0$.

Exercise 33.7 (Non-Homogeneous ODE). Suppose that $U \subset_o X$ is open and $Z : \mathbb{R} \times U \to X$ is a continuous function. Let $J = (a, b)$ be an interval and $t_0 \in J$. Suppose that $y \in C^1(J, U)$ is a solution to the "non-homogeneous" differential equation:

$$\dot{y}(t) = Z(t, y(t)) \text{ with } y(t_o) = x \in U. \tag{33.28}$$

Define $Y \in C^1(J - t_0, \mathbb{R} \times U)$ by $Y(t) := (t + t_0, y(t + t_0))$. Show that Y solves the "homogeneous" differential equation

$$\dot{Y}(t) = \tilde{Z}(Y(t)) \text{ with } Y(0) = (t_0, y_0), \tag{33.29}$$

where $\tilde{Z}(t, x) := (1, Z(x))$. Conversely, suppose that $Y \in C^1(J - t_0, \mathbb{R} \times U)$ is a solution to Eq. (33.29). Show that $Y(t) = (t + t_0, y(t + t_0))$ for some $y \in C^1(J, U)$ satisfying Eq. (33.28). (In this way the theory of non-homogeneous O.D.E.'s may be reduced to the theory of homogeneous O.D.E.'s.)

Exercise 33.8 (Differential Equations with Parameters). Let W be another Banach space, $U \times V \subset_o X \times W$ and $Z \in C(U \times V, X)$ be a locally Lipschitz function on $U \times V$. For each $(x, w) \in U \times V$, let $t \in J_{x,w} \to \varphi(t, x, w)$ denote the maximal solution to the ODE

$$\dot{y}(t) = Z(y(t), w) \text{ with } y(0) = x. \tag{33.30}$$

Prove

$$\mathcal{D} := \{(t, x, w) \in \mathbb{R} \times U \times V : t \in J_{x,w}\} \tag{33.31}$$

is open in $\mathbb{R} \times U \times V$ and φ and $\dot{\varphi}$ are continuous functions on $\mathcal{D}$.

Hint: If $y(t)$ solves the differential equation in (33.30), then $v(t) := (y(t), w)$ solves the differential equation,

$$\dot{v}(t) = \tilde{Z}(v(t)) \text{ with } v(0) = (x, w), \tag{33.32}$$

where $\tilde{Z}(x, w) := (Z(x, w), 0) \in X \times W$ and let $\psi(t, (x, w)) := v(t)$. Now apply the Theorem 33.16 to the differential equation (33.32).

Exercise 33.9 (Abstract Wave Equation). For $A \in L(X)$ and $t \in \mathbb{R}$, let

$$\cos(tA) := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n} A^{2n} \text{ and}$$
$$\frac{\sin(tA)}{A} := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1} A^{2n}.$$

Show that the unique solution $y \in C^2(\mathbb{R}, X)$ to

$$\ddot{y}(t) + A^2 y(t) = 0 \text{ with } y(0) = y_0 \text{ and } \dot{y}(0) = \dot{y}_0 \in X \tag{33.33}$$

is given by

$$y(t) = \cos(tA) y_0 + \frac{\sin(tA)}{A} \dot{y}_0.$$

Remark 33.21. Exercise 33.9 can be done by direct verification. Alternatively and more instructively, rewrite Eq. (33.33) as a first order ODE using Exercise 33.5. In doing so you will be lead to compute e^{tB} where $B \in L(X \times X)$ is given by

$$B = \begin{pmatrix} 0 & I \\ -A^2 & 0 \end{pmatrix},$$

where we are writing elements of $X \times X$ as column vectors, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. You should then show

$$e^{tB} = \begin{pmatrix} \cos(tA) & \frac{\sin(tA)}{A} \\ -A\sin(tA) & \cos(tA) \end{pmatrix}$$

where

$$A \sin(tA) := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1} A^{2(n+1)}.$$

Exercise 33.10 (Duhamel's Principle for the Abstract Wave Equation). Continue the notation in Exercise 33.9, but now consider the ODE,

$$\ddot{y}(t) + A^2 y(t) = f(t) \text{ with } y(0) = y_0 \text{ and } \dot{y}(0) = \dot{y}_0 \in X \tag{33.34}$$

where $f \in C(\mathbb{R}, X)$. Show the unique solution to Eq. (33.34) is given by

$$y(t) = \cos(tA) y_0 + \frac{\sin(tA)}{A} \dot{y}_0 + \int_0^t \frac{\sin((t - \tau) A)}{A} f(\tau) d\tau \tag{33.35}$$

Hint: Again this could be proved by direct calculation. However it is more instructive to deduce Eq. (33.35) from Exercise 32.21 and the comments in Remark 33.21.

Banach Space Calculus

In this section, X and Y will be Banach space and U will be an open subset of X.

Notation 34.1 (ε, O, and o notation) *Let $0 \in U \subset_o X$, and $f : U \to Y$ be a function. We will write:*

1. $f(x) = \varepsilon(x)$ *if* $\lim_{x\to 0} \|f(x)\| = 0$.
2. $f(x) = O(x)$ *if there are constants $C < \infty$ and $r > 0$ such that $\|f(x)\| \le C\|x\|$ for all $x \in B(0,r)$. This is equivalent to the condition that* $\limsup_{x\to 0} \left(\|x\|^{-1}\|f(x)\|\right) < \infty$, *where*
$$\limsup_{x\to 0} \frac{\|f(x)\|}{\|x\|} := \lim_{r\downarrow 0} \sup\{\|f(x)\| : 0 < \|x\| \le r\}.$$
3. $f(x) = o(x)$ *if* $f(x) = \varepsilon(x)O(x)$, *i.e.* $\lim_{x\to 0} \|f(x)\|/\|x\| = 0$.

Example 34.2. Here are some examples of properties of these symbols.

1. A function $f : U \subset_o X \to Y$ is continuous at $x_0 \in U$ if $f(x_0 + h) = f(x_0) + \varepsilon(h)$.
2. If $f(x) = \varepsilon(x)$ and $g(x) = \varepsilon(x)$ then $f(x) + g(x) = \varepsilon(x)$.
   Now let $g : Y \to Z$ be another function where Z is another Banach space.
3. If $f(x) = O(x)$ and $g(y) = o(y)$ then $g \circ f(x) = o(x)$.
4. If $f(x) = \varepsilon(x)$ and $g(y) = \varepsilon(y)$ then $g \circ f(x) = \varepsilon(x)$.

34.1 The Differential

Definition 34.3. *A function $f : U \subset_o X \to Y$ is **differentiable** at $x_0 \in U$ if there exists a linear transformation $\Lambda \in L(X,Y)$ such that*
$$f(x_0 + h) - f(x_0) - \Lambda h = o(h). \tag{34.1}$$
*We denote Λ by $f'(x_0)$ or $Df(x_0)$ if it exists. As with continuity, f is **differentiable on** U if f is differentiable at all points in U.*

Remark 34.4. The linear transformation Λ in Definition 34.3 is necessarily unique. Indeed if Λ_1 is another linear transformation such that Eq. (34.1) holds with Λ replaced by Λ_1, then
$$(\Lambda - \Lambda_1)h = o(h),$$
i.e.
$$\limsup_{h\to 0} \frac{\|(\Lambda - \Lambda_1)h\|}{\|h\|} = 0.$$
On the other hand, by definition of the operator norm,
$$\limsup_{h\to 0} \frac{\|(\Lambda - \Lambda_1)h\|}{\|h\|} = \|\Lambda - \Lambda_1\|.$$
The last two equations show that $\Lambda = \Lambda_1$.

Exercise 34.1. Show that a function $f : (a,b) \to X$ is a differentiable at $t \in (a,b)$ in the sense of Definition 32.9 iff it is differentiable in the sense of Definition 34.3. Also show $Df(t)v = v\dot{f}(t)$ for all $v \in \mathbb{R}$.

Example 34.5. If $T \in L(X,Y)$ and $x, h \in X$, then
$$T(x + h) - T(x) - Th = 0$$
which shows $T'(x) = T$ for all $x \in X$.

Example 34.6. Assume that $GL(X,Y)$ is non-empty. Then by Corollary 14.27, $GL(X,Y)$ is an open subset of $L(X,Y)$ and the inverse map $f : GL(X,Y) \to GL(Y,X)$, defined by $f(A) := A^{-1}$, is continuous. We will now show that f is differentiable and
$$f'(A)B = -A^{-1}BA^{-1} \text{ for all } B \in L(X,Y).$$
This is a consequence of the identity,
$$f(A+H) - f(A) = (A+H)^{-1}(A - (A+H))A^{-1} = -(A+H)^{-1}HA^{-1}$$
which may be used to find the estimate,

$$\begin{aligned}\left\|f(A+H)-f(A)+A^{-1}HA^{-1}\right\| &= \left\|\left[A^{-1}-(A+H)^{-1}\right]HA^{-1}\right\| \\ &\le \left\|A^{-1}-(A+H)^{-1}\right\|\|H\|\left\|A^{-1}\right\| \\ &\le \frac{\|A^{-1}\|^3\|H\|^2}{1-\|A^{-1}\|\|H\|} = O\left(\|H\|^2\right)\end{aligned}$$

wherein we have used the bound in Eq. (14.13) of Corollary 14.27 for the last inequality.

34.2 Product and Chain Rules

The following theorem summarizes some basic properties of the differential.

Theorem 34.7. *The differential D has the following properties:*

1. ***Linearity:*** *D is linear, i.e. $D(f+\lambda g) = Df + \lambda Dg$.*
2. ***Product Rule:*** *If $f : U \subset_o X \to Y$ and $A : U \subset_o X \to L(Y,Z)$ are differentiable at x_0 then so is $x \to (Af)(x) := A(x)f(x)$ and*
$$D(Af)(x_0)h = (DA(x_0)h)f(x_0) + A(x_0)Df(x_0)h.$$
3. ***Chain Rule:*** *If $f : U \subset_o X \to V \subset_o Y$ is differentiable at $x_0 \in U$, and $g : V \subset_o Y \to Z$ is differentiable at $y_0 := f(x_0)$, then $g \circ f$ is differentiable at x_0 and $(g\circ f)'(x_0) = g'(y_0)f'(x_0)$.*
4. ***Converse Chain Rule:*** *Suppose that $f : U \subset_o X \to V \subset_o Y$ is* ***continuous*** *at $x_0 \in U$, $g : V \subset_o Y \to Z$ is differentiable $y_0 := f(h_o)$, $g'(y_0)$ is invertible, and $g\circ f$ is differentiable at x_0, then f is differentiable at x_0 and*
$$f'(x_0) := [g'(y_0)]^{-1}(g\circ f)'(x_0). \tag{34.2}$$

Proof. Linearity. Let $f, g : U \subset_o X \to Y$ be two functions which are differentiable at $x_0 \in U$ and $\lambda \in \mathbb{R}$, then

$$\begin{aligned}&(f+\lambda g)(x_0+h) \\ &\quad= f(x_0) + Df(x_0)h + o(h) + \lambda(g(x_0) + Dg(x_0)h + o(h) \\ &\quad= (f+\lambda g)(x_0) + (Df(x_0) + \lambda Dg(x_0))h + o(h),\end{aligned}$$

which implies that $(f+\lambda g)$ is differentiable at x_0 and that

$$D(f+\lambda g)(x_0) = Df(x_0) + \lambda Dg(x_0).$$

Product Rule. The computation,

$$\begin{aligned}&A(x_0+h)f(x_0+h) \\ &\quad= (A(x_0) + DA(x_0)h + o(h))(f(x_0) + f'(x_0)h + o(h)) \\ &\quad= A(x_0)f(x_0) + A(x_0)f'(x_0)h + [DA(x_0)h]f(x_0) + o(h),\end{aligned}$$

verifies the product rule holds. This may also be considered as a special case of Proposition 34.9.

Chain Rule. Using $f(x_0+h)-f(x_0) = O(h)$ (see Eq. (34.1)) and $o(O(h)) = o(h)$,

$$\begin{aligned}&(g\circ f)(x_0+h) \\ &\quad= g(f(x_0)) + g'(f(x_0))(f(x_0+h) - f(x_0)) + o(f(x_0+h) - f(x_0)) \\ &\quad= g(f(x_0)) + g'(f(x_0))(Df(x_0)x_0 + o(h)) + o(f(x_0+h) - f(x_0) \\ &\quad= g(f(x_0)) + g'(f(x_0))Df(x_0)h + o(h).\end{aligned}$$

Converse Chain Rule. Since g is differentiable at $y_0 = f(x_0)$ and $g'(y_0)$ is invertible,

$$\begin{aligned}&g(f(x_0+h)) - g(f(x_0)) \\ &\quad= g'(f(x_0))(f(x_0+h) - f(x_0)) + o(f(x_0+h) - f(x_0)) \\ &\quad= g'(f(x_0))\left[f(x_0+h) - f(x_0) + o(f(x_0+h) - f(x_0))\right].\end{aligned}$$

And since $g\circ f$ is differentiable at x_0,

$$(g\circ f)(x_0+h) - g(f(x_0)) = (g\circ f)'(x_0)h + o(h).$$

Comparing these two equations shows that

$$\begin{aligned}&f(x_0+h) - f(x_0) + o(f(x_0+h) - f(x_0)) \\ &\quad= g'(f(x_0))^{-1}\left[(g\circ f)'(x_0)h + o(h)\right]\end{aligned}$$

which is equivalent to

$$\begin{aligned}&f(x_0+h) - f(x_0) + o(f(x_0+h) - f(x_0)) \\ &\quad= g'(f(x_0))^{-1}\left[(g\circ f)'(x_0)h + o(h)\right] \\ &\quad= g'(f(x_0))^{-1}\{(g\circ f)'(x_0)h + o(h) - o(f(x_0+h) - f(x_0))\} \\ &\quad= g'(f(x_0))^{-1}(g\circ f)'(x_0)h + o(h) + o(f(x_0+h) - f(x_0)).\end{aligned} \tag{34.3}$$

Using the continuity of f, $f(x_0+h) - f(x_0)$ is close to 0 if h is close to zero, and hence

$$\|o(f(x_0+h) - f(x_0))\| \le \frac{1}{2}\|f(x_0+h) - f(x_0)\| \tag{34.4}$$

for all h sufficiently close to 0. (We may replace $\frac{1}{2}$ by any number $\alpha > 0$ above.) Taking the norm of both sides of Eq. (34.3) and making use of Eq. (34.4) shows, for h close to 0, that

$$\|f(x_0+h)-f(x_0)\| \\ \le \|g'(f(x_0))^{-1}(g\circ f)'(x_0)\|\|h\| + o(\|h\|) + \frac{1}{2}\|f(x_0+h)-f(x_0)\|.$$

Solving for $\|f(x_0+h)-f(x_0)\|$ in this last equation shows that

$$f(x_0+h)-f(x_0)=O(h). \tag{34.5}$$

(This is an improvement, since the continuity of f only guaranteed that $f(x_0+h)-f(x_0)=\varepsilon(h)$.) Because of Eq. (34.5), we now know that $o(f(x_0+h)-f(x_0))=o(h)$, which combined with Eq. (34.3) shows that

$$f(x_0+h)-f(x_0)=g'(f(x_0))^{-1}(g\circ f)'(x_0)h+o(h),$$

i.e. f is differentiable at x_0 and $f'(x_0)=g'(f(x_0))^{-1}(g\circ f)'(x_0)$. ■

Corollary 34.8 (Chain Rule). *Suppose that $\sigma:(a,b)\to U\subset_o X$ is differentiable at $t\in(a,b)$ and $f:U\subset_o X\to Y$ is differentiable at $\sigma(t)\in U$. Then $f\circ\sigma$ is differentiable at t and*

$$d(f\circ\sigma)(t)/dt=f'(\sigma(t))\dot{\sigma}(t).$$

Proposition 34.9 (Product Rule II). *Suppose that $X:=X_1\times\cdots\times X_n$ with each X_i being a Banach space and $T:X_1\times\cdots\times X_n\to Y$ is a multilinear map, i.e.*

$$x_i\in X_i\to T(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n)\in Y$$

is linear when $x_1,\dots,x_{i-1},x_{i+1},\dots,x_n$ are held fixed. Then the following are equivalent:

1. *T is continuous.*
2. *T is continuous at $0\in X$.*
3. *There exists a constant $C<\infty$ such that*

$$\|T(x)\|_Y\le C\prod_{i=1}^n\|x_i\|_{X_i} \tag{34.6}$$

   for all $x=(x_1,\dots,x_n)\in X$.
4. *T is differentiable at all $x\in X_1\times\cdots\times X_n$.*

Moreover if T the differential of T is given by

$$T'(x)h=\sum_{i=1}^n T(x_1,\dots,x_{i-1},h_i,x_{i+1},\dots,x_n) \tag{34.7}$$

where $h=(h_1,\dots,h_n)\in X$.

Proof. Let us equip X with the norm

$$\|x\|_X:=\max\left\{\|x_i\|_{X_i}\right\}.$$

If T is continuous then T is continuous at 0. If T is continuous at 0, using $T(0)=0$, there exists a $\delta>0$ such that $\|T(x)\|_Y\le 1$ whenever $\|x\|_X\le\delta$. Now if $x\in X$ is arbitrary, let $x':=\delta\left(\|x_1\|_{X_1}^{-1}x_1,\dots,\|x_n\|_{X_n}^{-1}x_n\right)$. Then $\|x'\|_X\le\delta$ and hence

$$\left\|\left(\delta^n\prod_{i=1}^n\|x_i\|_{X_i}^{-1}\right)T(x_1,\dots,x_n)\right\|_Y=\|T(x')\|\le 1$$

from which Eq. (34.6) follows with $C=\delta^{-n}$.

Now suppose that Eq. (34.6) holds. For $x,h\in X$ and $\varepsilon\in\{0,1\}^n$ let $|\varepsilon|=\sum_{i=1}^n\varepsilon_i$ and

$$x^\varepsilon(h):=((1-\varepsilon_1)x_1+\varepsilon_1h_1,\dots,(1-\varepsilon_n)x_n+\varepsilon_nh_n)\in X.$$

By the multi-linearity of T,

$$\begin{aligned}T(x+h)=T(x_1+h_1,\dots,x_n+h_n)&=\sum_{\varepsilon\in\{0,1\}^n}T(x^\varepsilon(h))\\&=T(x)+\sum_{i=1}^n T(x_1,\dots,x_{i-1},h_i,x_{i+1},\dots,x_n)\\&\quad+\sum_{\varepsilon\in\{0,1\}^n:|\varepsilon|\ge 2}T(x^\varepsilon(h)).\end{aligned} \tag{34.8}$$

From Eq. (34.6),

$$\left\|\sum_{\varepsilon\in\{0,1\}^n:|\varepsilon|\ge 2}T(x^\varepsilon(h))\right\|=O\left(\|h\|^2\right),$$

and so it follows from Eq. (34.8) that $T'(x)$ exists and is given by Eq. (34.7). This completes the proof since it is trivial to check that T being differentiable at $x\in X$ implies continuity of T at $x\in X$. ■

Exercise 34.2. Let $\det:L(\mathbb{R}^n)\to\mathbb{R}$ be the determinant function on $n\times n$ matrices and for $A\in L(\mathbb{R}^n)$ we will let A_i denote the i^{th} – column of A and write $A=(A_1|A_2|\dots|A_n)$.

1. Show $\det'(A)$ exists for all $A\in L(\mathbb{R}^n)$ and

$$\det{}'(A)\,H = \sum_{i=1}^{n} \det(A_1|\dots|A_{i-1}|H_i|A_{i+1}|\dots|A_n) \tag{34.9}$$

for all $H \in L(\mathbb{R}^n)$. **Hint**: recall that $\det(A)$ is a multilinear function of its columns.

2. Use Eq. (34.9) along with basic properties of the determinant to show $\det'(I)\,H = \mathrm{tr}(H)$.
3. Suppose now that $A \in GL(\mathbb{R}^n)$, show

$$\det{}'(A)\,H = \det(A)\,\mathrm{tr}(A^{-1}H).$$

Hint: Notice that $\det(A+H) = \det(A)\det\left(I + A^{-1}H\right)$.

4. If $A \in L(\mathbb{R}^n)$, show $\det\left(e^A\right) = e^{\mathrm{tr}(A)}$. **Hint:** use the previous item and Corollary 34.8 to show

$$\frac{d}{dt}\det\left(e^{tA}\right) = \det\left(e^{tA}\right)\mathrm{tr}(A).$$

Definition 34.10. *Let X and Y be Banach spaces and let $\mathcal{L}^1(X,Y) := L(X,Y)$ and for $k \ge 2$ let $\mathcal{L}^k(X,Y)$ be defined inductively by $\mathcal{L}^{k+1}(X,Y) = L(X, \mathcal{L}^k(X,Y))$. For example $\mathcal{L}^2(X,Y) = L(X, L(X,Y))$ and $\mathcal{L}^3(X,Y) = L(X, L(X, L(X,Y)))$.*

Suppose $f : U \subset_o X \to Y$ is a function. If f is differentiable on U, then it makes sense to ask if $f' = Df : U \to L(X,Y) = \mathcal{L}^1(X,Y)$ is differentiable. If Df is differentiable on U then $f'' = D^2 f := DDf : U \to \mathcal{L}^2(X,Y)$. Similarly we define $f^{(n)} = D^n f : U \to \mathcal{L}^n(X,Y)$ inductively.

Definition 34.11. *Given $k \in \mathbb{N}$, let $C^k(U,Y)$ denote those functions $f : U \to Y$ such that $f^{(j)} := D^j f : U \to \mathcal{L}^j(X,Y)$ exists and is continuous for $j = 1, 2, \dots, k$.*

Example 34.12. Let us continue on with Example 34.6 but now let $X = Y$ to simplify the notation. So $f : GL(X) \to GL(X)$ is the map $f(A) = A^{-1}$ and

$$f'(A) = -L_{A^{-1}}R_{A^{-1}}, \text{ i.e. } f' = -L_f R_f.$$

where $L_A B = AB$ and $R_A B = BA$ for all $A, B \in L(X)$. As the reader may easily check, the maps

$$A \in L(X) \to L_A, R_A \in L(L(X))$$

are linear and bounded. So by the chain and the product rule we find $f''(A)$ exists for all $A \in L(X)$ and

$$f''(A)B = -L_{f'(A)B}R_f - L_f R_{f'(A)B}.$$

More explicitly

$$[f''(A)B]\,C = A^{-1}BA^{-1}CA^{-1} + A^{-1}CA^{-1}BA^{-1}. \tag{34.10}$$

Working inductively one shows $f : GL(X) \to GL(X)$ defined by $f(A) := A^{-1}$ is C^∞.

34.3 Partial Derivatives

Definition 34.13 (Partial or Directional Derivative). *Let $f : U \subset_o X \to Y$ be a function, $x_0 \in U$, and $v \in X$. We say that f is differentiable at x_0 in the direction v iff $\frac{d}{dt}|_0(f(x_0+tv)) =: (\partial_v f)(x_0)$ exists. We call $(\partial_v f)(x_0)$ the directional or partial derivative of f at x_0 in the direction v.*

Notice that if f is differentiable at x_0, then $\partial_v f(x_0)$ exists and is equal to $f'(x_0)v$, see Corollary 34.8.

Proposition 34.14. *Let $f : U \subset_o X \to Y$ be a continuous function and $D \subset X$ be a dense subspace of X. Assume $\partial_v f(x)$ exists for all $x \in U$ and $v \in D$, and there exists a continuous function $A : U \to L(X,Y)$ such that $\partial_v f(x) = A(x)v$ for all $v \in D$ and $x \in U \cap D$. Then $f \in C^1(U,Y)$ and $Df = A$.*

Proof. Let $x_0 \in U$, $\varepsilon > 0$ such that $B(x_0, 2\varepsilon) \subset U$ and $M := \sup\{\|A(x)\| : x \in B(x_0, 2\varepsilon)\} < \infty$[1]. For $x \in B(x_0,\varepsilon) \cap D$ and $v \in D \cap B(0,\varepsilon)$, by the fundamental theorem of calculus,

$$\begin{aligned} f(x+v) - f(x) &= \int_0^1 \frac{df(x+tv)}{dt}\,dt \\ &= \int_0^1 (\partial_v f)(x+tv)\,dt = \int_0^1 A(x+tv)\,v\,dt. \end{aligned} \tag{34.11}$$

For general $x \in B(x_0,\varepsilon)$ and $v \in B(0,\varepsilon)$, choose $x_n \in B(x_0,\varepsilon) \cap D$ and $v_n \in D \cap B(0,\varepsilon)$ such that $x_n \to x$ and $v_n \to v$. Then

[1] It should be noted well, unlike in finite dimensions closed and bounded sets need not be compact, so it is not sufficient to choose ϵ sufficiently small so that $\overline{B(x_0, 2\epsilon)} \subset U$. Here is a counterexample. Let $X \equiv H$ be a Hilbert space, $\{e_n\}_{n=1}^\infty$ be an orthonormal set. Define $f(x) \equiv \sum_{n=1}^\infty n\phi(\|x - e_n\|)$, where ϕ is any continuous function on $\mathbb{R}$ such that $\phi(0) = 1$ and ϕ is supported in $(-1,1)$. Notice that $\|e_n - e_m\|^2 = 2$ for all $m \ne n$, so that $\|e_n - e_m\| = \sqrt{2}$. Using this fact it is rather easy to check that for any $x_0 \in H$, there is an $\epsilon > 0$ such that for all $x \in B(x_0, \epsilon)$, only one term in the sum defining f is non-zero. Hence, f is continuous. However, $f(e_n) = n \to \infty$ as $n \to \infty$.

$$f(x_n+v_n)-f(x_n)=\int_0^1 A(x_n+tv_n)\,v_n\,dt \tag{34.12}$$

holds for all n. The left side of this last equation tends to $f(x+v)-f(x)$ by the continuity of f. For the right side of Eq. (34.12) we have

$$\begin{aligned}&\|\int_0^1 A(x+tv)\,v\,dt-\int_0^1 A(x_n+tv_n)\,v_n\,dt\|\\&\le\int_0^1\|A(x+tv)-A(x_n+tv_n)\|\|v\|\,dt+M\|v-v_n\|.\end{aligned}$$

It now follows by the continuity of A, the fact that $\|A(x+tv)-A(x_n+tv_n)\|\le M$, and the dominated convergence theorem that right side of Eq. (34.12) converges to $\int_0^1 A(x+tv)\,v\,dt$. Hence Eq. (34.11) is valid for all $x\in B(x_0,\varepsilon)$ and $v\in B(0,\varepsilon)$. We also see that

$$f(x+v)-f(x)-A(x)v=\varepsilon(v)v, \tag{34.13}$$

where $\varepsilon(v):=\int_0^1[A(x+tv)\,-A(x)]\,dt$. Now

$$\begin{aligned}\|\varepsilon(v)\|&\le\int_0^1\|A(x+tv)\,-A(x)\|\,dt\\&\le\max_{t\in[0,1]}\|A(x+tv)\,-A(x)\|\to 0\text{ as }v\to 0,\end{aligned}$$

by the continuity of A. Thus, we have shown that f is differentiable and that $Df(x)=A(x)$. ■

Corollary 34.15. *Suppose now that $X=\mathbb{R}^d$, $f:U\subset_o X\to Y$ be a continuous function such that $\partial_i f(x):=\partial_{e_i}f(x)$ exists and is continuous on U for $i=1,2,\dots,d$, where $\{e_i\}_{i=1}^d$ is the standard basis for $\mathbb{R}^d$. Then $f\in C^1(U,Y)$ and $Df(x)\,e_i=\partial_i f(x)$ for all i.*

Proof. For $x\in U$, let $A(x):\mathbb{R}^d\to Y$ be the unique linear map such that $A(x)\,e_i=\partial_i f(x)$ for $i=1,2,\dots,d$. Then $A:U\to L(\mathbb{R}^d,Y)$ is a continuous map. Now let $v\in\mathbb{R}^d$ and $v^{(i)}:=(v_1,v_2,\dots,v_i,0,\dots,0)$ for $i=1,2,\dots,d$ and $v^{(0)}:=0$. Then for $t\in\mathbb{R}$ near 0, using the fundamental theorem of calculus and the definition of $\partial_i f(x)$,

$$\begin{aligned}f(x+tv)-f(x)&=\sum_{i=1}^d\left[f\left(x+tv^{(i)}\right)-f\left(x+tv^{(i-1)}\right)\right]\\&=\sum_{i=1}^d\int_0^1\frac{d}{ds}f\left(x+tv^{(i-1)}+stv_ie_i\right)ds\\&=\sum_{i=1}^d tv_i\int_0^1\partial_i f\left(x+tv^{(i-1)}+stv_ie_i\right)ds\\&=\sum_{i=1}^d tv_i\int_0^1 A\left(x+tv^{(i-1)}+stv_ie_i\right)e_ids.\end{aligned}$$

Using the continuity of A, it now follows that

$$\begin{aligned}\lim_{t\to 0}\frac{f(x+tv)-f(x)}{t}&=\sum_{i=1}^d v_i\lim_{t\to 0}\int_0^1 A\left(x+tv^{(i-1)}+stv_ie_i\right)e_ids\\&=\sum_{i=1}^d v_i\int_0^1 A(x)\,e_ids=A(x)\,v\end{aligned}$$

which shows $\partial_v f(x)$ exists and $\partial_v f(x)=A(x)\,v$. The result now follows from an application of Proposition 34.14. ■

34.4 Higher Order Derivatives

It is somewhat inconvenient to work with the Banach spaces $\mathcal{L}^k(X,Y)$ in Definition 34.10. For this reason we will introduce an isomorphic Banach space, $M_k(X,Y)$.

Definition 34.16. *For $k\in\{1,2,3,\dots\}$, let $M_k(X,Y)$ denote the set of functions $f:X^k\to Y$ such that*

1. *For $i\in\{1,2,\dots,k\}$, $v\in X\to f\langle v_1,v_2,\dots,v_{i-1},v,v_{i+1},\dots,v_k\rangle\in Y$ is linear [2] for all $\{v_i\}_{i=1}^n\subset X$.*
2. *The norm $\|f\|_{M_k(X,Y)}$ should be finite, where*

$$\|f\|_{M_k(X,Y)}:=\sup\{\frac{\|f\langle v_1,v_2,\dots,v_k\rangle\|_Y}{\|v_1\|\|v_2\|\cdots\|v_k\|}:\{v_i\}_{i=1}^k\subset X\setminus\{0\}\}.$$

[2] I will routinely write $f\langle v_1,v_2,\dots,v_k\rangle$ rather than $f(v_1,v_2,\dots,v_k)$ when the function f depends on each of variables linearly, i.e. f is a multi-linear function.

Lemma 34.17. *There are linear operators* $j_k : \mathcal{L}^k(X,Y) \to M_k(X,Y)$ *defined inductively as follows:* $j_1 = Id_{L(X,Y)}$ *(notice that* $M_1(X,Y) = \mathcal{L}^1(X,Y) = L(X,Y)$*) and*

$$(j_{k+1}A)\langle v_0, v_1, \ldots, v_k\rangle = (j_k(Av_0))\langle v_1, v_2, \ldots, v_k\rangle \quad \forall v_i \in X.$$

(Notice that $Av_0 \in \mathcal{L}^k(X,Y)$*.) Moreover, the maps* j_k *are isometric isomorphisms.*

Proof. To get a feeling for what j_k is let us write out j_2 and j_3 explicitly. If $A \in \mathcal{L}^2(X,Y) = L(X, L(X,Y))$, then $(j_2A)\langle v_1, v_2\rangle = (Av_1)v_2$ and if $A \in \mathcal{L}^3(X,Y) = L(X, L(X, L(X,Y)))$, $(j_3A)\langle v_1, v_2, v_3\rangle = ((Av_1)v_2)v_3$ for all $v_i \in X$. It is easily checked that j_k is linear for all k. We will now show by induction that j_k is an isometry and in particular that j_k is injective. Clearly this is true if $k = 1$ since j_1 is the identity map. For $A \in \mathcal{L}^{k+1}(X,Y)$,

$$\begin{aligned}
&\|j_{k+1}A\|_{M_{k+1}(X,Y)} \\
&\quad := \sup\{\frac{\|(j_k(Av_0))\langle v_1, v_2, \ldots, v_k\rangle\|_Y}{\|v_0\|\|v_1\|\|v_2\| \cdots \|v_k\|} : \{v_i\}_{i=0}^k \subset X \setminus \{0\}\} \\
&\quad = \sup\{\frac{\|(j_k(Av_0))\|_{M_k(X,Y)}}{\|v_0\|} : v_0 \in X \setminus \{0\}\} \\
&\quad = \sup\{\frac{\|Av_0\|_{\mathcal{L}^k(X,Y)}}{\|v_0\|} : v_0 \in X \setminus \{0\}\} \\
&\quad = \|A\|_{L(X,\mathcal{L}^k(X,Y))} := \|A\|_{\mathcal{L}^{k+1}(X,Y)},
\end{aligned}$$

wherein the second to last inequality we have used the induction hypothesis. This shows that j_{k+1} is an isometry provided j_k is an isometry. To finish the proof it suffices to show that j_k is surjective for all k. Again this is true for $k = 1$. Suppose that j_k is invertible for some $k \ge 1$. Given $f \in M_{k+1}(X,Y)$ we must produce $A \in \mathcal{L}^{k+1}(X,Y) = L(X, \mathcal{L}^k(X,Y))$ such that $j_{k+1}A = f$. If such an equation is to hold, then for $v_0 \in X$, we would have $j_k(Av_0) = f\langle v_0, \cdots\rangle$. That is $Av_0 = j_k^{-1}(f\langle v_0, \cdots\rangle)$. It is easily checked that A so defined is linear, bounded, and $j_{k+1}A = f$. ∎

From now on we will identify $\mathcal{L}^k$ with M_k without further mention. In particular, we will view $D^k f$ as function on U with values in $M_k(X,Y)$.

Theorem 34.18 (Differentiability). *Suppose* $k \in \{1, 2, \ldots\}$ *and* D *is a dense subspace of* X, $f : U \subset_o X \to Y$ *is a function such that* $(\partial_{v_1}\partial_{v_2} \cdots \partial_{v_l} f)(x)$ *exists for all* $x \in D \cap U$, $\{v_i\}_{i=1}^l \subset D$, *and* $l = 1, 2, \ldots k$. *Further assume there exists continuous functions* $A_l : U \subset_o X \to M_l(X,Y)$ *such that such that* $(\partial_{v_1}\partial_{v_2} \cdots \partial_{v_l} f)(x) = A_l(x)\langle v_1, v_2, \ldots, v_l\rangle$ *for all* $x \in D \cap U$, $\{v_i\}_{i=1}^l \subset D$, *and* $l = 1, 2, \ldots k$. *Then* $D^l f(x)$ *exists and is equal to* $A_l(x)$ *for all* $x \in U$ *and* $l = 1, 2, \ldots, k$.

Proof. We will prove the theorem by induction on k. We have already proved the theorem when $k = 1$, see Proposition 34.14. Now suppose that $k > 1$ and that the statement of the theorem holds when k is replaced by $k - 1$. Hence we know that $D^l f(x) = A_l(x)$ for all $x \in U$ and $l = 1, 2, \ldots, k-1$. We are also given that

$$(\partial_{v_1}\partial_{v_2} \cdots \partial_{v_k} f)(x) = A_k(x)\langle v_1, v_2, \ldots, v_k\rangle \quad \forall x \in U \cap D, \{v_i\} \subset D. \tag{34.14}$$

Now we may write $(\partial_{v_2} \cdots \partial_{v_k} f)(x)$ as $(D^{k-1}f)(x)\langle v_2, v_3, \ldots, v_k\rangle$ so that Eq. (34.14) may be written as

$$\begin{aligned}
&\partial_{v_1}(D^{k-1}f)(x)\langle v_2, v_3, \ldots, v_k\rangle) \\
&\qquad = A_k(x)\langle v_1, v_2, \ldots, v_k\rangle \quad \forall x \in U \cap D, \{v_i\} \subset D.
\end{aligned} \tag{34.15}$$

So by the fundamental theorem of calculus, we have that

$$\begin{aligned}
&((D^{k-1}f)(x + v_1) - (D^{k-1}f)(x))\langle v_2, v_3, \ldots, v_k\rangle \\
&\qquad = \int_0^1 A_k(x + tv_1)\langle v_1, v_2, \ldots, v_k\rangle\, dt
\end{aligned} \tag{34.16}$$

for all $x \in U \cap D$ and $\{v_i\} \subset D$ with v_1 sufficiently small. By the same argument given in the proof of Proposition 34.14, Eq. (34.16) remains valid for all $x \in U$ and $\{v_i\} \subset X$ with v_1 sufficiently small. We may write this last equation alternatively as,

$$(D^{k-1}f)(x + v_1) - (D^{k-1}f)(x) = \int_0^1 A_k(x + tv_1)\langle v_1, \cdots\rangle\, dt. \tag{34.17}$$

Hence

$$\begin{aligned}
&(D^{k-1}f)(x + v_1) - (D^{k-1}f)(x) - A_k(x)\langle v_1, \cdots\rangle \\
&\qquad = \int_0^1 [A_k(x + tv_1) - A_k(x)]\langle v_1, \cdots\rangle\, dt
\end{aligned}$$

from which we get the estimate,

$$\|(D^{k-1}f)(x + v_1) - (D^{k-1}f)(x) - A_k(x)\langle v_1, \cdots\rangle\| \le \varepsilon(v_1)\|v_1\| \tag{34.18}$$

where $\varepsilon(v_1) := \int_0^1 \|A_k(x + tv_1) - A_k(x)\|\, dt$. Notice by the continuity of A_k that $\varepsilon(v_1) \to 0$ as $v_1 \to 0$. Thus it follow from Eq. (34.18) that $D^{k-1}f$ is differentiable and that $(D^k f)(x) = A_k(x)$. ∎

Example 34.19. Let $f : GL(X,Y) \to GL(Y,X)$ be defined by $f(A) := A^{-1}$. We assume that $GL(X,Y)$ is not empty. Then f is infinitely differentiable and

$$(D^k f)(A)\langle V_1, V_2, \dots, V_k\rangle = (-1)^k \sum_\sigma \{B^{-1}V_{\sigma(1)}B^{-1}V_{\sigma(2)}B^{-1}\cdots B^{-1}V_{\sigma(k)}B^{-1}\}, \quad (34.19)$$

where sum is over all permutations of σ of $\{1, 2, \dots, k\}$.

Let me check Eq. (34.19) in the case that $k = 2$. Notice that we have already shown that $(\partial_{V_1} f)(B) = Df(B)V_1 = -B^{-1}V_1B^{-1}$. Using the product rule we find that

$$(\partial_{V_2}\partial_{V_1} f)(B) = B^{-1}V_2B^{-1}V_1B^{-1} + B^{-1}V_1B^{-1}V_2B^{-1} =: A_2(B)\langle V_1, V_2\rangle.$$

Notice that $\|A_2(B)\langle V_1, V_2\rangle\| \le 2\|B^{-1}\|^3\|V_1\| \cdot \|V_2\|$, so that $\|A_2(B)\| \le 2\|B^{-1}\|^3 < \infty$. Hence $A_2 : GL(X,Y) \to M_2(L(X,Y), L(Y,X))$. Also

$$\begin{aligned}\|(A_2(B) - A_2(C))\langle V_1, V_2\rangle\| &\le 2\|B^{-1}V_2B^{-1}V_1B^{-1} - C^{-1}V_2C^{-1}V_1C^{-1}\| \\ &\le 2\|B^{-1}V_2B^{-1}V_1B^{-1} - B^{-1}V_2B^{-1}V_1C^{-1}\| \\ &\quad + 2\|B^{-1}V_2B^{-1}V_1C^{-1} - B^{-1}V_2C^{-1}V_1C^{-1}\| \\ &\quad + 2\|B^{-1}V_2C^{-1}V_1C^{-1} - C^{-1}V_2C^{-1}V_1C^{-1}\| \\ &\le 2\|B^{-1}\|^2\|V_2\|\|V_1\|\|B^{-1} - C^{-1}\| \\ &\quad + 2\|B^{-1}\|\|C^{-1}\|\|V_2\|\|V_1\|\|B^{-1} - C^{-1}\| \\ &\quad + 2\|C^{-1}\|^2\|V_2\|\|V_1\|\|B^{-1} - C^{-1}\|.\end{aligned}$$

This shows that

$$\|A_2(B) - A_2(C)\| \le 2\|B^{-1} - C^{-1}\|\{\|B^{-1}\|^2 + \|B^{-1}\|\|C^{-1}\| + \|C^{-1}\|^2\}.$$

Since $B \to B^{-1}$ is differentiable and hence continuous, it follows that $A_2(B)$ is also continuous in B. Hence by Theorem 34.18 $D^2 f(A)$ exists and is given as in Eq. (34.19)

Example 34.20. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a C^∞– function and $F(x) := \int_0^1 f(x(t))\,dt$ for $x \in X := C([0,1], \mathbb{R})$ equipped with the norm $\|x\| := \max_{t\in[0,1]} |x(t)|$. Then $F : X \to \mathbb{R}$ is also infinitely differentiable and

$$(D^k F)(x)\langle v_1, v_2, \dots, v_k\rangle = \int_0^1 f^{(k)}(x(t))v_1(t)\cdots v_k(t)\,dt, \quad (34.20)$$

for all $x \in X$ and $\{v_i\} \subset X$.

To verify this example, notice that

$$\begin{aligned}(\partial_v F)(x) &:= \frac{d}{ds}|_0 F(x + sv) = \frac{d}{ds}|_0 \int_0^1 f(x(t) + sv(t))\,dt \\ &= \int_0^1 \frac{d}{ds}|_0 f(x(t) + sv(t))\,dt = \int_0^1 f'(x(t))v(t)\,dt.\end{aligned}$$

Similar computations show that

$$(\partial_{v_1}\partial_{v_2}\cdots\partial_{v_k} f)(x) = \int_0^1 f^{(k)}(x(t))v_1(t)\cdots v_k(t)\,dt =: A_k(x)\langle v_1, v_2, \dots, v_k\rangle.$$

Now for $x, y \in X$,

$$\begin{aligned}&|A_k(x)\langle v_1, v_2, \dots, v_k\rangle - A_k(y)\langle v_1, v_2, \dots, v_k\rangle| \\ &\quad\le \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))| \cdot |v_1(t)\cdots v_k(t)|\,dt \\ &\quad\le \prod_{i=1}^k \|v_i\| \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))|dt,\end{aligned}$$

which shows that

$$\|A_k(x) - A_k(y)\| \le \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))|dt.$$

This last expression is easily seen to go to zero as $y \to x$ in X. Hence A_k is continuous. Thus we may apply Theorem 34.18 to conclude that Eq. (34.20) is valid.

34.5 Inverse and Implicit Function Theorems

In this section, let X be a Banach space, $R > 0$, $U = B = B(0,R) \subset X$ and $\varepsilon : U \to X$ be a continuous function such that $\varepsilon(0) = 0$. Our immediate goal is to give a sufficient condition on ε so that $F(x) := x + \varepsilon(x)$ is a homeomorphism from U to $F(U)$ with $F(U)$ being an open subset of X. Let's start by looking at the one dimensional case first. So for the moment assume that $X = \mathbb{R}$, $U = (-1, 1)$, and $\varepsilon : U \to \mathbb{R}$ is C^1. Then F will be injective iff F is either strictly increasing or decreasing. Since we are thinking that F is a "small" perturbation of the identity function we will assume that F is strictly increasing, i.e. $F' = 1 + \varepsilon' > 0$. This positivity condition is not so easily interpreted for operators on a Banach space. However the condition that $|\varepsilon'| \le \alpha < 1$ is easily interpreted in the Banach space setting and it implies $1 + \varepsilon' > 0$.

Lemma 34.21. *Suppose that $U = B = B(0,R)$ $(R > 0)$ is a ball in X and $\varepsilon : B \to X$ is a C^1 function such that $\|D\varepsilon\| \le \alpha < \infty$ on U. Then*

$$\|\varepsilon(x) - \varepsilon(y)\| \le \alpha\|x - y\| \text{ for all } x, y \in U. \quad (34.21)$$

Proof. By the fundamental theorem of calculus and the chain rule:

$$\begin{aligned}\varepsilon(y)-\varepsilon(x) &= \int_0^1 \frac{d}{dt}\varepsilon(x+t(y-x))dt \\ &= \int_0^1 [D\varepsilon(x+t(y-x))](y-x)dt.\end{aligned}$$

Therefore, by the triangle inequality and the assumption that $\|D\varepsilon(x)\| \le \alpha$ on B,

$$\|\varepsilon(y)-\varepsilon(x)\| \le \int_0^1 \|D\varepsilon(x+t(y-x))\|dt \cdot \|(y-x)\| \le \alpha\|(y-x)\|.$$

■

Remark 34.22. It is easily checked that if $\varepsilon : U = B(0,R) \to X$ is C^1 and satisfies (34.21) then $\|D\varepsilon\| \le \alpha$ on U.

Using the above remark and the analogy to the one dimensional example, one is lead to the following proposition.

Proposition 34.23. *Suppose $\alpha \in (0,1)$, $R > 0$, $U = B(0,R) \subset_o X$ and $\varepsilon : U \to X$ is a continuous function such that $\varepsilon(0) = 0$ and*

$$\|\varepsilon(x)-\varepsilon(y)\| \le \alpha\|x-y\| \quad \forall\, x,y \in U. \tag{34.22}$$

Then $F : U \to X$ defined by $F(x) := x + \varepsilon(x)$ for $x \in U$ satisfies:

1. *F is an injective map and $G = F^{-1} : V \to U$ is continuous where $V := F(U)$.*
2. *If $x_0 \in U$, $z_0 = F(x_0)$ and $r > 0$ such that $B(x_0,r) \subset U$, then*
$$B(z_0,(1-\alpha)r) \subset F(B(x_0,r)) \subset B(z_0,(1+\alpha)r). \tag{34.23}$$
In particular, for all $r \le R$,
$$B(0,(1-\alpha)r) \subset F(B(0,r)) \subset B(0,(1+\alpha)r), \tag{34.24}$$
see Figure 34.1 below.
3. *$V := F(U)$ is open subset of X and $F : U \to V$ is a homeomorphism.*

Proof.

1. Using the definition of F and the estimate in Eq. (34.22),
$$\begin{aligned}\|x-y\| &= \|(F(x)-F(y)) - (\varepsilon(x)-\varepsilon(y))\| \\ &\le \|F(x)-F(y)\| + \|\varepsilon(x)-\varepsilon(y)\| \\ &\le \|F(x)-F(y)\| + \alpha\|(x-y)\|\end{aligned}$$

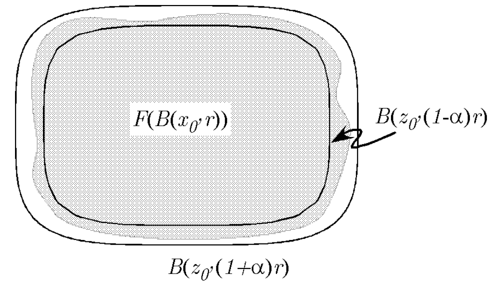


Fig. 34.1. Nesting of $F(B(x_0,r))$ between $B(z_0,(1-\alpha)r)$ and $B(z_0,(1+\alpha)r)$.

for all $x,y \in U$. This implies
$$\|x-y\| \le (1-\alpha)^{-1}\|F(x)-F(y)\| \tag{34.25}$$
which shows F is injective on U and hence shows the inverse function $G = F^{-1} : V := F(U) \to U$ is well defined. Moreover, replacing x,y in Eq. (34.25) by $G(x)$ and $G(y)$ respectively with $x,y \in V$ shows
$$\|G(x)-G(y)\| \le (1-\alpha)^{-1}\|x-y\| \text{ for all } x,y \in V. \tag{34.26}$$
Hence G is Lipschitz on V and hence continuous.

2. Let $x_0 \in U$, $r > 0$ and $z_0 = F(x_0) = x_0 + \varepsilon(x_0)$ be as in item 2. The second inclusion in Eq. (34.23) follows from the simple computation:
$$\begin{aligned}\|F(x_0+h)-z_0\| &= \|h+\varepsilon(x_0+h)-\varepsilon(x_0)\| \\ &\le \|h\| + \|\varepsilon(x_0+h)-\varepsilon(x_0)\| \\ &\le (1+\alpha)\|h\| < (1+\alpha)r\end{aligned}$$
for all $h \in B(0,r)$. To prove the first inclusion in Eq. (34.23) we must find, for every $z \in B(z_0,(1-\alpha)r)$, an $h \in B(0,r)$ such that $z = F(x_0+h)$ or equivalently an $h \in B(0,r)$ solving
$$z - z_0 = F(x_0+h) - F(x_0) = h + \varepsilon(x_0+h) - \varepsilon(x_0).$$
Let $k := z - z_0$ and for $h \in B(0,r)$, let $\delta(h) := \varepsilon(x_0+h) - \varepsilon(x_0)$. With this notation it suffices to show for each $k \in B(z_0,(1-\alpha)r)$ there exists $h \in B(0,r)$ such that $k = h + \delta(h)$. Notice that $\delta(0) = 0$ and
$$\|\delta(h_1)-\delta(h_2)\| = \|\varepsilon(x_0+h_1)-\varepsilon(x_0+h_2)\| \le \alpha\|h_1-h_2\| \tag{34.27}$$
for all $h_1,h_2 \in B(0,r)$. We are now going to solve the equation $k = h+\delta(h)$ for h by the method of successive approximations starting with $h_0 = 0$ and then defining h_n inductively by

$$h_{n+1} = k - \delta(h_n). \tag{34.28}$$

A simple induction argument using Eq. (34.27) shows that

$$\|h_{n+1} - h_n\| \le \alpha^n \|k\| \text{ for all } n \in \mathbb{N}_0$$

and in particular that

$$\begin{aligned} \|h_N\| &= \left\| \sum_{n=0}^{N-1} (h_{n+1} - h_n) \right\| \le \sum_{n=0}^{N-1} \|h_{n+1} - h_n\| \\ &\le \sum_{n=0}^{N-1} \alpha^n \|k\| = \frac{1-\alpha^N}{1-\alpha} \|k\|. \end{aligned} \tag{34.29}$$

Since $\|k\| < (1-\alpha) r$, this implies that $\|h_N\| < r$ for all N showing the approximation procedure is well defined. Let

$$h := \lim_{N\to\infty} h_n = \sum_{n=0}^{\infty} (h_{n+1} - h_n) \in X$$

which exists since the sum in the previous equation is absolutely convergent. Passing to the limit in Eqs. (34.29) and (34.28) shows that $\|h\| \le (1-\alpha)^{-1}\|k\| < r$ and $h = k - \delta(h)$, i.e. $h \in B(0,r)$ solves $k = h + \delta(h)$ as desired.

3. Given $x_0 \in U$, the first inclusion in Eq. (34.23) shows that $z_0 = F(x_0)$ is in the interior of $F(U)$. Since $z_0 \in F(U)$ was arbitrary, it follows that $V = F(U)$ is open. The continuity of the inverse function has already been proved in item 1.

■

For the remainder of this section let X and Y be two Banach spaces, $U \subset_o X$, $k \ge 1$, and $f \in C^k(U,Y)$.

Lemma 34.24. *Suppose $x_0 \in U$, $R > 0$ is such that $B^X(x_0,R) \subset U$ and $T : B^X(x_0,R) \to Y$ is a C^1 – function such that $T'(x_0)$ is invertible. Let*

$$\alpha(R) := \sup_{x\in B^X(x_0,R)} \left\|T'(x_0)^{-1}T'(x) - I\right\|_{L(X)} \tag{34.30}$$

and $\varepsilon \in C^1\left(B^X(0,R), X\right)$ be defined by

$$\varepsilon(h) = T'(x_0)^{-1}\left[T(x_0+h) - T(x_0)\right] - h \tag{34.31}$$

so that

$$T(x_0+h) = T(x_0) + T'(x_0)\left(h + \varepsilon(h)\right). \tag{34.32}$$

Then $\varepsilon(h) = o(h)$ as $h \to 0$ and

$$\|\varepsilon(h') - \varepsilon(h)\| \le \alpha(R)\|h' - h\| \text{ for all } h, h' \in B^X(0,R). \tag{34.33}$$

If $\alpha(R) < 1$ (which may be achieved by shrinking R if necessary), then $T'(x)$ is invertible for all $x \in B^X(x_0,R)$ and

$$\sup_{x\in B^X(x_0,R)} \left\|T'(x)^{-1}\right\|_{L(Y,X)} \le \frac{1}{1-\alpha(R)} \left\|T'(x_0)^{-1}\right\|_{L(Y,X)}. \tag{34.34}$$

Proof. By definition of $T'(x_0)$ and using $T'(x_0)^{-1}$ exists,

$$T(x_0+h) - T(x_0) = T'(x_0)h + o(h)$$

from which it follows that $\varepsilon(h) = o(h)$. In fact by the fundamental theorem of calculus,

$$\varepsilon(h) = \int_0^1 \left(T'(x_0)^{-1}T'(x_0+th) - I\right) h dt$$

but we will not use this here. Let $h, h' \in B^X(0,R)$ and apply the fundamental theorem of calculus to $t \to T(x_0 + t(h'-h))$ to conclude

$$\begin{aligned} \varepsilon(h') - \varepsilon(h) &= T'(x_0)^{-1}\left[T(x_0+h') - T(x_0+h)\right] - (h'-h) \\ &= \left[\int_0^1 \left(T'(x_0)^{-1}T'(x_0+t(h'-h)) - I\right) dt\right](h'-h). \end{aligned}$$

Taking norms of this equation gives

$$\begin{aligned} \|\varepsilon(h') - \varepsilon(h)\| &\le \left[\int_0^1 \left\|T'(x_0)^{-1}T'(x_0+t(h'-h)) - I\right\| dt\right] \|h'-h\| \\ &\le \alpha(R)\|h'-h\|. \end{aligned}$$

It only remains to prove Eq. (34.34), so suppose now that $\alpha(R) < 1$. Then by Proposition 14.26, $T'(x_0)^{-1}T'(x) = I - \left(I - T'(x_0)^{-1}T'(x)\right)$ is invertible and

$$\left\|\left[T'(x_0)^{-1}T'(x)\right]^{-1}\right\| \le \frac{1}{1-\alpha(R)} \text{ for all } x \in B^X(x_0,R).$$

Since $T'(x) = T'(x_0)\left[T'(x_0)^{-1}T'(x)\right]$ this implies $T'(x)$ is invertible and

$$\left\|T'(x)^{-1}\right\| = \left\|\left[T'(x_0)^{-1}T'(x)\right]^{-1} T'(x_0)^{-1}\right\| \le \frac{1}{1-\alpha(R)} \left\|T'(x_0)^{-1}\right\|$$

for all $x \in B^X(x_0,R)$. ■

Theorem 34.25 (Inverse Function Theorem). *Suppose $U \subset_o X$, $k \geq 1$ and $T \in C^k(U, Y)$ such that $T'(x)$ is invertible for all $x \in U$. Further assume $x_0 \in U$ and $R > 0$ such that $B^X(x_0, R) \subset U$.*

1. *For all $r \leq R$,*
$$T(B^X(x_0,r)) \subset T(x_0) + T'(x_0)\, B^X(0,(1+\alpha(r))r)\,. \tag{34.35}$$
2. *If we further assume that*
$$\alpha(R) := \sup_{x \in B^X(x_0,R)} \left\| T'(x_0)^{-1} T'(x) - I \right\| < 1,$$
which may always be achieved by taking R sufficiently small, then
$$T(x_0) + T'(x_0)\, B^X(0,(1-\alpha(r))r) \subset T(B^X(x_0,r)) \tag{34.36}$$
for all $r \leq R$, see Figure 34.2.
3. *$T : U \to Y$ is an open mapping, in particular $V := T(U) \subset_o Y$.*
4. *Again if R is sufficiently small so that $\alpha(R) < 1$, then $T|_{B^X(x_0,R)} : B^X(x_0,R) \to T(B^X(x_0,R))$ is invertible and $T|_{B^X(x_0,R)}^{-1} : T\left(B^X(x_0,R)\right) \to B^X(x_0,R)$ is a C^k – map.*
5. *If T is injective, then $T^{-1} : V \to U$ is also a C^k – map and*
$$\left(T^{-1}\right)'(y) = \left[T'(T^{-1}(y))\right]^{-1} \text{ for all } y \in V.$$

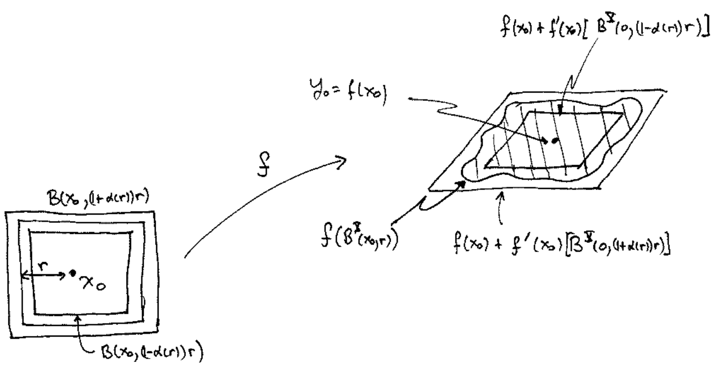


Fig. 34.2. The nesting of $T(B^X(x_0,r))$ between $T(x_0) + T'(x_0)\, B^X(0,(1-\alpha(r))r)$ and$T(x_0) + T'(x_0)\, B^X(0,(1+\alpha(r))r)$.

Proof. Let $\varepsilon \in C^1\left(B^X(0,R), X\right)$ be as defined in Eq. (34.31).

1. Using Eqs. (34.32) and (34.24),
$$\begin{aligned} T\left(B^X(x_0,r)\right) &= T(x_0) + T'(x_0)(I+\varepsilon)\left(B^X(0,r)\right) \\ &\subset T(x_0) + T'(x_0)\, B^X(0,(1+\alpha(r))\,r) \end{aligned} \tag{34.37}$$
which proves Eq. (34.35).
2. Now assume $\alpha(R) < 1$, then by Eqs. (34.37) and (34.24),
$$\begin{aligned} &T(x_0) + T'(x_0)\, B^X(0,(1-\alpha(r))\,r) \\ &\quad \subset T(x_0) + T'(x_0)(I+\varepsilon)\left(B^X(0,r)\right) = T\left(B^X(x_0,r)\right) \end{aligned}$$
which proves Eq. (34.36).
3. Notice that $h \in X \to T(x_0) + T'(x_0)\, h \in Y$ is a homeomorphism. The fact that T is an open map follows easily from Eq. (34.36) which shows that $T(x_0)$ is interior of $T(W)$ for any $W \subset_o X$ with $x_0 \in W$.
4. The fact that $T|_{B^X(x_0,R)} : B^X(x_0,R) \to T(B^X(x_0,R))$ is invertible with a continuous inverse follows from Eq. (34.32) and Proposition 34.23. It now follows from the converse to the chain rule, Theorem 34.7, that $g := T|_{B^X(x_0,R)}^{-1} : T\left(B^X(x_0,R)\right) \to B^X(x_0,R)$ is differentiable and
$$g'(y) = \left[T'(g(y))\right]^{-1} \text{ for all } y \in T\left(B^X(x_0,R)\right).$$
This equation shows g is C^1. Now suppose that $k \geq 2$. Since $T' \in C^{k-1}(B, L(X))$ and $i(A) := A^{-1}$ is a smooth map by Example 34.19, $g' = i \circ T' \circ g$ is C^1, i.e. g is C^2. If $k \geq 2$, we may use the same argument to now show g is C^3. Continuing this way inductively, we learn g is C^k.
5. Since differentiability and smoothness is local, the assertion in item 5. follows directly from what has already been proved.

■

Theorem 34.26 (Implicit Function Theorem). *Suppose that X, Y, and W are three Banach spaces, $k \geq 1$, $A \subset X \times Y$ is an open set, (x_0, y_0) is a point in A, and $f : A \to W$ is a C^k – map such $f(x_0, y_0) = 0$. Assume that $D_2 f(x_0, y_0) := D(f(x_0, \cdot))(y_0) : Y \to W$ is a bounded invertible linear transformation. Then there is an open neighborhood U_0 of x_0 in X such that for all connected open neighborhoods U of x_0 contained in U_0, there is a unique continuous function $u : U \to Y$ such that $u(x_0) = y_o$, $(x, u(x)) \in A$ and $f(x, u(x)) = 0$ for all $x \in U$. Moreover u is necessarily C^k and*
$$Du(x) = -D_2 f(x, u(x))^{-1} D_1 f(x, u(x)) \text{ for all } x \in U. \tag{34.38}$$

Proof. By replacing f by $(x,y) \to D_2 f(x_0,y_0)^{-1} f(x,y)$ if necessary, we may assume with out loss of generality that $W = Y$ and $D_2 f(x_0,y_0) = I_Y$. Define $F : A \to X \times Y$ by $F(x,y) := (x, f(x,y))$ for all $(x,y) \in A$. Notice that

$$DF(x,y) = \begin{bmatrix} I & D_1 f(x,y) \\ 0 & D_2 f(x,y) \end{bmatrix}$$

which is invertible iff $D_2 f(x,y)$ is invertible and if $D_2 f(x,y)$ is invertible then

$$DF(x,y)^{-1} = \begin{bmatrix} I & -D_1 f(x,y) D_2 f(x,y)^{-1} \\ 0 & D_2 f(x,y)^{-1} \end{bmatrix}.$$

Since $D_2 f(x_0,y_0) = I$ is invertible, the inverse function theorem guarantees that there exists a neighborhood U_0 of x_0 and V_0 of y_0 such that $U_0 \times V_0 \subset A$, $F(U_0 \times V_0)$ is open in $X \times Y$, $F|_{(U_0 \times V_0)}$ has a C^k–inverse which we call F^{-1}. Let $\pi_2(x,y) := y$ for all $(x,y) \in X \times Y$ and define C^k – function u_0 on U_0 by $u_0(x) := \pi_2 \circ F^{-1}(x,0)$. Since $F^{-1}(x,0) = (\tilde{x}, u_0(x))$ iff

$$(x,0) = F(\tilde{x}, u_0(x)) = (\tilde{x}, f(\tilde{x}, u_0(x))),$$

it follows that $x = \tilde{x}$ and $f(x, u_0(x)) = 0$. Thus

$$(x, u_0(x)) = F^{-1}(x,0) \in U_0 \times V_0 \subset A$$

and $f(x, u_0(x)) = 0$ for all $x \in U_0$. Moreover, u_0 is C^k being the composition of the C^k– functions, $x \to (x,0)$, F^{-1}, and π_2. So if $U \subset U_0$ is a connected set containing x_0, we may define $u := u_0|_U$ to show the existence of the functions u as described in the statement of the theorem. The only statement left to prove is the uniqueness of such a function u. Suppose that $u_1 : U \to Y$ is another continuous function such that $u_1(x_0) = y_0$, and $(x, u_1(x)) \in A$ and $f(x, u_1(x)) = 0$ for all $x \in U$. Let

$$O := \{x \in U | u(x) = u_1(x)\} = \{x \in U | u_0(x) = u_1(x)\}.$$

Clearly O is a (relatively) closed subset of U which is not empty since $x_0 \in O$. Because U is connected, if we show that O is also an open set we will have shown that $O = U$ or equivalently that $u_1 = u_0$ on U. So suppose that $x \in O$, i.e. $u_0(x) = u_1(x)$. For $\tilde{x}$ near $x \in U$,

$$0 = 0 - 0 = f(\tilde{x}, u_0(\tilde{x})) - f(\tilde{x}, u_1(\tilde{x})) = R(\tilde{x})(u_1(\tilde{x}) - u_0(\tilde{x})) \tag{34.39}$$

where

$$R(\tilde{x}) := \int_0^1 D_2 f((\tilde{x}, u_0(\tilde{x}) + t(u_1(\tilde{x}) - u_0(\tilde{x}))) dt. \tag{34.40}$$

From Eq. (34.40) and the continuity of u_0 and u_1, $\lim_{\tilde{x} \to x} R(\tilde{x}) = D_2 f(x, u_0(x))$ which is invertible.[3] Thus $R(\tilde{x})$ is invertible for all $\tilde{x}$ sufficiently close to x which combined with Eq. (34.39) implies that $u_1(\tilde{x}) = u_0(\tilde{x})$ for all $\tilde{x}$ sufficiently close to x. Since $x \in O$ was arbitrary, we have shown that O is open. ■

34.6 Smooth Dependence of ODE's on Initial Conditions*

In this subsection, let X be a Banach space, $U \subset_o X$ and J be an open interval with $0 \in J$.

Lemma 34.27. *If $Z \in C(J \times U, X)$ such that $D_x Z(t,x)$ exists for all $(t,x) \in J \times U$ and $D_x Z(t,x) \in C(J \times U, X)$ then Z is locally Lipschitz in x, see Definition 33.7.*

Proof. Suppose $I \sqsubset\sqsubset J$ and $x \in U$. By the continuity of DZ, for every $t \in I$ there an open neighborhood N_t of $t \in I$ and $\varepsilon_t > 0$ such that $B(x, \varepsilon_t) \subset U$ and

$$\sup \{\|D_x Z(t', x')\| : (t', x') \in N_t \times B(x, \varepsilon_t)\} < \infty.$$

By the compactness of I, there exists a finite subset $\Lambda \subset I$ such that $I \subset \cup_{t \in I} N_t$. Let $\varepsilon(x, I) := \min \{\varepsilon_t : t \in \Lambda\}$ and

$$K(x, I) := \sup \{\|DZ(t, x')\| (t, x') \in I \times B(x, \varepsilon(x, I))\} < \infty.$$

Then by the fundamental theorem of calculus and the triangle inequality,

$$\begin{aligned} \|Z(t, x_1) - Z(t, x_0)\| &\le \left(\int_0^1 \|D_x Z(t, x_0 + s(x_1 - x_0)\| \, ds \right) \|x_1 - x_0\| \\ &\le K(x, I) \|x_1 - x_0\| \end{aligned}$$

for all $x_0, x_1 \in B(x, \varepsilon(x, I))$ and $t \in I$. ■

Theorem 34.28 (Smooth Dependence of ODE's on Initial Conditions). *Let X be a Banach space, $U \subset_o X$, $Z \in C(\mathbb{R} \times U, X)$ such that $D_x Z \in C(\mathbb{R} \times U, X)$ and $\varphi : \mathcal{D}(Z) \subset \mathbb{R} \times X \to X$ denote the maximal solution operator to the ordinary differential equation*

$$\dot{y}(t) = Z(t, y(t)) \text{ with } y(0) = x \in U, \tag{34.41}$$

see Notation 33.10 and Theorem 33.16. Then $\varphi \in C^1(\mathcal{D}(Z), U)$, $\partial_t D_x \varphi(t,x)$ exists and is continuous for $(t,x) \in \mathcal{D}(Z)$ and $D_x \varphi(t,x)$ satisfies the linear differential equation,

[3] Notice that $DF(x, u_0(x))$ is invertible for all $x \in U_0$ since $F|_{U_0 \times V_0}$ has a C^1 inverse. Therefore $D_2 f(x, u_0(x))$ is also invertible for all $x \in U_0$.

$$\frac{d}{dt}D_x\varphi(t,x) = [(D_xZ)(t,\varphi(t,x))]D_x\varphi(t,x) \text{ with } D_x\varphi(0,x) = I_X \tag{34.42}$$

for $t \in J_x$.

Proof. Let $x_0 \in U$ and J be an open interval such that $0 \in J \subset \bar{J} \sqsubset\sqsubset J_{x_0}$, $y_0 := y(\cdot, x_0)|_J$ and

$$\mathcal{O}_\varepsilon := \{y \in BC(J,U) : \|y - y_0\|_\infty < \varepsilon\} \subset_o BC(J,X).$$

By Lemma 34.27, Z is locally Lipschitz and therefore Theorem 33.16 is applicable. By Eq. (33.25) of Theorem 33.16, there exists $\varepsilon > 0$ and $\delta > 0$ such that $G : B(x_0,\delta) \to \mathcal{O}_\varepsilon$ defined by $G(x) := \varphi(\cdot, x)|_J$ is continuous. By Lemma 34.29 below, for $\varepsilon > 0$ sufficiently small the function $F : \mathcal{O}_\varepsilon \to BC(J,X)$ defined by

$$F(y) := y - \int_0^\cdot Z(t,y(t))dt \tag{34.43}$$

is C^1 and

$$DF(y)v = v - \int_0^\cdot D_yZ(t,y(t))v(t)dt. \tag{34.44}$$

By the existence and uniqueness for linear ordinary differential equations, Theorem 32.22, $DF(y)$ is invertible for any $y \in BC(J,U)$. By the definition of φ, $F(G(x)) = h(x)$ for all $x \in B(x_0,\delta)$ where $h : X \to BC(J,X)$ is defined by $h(x)(t) = x$ for all $t \in J$, i.e. $h(x)$ is the constant path at x. Since h is a bounded linear map, h is smooth and $Dh(x) = h$ for all $x \in X$. We may now apply the converse to the chain rule in Theorem 34.7 to conclude $G \in C^1(B(x_0,\delta),\mathcal{O})$ and $DG(x) = [DF(G(x))]^{-1}Dh(x)$ or equivalently, $DF(G(x))DG(x) = h$ which in turn is equivalent to

$$D_x\varphi(t,x) - \int_0^t [DZ(\varphi(\tau,x)]D_x\varphi(\tau,x)\,d\tau = I_X.$$

As usual this equation implies $D_x\varphi(t,x)$ is differentiable in t, $D_x\varphi(t,x)$ is continuous in (t,x) and $D_x\varphi(t,x)$ satisfies Eq. (34.42). ∎

Lemma 34.29. *Continuing the notation used in the proof of Theorem 34.28 and further let*

$$f(y) := \int_0^\cdot Z(\tau,y(\tau))\,d\tau \text{ for } y \in \mathcal{O}_\varepsilon.$$

Then $f \in C^1(\mathcal{O}_\varepsilon, Y)$ *and for all* $y \in \mathcal{O}_\varepsilon$,

$$f'(y)h = \int_0^\cdot D_xZ(\tau,y(\tau))h(\tau)\,d\tau =: \Lambda_y h.$$

Proof. Let $h \in Y$ be sufficiently small and $\tau \in J$, then by fundamental theorem of calculus,

$$\begin{aligned} Z(\tau,&y(\tau)+h(\tau)) - Z(\tau,y(\tau)) \\ &= \int_0^1 [D_xZ(\tau,y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))]dr \end{aligned}$$

and therefore,

$$\begin{aligned} f(y+&h) - f(y) - \Lambda_y h(t) \\ &= \int_0^t [Z(\tau,y(\tau)+h(\tau)) - Z(\tau,y(\tau)) - D_xZ(\tau,y(\tau))h(\tau)]\,d\tau \\ &= \int_0^t d\tau \int_0^1 dr[D_xZ(\tau,y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))]h(\tau). \end{aligned}$$

Therefore,

$$\|(f(y+h) - f(y) - \Lambda_y h)\|_\infty \le \|h\|_\infty \delta(h) \tag{34.45}$$

where

$$\delta(h) := \int_J d\tau \int_0^1 dr\,\|D_xZ(\tau,y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))\|.$$

With the aid of Lemmas 34.27 and Lemma 33.8,

$$(r,\tau,h) \in [0,1]\times J\times Y \to \|D_xZ(\tau,y(\tau)+rh(\tau))\|$$

is bounded for small h provided $\varepsilon > 0$ is sufficiently small. Thus it follows from the dominated convergence theorem that $\delta(h) \to 0$ as $h \to 0$ and hence Eq. (34.45) implies $f'(y)$ exists and is given by Λ_y. Similarly,

$$\begin{aligned} \|f'(&y+h) - f'(y)\|_{op} \\ &\le \int_J \|D_xZ(\tau,y(\tau)+h(\tau)) - D_xZ(\tau,y(\tau))\|\,d\tau \to 0 \text{ as } h \to 0 \end{aligned}$$

showing f' is continuous. ∎

Remark 34.30. If $Z \in C^k(U,X)$, then an inductive argument shows that $\varphi \in C^k(\mathcal{D}(Z),X)$. For example if $Z \in C^2(U,X)$ then $(y(t),u(t)) := (\varphi(t,x), D_x\varphi(t,x))$ solves the ODE,

$$\frac{d}{dt}(y(t),u(t)) = \tilde{Z}((y(t),u(t))) \text{ with } (y(0),u(0)) = (x, Id_X)$$

where $\tilde{Z}$ is the C^1 – vector field defined by

$$\tilde{Z}(x,u) = (Z(x), D_x Z(x) u).$$

Therefore Theorem 34.28 may be applied to this equation to deduce: $D_x^2\varphi(t,x)$ and $D_x^2\varphi(t,x)$ exist and are continuous. We may now differentiate Eq. (34.42) to find $D_x^2\varphi(t,x)$ satisfies the ODE,

$$\begin{aligned}\frac{d}{dt}D_x^2\varphi(t,x) &= [\left(\partial_{D_x\varphi(t,x)}D_x Z\right)(t,\varphi(t,x))]D_x\varphi(t,x)\\ &\quad + [(D_x Z)(t,\varphi(t,x))]D_x^2\varphi(t,x)\end{aligned}$$

with $D_x^2\varphi(0,x) = 0.$

34.7 Existence of Periodic Solutions

A detailed discussion of the inverse function theorem on Banach and Frechét spaces may be found in Richard Hamilton's, "The Inverse Function Theorem of Nash and Moser." The applications in this section are taken from this paper. In what follows we say $f \in C_{2\pi}^k(\mathbb{R},(c,d))$ if $f \in C_{2\pi}^k(\mathbb{R},(c,d))$ and f is 2π – periodic, i.e. $f(x+2\pi) = f(x)$ for all $x \in \mathbb{R}$.

Theorem 34.31 (Taken from Hamilton, p. 110.). *Let $p : U := (a,b) \to V := (c,d)$ be a smooth function with $p' > 0$ on (a,b). For every $g \in C_{2\pi}^\infty(\mathbb{R},(c,d))$ there exists a unique function $y \in C_{2\pi}^\infty(\mathbb{R},(a,b))$ such that*

$$\dot{y}(t) + p(y(t)) = g(t).$$

Proof. Let $\tilde{V} := C_{2\pi}^0(\mathbb{R},(c,d)) \subset_o C_{2\pi}^0(\mathbb{R},\mathbb{R})$ and $\tilde{U} \subset_o C_{2\pi}^1(\mathbb{R},(a,b))$ be given by

$$\tilde{U} := \left\{y \in C_{2\pi}^1(\mathbb{R},\mathbb{R}) : a < y(t) < b \;\&\; c < \dot{y}(t) + p(y(t)) < d \;\forall\; t\right\}.$$

The proof will be completed by showing $P : \tilde{U} \to \tilde{V}$ defined by

$$P(y)(t) = \dot{y}(t) + p(y(t)) \text{ for } y \in \tilde{U} \text{ and } t \in \mathbb{R}$$

is bijective. Note that if $P(y)$ is smooth then so is y.

Step 1. The differential of P is given by $P'(y)h = \dot{h} + p'(y)h$, see Exercise 34.8. We will now show that the linear mapping $P'(y)$ is invertible. Indeed let $f = p'(y) > 0$, then the general solution to the Eq. $\dot{h} + fh = k$ is given by

$$h(t) = e^{-\int_0^t f(\tau)d\tau}h_0 + \int_0^t e^{-\int_\tau^t f(s)ds}k(\tau)d\tau$$

where h_0 is a constant. We wish to choose h_0 so that $h(2\pi) = h_0$, i.e. so that

$$h_0\left(1 - e^{-c(f)}\right) = \int_0^{2\pi} e^{-\int_\tau^t f(s)ds}k(\tau)d\tau$$

where

$$c(f) = \int_0^{2\pi} f(\tau)d\tau = \int_0^{2\pi} p'(y(\tau))d\tau > 0.$$

The unique solution $h \in C_{2\pi}^1(\mathbb{R},\mathbb{R})$ to $P'(y)h = k$ is given by

$$\begin{aligned}h(t) &= \left(1 - e^{-c(f)}\right)^{-1} e^{-\int_0^t f(\tau)d\tau}\int_0^{2\pi} e^{-\int_\tau^t f(s)ds}k(\tau)d\tau + \int_0^t e^{-\int_\tau^t f(s)ds}k(\tau)d\tau\\ &= \left(1 - e^{-c(f)}\right)^{-1} e^{-\int_0^t f(s)ds}\int_0^{2\pi} e^{-\int_\tau^t f(s)ds}k(\tau)d\tau + \int_0^t e^{-\int_\tau^t f(s)ds}k(\tau)d\tau.\end{aligned}$$

Therefore $P'(y)$ is invertible for all y. Hence by the inverse function Theorem (Theorem 34.25), $P : \tilde{U} \to \tilde{V}$ is an open mapping which is locally invertible.

Step 2. Let us now prove $P : \tilde{U} \to \tilde{V}$ is injective. For this suppose $y_1, y_2 \in \tilde{U}$ such that $P(y_1) = g = P(y_2)$ and let $z = y_2 - y_1$. Since

$$\dot{z}(t) + p(y_2(t)) - p(y_1(t)) = g(t) - g(t) = 0,$$

if $t_m \in \mathbb{R}$ is point where $z(t_m)$ takes on its maximum, then $\dot{z}(t_m) = 0$ and hence

$$p(y_2(t_m)) - p(y_1(t_m)) = 0.$$

Since p is increasing this implies $y_2(t_m) = y_1(t_m)$ and hence $z(t_m) = 0$. This shows $z(t) \le 0$ for all t and a similar argument using a minimizer of z shows $z(t) \ge 0$ for all t. So we conclude $y_1 = y_2$.

Step 3. Let $W := P(\tilde{U})$, we wish to show $W = \tilde{V}$. By step 1., we know W is an open subset of $\tilde{V}$ and since $\tilde{V}$ is connected, to finish the proof it suffices to show W is relatively closed in $\tilde{V}$. So suppose $y_j \in \tilde{U}$ such that $g_j := P(y_j) \to g \in \tilde{V}$. We must now show $g \in W$, i.e. $g = P(y)$ for some $y \in W$. If t_m is a maximizer of y_j, then $\dot{y}_j(t_m) = 0$ and hence $g_j(t_m) = p(y_j(t_m)) < d$ and therefore $y_j(t_m) < b$ because p is increasing. A similar argument works for the minimizers then allows us to conclude $\operatorname{Ran}(p \circ y_j) \subset \operatorname{Ran}(g_j) \sqsubset\sqsubset (c,d)$ for all j. Since g_j is converging uniformly to g, there exists $c < \gamma < \delta < d$ such that $\operatorname{Ran}(p \circ y_j) \subset \operatorname{Ran}(g_j) \subset [\gamma,\delta]$ for all j. Again since $p' > 0$,

$$\operatorname{Ran}(y_j) \subset p^{-1}([\gamma,\delta]) = [\alpha,\beta] \sqsubset\sqsubset (a,b) \text{ for all } j.$$

In particular $\sup\{|\dot{y}_j(t)| : t \in \mathbb{R} \text{ and } j\} < \infty$ since

$$\dot{y}_j(t) = g_j(t) - p(y_j(t)) \subset [\gamma,\delta] - [\gamma,\delta] \tag{34.46}$$

which is a compact subset of $\mathbb{R}$. The Arzela-Ascoli Theorem (see Theoerem 36.11 below) now allows us to assume, by passing to a subsequence if necessary, that y_j is converging uniformly to $y \in C_{2\pi}^0(\mathbb{R},[\alpha,\beta])$. It now follows that

$$\dot{y}_j(t) = g_j(t) - p(y_j(t)) \to g - p(y)$$

uniformly in t. Hence we conclude that $y \in C^1_{2\pi}(\mathbb{R},\mathbb{R}) \cap C^0_{2\pi}(\mathbb{R},[\alpha,\beta])$, $\dot{y}_j \to y$ and $P(y) = g$. This has proved that $g \in W$ and hence that W is relatively closed in $\tilde{V}$. ■

34.8 Contraction Mapping Principle

Some of the arguments used in Chapter 33 as well as this chapter may be abstracted to a general principle of finding fixed points on a complete metric space. This is the content of this section.

Theorem 34.32 (Contraction Mapping Principle). *Suppose that (X,ρ) is a complete metric space and $S : X \to X$ is a contraction, i.e. there exists $\alpha \in (0,1)$ such that $\rho(S(x),S(y)) \le \alpha\rho(x,y)$ for all $x,y \in X$. Then S has a unique fixed point in X, i.e. there exists a unique point $x \in X$ such that $S(x) = x$.*

Proof. For uniqueness suppose that x and x' are two fixed points of S, then

$$\rho(x,x') = \rho(S(x),S(x')) \le \alpha\rho(x,x').$$

Therefore $(1-\alpha)\rho(x,x') \le 0$ which implies that $\rho(x,x') = 0$ since $1-\alpha > 0$. Thus $x = x'$. For existence, let $x_0 \in X$ be any point in X and define $x_n \in X$ inductively by $x_{n+1} = S(x_n)$ for $n \ge 0$. We will show that $x := \lim_{n\to\infty} x_n$ exists in X and because S is continuous this will imply,

$$x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} S(x_n) = S(\lim_{n\to\infty} x_n) = S(x),$$

showing x is a fixed point of S. So to finish the proof, because X is complete, it suffices to show $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence in X. An easy inductive computation shows, for $n \ge 0$, that

$$\rho(x_{n+1},x_n) = \rho(S(x_n),S(x_{n-1})) \le \alpha\rho(x_n,x_{n-1}) \le \cdots \le \alpha^n\rho(x_1,x_0).$$

Another inductive argument using the triangle inequality shows, for $m > n$, that,

$$\rho(x_m,x_n) \le \rho(x_m,x_{m-1}) + \rho(x_{m-1},x_n) \le \cdots \le \sum_{k=n}^{m-1}\rho(x_{k+1},x_k).$$

Combining the last two inequalities gives (using again that $\alpha \in (0,1)$),

$$\rho(x_m,x_n) \le \sum_{k=n}^{m-1}\alpha^k\rho(x_1,x_0) \le \rho(x_1,x_0)\alpha^n\sum_{l=0}^{\infty}\alpha^l = \rho(x_1,x_0)\frac{\alpha^n}{1-\alpha}.$$

This last equation shows that $\rho(x_m,x_n) \to 0$ as $m,n \to \infty$, i.e. $\{x_n\}_{n=0}^\infty$ is a Cauchy sequence. ■

Corollary 34.33 (Contraction Mapping Principle II). *Suppose that (X,ρ) is a complete metric space and $S : X \to X$ is a continuous map such that $S^{(n)}$ is a contraction for some $n \in \mathbb{N}$. Here*

$$S^{(n)} := \overbrace{S \circ S \circ \ldots \circ S}^{n \text{ times}}$$

and we are assuming there exists $\alpha \in (0,1)$ such that $\rho(S^{(n)}(x),S^{(n)}(y)) \le \alpha\rho(x,y)$ for all $x,y \in X$. Then S has a unique fixed point in X.

Proof. Let $T := S^{(n)}$, then $T : X \to X$ is a contraction and hence T has a unique fixed point $x \in X$. Since any fixed point of S is also a fixed point of T, we see if S has a fixed point then it must be x. Now

$$T(S(x)) = S^{(n)}(S(x)) = S(S^{(n)}(x)) = S(T(x)) = S(x),$$

which shows that $S(x)$ is also a fixed point of T. Since T has only one fixed point, we must have that $S(x) = x$. So we have shown that x is a fixed point of S and this fixed point is unique. ■

Lemma 34.34. *Suppose that (X,ρ) is a complete metric space, $n \in \mathbb{N}$, Z is a topological space, and $\alpha \in (0,1)$. Suppose for each $z \in Z$ there is a map $S_z : X \to X$ with the following properties:*

Contraction property $\rho(S_z^{(n)}(x),S_z^{(n)}(y)) \le \alpha\rho(x,y)$ for all $x,y \in X$ and $z \in Z$.
Continuity in z For each $x \in X$ the map $z \in Z \to S_z(x) \in X$ is continuous.

By Corollary 34.33 above, for each $z \in Z$ there is a unique fixed point $G(z) \in X$ of S_z.
Conclusion: *The map $G : Z \to X$ is continuous.*

Proof. Let $T_z := S_z^{(n)}$. If $z,w \in Z$, then

$$\begin{aligned}\rho(G(z),G(w)) &= \rho(T_z(G(z)),T_w(G(w))) \\ &\le \rho(T_z(G(z)),T_w(G(z))) + \rho(T_w(G(z)),T_w(G(w))) \\ &\le \rho(T_z(G(z)),T_w(G(z))) + \alpha\rho(G(z),G(w)).\end{aligned}$$

Solving this inequality for $\rho(G(z),G(w))$ gives

$$\rho(G(z),G(w)) \le \frac{1}{1-\alpha}\rho(T_z(G(z)),T_w(G(z))).$$

Since $w \to T_w(G(z))$ is continuous it follows from the above equation that $G(w) \to G(z)$ as $w \to z$, i.e. G is continuous. ■

34.9 Exercises

Exercise 34.3. Suppose that $A : \mathbb{R} \to L(X)$ is a continuous function and $V : \mathbb{R} \to L(X)$ is the unique solution to the linear differential equation

$$\dot{V}(t) = A(t)V(t) \text{ with } V(0) = I. \tag{34.47}$$

Assuming that $V(t)$ is invertible for all $t \in \mathbb{R}$, show that $V^{-1}(t) := [V(t)]^{-1}$ must solve the differential equation

$$\frac{d}{dt}V^{-1}(t) = -V^{-1}(t)A(t) \text{ with } V^{-1}(0) = I. \tag{34.48}$$

See Exercise 32.12 as well.

Exercise 34.4 (Differential Equations with Parameters). Let W be another Banach space, $U \times V \subset_o X \times W$ and $Z \in C^1(U \times V, X)$. For each $(x, w) \in U \times V$, let $t \in J_{x,w} \to \varphi(t, x, w)$ denote the maximal solution to the ODE

$$\dot{y}(t) = Z(y(t), w) \text{ with } y(0) = x \tag{34.49}$$

and

$$\mathcal{D} := \{(t, x, w) \in \mathbb{R} \times U \times V : t \in J_{x,w}\}$$

as in Exercise 33.8.

1. Prove that φ is C^1 and that $D_w\varphi(t, x, w)$ solves the differential equation:
$$\frac{d}{dt}D_w\varphi(t, x, w) = (D_xZ)(\varphi(t, x, w), w)D_w\varphi(t, x, w) + (D_wZ)(\varphi(t, x, w), w)$$
with $D_w\varphi(0, x, w) = 0 \in L(W, X)$. **Hint:** See the hint for Exercise 33.8 with the reference to Theorem 33.16 being replace by Theorem 34.28.
2. Also show with the aid of Duhamel's principle (Exercise 32.22) and Theorem 34.28 that
$$D_w\varphi(t, x, w) = D_x\varphi(t, x, w)\int_0^t D_x\varphi(\tau, x, w)^{-1}(D_wZ)(\varphi(\tau, x, w), w)d\tau$$

Exercise 34.5. (Differential of e^A) Let $f : L(X) \to GL(X)$ be the exponential function $f(A) = e^A$. Prove that f is differentiable and that

$$Df(A)B = \int_0^1 e^{(1-t)A}Be^{tA}\,dt. \tag{34.50}$$

Hint: Let $B \in L(X)$ and define $w(t, s) = e^{t(A+sB)}$ for all $t, s \in \mathbb{R}$. Notice that

$$dw(t, s)/dt = (A + sB)w(t, s) \text{ with } w(0, s) = I \in L(X). \tag{34.51}$$

Use Exercise 34.4 to conclude that w is C^1 and that $w'(t, 0) := dw(t, s)/ds|_{s=0}$ satisfies the differential equation,

$$\frac{d}{dt}w'(t, 0) = Aw'(t, 0) + Be^{tA} \text{ with } w(0, 0) = 0 \in L(X). \tag{34.52}$$

Solve this equation by Duhamel's principle (Exercise 32.22) and then apply Proposition 34.14 to conclude that f is differentiable with differential given by Eq. (34.50).

Exercise 34.6 (Local ODE Existence). Let S_x be defined as in Eq. (33.17) from the proof of Theorem 33.5. Verify that S_x satisfies the hypothesis of Corollary 34.33. In particular we could have used Corollary 34.33 to prove Theorem 33.5.

Exercise 34.7 (Local ODE Existence Again). Let $J = (-1, 1)$, $Z \in C^1(X, X)$, $Y := BC(J, X)$ and for $y \in Y$ and $s \in J$ let $y_s \in Y$ be defined by $y_s(t) := y(st)$. Use the following outline to prove the ODE

$$\dot{y}(t) = Z(y(t)) \text{ with } y(0) = x \tag{34.53}$$

has a unique solution for small t and this solution is C^1 in x.

1. If y solves Eq. (34.53) then y_s solves
$$\dot{y}_s(t) = sZ(y_s(t)) \text{ with } y_s(0) = x$$
or equivalently
$$y_s(t) = x + s\int_0^t Z(y_s(\tau))d\tau. \tag{34.54}$$
Notice that when $s = 0$, the unique solution to this equation is $y_0(t) = x$.
2. Let $F : J \times Y \to J \times Y$ be defined by
$$F(s, y) := (s, y(t) - s\int_0^t Z(y(\tau))d\tau).$$
Show the differential of F is given by
$$F'(s, y)(a, v) = \left(a, t \to v(t) - s\int_0^t Z'(y(\tau))v(\tau)d\tau - a\int_0^{\cdot} Z(y(\tau))d\tau\right).$$
3. Verify $F'(0, y) : \mathbb{R} \times Y \to \mathbb{R} \times Y$ is invertible for all $y \in Y$ and notice that $F(0, y) = (0, y)$.
4. For $x \in X$, let $C_x \in Y$ be the constant path at x, i.e. $C_x(t) = x$ for all $t \in J$. Use the inverse function Theorem 34.25 to conclude there exists $\varepsilon > 0$ and a C^1 map $\varphi : (-\varepsilon, \varepsilon) \times B(x_0, \varepsilon) \to Y$ such that
$$F(s, \varphi(s, x)) = (s, C_x) \text{ for all } (s, x) \in (-\varepsilon, \varepsilon) \times B(x_0, \varepsilon).$$

5. Show, for $s \le \varepsilon$ that $y_s(t) := \varphi(s,x)(t)$ satisfies Eq. (34.54). Now define $y(t,x) = \varphi(\varepsilon/2, x)(2t/\varepsilon)$ and show $y(t,x)$ solve Eq. (34.53) for $|t| < \varepsilon/2$ and $x \in B(x_0, \varepsilon)$.

Exercise 34.8. Show P defined in Theorem 34.31 is continuously differentiable and $P'(y)h = \dot{h} + p'(y)h$.

Exercise 34.9. Embedded sub-manifold problems.

Exercise 34.10. Lagrange Multiplier problems.

34.9.1 Alternate construction of g. To be made into an exercise.

Suppose $U \subset_o X$ and $f : U \to Y$ is a C^2 – function. Then we are looking for a function $g(y)$ such that $f(g(y)) = y$. Fix an $x_0 \in U$ and $y_0 = f(x_0) \in Y$. Suppose such a g exists and let $x(t) = g(y_0 + th)$ for some $h \in Y$. Then differentiating $f(x(t)) = y_0 + th$ implies

$$\frac{d}{dt} f(x(t)) = f'(x(t))\dot{x}(t) = h$$

or equivalently that

$$\dot{x}(t) = [f'(x(t))]^{-1} h = Z(h, x(t)) \text{ with } x(0) = x_0 \tag{34.55}$$

where $Z(h,x) = [f'(x(t))]^{-1} h$. Conversely if x solves Eq. (34.55) we have $\frac{d}{dt} f(x(t)) = h$ and hence that

$$f(x(1)) = y_0 + h.$$

Thus if we define

$$g(y_0 + h) := e^{Z(h,\cdot)}(x_0),$$

then $f(g(y_0 + h)) = y_0 + h$ for all h sufficiently small. This shows f is an open mapping.

Part VIII

Topological Spaces

Topological Space Basics

Using the metric space results above as motivation we will axiomatize the notion of being an open set to more general settings. See [12,13] and many more references on point-set topology.

Definition 35.1. *A collection of subsets τ of X is a* ***topology*** *if*

1. $\emptyset, X \in \tau$.
2. *τ is closed under arbitrary unions, i.e. if $V_\alpha \in \tau$, for $\alpha \in I$ then $\bigcup_{\alpha \in I} V_\alpha \in \tau$.*
3. *τ is closed under finite intersections, i.e. if $V_1, \ldots, V_n \in \tau$ then $V_1 \cap \cdots \cap V_n \in \tau$.*

A pair (X, τ) where τ is a topology on X will be called a ***topological space****.*

Notation 35.2 *Let (X, τ) be a topological space.*

1. *The elements, $V \in \tau$, are called* ***open*** *sets. We will often write $V \subset_o X$ to indicate V is an open subset of X.*
2. *A subset $F \subset X$ is* ***closed*** *if F^c is open and we will write $F \sqsubset X$ if F is a closed subset of X.*
3. *An* ***open neighborhood*** *of a point $x \in X$ is an open set $V \subset X$ such that $x \in V$. Let $\tau_x = \{V \in \tau : x \in V\}$ denote the collection of open neighborhoods of x.*
4. *A subset $W \subset X$ is a* ***neighborhood*** *of x if there exists $V \in \tau_x$ such that $V \subset W$.*
5. *A collection $\eta \subset \tau_x$ is called a* ***neighborhood base*** *at $x \in X$ if for all $V \in \tau_x$ there exists $W \in \eta$ such that $W \subset V$.*

The notation τ_x should not be confused with

$$\tau_{\{x\}} := i_{\{x\}}^{-1}(\tau) = \{\{x\} \cap V : V \in \tau\} = \{\emptyset, \{x\}\}.$$

Example 35.3. 1. Let (X, d) be a metric space, we write τ_d for the collection of d – open sets in X. We have already seen that τ_d is a topology, see Exercise 13.2. The collection of sets $\eta = \{B_x(\varepsilon) : \varepsilon \in \mathbb{D}\}$ where $\mathbb{D}$ is any dense subset of $(0, 1]$ is a neighborhood base at x.
2. Let X be any set, then $\tau = 2^X$ is the discrete topology on X. In this topology all subsets of X are both open and closed. At the opposite extreme we have the **trivial topology**, $\tau = \{\emptyset, X\}$. In this topology only the empty set and X are open (closed).
3. Let $X = \{1, 2, 3\}$, then $\tau = \{\emptyset, X, \{2, 3\}\}$ is a topology on X which does not come from a metric.
4. Again let $X = \{1, 2, 3\}$. Then $\tau = \{\{1\}, \{2, 3\}, \emptyset, X\}$. is a topology, and the sets X, $\{1\}$, $\{2, 3\}, \emptyset$ are open and closed. The sets $\{1, 2\}$ and $\{1, 3\}$ are neither open nor closed.

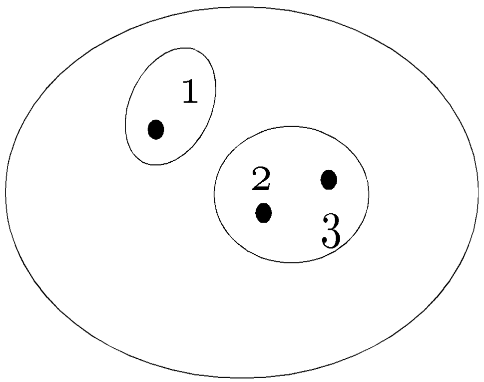


Fig. 35.1. A topology.

Definition 35.4. *Let (X, τ_X) and (Y, τ_Y) be topological spaces. A function $f : X \to Y$ is* ***continuous*** *if*

$$f^{-1}(\tau_Y) := \left\{f^{-1}(V) : V \in \tau_Y\right\} \subset \tau_X.$$

We will also say that f is τ_X/τ_Y –continuous or (τ_X, τ_Y) – continuous. Let $C(X, Y)$ denote the set of continuous functions from X to Y.

Exercise 35.1. Show $f : X \to Y$ is continuous iff $f^{-1}(C)$ is closed in X for all closed subsets C of Y.

Definition 35.5. *A map $f : X \to Y$ between topological spaces is called a* ***homeomorphism*** *provided that f is bijective, f is continuous and $f^{-1} : Y \to X$ is continuous. If there exists $f : X \to Y$ which is a homeomorphism, we say* *that X and Y are homeomorphic. (As topological spaces X and Y are essentially the same.)*

35.1 Constructing Topologies and Checking Continuity

Proposition 35.6. *Let $\mathcal{E}$ be any collection of subsets of X. Then there exists a unique smallest topology $\tau(\mathcal{E})$ which contains $\mathcal{E}$.*

Proof. Since 2^X is a topology and $\mathcal{E} \subset 2^X$, $\mathcal{E}$ is always a subset of a topology. It is now easily seen that

$$\tau(\mathcal{E}) := \bigcap\{\tau : \tau \text{ is a topology and } \mathcal{E} \subset \tau\}$$

is a topology which is clearly the smallest possible topology containing $\mathcal{E}$. ■

The following proposition gives an explicit descriptions of $\tau(\mathcal{E})$.

Proposition 35.7. *Let X be a set and $\mathcal{E} \subset 2^X$. For simplicity of notation, assume that $X, \emptyset \in \mathcal{E}$. (If this is not the case simply replace $\mathcal{E}$ by $\mathcal{E} \cup \{X, \emptyset\}$.) Then*

$$\tau(\mathcal{E}) := \{\textit{arbitrary unions of finite intersections of elements from } \mathcal{E}\}. \quad (35.1)$$

Proof. Let τ be given as in the right side of Eq. (35.1). From the definition of a topology any topology containing $\mathcal{E}$ must contain τ and hence $\mathcal{E} \subset \tau \subset \tau(\mathcal{E})$. The proof will be completed by showing τ is a topology. The validation of τ being a topology is routine except for showing that τ is closed under taking finite intersections. Let $V, W \in \tau$ which by definition may be expressed as

$$V = \cup_{\alpha \in A} V_\alpha \text{ and } W = \cup_{\beta \in B} W_\beta,$$

where V_α and W_β are sets which are finite intersection of elements from $\mathcal{E}$. Then

$$V \cap W = (\cup_{\alpha \in A} V_\alpha) \cap (\cup_{\beta \in B} W_\beta) = \bigcup_{(\alpha,\beta) \in A \times B} V_\alpha \cap W_\beta.$$

Since for each $(\alpha, \beta) \in A \times B$, $V_\alpha \cap W_\beta$ is still a finite intersection of elements from $\mathcal{E}$, $V \cap W \in \tau$ showing τ is closed under taking finite intersections. ■

Definition 35.8. *Let (X, τ) be a topological space. We say that $\mathcal{S} \subset \tau$ is a **sub-base** for the topology τ iff $\tau = \tau(\mathcal{S})$ and $X = \cup\mathcal{S} := \cup_{V \in \mathcal{S}} V$. We say $\mathcal{V} \subset \tau$ is a **base** for the topology τ iff $\mathcal{V}$ is a sub-base with the property that every element $V \in \tau$ may be written as*

$$V = \cup\{B \in \mathcal{V} : B \subset V\}.$$

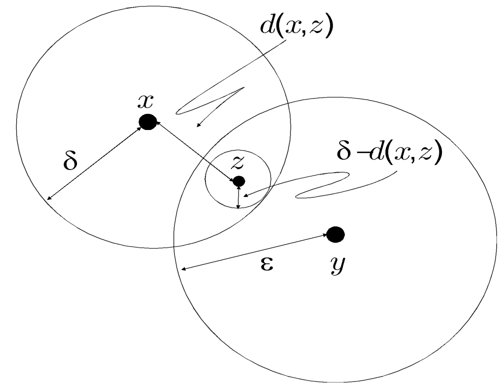


Fig. 35.2. Fitting balls in the intersection.

Exercise 35.2. Suppose that $\mathcal{S}$ is a sub-base for a topology τ on a set X.

1. Show $\mathcal{V} := \mathcal{S}_f$ ($\mathcal{S}_f$ is the collection of finite intersections of elements from $\mathcal{S}$) is a base for τ.
2. Show $\mathcal{S}$ is itself a base for τ iff
$$V_1 \cap V_2 = \cup\{S \in \mathcal{S} : S \subset V_1 \cap V_2\}.$$
for every pair of sets $V_1, V_2 \in \mathcal{S}$.

Remark 35.9. Let (X, d) be a metric space, then $\mathcal{E} = \{B_x(\delta) : x \in X \text{ and } \delta > 0\}$ is a base for τ_d – the topology associated to the metric d. This is the content of Exercise 13.3.

Let us check directly that $\mathcal{E}$ is a base for a topology. Suppose that $x, y \in X$ and $\varepsilon, \delta > 0$. If $z \in B(x, \delta) \cap B(y, \varepsilon)$, then

$$B(z, \alpha) \subset B(x, \delta) \cap B(y, \varepsilon) \quad (35.2)$$

where $\alpha = \min\{\delta - d(x, z), \varepsilon - d(y, z)\}$, see Figure 35.2. This is a formal consequence of the triangle inequality. For example let us show that $B(z, \alpha) \subset B(x, \delta)$. By the definition of α, we have that $\alpha \leq \delta - d(x, z)$ or that $d(x, z) \leq \delta - \alpha$. Hence if $w \in B(z, \alpha)$, then

$$d(x, w) \leq d(x, z) + d(z, w) \leq \delta - \alpha + d(z, w) < \delta - \alpha + \alpha = \delta$$

which shows that $w \in B(x, \delta)$. Similarly we show that $w \in B(y, \varepsilon)$ as well.

Owing to Exercise 35.2, this shows $\mathcal{E}$ is a base for a topology. We do not need to use Exercise 35.2 here since in fact Equation (35.2) may be generalized to finite intersection of balls. Namely if $x_i \in X$, $\delta_i > 0$ and $z \in \cap_{i=1}^n B(x_i, \delta_i)$, then

$$B(z,\alpha) \subset \cap_{i=1}^{n} B(x_i,\delta_i) \tag{35.3}$$

where now $\alpha := \min\{\delta_i - d(x_i,z) : i = 1,2,\dots,n\}$. By Eq. (35.3) it follows that any finite intersection of open balls may be written as a union of open balls.

Exercise 35.3. Suppose $f : X \to Y$ is a function and τ_X and τ_Y are topologies on X and Y respectively. Show

$$f^{-1}\tau_Y := \left\{f^{-1}(V) \subset X : V \in \tau_Y\right\} \text{ and } f_*\tau_X := \left\{V \subset Y : f^{-1}(V) \in \tau_X\right\}$$

(as in Notation 2.7) are also topologies on X and Y respectively.

Remark 35.10. Let $f : X \to Y$ be a function. Given a topology $\tau_Y \subset 2^Y$, the topology $\tau_X := f^{-1}(\tau_Y)$ is the smallest topology on X such that f is (τ_X,τ_Y) - continuous. Similarly, if τ_X is a topology on X then $\tau_Y = f_*\tau_X$ is the largest topology on Y such that f is (τ_X,τ_Y) - continuous.

Definition 35.11. *Let (X,τ) be a topological space and A subset of X. The* ***relative topology*** *or* ***induced topology*** *on A is the collection of sets*

$$\tau_A = i_A^{-1}(\tau) = \{A \cap V : V \in \tau\},$$

where $i_A : A \to X$ is the inclusion map as in Definition 2.8.

Lemma 35.12. *The relative topology, τ_A, is a topology on A. Moreover a subset $B \subset A$ is τ_A – closed iff there is a τ – closed subset, C, of X such that $B = C \cap A$.*

Proof. The first assertion is a consequence of Exercise 35.3. For the second, $B \subset A$ is τ_A – closed iff $A \setminus B = A \cap V$ for some $V \in \tau$ which is equivalent to $B = A \setminus (A \cap V) = A \cap V^c$ for some $V \in \tau$. ■

Exercise 35.4. Show if (X,d) is a metric space and $\tau = \tau_d$ is the topology coming from d, then $(\tau_d)_A$ is the topology induced by making A into a metric space using the metric $d|_{A\times A}$.

Lemma 35.13. *Suppose that (X,τ_X), (Y,τ_Y) and (Z,τ_Z) are topological spaces. If $f : (X,\tau_X) \to (Y,\tau_Y)$ and $g : (Y,\tau_Y) \to (Z,\tau_Z)$ are continuous functions then $g \circ f : (X,\tau_X) \to (Z,\tau_Z)$ is continuous as well.*

Proof. This is easy since by assumption $g^{-1}(\tau_Z) \subset \tau_Y$ and $f^{-1}(\tau_Y) \subset \tau_X$ so that

$$(g \circ f)^{-1}(\tau_Z) = f^{-1}\left(g^{-1}(\tau_Z)\right) \subset f^{-1}(\tau_Y) \subset \tau_X.$$

■

The following elementary lemma turns out to be extremely useful because it may be used to greatly simplify the verification that a given function is continuous.

Lemma 35.14. *Suppose that $f : X \to Y$ is a function, $\mathcal{E} \subset 2^Y$ and $A \subset Y$, then*

$$\tau\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\tau(\mathcal{E})) \text{ and} \tag{35.4}$$

$$\tau(\mathcal{E}_A) = (\tau(\mathcal{E}))_A. \tag{35.5}$$

Moreover, if $\tau_Y = \tau(\mathcal{E})$ and τ_X is a topology on X, then f is (τ_X,τ_Y) – continuous iff $f^{-1}(\mathcal{E}) \subset \tau_X$.

Proof. We will give two proof of Eq. (35.4). The first proof is more constructive than the second, but the second proof works in the context of σ – algebras, see Lemma 9.3.

First Proof. There is no harm (as the reader should verify) in replacing $\mathcal{E}$ by $\mathcal{E} \cup \{\emptyset, Y\}$ if necessary so that we may assume that $\emptyset, Y \in \mathcal{E}$. By Proposition 35.7, the general element V of $\tau(\mathcal{E})$ is an arbitrary unions of finite intersections of elements from $\mathcal{E}$. Since f^{-1} preserves all of the set operations, it follows that $f^{-1}\tau(\mathcal{E})$ consists of sets which are arbitrary unions of finite intersections of elements from $f^{-1}\mathcal{E}$, which is precisely $\tau\left(f^{-1}(\mathcal{E})\right)$ by another application of Proposition 35.7.

Second Proof. By Exercise 35.3, $f^{-1}(\tau(\mathcal{E}))$ is a topology and since $\mathcal{E} \subset \tau(\mathcal{E})$, $f^{-1}(\mathcal{E}) \subset f^{-1}(\tau(\mathcal{E}))$. It now follows that $\tau(f^{-1}(\mathcal{E})) \subset f^{-1}(\tau(\mathcal{E}))$. For the reverse inclusion notice that

$$f_*\tau\left(f^{-1}(\mathcal{E})\right) = \left\{B \subset Y : f^{-1}(B) \in \tau\left(f^{-1}(\mathcal{E})\right)\right\}$$

is a topology which contains $\mathcal{E}$ and thus $\tau(\mathcal{E}) \subset f_*\tau\left(f^{-1}(\mathcal{E})\right)$. Hence if $B \in \tau(\mathcal{E})$ we know that $f^{-1}(B) \in \tau\left(f^{-1}(\mathcal{E})\right)$, i.e. $f^{-1}(\tau(\mathcal{E})) \subset \tau\left(f^{-1}(\mathcal{E})\right)$ and Eq. (35.4) has been proved. Applying Eq. (35.4) with $X = A$ and $f = i_A$ being the inclusion map implies

$$(\tau(\mathcal{E}))_A = i_A^{-1}(\tau(\mathcal{E})) = \tau(i_A^{-1}(\mathcal{E})) = \tau(\mathcal{E}_A).$$

Lastly if $f^{-1}\mathcal{E} \subset \tau_X$, then $f^{-1}\tau(\mathcal{E}) = \tau\left(f^{-1}\mathcal{E}\right) \subset \tau_X$ which shows f is (τ_X,τ_Y) – continuous. ■

Corollary 35.15. *If (X,τ) is a topological space and $f : X \to \mathbb{R}$ is a function then the following are equivalent:*

1. *f is $(\tau,\tau_{\mathbb{R}})$ - continuous,*
2. *$f^{-1}((a,b)) \in \tau$ for all $-\infty < a < b < \infty$,*
3. *$f^{-1}((a,\infty)) \in \tau$ and $f^{-1}((-\infty,b)) \in \tau$ for all $a,b \in \mathbb{Q}$.*

(We are using $\tau_{\mathbb{R}}$ to denote the standard topology on $\mathbb{R}$ induced by the metric $d(x,y) = |x-y|$.)

Proof. Apply Lemma 35.14 with appropriate choices of $\mathcal{E}$. ■

Definition 35.16. *Let (X,τ_X) and (Y,τ_Y) be topological spaces. A function $f:X\to Y$ is* ***continuous at a point*** *$x\in X$ if for every open neighborhood V of $f(x)$ there is an open neighborhood U of x such that $U\subset f^{-1}(V)$. See Figure 35.3.*

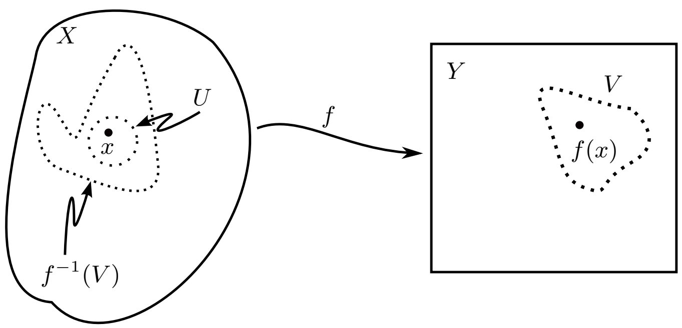


Fig. 35.3. Checking that a function is continuous at $x\in X$.

Exercise 35.5. Show $f:X\to Y$ is continuous (Definition 35.16) iff f is continuous at all points $x\in X$.

Definition 35.17. *Given topological spaces (X,τ) and (Y,τ') and a subset $A\subset X$. We say a function $f:A\to Y$ is* ***continuous*** *iff f is τ_A/τ' – continuous.*

Definition 35.18. *Let (X,τ) be a topological space and $A\subset X$. A collection of subsets $\mathcal{U}\subset\tau$ is an* ***open cover*** *of A if $A\subset\bigcup\mathcal{U}:=\bigcup_{U\in\mathcal{U}}U$.*

Proposition 35.19 (Localizing Continuity). *Let (X,τ) and (Y,τ') be topological spaces and $f:X\to Y$ be a function.*

1. *If f is continuous and $A\subset X$ then $f|_A:A\to Y$ is continuous.*
2. *Suppose there exists an open cover, $\mathcal{U}\subset\tau$, of X such that $f|_A$ is continuous for all $A\in\mathcal{U}$, then f is continuous.*

Proof. 1. If $f:X\to Y$ is continuous then $f|_A=f\circ i_A$ is the composition of continuous maps and hence continuous. Here $i_A:A\to X$ is the inclusion map.

2. Let $V\in\tau'$, then

$$f^{-1}(V)=\cup_{A\in\mathcal{U}}\left(f^{-1}(V)\cap A\right)=\cup_{A\in\mathcal{U}}f|_A^{-1}(V). \tag{35.6}$$

Since each $A\in\mathcal{U}$ is open, $\tau_A\subset\tau$ and by assumption, $f|_A^{-1}(V)\in\tau_A\subset\tau$. Hence Eq. (35.6) shows $f^{-1}(V)$ is a union of τ – open sets and hence is also τ – open. ■

Exercise 35.6 (A Baby Extension Theorem). Suppose $V\in\tau$ and $f:V\to\mathbb{C}$ is a continuous function. Further assume there is a closed subset C such that $\{x\in V:f(x)\neq 0\}\subset C\subset V$, then $F:X\to\mathbb{C}$ defined by

$$F(x)=\begin{cases} f(x) & \text{if } x\in V\\ 0 & \text{if } x\notin V\end{cases}$$

is continuous.

Exercise 35.7 (Building Continuous Functions). Prove the following variant of item 2. of Proposition 35.19. Namely, suppose there exists a **finite** collection $\mathcal{F}$ of closed subsets of X such that $X=\cup_{A\in\mathcal{F}}A$ and $f|_A$ is continuous for all $A\in\mathcal{F}$, then f is continuous. Given an example showing that the assumption that $\mathcal{F}$ is finite can not be eliminated. **Hint:** consider $f^{-1}(C)$ where C is a closed subset of Y.

35.2 Product Spaces I

Definition 35.20. *Let X be a set and suppose there is a collection of topological spaces $\{(Y_\alpha,\tau_\alpha):\alpha\in A\}$ and functions $f_\alpha:X\to Y_\alpha$ for all $\alpha\in A$. Let $\tau(f_\alpha:\alpha\in A)$ denote the smallest topology on X such that each f_α is continuous, i.e.*

$$\tau(f_\alpha:\alpha\in A)=\tau(\cup_\alpha f_\alpha^{-1}(\tau_\alpha)).$$

A neighborhood base for this topology about a point $x\in X$ can be taken to be all sets of the form

$$V=\cap_{\alpha\in\Lambda}f_\alpha^{-1}(V_\alpha),$$

where Λ is any finite subset of A and V_α are open neighborhood of $f_\alpha(x)\in Y_\alpha$ for all $\alpha\in\Lambda$.

Proposition 35.21 (Topologies Generated by Functions). *Assuming the notation in Definition 35.20 and additionally let (Z,τ_Z) be a topological space and $g:Z\to X$ be a function. Then g is $(\tau_Z,\tau(f_\alpha:\alpha\in A))$ – continuous iff $f_\alpha\circ g$ is (τ_Z,τ_α)–continuous for all $\alpha\in A$.*

Proof. ($\Rightarrow$) If g is $(\tau_Z,\tau(f_\alpha:\alpha\in A))$ – continuous, then the composition $f_\alpha\circ g$ is (τ_Z,τ_α) – continuous by Lemma 35.13. ($\Leftarrow$) Let

$$\tau_X=\tau(f_\alpha:\alpha\in A)=\tau\left(\cup_{\alpha\in A}f_\alpha^{-1}(\tau_\alpha)\right).$$

If $f_\alpha\circ g$ is (τ_Z,τ_α) – continuous for all α, then

$$g^{-1}f_\alpha^{-1}(\tau_\alpha)\subset\tau_Z\,\forall\,\alpha\in A$$

and therefore

$$g^{-1}\left(\cup_{\alpha\in A}f_\alpha^{-1}(\tau_\alpha)\right)=\cup_{\alpha\in A}g^{-1}f_\alpha^{-1}(\tau_\alpha)\subset\tau_Z$$

Hence

$$g^{-1}(\tau_X)=g^{-1}\left(\tau\left(\cup_{\alpha\in A}f_\alpha^{-1}(\tau_\alpha)\right)\right)=\tau(g^{-1}\left(\cup_{\alpha\in A}f_\alpha^{-1}(\tau_\alpha)\right)\subset\tau_Z$$

which shows that g is (τ_Z,τ_X) – continuous. ■

Let $\{(X_\alpha,\tau_\alpha)\}_{\alpha\in A}$ be a collection of topological spaces, $X=X_A=\prod_{\alpha\in A}X_\alpha$ and $\pi_\alpha:X_A\to X_\alpha$ be the canonical projection map as in Notation 2.2.

Definition 35.22. *The* ***product topology*** $\tau=\otimes_{\alpha\in A}\tau_\alpha$ *is the smallest topology on* X_A *such that each projection* π_α *is continuous. Explicitly,* τ *is the topology generated by the collection of sets,*

$$\mathcal{E}=\{\pi_\alpha^{-1}(V_\alpha):\alpha\in A,V_\alpha\in\tau_\alpha\}=\cup_{\alpha\in A}\pi^{-1}\tau_\alpha. \tag{35.7}$$

Applying Proposition 35.21 in this setting implies the following proposition.

Proposition 35.23. *Suppose* Y *is a topological space and* $f:Y\to X_A$ *is a map. Then* f *is continuous iff* $\pi_\alpha\circ f:Y\to X_\alpha$ *is continuous for all* $\alpha\in A$. *In particular if* $A=\{1,2,\dots,n\}$ *so that* $X_A=X_1\times X_2\times\cdots\times X_n$ *and* $f(y)=(f_1(y),f_2(y),\dots,f_n(y))\in X_1\times X_2\times\cdots\times X_n$, *then* $f:Y\to X_A$ *is continuous iff* $f_i:Y\to X_i$ *is continuous for all* i.

Proposition 35.24. *Suppose that* (X,τ) *is a topological space and* $\{f_n\}\subset X^A$ *(see Notation 2.2) is a sequence. Then* $f_n\to f$ *in the product topology of* X^A *iff* $f_n(\alpha)\to f(\alpha)$ *for all* $\alpha\in A$.

Proof. Since π_α is continuous, if $f_n\to f$ then $f_n(\alpha)=\pi_\alpha(f_n)\to\pi_\alpha(f)=f(\alpha)$ for all $\alpha\in A$. Conversely, $f_n(\alpha)\to f(\alpha)$ for all $\alpha\in A$ iff $\pi_\alpha(f_n)\to\pi_\alpha(f)$ for all $\alpha\in A$. Therefore if $V=\pi_\alpha^{-1}(V_\alpha)\in\mathcal{E}$ (with $\mathcal{E}$ as in Eq. (35.7)) and $f\in V$, then $\pi_\alpha(f)\in V_\alpha$ and $\pi_\alpha(f_n)\in V_\alpha$ for a.a. n and hence $f_n\in V$ for a.a. n. This shows that $f_n\to f$ as $n\to\infty$. ■

Proposition 35.25. *Suppose that* $(X_\alpha,\tau_\alpha)_{\alpha\in A}$ *is a collection of topological spaces and* $\otimes_{\alpha\in A}\tau_\alpha$ *is the product topology on* $X:=\prod_{\alpha\in A}X_\alpha$.

1. *If* $\mathcal{E}_\alpha\subset\tau_\alpha$ *generates* τ_α *for each* $\alpha\in A$, *then*
$$\otimes_{\alpha\in A}\tau_\alpha=\tau\left(\cup_{\alpha\in A}\pi_\alpha^{-1}(\mathcal{E}_\alpha)\right) \tag{35.8}$$
2. *If* $\mathcal{B}_\alpha\subset\tau_\alpha$ *is a base for* τ_α *for each* α, *then the collection of sets,* $\mathcal{V}$, *of the form*
$$V=\cap_{\alpha\in\Lambda}\pi_\alpha^{-1}V_\alpha=\prod_{\alpha\in\Lambda}V_\alpha\times\prod_{\alpha\notin\Lambda}X_\alpha=:V_\Lambda\times X_{A\setminus\Lambda}, \tag{35.9}$$
where $\Lambda\subset\subset A$ *and* $V_\alpha\in\mathcal{B}_\alpha$ *for all* $\alpha\in\Lambda$ *is base for* $\otimes_{\alpha\in A}\tau_\alpha$.

Proof. 1. Since

$$\begin{aligned}\cup_\alpha\pi_\alpha^{-1}\mathcal{E}_\alpha\subset\cup_\alpha\pi_\alpha^{-1}\tau_\alpha&=\cup_\alpha\pi_\alpha^{-1}(\tau(\mathcal{E}_\alpha))\\&=\cup_\alpha\tau(\pi_\alpha^{-1}\mathcal{E}_\alpha)\subset\tau\left(\cup_\alpha\pi_\alpha^{-1}\mathcal{E}_\alpha\right),\end{aligned}$$

it follows that

$$\tau\left(\cup_\alpha\pi_\alpha^{-1}\mathcal{E}_\alpha\right)\subset\otimes_\alpha\tau_\alpha\subset\tau\left(\cup_\alpha\pi_\alpha^{-1}\mathcal{E}_\alpha\right).$$

2. Now let $\mathcal{U}=\left[\cup_\alpha\pi_\alpha^{-1}\tau_\alpha\right]_f$ denote the collection of sets consisting of finite intersections of elements from $\cup_\alpha\pi_\alpha^{-1}\tau_\alpha$. Notice that $\mathcal{U}$ may be described as those sets in Eq. (35.9) where $V_\alpha\in\tau_\alpha$ for all $\alpha\in\Lambda$. By Exercise 35.2, $\mathcal{U}$ is a base for the product topology, $\otimes_{\alpha\in A}\tau_\alpha$. Hence for $W\in\otimes_{\alpha\in A}\tau_\alpha$ and $x\in W$, there exists a $V\in\mathcal{U}$ of the form in Eq. (35.9) such that $x\in V\subset W$. Since $\mathcal{B}_\alpha$ is a base for τ_α, there exists $U_\alpha\in\mathcal{B}_\alpha$ such that $x_\alpha\in U_\alpha\subset V_\alpha$ for each $\alpha\in\Lambda$. With this notation, the set $U_\Lambda\times X_{A\setminus\Lambda}\in\mathcal{V}$ and $x\in U_\Lambda\times X_{A\setminus\Lambda}\subset V\subset W$. This shows that every open set in X may be written as a union of elements from $\mathcal{V}$, i.e. $\mathcal{V}$ is a base for the product topology. ■

Notation 35.26 *Let* $\mathcal{E}_i\subset 2^{X_i}$ *be a collection of subsets of a set* X_i *for each* $i=1,2,\dots,n$. *We will write, by abuse of notation,* $\mathcal{E}_1\times\mathcal{E}_2\times\cdots\times\mathcal{E}_n$ *for the collection of subsets of* $X_1\times\cdots\times X_n$ *of the form* $A_1\times A_2\times\cdots\times A_n$ *with* $A_i\in\mathcal{E}_i$ *for all* i. *That is we are identifying* $(A_1,A_2,\dots,A_n)$ *with* $A_1\times A_2\times\cdots\times A_n$.

Corollary 35.27. *Suppose* $A=\{1,2,\dots,n\}$ *so* $X=X_1\times X_2\times\cdots\times X_n$.

1. *If* $\mathcal{E}_i\subset 2^{X_i}$, $\tau_i=\tau(\mathcal{E}_i)$ *and* $X_i\in\mathcal{E}_i$ *for each* i, *then*
$$\tau_1\otimes\tau_2\otimes\cdots\otimes\tau_n=\tau(\mathcal{E}_1\times\mathcal{E}_2\times\cdots\times\mathcal{E}_n) \tag{35.10}$$
and in particular
$$\tau_1\otimes\tau_2\otimes\cdots\otimes\tau_n=\tau(\tau_1\times\cdots\times\tau_n). \tag{35.11}$$
2. *Furthermore if* $\mathcal{B}_i\subset\tau_i$ *is a base for the topology* τ_i *for each* i, *then* $\mathcal{B}_1\times\cdots\times\mathcal{B}_n$ *is a base for the product topology,* $\tau_1\otimes\tau_2\otimes\cdots\otimes\tau_n$.

Proof. (The proof is a minor variation on the proof of Proposition 35.25.) 1. Let $\left[\cup_{i\in A}\pi_i^{-1}(\mathcal{E}_i)\right]_f$ denotes the collection of sets which are finite intersections from $\cup_{i\in A}\pi_i^{-1}(\mathcal{E}_i)$, then, using $X_i\in\mathcal{E}_i$ for all i,

$$\cup_{i\in A}\pi_i^{-1}(\mathcal{E}_i)\subset\mathcal{E}_1\times\mathcal{E}_2\times\cdots\times\mathcal{E}_n\subset\left[\cup_{i\in A}\pi_i^{-1}(\mathcal{E}_i)\right]_f.$$

Therefore

$$\tau=\tau\left(\cup_{i\in A}\pi_i^{-1}(\mathcal{E}_i)\right)\subset\tau(\mathcal{E}_1\times\mathcal{E}_2\times\cdots\times\mathcal{E}_n)\subset\tau\left(\left[\cup_{i\in A}\pi_i^{-1}(\mathcal{E}_i)\right]_f\right)=\tau.$$

2. Observe that $\tau_1 \times \cdots \times \tau_n$ is closed under finite intersections and generates $\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n$, therefore $\tau_1 \times \cdots \times \tau_n$ is a base for the product topology. The proof that $\mathcal{B}_1 \times \cdots \times \mathcal{B}_n$ is also a base for $\tau_1 \otimes \tau_2 \otimes \cdots \otimes \tau_n$ follows the same method used to prove item 2. in Proposition 35.25. ■

Lemma 35.28. *Let (X_i, d_i) for $i = 1, \dots, n$ be metric spaces, $X := X_1 \times \cdots \times X_n$ and for $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ in X let*

$$d(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i). \tag{35.12}$$

Then the topology, τ_d, associated to the metric d is the product topology on X, i.e.

$$\tau_d = \tau_{d_1} \otimes \tau_{d_2} \otimes \cdots \otimes \tau_{d_n}.$$

Proof. Let $\rho(x, y) = \max\{d_i(x_i, y_i) : i = 1, 2, \dots, n\}$. Then ρ is equivalent to d and hence $\tau_\rho = \tau_d$. Moreover if $\varepsilon > 0$ and $x = (x_1, x_2, \dots, x_n) \in X$, then

$$B_x^\rho(\varepsilon) = B_{x_1}^{d_1}(\varepsilon) \times \cdots \times B_{x_n}^{d_n}(\varepsilon).$$

By Remark 35.9,

$$\mathcal{E} := \{B_x^\rho(\varepsilon) : x \in X \text{ and } \varepsilon > 0\}$$

is a base for τ_ρ and by Proposition 35.25 $\mathcal{E}$ is also a base for $\tau_{d_1} \otimes \tau_{d_2} \otimes \cdots \otimes \tau_{d_n}$. Therefore,

$$\tau_{d_1} \otimes \tau_{d_2} \otimes \cdots \otimes \tau_{d_n} = \tau(\mathcal{E}) = \tau_\rho = \tau_d.$$

■

35.3 Closure operations

Definition 35.29. *Let (X, τ) be a topological space and A be a subset of X.*

1. *The **closure** of A is the smallest closed set $\bar{A}$ containing A, i.e.*
$$\bar{A} := \cap \{F : A \subset F \sqsubset X\}.$$
(Because of Exercise 13.4 this is consistent with Definition 13.10 for the closure of a set in a metric space.)
2. *The **interior** of A is the largest open set A^o contained in A, i.e.*
$$A^o = \cup \{V \in \tau : V \subset A\}.$$
(With this notation the definition of a neighborhood of $x \in X$ may be stated as: $A \subset X$ is a neighborhood of a point $x \in X$ if $x \in A^o$.)
3. *The **accumulation points** of A is the set*
$$\mathrm{acc}(A) = \{x \in X : V \cap [A \setminus \{x\}] \neq \emptyset \text{ for all } V \in \tau_x\}.$$
4. *The **boundary** of A is the set $\mathrm{bd}(A) := \bar{A} \setminus A^o$.*

Remark 35.30. The relationships between the interior and the closure of a set are:

$$(A^o)^c = \bigcap \{V^c : V \in \tau \text{ and } V \subset A\} = \bigcap \{C : C \text{ is closed } C \supset A^c\} = \overline{A^c}$$

and similarly, $(\bar{A})^c = (A^c)^o$. Hence the boundary of A may be written as

$$\mathrm{bd}(A) := \bar{A} \setminus A^o = \bar{A} \cap (A^o)^c = \bar{A} \cap \overline{A^c}, \tag{35.13}$$

which is to say $\mathrm{bd}(A)$ consists of the points in both the closures of A and A^c. Notice that $\bar{A} = A^o \cup \mathrm{bd}(A) = A \cup \mathrm{bd}(A)$.

Exercise 35.8. Show that $\mathrm{bd}(A) \setminus A = \mathrm{acc}(A) \setminus A$.

Proposition 35.31. *Let $A \subset X$ and $x \in X$.*

1. *If $V \subset_o X$ and $A \cap V = \emptyset$ then $\bar{A} \cap V = \emptyset$.*
2. *$x \in \bar{A}$ iff $V \cap A \neq \emptyset$ for all $V \in \tau_x$.*
3. *$x \in \mathrm{bd}(A)$ iff $V \cap A \neq \emptyset$ and $V \cap A^c \neq \emptyset$ for all $V \in \tau_x$.*
4. *$\bar{A} = A \cup \mathrm{acc}(A)$.*

Proof. 1. Since $A \cap V = \emptyset$, $A \subset V^c$ and since V^c is closed, $\bar{A} \subset V^c$. That is to say $\bar{A} \cap V = \emptyset$.

2. By Remark 35.30[1], $\bar{A} = ((A^c)^o)^c$ so $x \in \bar{A}$ iff $x \notin (A^c)^o$ which happens iff $V \not\subset A^c$ for all $V \in \tau_x$, i.e. iff $V \cap A \neq \emptyset$ for all $V \in \tau_x$.

3. This assertion easily follows from the Item 2. and Eq. (35.13).

4. Item 4. is an easy consequence of the definition of $\mathrm{acc}(A)$ and item 2. ■

Lemma 35.32. *Let $A \subset Y \subset X$, $\bar{A}^Y$ denote the closure of A in Y with its relative topology and $\bar{A} = \bar{A}^X$ be the closure of A in X, then $\bar{A}^Y = \bar{A}^X \cap Y$.*

Proof. Using Lemma 35.12,

$$\bar{A}^Y = \cap \{B \sqsubset Y : A \subset B\} = \cap \{C \cap Y : A \subset C \sqsubset X\}$$
$$= Y \cap (\cap \{C : A \subset C \sqsubset X\}) = Y \cap \bar{A}^X.$$

Alternative proof. Let $x \in Y$ then $x \in \bar{A}^Y$ iff $V \cap A \neq \emptyset$ for all $V \in \tau_Y$ such that $x \in V$. This happens iff for all $U \in \tau_x$, $U \cap Y \cap A = U \cap A \neq \emptyset$ which happens iff $x \in \bar{A}^X$. That is to say $\bar{A}^Y = \bar{A}^X \cap Y$. ■

The support of a function may now be defined as in Definition 32.30 above.

[1] Here is another direct proof of item 2. which goes by showing $x \notin \bar{A}$ iff there exists $V \in \tau_x$ such that $V \cap A = \emptyset$. If $x \notin \bar{A}$ then $V = (\bar{A})^c \in \tau_x$ and $V \cap A \subset V \cap \bar{A} = \emptyset$. Conversely if there exists $V \in \tau_x$ such that $A \cap V = \emptyset$ then by Item 1. $\bar{A} \cap V = \emptyset$.

Definition 35.33 (Support). *Let $f : X \to Y$ be a function from a topological space (X, τ_X) to a vector space Y. Then we define the support of f by*

$$\mathrm{supp}(f) := \overline{\{x \in X : f(x) \neq 0\}},$$

a closed subset of X.

The next result is included for completeness but will not be used in the sequel so may be omitted.

Lemma 35.34. *Suppose that $f : X \to Y$ is a map between topological spaces. Then the following are equivalent:*

1. *f is continuous.*
2. *$f(\bar{A}) \subset \overline{f(A)}$ for all $A \subset X$*
3. *$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B})$ for all $B \subset Y$.*

Proof. If f is continuous, then $f^{-1}\left(\overline{f(A)}\right)$ is closed and since $A \subset f^{-1}(f(A)) \subset f^{-1}\left(\overline{f(A)}\right)$ it follows that $\bar{A} \subset f^{-1}\left(\overline{f(A)}\right)$. From this equation we learn that $f(\bar{A}) \subset \overline{f(A)}$ so that 1. implies 2. Now assume 2., then for $B \subset Y$ (taking $A = f^{-1}(\bar{B})$) we have

$$f(\overline{f^{-1}(B)}) \subset f(\overline{f^{-1}(\bar{B})}) \subset \overline{f(f^{-1}(\bar{B}))} \subset \bar{B}$$

and therefore

$$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B}). \tag{35.14}$$

This shows that 2. implies 3. Finally if Eq. (35.14) holds for all B, then when B is closed this shows that

$$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B}) = f^{-1}(B) \subset \overline{f^{-1}(B)}$$

which shows that

$$f^{-1}(B) = \overline{f^{-1}(B)}.$$

Therefore $f^{-1}(B)$ is closed whenever B is closed which implies that f is continuous. ■

35.4 Countability Axioms

Definition 35.35. *Let (X, τ) be a topological space. A sequence $\{x_n\}_{n=1}^\infty \subset X$* ***converges*** *to a point $x \in X$ if for all $V \in \tau_x$, $x_n \in V$ almost always (abbreviated a.a.), i.e. $\#(\{n : x_n \notin V\}) < \infty$. We will write $x_n \to x$ as $n \to \infty$ or $\lim_{n\to\infty} x_n = x$ when x_n converges to x.*

Example 35.36. Let $X = \{1, 2, 3\}$ and $\tau = \{X, \emptyset, \{1, 2\}, \{2, 3\}, \{2\}\}$ and $x_n = 2$ for all n. Then $x_n \to x$ for every $x \in X$. So limits need not be unique!

Definition 35.37 (First Countable). *A topological space, (X, τ), is* ***first countable*** *iff every point $x \in X$ has a countable neighborhood base as defined in Notation 35.2*

Example 35.38. All metric spaces, (X, d), are first countable. Indeed, if $x \in X$ then $\left\{B\left(x, \frac{1}{n}\right) : n \in \mathbb{N}\right\}$ is a countable neighborhood base at $x \in X$.

Exercise 35.9. Suppose X is an uncountable set and define $\tau \subset 2^X$ so that $V \in \tau$ iff V^c is finite or countable or $V = \emptyset$. Show τ is a topology on X which is closed under countable intersections and that (X, τ) is **not** first countable.

Exercise 35.10. Let $\{0, 1\}$ be equipped with the discrete topology and $X = \{0, 1\}^{\mathbb{R}}$ be equipped with the product topology, τ. Show (X, τ) is **not** first countable.

The spaces described in Exercises 35.9 and 35.10 are examples of topological spaces which are not metrizable, i.e. the topology is not induced by any metric on X. Like for metric spaces, when τ is first countable, we may formulate many topological notions in terms of sequences.

Proposition 35.39. *If $f : X \to Y$ is continuous at $x \in X$ and $\lim_{n\to\infty} x_n = x \in X$, then $\lim_{n\to\infty} f(x_n) = f(x) \in Y$. Moreover, if there exists a countable neighborhood base η of $x \in X$, then f is continuous at x iff $\lim_{n\to\infty} f(x_n) = f(x)$ for all sequences $\{x_n\}_{n=1}^\infty \subset X$ such that $x_n \to x$ as $n \to \infty$.*

Proof. If $f : X \to Y$ is continuous and $W \in \tau_Y$ is a neighborhood of $f(x) \in Y$, then there exists a neighborhood V of $x \in X$ such that $f(V) \subset W$. Since $x_n \to x$, $x_n \in V$ a.a. and therefore $f(x_n) \in f(V) \subset W$ a.a., i.e. $f(x_n) \to f(x)$ as $n \to \infty$. Conversely suppose that $\eta := \{W_n\}_{n=1}^\infty$ is a countable neighborhood base at x and $\lim_{n\to\infty} f(x_n) = f(x)$ for all sequences $\{x_n\}_{n=1}^\infty \subset X$ such that $x_n \to x$. By replacing W_n by $W_1 \cap \dots \cap W_n$ if necessary, we may assume that $\{W_n\}_{n=1}^\infty$ is a decreasing sequence of sets. If f were **not** continuous at x then there exists $V \in \tau_{f(x)}$ such that $x \notin \left[f^{-1}(V)\right]^o$. Therefore, W_n is not a subset of $f^{-1}(V)$ for all n. Hence for each n, we may choose $x_n \in W_n \setminus f^{-1}(V)$. This sequence then has the property that $x_n \to x$ as $n \to \infty$ while $f(x_n) \notin V$ for all n and hence $\lim_{n\to\infty} f(x_n) \neq f(x)$. ■

Lemma 35.40. *Suppose there exists $\{x_n\}_{n=1}^\infty \subset A$ such that $x_n \to x$, then $x \in \bar{A}$. Conversely if (X, τ) is a first countable space (like a metric space) then if $x \in \bar{A}$ there exists $\{x_n\}_{n=1}^\infty \subset A$ such that $x_n \to x$.*

Proof. Suppose $\{x_n\}_{n=1}^{\infty} \subset A$ and $x_n \to x \in X$. Since $\bar{A}^c$ is an open set, if $x \in \bar{A}^c$ then $x_n \in \bar{A}^c \subset A^c$ a.a. contradicting the assumption that $\{x_n\}_{n=1}^{\infty} \subset A$. Hence $x \in \bar{A}$. For the converse we now assume that (X, τ) is first countable and that $\{V_n\}_{n=1}^{\infty}$ is a countable neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By Proposition 35.31, $x \in \bar{A}$ iff $V \cap A \neq \emptyset$ for all $V \in \tau_x$. Hence $x \in \bar{A}$ implies there exists $x_n \in V_n \cap A$ for all n. It is now easily seen that $x_n \to x$ as $n \to \infty$. ∎

Definition 35.41. *A topological space, (X, τ), is* ***second countable*** *if there exists a countable base $\mathcal{V}$ for τ, i.e. $\mathcal{V} \subset \tau$ is a countable set such that for every $W \in \tau$,*

$$W = \cup\{V : V \in \mathcal{V} \text{ such that } V \subset W\}.$$

Definition 35.42. *A subset D of a topological space X is* ***dense*** *if $\bar{D} = X$. A topological space is said to be* ***separable*** *if it contains a countable dense subset, D.*

Example 35.43. The following are examples of countable dense sets.

1. The rational numbers, $\mathbb{Q}$, are dense in $\mathbb{R}$ equipped with the usual topology.
2. More generally, $\mathbb{Q}^d$ is a countable dense subset of $\mathbb{R}^d$ for any $d \in \mathbb{N}$.
3. Even more generally, for any function $\mu : \mathbb{N} \to (0, \infty)$, $\ell^p(\mu)$ is separable for all $1 \leq p < \infty$. For example, let $\Gamma \subset \mathbb{F}$ be a countable dense set, then
$$D := \{x \in \ell^p(\mu) : x_i \in \Gamma \text{ for all } i \text{ and } \#\{j : x_j \neq 0\} < \infty\}.$$
The set Γ can be taken to be $\mathbb{Q}$ if $\mathbb{F} = \mathbb{R}$ or $\mathbb{Q} + i\mathbb{Q}$ if $\mathbb{F} = \mathbb{C}$.
4. If (X, d) is a metric space which is separable then every subset $Y \subset X$ is also separable in the induced topology.

To prove 4. above, let $A = \{x_n\}_{n=1}^{\infty} \subset X$ be a countable dense subset of X. Let $d_Y(x) = \inf\{d(x, y) : y \in Y\}$ be the distance from x to Y and recall that $d_Y : X \to [0, \infty)$ is continuous. Let $\varepsilon_n = \max\left\{d_Y(x_n), \frac{1}{n}\right\} \geq 0$ and for each n let $y_n \in B_{x_n}(2\varepsilon_n)$. Then if $y \in Y$ and $\varepsilon > 0$ we may choose $n \in \mathbb{N}$ such that $d(y, x_n) \leq \varepsilon_n < \varepsilon/3$. Then $d(y_n, x_n) \leq 2\varepsilon_n < 2\varepsilon/3$ and therefore

$$d(y, y_n) \leq d(y, x_n) + d(x_n, y_n) < \varepsilon.$$

This shows that $B := \{y_n\}_{n=1}^{\infty}$ is a countable dense subset of Y.

Exercise 35.11. Show $\ell^{\infty}(\mathbb{N})$ is not separable.

Exercise 35.12. Show every second countable topological space (X, τ) is separable. Show the converse is not true by showing $X := \mathbb{R}$ with $\tau = \{\emptyset\} \cup \{V \subset \mathbb{R} : 0 \in V\}$ is a separable, first countable but not a second countable topological space.

Exercise 35.13. Every separable metric space, (X, d) is second countable.

Exercise 35.14. Suppose $\mathcal{E} \subset 2^X$ is a countable collection of subsets of X, then $\tau = \tau(\mathcal{E})$ is a second countable topology on X.

35.5 Connectedness

Definition 35.44. *(X, τ) is* ***disconnected*** *if there exist non-empty open sets U and V of X such that $U \cap V = \emptyset$ and $X = U \cup V$. We say $\{U, V\}$ is a* ***disconnection*** *of X. The topological space (X, τ) is called* ***connected*** *if it is not disconnected, i.e. if there is no disconnection of X. If $A \subset X$ we say A is connected iff (A, τ_A) is connected where τ_A is the relative topology on A. Explicitly, A is disconnected in (X, τ) iff there exists $U, V \in \tau$ such that $U \cap A \neq \emptyset$, $U \cap A \neq \emptyset$, $A \cap U \cap V = \emptyset$ and $A \subset U \cup V$.*

The reader should check that the following statement is an equivalent definition of connectivity. A topological space (X, τ) is connected iff the only sets $A \subset X$ which are both open and closed are the sets X and $\emptyset$. This version of the definition is often used in practice.

Remark 35.45. Let $A \subset Y \subset X$. Then A is connected in X iff A is connected in Y.

Proof. Since

$$\tau_A := \{V \cap A : V \subset X\} = \{V \cap A \cap Y : V \subset X\} = \{U \cap A : U \subset_o Y\},$$

the relative topology on A inherited from X is the same as the relative topology on A inherited from Y. Since connectivity is a statement about the relative topologies on A, A is connected in X iff A is connected in Y. ∎

Theorem 35.46 (The Connected Subsets of $\mathbb{R}$). *The connected subsets of $\mathbb{R}$ are intervals.*

Proof. Suppose that $A \subset \mathbb{R}$ is a connected subset and that $a, b \in A$ with $a < b$. If there exists $c \in (a, b)$ such that $c \notin A$, then $U := (-\infty, c) \cap A$ and $V := (c, \infty) \cap A$ would form a disconnection of A. Hence $(a, b) \subset A$. Let $\alpha := \inf(A)$ and $\beta := \sup(A)$ and choose $\alpha_n, \beta_n \in A$ such that $\alpha_n < \beta_n$ and $\alpha_n \downarrow \alpha$ and $\beta_n \uparrow \beta$ as $n \to \infty$. By what we have just shown, $(\alpha_n, \beta_n) \subset A$ for all n and hence $(\alpha, \beta) = \cup_{n=1}^{\infty}(\alpha_n, \beta_n) \subset A$. From this it follows that $A = (\alpha, \beta)$, $[\alpha, \beta)$, $(\alpha, \beta]$ or $[\alpha, \beta]$, i.e. A is an interval.

Conversely suppose that A is a sub-interval of $\mathbb{R}$. For the sake of contradiction, suppose that $\{U, V\}$ is a disconnection of A with $a \in U$, $b \in V$. After relabelling U and V if necessary we may assume that $a < b$. Since A is an interval $[a,b] \subset A$. Let $p = \sup([a,b] \cap U)$, then because U and V are open, $a < p < b$. Now p can not be in U for otherwise $\sup([a,b] \cap U) > p$ and p can not be in V for otherwise $p < \sup([a,b] \cap U)$. From this it follows that $p \notin U \cup V$ and hence $A \neq U \cup V$ contradicting the assumption that $\{U, V\}$ is a disconnection.

Alternative proof of the converse. Because of Proposition 35.49 below, it suffices to assume that A is an open interval. For the sake of contradiction, suppose that $\{U, V\}$ is a disconnection of A with $a \in U$, $b \in V$. After relabelling U and V if necessary we may assume that $a < b$. Let $J_a = (\alpha, \beta)$ be the maximal open interval in U which contains a. (See Exercise 35.22 or the footnote to Proposition 25.48 for the structure of open subsets of $\mathbb{R}$.) If $\beta \in U$ we could extend J_a to the right and still be in U violating the definition of β. Moreover we can not have $\beta \in V$ because in this case J_a would not be in U. Therefore $\beta \notin U \cup V = A$ while on the other hand $a < \beta < b$ and so $\beta \in A$ as A is an interval and we have reached the desired contradiction. ∎

For $x \in V$, let $a_x := \inf\{a : (a, x] \subset V\}$ and $b_x := \sup\{b : [x, b) \subset V\}$. Since V is open, $a_x < x < b_x$ and it is easily seen that $J_x := (a_x, b_x) \subset V$. Moreover if $y \in V$ and $J_x \cap J_y \neq \emptyset$, then $J_x = J_y$. The collection, $\{J_x : x \in V\}$, is at most countable since we may label each $J \in \{J_x : x \in V\}$ by choosing a rational number $r \in J$. Letting $\{J_n : n < N\}$, with $N = \infty$ allowed, be an enumeration of $\{J_x : x \in V\}$, we have $V = \coprod_{n<N} J_n$ as desired.

The following elementary but important lemma is left as an exercise to the reader.

Lemma 35.47. *Suppose that $f : X \to Y$ is a continuous map between topological spaces. Then $f(X) \subset Y$ is connected if X is connected.*

Here is a typical way these connectedness ideas are used.

Example 35.48. Suppose that $f : X \to Y$ is a continuous map between two topological spaces, the space X is connected and the space Y is "T_1," i.e. $\{y\}$ is a closed set for all $y \in Y$ as in Definition 37.37 below. Further assume f is locally constant, i.e. for all $x \in X$ there exists an open neighborhood V of x in X such that $f|_V$ is constant. Then f is constant, i.e. $f(X) = \{y_0\}$ for some $y_0 \in Y$. To prove this, let $y_0 \in f(X)$ and let $W := f^{-1}(\{y_0\})$. Since $\{y_0\} \subset Y$ is a closed set and since f is continuous $W \subset X$ is also closed. Since f is locally constant, W is open as well and since X is connected it follows that $W = X$, i.e. $f(X) = \{y_0\}$.

As a concrete application of this result, suppose that X is a connected open subset of $\mathbb{R}^d$ and $f : X \to \mathbb{R}$ is a C^1 – function such that $\nabla f \equiv 0$. If $x \in X$ and $\varepsilon > 0$ such that $B(x, \varepsilon) \subset X$, we have, for any $|v| < \varepsilon$ and $t \in [-1, 1]$, that

$$\frac{d}{dt} f(x + tv) = \nabla f(x + tv) \cdot v = 0.$$

Therefore $f(x + v) = f(x)$ for all $|v| < \varepsilon$ and this shows f is locally constant. Hence, by what we have just proved, f is constant on X.

Theorem 35.49 (Properties of Connected Sets). *Let (X, τ) be a topological space.*

1. *If $B \subset X$ is a connected set and X is the disjoint union of two open sets U and V, then either $B \subset U$ or $B \subset V$.*
2. *If $A \subset X$ is connected,*
   a) *then $\bar{A}$ is connected.*
   b) *More generally, if A is connected and $B \subset \mathrm{acc}(A)$ or $B \subset \mathrm{bd}(A)$, then $A \cup B$ is connected as well. (Recall that $\mathrm{acc}(A)$ – the set of accumulation points of A was defined in Definition 35.29 above. Moreover by Exercise 35.8, we know that $\mathrm{acc}(A) \setminus A = \mathrm{bd}(A) \setminus A$. What we are really showing here is that for any B such that $A \subset B \subset \bar{A}$, then B is connected.)*
3. *If $\{E_\alpha\}_{\alpha \in A}$ is a collection of connected sets such that $E_\alpha \cap E_\beta \neq \emptyset$ for all $\alpha, \beta \in A$,[2] then $Y := \bigcup_{\alpha \in A} E_\alpha$ is connected as well.*
4. *Suppose $A, B \subset X$ are non-empty connected subsets of X such that $\bar{A} \cap B \neq \emptyset$, then $A \cup B$ is connected in X.*
5. *Every point $x \in X$ is contained in a unique maximal connected subset C_x of X and this subset is closed. The set C_x is called the **connected component** of x.*

Proof.

1. Since B is the disjoint union of the relatively open sets $B \cap U$ and $B \cap V$, we must have $B \cap U = B$ or $B \cap V = B$ for otherwise $\{B \cap U, B \cap V\}$ would be a disconnection of B.
2. a) Let $Y = \bar{A}$ be equipped with the relative topology from X. Suppose that $U, V \subset_o Y$ form a disconnection of $Y = \bar{A}$. Then by 1. either $A \subset U$ or $A \subset V$. Say that $A \subset U$. Since U is both open and closed in Y, it follows that $Y = \bar{A} \subset U$. Therefore $V = \emptyset$ and we have a contradiction to the assumption that $\{U, V\}$ is a disconnection of $Y = \bar{A}$. Hence we must conclude that $Y = \bar{A}$ is connected as well.
   b) Now let $Y = A \cup B$ with $B \subset \mathrm{acc}(A)$, then
$$\bar{A}^Y = \bar{A} \cap Y = (A \cup \mathrm{acc}(A)) \cap Y = A \cup B.$$
   Because A is connected in Y, by (2a) $Y = A \cup B = \bar{A}^Y$ is also connected.

[2] One may assume much less here. What we really need is for any $\alpha, \beta \in A$ there exists $\{\alpha_i\}_{i=0}^n$ in A such that $\alpha_0 = \alpha$, $\alpha_n = \beta$, and $E_{\alpha_i} \cap E_{\alpha_{i+1}} \neq \emptyset$ for all $0 \le i < n$. Moreover if we make use of item 4. it suffices to assume that
$$\bar{E}_{\alpha_i} \cap E_{\alpha_{i+1}} \cup E_{\alpha_i} \cap \bar{E}_{\alpha_{i+1}} \neq \emptyset \text{ for all } 0 \le i < n.$$

3. Let $Y := \bigcup_{\alpha\in A} E_\alpha$. By Remark 35.45, we know that E_α is connected in Y for each $\alpha \in A$. If $\{U,V\}$ were a disconnection of Y, by item (1), either $E_\alpha \subset U$ or $E_\alpha \subset V$ for all α. Let $\Lambda = \{\alpha \in A : E_\alpha \subset U\}$ then $U = \cup_{\alpha\in\Lambda} E_\alpha$ and $V = \cup_{\alpha\in A\setminus\Lambda} E_\alpha$. (Notice that neither Λ or $A \setminus \Lambda$ can be empty since U and V are not empty.) Since
$$\emptyset = U \cap V = \bigcup\nolimits_{\alpha\in\Lambda,\beta\in\Lambda^c} (E_\alpha \cap E_\beta) \neq \emptyset.$$
we have reached a contradiction and hence no such disconnection exists.
4. (A good example to keep in mind here is $X = \mathbb{R}$, $A = (0, 1)$ and $B = [1, 2)$.) For sake of contradiction suppose that $\{U,V\}$ were a disconnection of $Y = A\cup B$. By item (1) either $A \subset U$ or $A \subset V$, say $A \subset U$ in which case $B \subset V$. Since $Y = A\cup B$ we must have $A = U$ and $B = V$ and so we may conclude: A and B are disjoint subsets of Y which are both open and closed. This implies
$$A = \bar{A}^Y = \bar{A} \cap Y = \bar{A} \cap (A \cup B) = A \cup (\bar{A} \cap B)$$
and therefore
$$\emptyset = A \cap B = \left[A \cup (\bar{A} \cap B)\right] \cap B = \bar{A} \cap B \neq \emptyset$$
which gives us the desired contradiction.
Alternative proof. Let $A' := A \cup \left[B \cap \bar{A}\right]$ so that $A \cup B = A' \cup B$. By item 2b. we know A' is still connected and since $A' \cap B \neq \emptyset$ we may now apply item 3. to finish the proof.
5. Let $\mathcal{C}$ denote the collection of connected subsets $C \subset X$ such that $x \in C$. Then by item 3., the set $C_x := \cup\mathcal{C}$ is also a connected subset of X which contains x and clearly this is the unique maximal connected set containing x. Since $\bar{C}_x$ is also connected by item (2) and C_x is maximal, $C_x = \bar{C}_x$, i.e. C_x is closed.

■

Theorem 35.50 (Intermediate Value Theorem). *Suppose that (X, τ) is a connected topological space and $f : X \to \mathbb{R}$ is a continuous map. Then f satisfies the intermediate value property. Namely, for every pair $x, y \in X$ such that $f(x) < f(y)$ and $c \in (f(x), f(y))$, there exists $z \in X$ such that $f(z) = c$.*

Proof. By Lemma 35.47, $f(X)$ is a connected subset of $\mathbb{R}$. So by Theorem 35.46, $f(X)$ is a subinterval of $\mathbb{R}$ and this completes the proof. ■

Definition 35.51. *A topological space X is **path connected** if to every pair of points $\{x_0, x_1\} \subset X$ there exists a continuous **path,** $\sigma \in C([0, 1], X)$, such that $\sigma(0) = x_0$ and $\sigma(1) = x_1$. The space X is said to be **locally path connected** if for each $x \in X$, there is an open neighborhood $V \subset X$ of x which is path connected.*

Proposition 35.52. *Let X be a topological space.*

1. *If X is path connected then X is connected.*
2. *If X is connected and locally path connected, then X is path connected.*
3. *If X is any connected open subset of $\mathbb{R}^n$, then X is path connected.*

Proof. The reader is asked to prove this proposition in Exercises 35.28 – 35.30 below. ■

Proposition 35.53 (Stability of Connectedness Under Products). *Let (X_α, τ_α) be connected topological spaces. Then the product space $X_A = \prod_{\alpha\in A} X_\alpha$ equipped with the product topology is connected.*

Proof. Let us begin with the case of two factors, namely assume that X and Y are connected topological spaces, then we will show that $X \times Y$ is connected as well. Given $x \in X$, let $f_x : Y \to X \times Y$ be the map $f_x(y) = (x, y)$ and notice that f_x is continuous since $\pi_X \circ f_x(y) = x$ and $\pi_Y \circ f_x(y) = y$ are continuous maps. From this we conclude that $\{x\} \times Y = f_x(Y)$ is connected by Lemma 35.47. A similar argument shows $X \times \{y\}$ is connected for all $y \in Y$.

Let $p = (x_0, y_0) \in X \times Y$ and C_p denote the connected component of p. Since $\{x_0\} \times Y$ is connected and $p \in \{x_0\} \times Y$ it follows that $\{x_0\} \times Y \subset C_p$ and hence C_p is also the connected component (x_0, y) for all $y \in Y$. Similarly, $X \times \{y\} \subset C_{(x_0,y)} = C_p$ is connected, and therefore $X \times \{y\} \subset C_p$. So we have shown $(x, y) \in C_p$ for all $x \in X$ and $y \in Y$, see Figure 35.4. By induction the theorem holds whenever A is a finite set, i.e. for products of a finite number of connected spaces.

For the general case, again choose a point $p \in X_A = X^A$ and again let $C = C_p$ be the connected component of p. Recall that C_p is closed and therefore if C_p is a proper subset of X_A, then $X_A \setminus C_p$ is a non-empty open set. By the definition of the product topology, this would imply that $X_A \setminus C_p$ contains a non-empty open set of the form
$$V := \cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(V_\alpha) = V_\Lambda \times X_{A\setminus\Lambda} \subset X_A \setminus C_p \tag{35.15}$$
where $\Lambda \subset\subset A$ and $V_\alpha \in \tau_\alpha$ for all $\alpha \in \Lambda$.

On the other hand, let $\varphi : X_\Lambda \to X_A$ by $\varphi(y) = x$ where
$$x_\alpha = \begin{cases} y_\alpha & \text{if } \alpha \in \Lambda \\ p_\alpha & \text{if } \alpha \notin \Lambda. \end{cases}$$
If $\alpha \in \Lambda$, $\pi_\alpha \circ \varphi(y) = y_\alpha = \pi_\alpha(y)$ and if $\alpha \in A \setminus \Lambda$ then $\pi_\alpha \circ \varphi(y) = p_\alpha$ so that in every case $\pi_\alpha \circ \varphi : X_\Lambda \to X_\alpha$ is continuous and therefore φ is continuous. Since X_Λ is a product of a finite number of connected spaces and so is connected and thus so is the continuous image, $\varphi(X_\Lambda) = X_\Lambda \times \{p_\alpha\}_{\alpha\in A\setminus\Lambda} \subset X_A$. Since $p \in \varphi(X_\Lambda)$ we must have

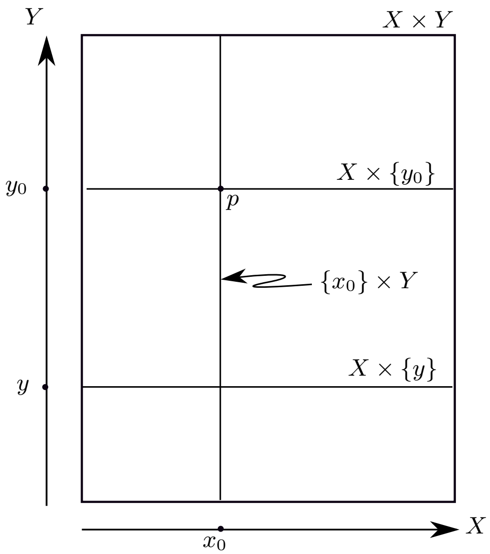


Fig. 35.4. This picture illustrates why the connected component of p in $X \times Y$ must contain all points of $X \times Y$.

$$X_\Lambda \times \{p_\alpha\}_{\alpha \in A \setminus \Lambda} \subset C_p. \tag{35.16}$$

Hence it follows from Eqs. (35.15) and (35.16) that

$$V_\Lambda \times \{p_\alpha\}_{\alpha \in A \setminus \Lambda} = \left(X_\Lambda \times \{p_\alpha\}_{\alpha \in A \setminus \Lambda} \right) \cap V \subset C_p \cap [X_A \setminus C_p] = \emptyset$$

which is a contradiction since $V_\Lambda \times \{p_\alpha\}_{\alpha \in A \setminus \Lambda} \neq \emptyset$. ∎

35.6 Compactness

Definition 35.54. *The subset A of a topological space (X, τ) is said to be* ***compact*** *if every open cover (Definition 35.18) of A has finite a sub-cover, i.e. if $\mathcal{U}$ is an open cover of A there exists $\mathcal{U}_0 \subset\subset \mathcal{U}$ such that $\mathcal{U}_0$ is a cover of A. (We will write $A \sqsubset\sqsubset X$ to denote that $A \subset X$ and A is compact.) A subset $A \subset X$ is* ***precompact*** *if $\bar{A}$ is compact.*

Proposition 35.55. *Suppose that $K \subset X$ is a compact set and $F \subset K$ is a closed subset. Then F is compact. If $\{K_i\}_{i=1}^n$ is a finite collections of compact subsets of X then $K = \cup_{i=1}^n K_i$ is also a compact subset of X.*

Proof. Let $\mathcal{U} \subset \tau$ be an open cover of F, then $\mathcal{U} \cup \{F^c\}$ is an open cover of K. The cover $\mathcal{U} \cup \{F^c\}$ of K has a finite subcover which we denote by $\mathcal{U}_0 \cup \{F^c\}$ where $\mathcal{U}_0 \subset\subset \mathcal{U}$. Since $F \cap F^c = \emptyset$, it follows that $\mathcal{U}_0$ is the desired subcover of F. For the second assertion suppose $\mathcal{U} \subset \tau$ is an open cover of K. Then $\mathcal{U}$ covers each compact set K_i and therefore there exists a finite subset $\mathcal{U}_i \subset\subset \mathcal{U}$ for each i such that $K_i \subset \cup \mathcal{U}_i$. Then $\mathcal{U}_0 := \cup_{i=1}^n \mathcal{U}_i$ is a finite cover of K. ∎

Exercise 35.15 (Suggested by Michael Gurvich). Show by example that the intersection of two compact sets need not be compact. (This pathology disappears if one assumes the topology is Hausdorff, see Definition 37.2 below.)

Exercise 35.16. Suppose $f : X \to Y$ is continuous and $K \subset X$ is compact, then $f(K)$ is a compact subset of Y. Give an example of continuous map, $f : X \to Y$, and a compact subset K of Y such that $f^{-1}(K)$ is not compact.

Exercise 35.17 (Dini's Theorem). Let X be a compact topological space and $f_n : X \to [0, \infty)$ be a sequence of continuous functions such that $f_n(x) \downarrow 0$ as $n \to \infty$ for each $x \in X$. Show that in fact $f_n \downarrow 0$ uniformly in x, i.e. $\sup_{x \in X} f_n(x) \downarrow 0$ as $n \to \infty$. **Hint:** Given $\varepsilon > 0$, consider the open sets $V_n := \{x \in X : f_n(x) < \varepsilon\}$.

Definition 35.56. *A collection $\mathcal{F}$ of closed subsets of a topological space (X, τ) has the* ***finite intersection property*** *if $\cap \mathcal{F}_0 \neq \emptyset$ for all $\mathcal{F}_0 \subset\subset \mathcal{F}$.*

The notion of compactness may be expressed in terms of closed sets as follows.

Proposition 35.57. *A topological space X is compact iff every family of closed sets $\mathcal{F} \subset 2^X$ having the* ***finite intersection property*** *satisfies $\bigcap \mathcal{F} \neq \emptyset$.*

Proof. The basic point here is that complementation interchanges open and closed sets and open covers go over to collections of closed sets with empty intersection. Here are the details.

($\Rightarrow$) Suppose that X is compact and $\mathcal{F} \subset 2^X$ is a collection of closed sets such that $\bigcap \mathcal{F} = \emptyset$. Let

$$\mathcal{U} = \mathcal{F}^c := \{C^c : C \in \mathcal{F}\} \subset \tau,$$

then $\mathcal{U}$ is a cover of X and hence has a finite subcover, $\mathcal{U}_0$. Let $\mathcal{F}_0 = \mathcal{U}_0^c \subset\subset \mathcal{F}$, then $\cap \mathcal{F}_0 = \emptyset$ so that $\mathcal{F}$ does not have the finite intersection property.

($\Leftarrow$) If X is not compact, there exists an open cover $\mathcal{U}$ of X with no finite subcover. Let

$$\mathcal{F} = \mathcal{U}^c := \{U^c : U \in \mathcal{U}\},$$

then $\mathcal{F}$ is a collection of closed sets with the finite intersection property while $\bigcap \mathcal{F} = \emptyset$. ∎

Exercise 35.18. Let (X, τ) be a topological space. Show that $A \subset X$ is compact iff (A, τ_A) is a compact topological space.

Metric Space Compactness Criteria

Let (X,d) be a metric space and for $x\in X$ and $\varepsilon>0$ let

$$B_x'(\varepsilon):=B_x(\varepsilon)\setminus\{x\}$$

be the ball centered at x of radius $\varepsilon>0$ with x deleted. Recall from Definition 35.29 that a point $x\in X$ is an accumulation point of a subset $E\subset X$ if $\emptyset\neq E\cap V\setminus\{x\}$ for all open neighborhoods, V, of x. The proof of the following elementary lemma is left to the reader.

Lemma 35.58. *Let $E\subset X$ be a subset of a metric space (X,d). Then the following are equivalent:*

1. *$x\in X$ is an accumulation point of E.*
2. *$B_x'(\varepsilon)\cap E\neq\emptyset$ for all $\varepsilon>0$.*
3. *$B_x(\varepsilon)\cap E$ is an infinite set for all $\varepsilon>0$.*
4. *There exists $\{x_n\}_{n=1}^\infty\subset E\setminus\{x\}$ with $\lim_{n\to\infty}x_n=x$.*

Definition 35.59. *A metric space (X,d) is $\varepsilon-$ **bounded** $(\varepsilon>0)$ if there exists a finite cover of X by balls of radius ε and it is **totally bounded** if it is $\varepsilon-$ bounded for all $\varepsilon>0$.*

Theorem 35.60. *Let (X,d) be a metric space. The following are equivalent.*

(a) X is compact.
(b) Every infinite subset of X has an accumulation point.
(c) Every sequence $\{x_n\}_{n=1}^\infty\subset X$ has a convergent subsequence.
(d) X is totally bounded and complete.

Proof. The proof will consist of showing that $a\Rightarrow b\Rightarrow c\Rightarrow d\Rightarrow a$.

$(a\Rightarrow b)$ We will show that **not** $b\Rightarrow$ **not** a. Suppose there exists an infinite subset $E\subset X$ which has no accumulation points. Then for all $x\in X$ there exists $\delta_x>0$ such that $V_x:=B_x(\delta_x)$ satisfies $(V_x\setminus\{x\})\cap E=\emptyset$. Clearly $\mathcal{V}=\{V_x\}_{x\in X}$ is a cover of X, yet $\mathcal{V}$ has no finite sub cover. Indeed, for each $x\in X$, $V_x\cap E\subset\{x\}$ and hence if $\Lambda\subset\subset X$, $\cup_{x\in\Lambda}V_x$ can only contain a finite number of points from E (namely $\Lambda\cap E$). Thus for any $\Lambda\subset\subset X$, $E\not\subseteq\cup_{x\in\Lambda}V_x$ and in particular $X\neq\cup_{x\in\Lambda}V_x$. (See Figure 35.5.)

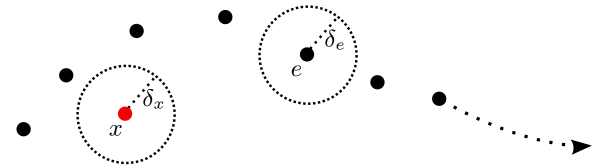


Fig. 35.5. The black dots represents an infinite set, E, with not accumulation points. For each $x\in X\setminus E$ we choose $\delta_x>0$ so that $B_x(\delta_x)\cap E=\emptyset$ and for $x\in E$ so that $B_x(\delta_x)\cap E=\{x\}$.

$(b\Rightarrow c)$ Let $\{x_n\}_{n=1}^\infty\subset X$ be a sequence and $E:=\{x_n:n\in\mathbb{N}\}$. If $\#(E)<\infty$, then $\{x_n\}_{n=1}^\infty$ has a subsequence $\{x_{n_k}\}_{k=1}^\infty$ which is constant and hence convergent. On the other hand if $\#(E)=\infty$ then by assumption E has an accumulation point and hence by Lemma 35.58, $\{x_n\}_{n=1}^\infty$ has a convergent subsequence.

$(c\Rightarrow d)$ Suppose $\{x_n\}_{n=1}^\infty\subset X$ is a Cauchy sequence. By assumption there exists a subsequence $\{x_{n_k}\}_{k=1}^\infty$ which is convergent to some point $x\in X$. Since $\{x_n\}_{n=1}^\infty$ is Cauchy it follows that $x_n\to x$ as $n\to\infty$ showing X is complete. We now show that X is totally bounded. Let $\varepsilon>0$ be given and choose an arbitrary point $x_1\in X$. If possible choose $x_2\in X$ such that $d(x_2,x_1)\geq\varepsilon$, then if possible choose $x_3\in X$ such that $d_{\{x_1,x_2\}}(x_3)\geq\varepsilon$ and continue inductively choosing points $\{x_j\}_{j=1}^n\subset X$ such that $d_{\{x_1,\dots,x_{n-1}\}}(x_n)\geq\varepsilon$. (See Figure 35.6.) This process must terminate, for otherwise we would produce a sequence $\{x_n\}_{n=1}^\infty\subset X$ which can have no convergent subsequences. Indeed, the x_n have been chosen so that $d(x_n,x_m)\geq\varepsilon>0$ for every $m\neq n$ and hence no subsequence of $\{x_n\}_{n=1}^\infty$ can be Cauchy.

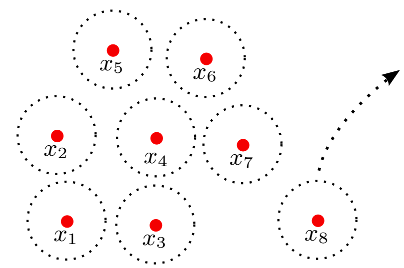


Fig. 35.6. Constructing a set with out an accumulation point.

$(d\Rightarrow a)$ For sake of contradiction, assume there exists an open cover $\mathcal{V}=\{V_\alpha\}_{\alpha\in A}$ of X with no finite subcover. Since X is totally bounded for each $n\in\mathbb{N}$ there exists $\Lambda_n\subset\subset X$ such that

$$X=\bigcup_{x\in\Lambda_n}B_x(1/n)\subset\bigcup_{x\in\Lambda_n}C_x(1/n).$$

Choose $x_1\in\Lambda_1$ such that no finite subset of $\mathcal{V}$ covers $K_1:=C_{x_1}(1)$. Since $K_1=\cup_{x\in\Lambda_2}K_1\cap C_x(1/2)$, there exists $x_2\in\Lambda_2$ such that $K_2:=K_1\cap C_{x_2}(1/2)$ can not be covered by a finite subset of $\mathcal{V}$, see Figure 35.7. Continuing this way inductively, we construct sets $K_n=K_{n-1}\cap C_{x_n}(1/n)$ with $x_n\in\Lambda_n$ such that no

K_n can be covered by a finite subset of $\mathcal{V}$. Now choose $y_n \in K_n$ for each n. Since $\{K_n\}_{n=1}^\infty$ is a decreasing sequence of closed sets such that $\operatorname{diam}(K_n) \leq 2/n$, it follows that $\{y_n\}$ is a Cauchy and hence convergent with

$$y = \lim_{n\to\infty} y_n \in \cap_{m=1}^\infty K_m.$$

Since $\mathcal{V}$ is a cover of X, there exists $V \in \mathcal{V}$ such that $y \in V$. Since $K_n \downarrow \{y\}$ and $\operatorname{diam}(K_n) \to 0$, it now follows that $K_n \subset V$ for some n large. But this violates the assertion that K_n can not be covered by a finite subset of $\mathcal{V}$.

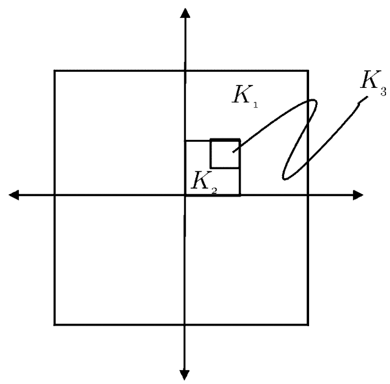


Fig. 35.7. Nested Sequence of cubes.

■

Corollary 35.61. *Any compact metric space (X, d) is second countable and hence also separable by Exercise 35.12. (See Example 37.25 below for an example of a compact topological space which is not separable.)*

Proof. To each integer n, there exists $\Lambda_n \subset\subset X$ such that $X = \cup_{x\in\Lambda_n} B(x, 1/n)$. The collection of open balls,

$$\mathcal{V} := \cup_{n\in\mathbb{N}} \cup_{x\in\Lambda_n} \{B(x, 1/n)\}$$

forms a countable basis for the metric topology on X. To check this, suppose that $x_0 \in X$ and $\varepsilon > 0$ are given and choose $n \in \mathbb{N}$ such that $1/n < \varepsilon/2$ and $x \in \Lambda_n$ such that $d(x_0, x) < 1/n$. Then $B(x, 1/n) \subset B(x_0, \varepsilon)$ because for $y \in B(x, 1/n)$,

$$d(y, x_0) \leq d(y, x) + d(x, x_0) < 2/n < \varepsilon.$$

■

Corollary 35.62. *The compact subsets of $\mathbb{R}^n$ are the closed and bounded sets.*

Proof. This is a consequence of Theorem 32.2 and Theorem 35.60. Here is another proof. If K is closed and bounded then K is complete (being the closed subset of a complete space) and K is contained in $[-M, M]^n$ for some positive integer M. For $\delta > 0$, let

$$\Lambda_\delta = \delta\mathbb{Z}^n \cap [-M, M]^n := \{\delta x : x \in \mathbb{Z}^n \text{ and } \delta|x_i| \leq M \text{ for } i = 1, 2, \ldots, n\}.$$

We will show, by choosing $\delta > 0$ sufficiently small, that

$$K \subset [-M, M]^n \subset \cup_{x\in\Lambda_\delta} B(x, \varepsilon) \tag{35.17}$$

which shows that K is totally bounded. Hence by Theorem 35.60, K is compact. Suppose that $y \in [-M, M]^n$, then there exists $x \in \Lambda_\delta$ such that $|y_i - x_i| \leq \delta$ for $i = 1, 2, \ldots, n$. Hence

$$d^2(x, y) = \sum_{i=1}^n (y_i - x_i)^2 \leq n\delta^2$$

which shows that $d(x, y) \leq \sqrt{n}\delta$. Hence if choose $\delta < \varepsilon/\sqrt{n}$ we have shows that $d(x, y) < \varepsilon$, i.e. Eq. (35.17) holds. ■

Example 35.63. Let $X = \ell^p(\mathbb{N})$ with $p \in [1, \infty)$ and $\mu \in \ell^p(\mathbb{N})$ such that $\mu(k) \geq 0$ for all $k \in \mathbb{N}$. The set

$$K := \{x \in X : |x(k)| \leq \mu(k) \text{ for all } k \in \mathbb{N}\}$$

is compact. To prove this, let $\{x_n\}_{n=1}^\infty \subset K$ be a sequence. By compactness of closed bounded sets in $\mathbb{C}$, for each $k \in \mathbb{N}$ there is a subsequence of $\{x_n(k)\}_{n=1}^\infty \subset \mathbb{C}$ which is convergent. By Cantor's diagonalization trick, we may choose a subsequence $\{y_n\}_{n=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ such that $y(k) := \lim_{n\to\infty} y_n(k)$ exists for all $k \in \mathbb{N}$.[3] Since $|y_n(k)| \leq \mu(k)$ for all n it follows that $|y(k)| \leq \mu(k)$, i.e. $y \in K$. Finally

[3] The argument is as follows. Let $\{n_j^1\}_{j=1}^\infty$ be a subsequence of $\mathbb{N} = \{n\}_{n=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^1}(1)$ exists. Now choose a subsequence $\{n_j^2\}_{j=1}^\infty$ of $\{n_j^1\}_{j=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^2}(2)$ exists and similarly $\{n_j^3\}_{j=1}^\infty$ of $\{n_j^2\}_{j=1}^\infty$ such that $\lim_{j\to\infty} x_{n_j^3}(3)$ exists. Continue on this way inductively to get

$$\{n\}_{n=1}^\infty \supset \{n_j^1\}_{j=1}^\infty \supset \{n_j^2\}_{j=1}^\infty \supset \{n_j^3\}_{j=1}^\infty \supset \ldots$$

such that $\lim_{j\to\infty} x_{n_j^k}(k)$ exists for all $k \in \mathbb{N}$. Let $m_j := n_j^j$ so that eventually $\{m_j\}_{j=1}^\infty$ is a subsequence of $\{n_j^k\}_{j=1}^\infty$ for all k. Therefore, we may take $y_j := x_{m_j}$.

$$\lim_{n\to\infty}\|y-y_n\|_p^p=\lim_{n\to\infty}\sum_{k=1}^{\infty}|y(k)-y_n(k)|^p=\sum_{k=1}^{\infty}\lim_{n\to\infty}|y(k)-y_n(k)|^p=0$$

wherein we have used the Dominated convergence theorem. (Note

$$|y(k)-y_n(k)|^p\le 2^p\mu^p(k)$$

and μ^p is summable.) Therefore $y_n\to y$ and we are done.

Alternatively, we can prove K is compact by showing that K is closed and totally bounded. It is simple to show K is closed, for if $\{x_n\}_{n=1}^{\infty}\subset K$ is a convergent sequence in X, $x:=\lim_{n\to\infty}x_n$, then

$$|x(k)|\le\lim_{n\to\infty}|x_n(k)|\le\mu(k)\ \forall\ k\in\mathbb{N}.$$

This shows that $x\in K$ and hence K is closed. To see that K is totally bounded, let $\varepsilon>0$ and choose N such that $\left(\sum_{k=N+1}^{\infty}|\mu(k)|^p\right)^{1/p}<\varepsilon$. Since $\prod_{k=1}^{N}C_{\mu(k)}(0)\subset\mathbb{C}^N$ is closed and bounded, it is compact. Therefore there exists a finite subset $\Lambda\subset\prod_{k=1}^{N}C_{\mu(k)}(0)$ such that

$$\prod_{k=1}^{N}C_{\mu(k)}(0)\subset\cup_{z\in\Lambda}B_z^N(\varepsilon)$$

where $B_z^N(\varepsilon)$ is the open ball centered at $z\in\mathbb{C}^N$ relative to the $\ell^p(\{1,2,3,\dots,N\})$ – norm. For each $z\in\Lambda$, let $\tilde z\in X$ be defined by $\tilde z(k)=z(k)$ if $k\le N$ and $\tilde z(k)=0$ for $k\ge N+1$. I now claim that

$$K\subset\cup_{z\in\Lambda}B_{\tilde z}(2\varepsilon) \tag{35.18}$$

which, when verified, shows K is totally bounded. To verify Eq. (35.18), let $x\in K$ and write $x=u+v$ where $u(k)=x(k)$ for $k\le N$ and $u(k)=0$ for $k<N$. Then by construction $u\in B_{\tilde z}(\varepsilon)$ for some $\tilde z\in\Lambda$ and

$$\|v\|_p\le\left(\sum_{k=N+1}^{\infty}|\mu(k)|^p\right)^{1/p}<\varepsilon.$$

So we have

$$\|x-\tilde z\|_p=\|u+v-\tilde z\|_p\le\|u-\tilde z\|_p+\|v\|_p<2\varepsilon.$$

Exercise 35.19 (Extreme value theorem). Let (X,τ) be a compact topological space and $f:X\to\mathbb{R}$ be a continuous function. Show $-\infty<\inf f\le\sup f<\infty$ and there exists $a,b\in X$ such that $f(a)=\inf f$ and $f(b)=\sup f$[4]. **Hint:** use Exercise 35.16 and Corollary 35.62.

[4] Here is a proof if X is a metric space. Let $\{x_n\}_{n=1}^{\infty}\subset X$ be a sequence such that $f(x_n)\uparrow\sup f$. By compactness of X we may assume, by passing to a subsequence if necessary that $x_n\to b\in X$ as $n\to\infty$. By continuity of f, $f(b)=\sup f$.

Exercise 35.20 (Uniform Continuity). Let (X,d) be a compact metric space, (Y,ρ) be a metric space and $f:X\to Y$ be a continuous function. Show that f is uniformly continuous, i.e. if $\varepsilon>0$ there exists $\delta>0$ such that $\rho(f(y),f(x))<\varepsilon$ if $x,y\in X$ with $d(x,y)<\delta$. **Hint:** you could follow the argument in the proof of Theorem 32.2.

Definition 35.64. *Let L be a vector space. We say that two norms, $|\cdot|$ and $\|\cdot\|$, on L are equivalent if there exists constants $\alpha,\beta\in(0,\infty)$ such that*

$$\|f\|\le\alpha|f|\ \text{ and }\ |f|\le\beta\|f\|\ \text{ for all } f\in L.$$

Theorem 35.65. *Let L be a finite dimensional vector space. Then any two norms $|\cdot|$ and $\|\cdot\|$ on L are equivalent. (This is typically not true for norms on infinite dimensional spaces, see for example Exercise 14.8.)*

Proof. Let $\{f_i\}_{i=1}^n$ be a basis for L and define a new norm on L by

$$\left\|\sum_{i=1}^{n}a_if_i\right\|_2:=\sqrt{\sum_{i=1}^{n}|a_i|^2}\text{ for }a_i\in\mathbb{F}.$$

By the triangle inequality for the norm $|\cdot|$, we find

$$\left|\sum_{i=1}^{n}a_if_i\right|\le\sum_{i=1}^{n}|a_i|\,|f_i|\le\sqrt{\sum_{i=1}^{n}|f_i|^2}\sqrt{\sum_{i=1}^{n}|a_i|^2}\le M\left\|\sum_{i=1}^{n}a_if_i\right\|_2$$

where $M=\sqrt{\sum_{i=1}^{n}|f_i|^2}$. Thus we have $|f|\le M\|f\|_2$ for all $f\in L$ and this inequality shows that $|\cdot|$ is continuous relative to $\|\cdot\|_2$. Since the normed space $(L,\|\cdot\|_2)$ is homeomorphic and isomorphic to $\mathbb{F}^n$ with the standard euclidean norm, the closed bounded set, $S:=\{f\in L:\|f\|_2=1\}\subset L$, is a compact subset of L relative to $\|\cdot\|_2$. Therefore by Exercise 35.19 there exists $f_0\in S$ such that

$$m=\inf\{|f|:f\in S\}=|f_0|>0.$$

Hence given $0\ne f\in L$, then $\frac{f}{\|f\|_2}\in S$ so that

$$m\le\left|\frac{f}{\|f\|_2}\right|=|f|\frac{1}{\|f\|_2}$$

or equivalently

$$\|f\|_2\le\frac{1}{m}|f|.$$

This shows that $|\cdot|$ and $\|\cdot\|_2$ are equivalent norms. Similarly one shows that $\|\cdot\|$ and $\|\cdot\|_2$ are equivalent and hence so are $|\cdot|$ and $\|\cdot\|$. ■

Corollary 35.66. *If $(L, \|\cdot\|)$ is a finite dimensional normed space, then $A \subset L$ is compact iff A is closed and bounded relative to the given norm, $\|\cdot\|$.*

Corollary 35.67. *Every finite dimensional normed vector space $(L, \|\cdot\|)$ is complete. In particular any finite dimensional subspace of a normed vector space is automatically closed.*

Proof. If $\{f_n\}_{n=1}^{\infty} \subset L$ is a Cauchy sequence, then $\{f_n\}_{n=1}^{\infty}$ is bounded and hence has a convergent subsequence, $g_k = f_{n_k}$, by Corollary 35.66. It is now routine to show $\lim_{n\to\infty} f_n = f := \lim_{k\to\infty} g_k$. ■

Theorem 35.68. *Suppose that $(X, \|\cdot\|)$ is a normed vector in which the unit ball, $V := B_0(1)$, is precompact. Then $\dim X < \infty$.*

An alternate proof is given in Proposition 35.70.. Since $\bar{V}$ is compact, we may choose $\Lambda \subset\subset X$ such that

$$\bar{V} \subset \cup_{x\in\Lambda}\left(x + \frac{1}{2}V\right) \tag{35.19}$$

where, for any $\delta > 0$,

$$\delta V := \{\delta x : x \in V\} = B_0(\delta).$$

Let $Y := \mathrm{span}(\Lambda)$, then Eq. (35.19) implies,

$$V \subset \bar{V} \subset Y + \frac{1}{2}V.$$

Multiplying this equation by $\frac{1}{2}$ then shows

$$\frac{1}{2}V \subset \frac{1}{2}Y + \frac{1}{4}V = Y + \frac{1}{4}V$$

and hence

$$V \subset Y + \frac{1}{2}V \subset Y + Y + \frac{1}{4}V = Y + \frac{1}{4}V.$$

Continuing this way inductively then shows that

$$V \subset Y + \frac{1}{2^n}V \text{ for all } n \in \mathbb{N}. \tag{35.20}$$

Indeed, if Eq. (35.20) holds, then

$$V \subset Y + \frac{1}{2}V \subset Y + \frac{1}{2}\left(Y + \frac{1}{2^n}V\right) = Y + \frac{1}{2^{n+1}}V.$$

Hence if $x \in V$, there exists $y_n \in Y$ and $z_n \in B_0(2^{-n})$ such that $y_n + z_n \to x$. Since $\lim_{n\to\infty} z_n = 0$, it follows that $x = \lim_{n\to\infty} y_n \in \bar{Y}$. Since $\dim Y \le \#(\Lambda) < \infty$, Corollary 35.67 implies $Y = \bar{Y}$ and so we have shown that $V \subset Y$. Since for any $x \in X$, $\frac{1}{2\|x\|}x \in V \subset Y$, we have $x \in Y$ for all $x \in X$, i.e. $X = Y$. ■

Lemma 35.69. *Let H be a normed linear space and H_0 a closed proper subspace. For any $\varepsilon > 0$, there exists $x_0 \in H$ such that $\|x_0\| = 1$ and $\|x - x_0\| \ge 1 - \varepsilon$ whenever $x \in H_0$.*

Proof. Can assume $\varepsilon < 1$. Take any $z_0 \notin H_0$. Let $d = \inf_{x\in H_0}\|x - z_0\|$. For any $\delta > 0$, there exists $z \in H_0$, such that $\|z - z_0\| \le d + \delta$. Take $\delta = \frac{\varepsilon d}{1-\varepsilon}$. Let $x_0 = (z - z_0)/\|z - z_0\|$, where z is determined for this δ. Then $\|x_0\| = 1$, and if $x \in H_0$,

$$\|x - x_0\| = \frac{\|\|z - z_0\|x - z + z_0\|}{\|z - z_0\|} \ge \frac{d}{\|z - z_0\|} \ge \frac{d}{d + \delta} = 1 - \varepsilon.$$

[Here is the proof again at a higher level. Choose $h \in H_0$ such that $d := \mathrm{dist}(z_0, H_0) \cong \|h - z_0\|$ and then take $x_0 := (h - z_0)/\|h - z_0\|$ as above. Then

$$\begin{aligned}\mathrm{dist}(x_0, H_0) &= \frac{1}{\|h - z_0\|}\mathrm{dist}(h - z_0, H_0) \\ &= \frac{1}{\|h - z_0\|}\mathrm{dist}(z_0, H_0) \cong 1,\end{aligned}$$

where we have used the easily verified fact that $\mathrm{dist}(ax + h, H) = |a|\,\mathrm{dist}(x, H)$ for all $a \in \mathbb{R}$ and $h \in H$.] ■

Proposition 35.70. *A locally compact Banach space is finite dimensional.*

Proof. We prove that an infinite dimensional Banach space is not locally compact. We construct a sequence $x_1, x_2, \ldots, x_n, \ldots$ such that $\|x_n\| = 1$, $\|x_i - x_j\| \ge 1/2$, $i \ne j$. Take x_1 to be any unit vector. Suppose vectors $x_1, \ldots, x_n$ are constructed. Let H_0 be the linear span of $x_1, \ldots, x_n$. By Corollary 35.67, H_0 is closed. By Lemma 35.69, there exists x_{n+1}such that $\|x_i - x_{n+1}\| \ge 1/2$, $i = 1, \ldots, n$. Now the sequence just constructed has no Cauchy subsequence. Hence the closed unit ball is not compact. Similarly the closed ball of radius $r > 0$ is also not compact. ■

Exercise 35.21. Suppose $(Y, \|\cdot\|_Y)$ is a normed space and $(X, \|\cdot\|_X)$ is a finite dimensional normed space. Show **every** linear transformation $T : X \to Y$ is necessarily bounded.

35.7 Exercises

35.7.1 General Topological Space Problems

Exercise 35.22. Let V be an open subset of $\mathbb{R}$. Show V may be written as a disjoint union of open intervals $J_n = (a_n, b_n)$, where $a_n, b_n \in \mathbb{R}\cup\{\pm\infty\}$ for $n = 1, 2, \cdots < N$ with $N = \infty$ possible.

Exercise 35.23. Let (X, τ) and (Y, τ') be a topological spaces, $f : X \to Y$ be a function, $\mathcal{U}$ be an open cover of X and $\{F_j\}_{j=1}^n$ be a finite cover of X by closed sets.

1. If $A \subset X$ is any set and $f : X \to Y$ is (τ, τ') – continuous then $f|_A : A \to Y$ is (τ_A, τ') – continuous.
2. Show $f : X \to Y$ is (τ, τ') – continuous iff $f|_U : U \to Y$ is (τ_U, τ') – continuous for all $U \in \mathcal{U}$.
3. Show $f : X \to Y$ is (τ, τ') – continuous iff $f|_{F_j} : F_j \to Y$ is (τ_{F_j}, τ') – continuous for all $j = 1, 2, \ldots, n$.

Exercise 35.24. Suppose that X is a set, $\{(Y_\alpha, \tau_\alpha) : \alpha \in A\}$ is a family of topological spaces and $f_\alpha : X \to Y_\alpha$ is a given function for all $\alpha \in A$. Assuming that $\mathcal{S}_\alpha \subset \tau_\alpha$ is a sub-base for the topology τ_α for each $\alpha \in A$, show $\mathcal{S} := \cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{S}_\alpha)$ is a sub-base for the topology $\tau := \tau(f_\alpha : \alpha \in A)$.

35.7.2 Connectedness Problems

Exercise 35.25. Show any non-trivial interval in $\mathbb{Q}$ is disconnected.

Exercise 35.26. Suppose $a < b$ and $f : (a, b) \to \mathbb{R}$ is a non-decreasing function. Show if f satisfies the intermediate value property (see Theorem 35.50), then f is continuous.

Exercise 35.27. Suppose $-\infty < a < b \le \infty$ and $f : [a, b) \to \mathbb{R}$ is a strictly increasing continuous function. Using the intermediate value theorem, one sees that $f([a, b))$ is an interval and since f is strictly increasing it must of the form $[c, d)$ for some $c \in \mathbb{R}$ and $d \in \bar{\mathbb{R}}$ with $c < d$. Show the inverse function $f^{-1} : [c, d) \to [a, b)$ is continuous and is strictly increasing. In particular if $n \in \mathbb{N}$, apply this result to $f(x) = x^n$ for $x \in [0, \infty)$ to construct the positive n^{th} – root of a real number. Compare with Exercise 3.8.

Exercise 35.28. Prove item 1. of Proposition 35.52, i.e. if X is path connected then X is connected.. **Hint:** show X is not connected implies X is not path connected.

Exercise 35.29. Prove item 2. of Proposition 35.52, i.e. if X is connected and locally path connected, then X is path connected. **Hint:** fix $x_0 \in X$ and let W denote the set of $x \in X$ such that there exists $\sigma \in C([0, 1], X)$ satisfying $\sigma(0) = x_0$ and $\sigma(1) = x$. Then show W is both open and closed.

Exercise 35.30. Prove item 3. of Proposition 35.52, i.e. if X is any connected open subset of $\mathbb{R}^n$, then X is path connected.

Exercise 35.31. Let

$$X := \left\{(x, y) \in \mathbb{R}^2 : y = \sin(x^{-1}) \text{ with } x \neq 0\right\} \cup \{(0, 0)\}$$

equipped with the relative topology induced from the standard topology on $\mathbb{R}^2$. Show X is connected but not path connected.

35.7.3 Metric Spaces as Topological Spaces

Definition 35.71. *Two metrics d and ρ on a set X are said to be **equivalent** if there exists a constant $c \in (0, \infty)$ such that $c^{-1}\rho \le d \le c\rho$.*

Exercise 35.32. Suppose that d and ρ are two metrics on X.

1. Show $\tau_d = \tau_\rho$ if d and ρ are equivalent.
2. Show by example that it is possible for $\tau_d = \tau_\rho$ even though d and ρ are inequivalent.

Exercise 35.33. Let (X_i, d_i) for $i = 1, \ldots, n$ be a finite collection of metric spaces and for $1 \le p \le \infty$ and $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $X := \prod_{i=1}^n X_i$, let

$$\rho_p(x, y) = \begin{cases} \left(\sum_{i=1}^n [d_i(x_i, y_i)]^p\right)^{1/p} & \text{if } p \neq \infty \\ \max_i d_i(x_i, y_i) & \text{if } p = \infty \end{cases}.$$

1. Show (X, ρ_p) is a metric space for $p \in [1, \infty]$. **Hint**: Minkowski's inequality.
2. Show for any $p, q \in [1, \infty]$, the metrics ρ_p and ρ_q are equivalent. **Hint:** This can be done with explicit estimates or you could use Theorem 35.65 below.

Notation 35.72 *Let X be a set and $\mathbf{p} := \{p_n\}_{n=0}^\infty$ be a family of semi-metrics on X, i.e. $p_n : X \times X \to [0, \infty)$ are functions satisfying the assumptions of metric except for the assertion that $p_n(x, y) = 0$ implies $x = y$. Further assume that $p_n(x, y) \le p_{n+1}(x, y)$ for all n and if $p_n(x, y) = 0$ for all $n \in \mathbb{N}$ then $x = y$. Given $n \in \mathbb{N}$ and $x \in X$ let*

$$B_n(x, \varepsilon) := \{y \in X : p_n(x, y) < \varepsilon\}.$$

We will write $\tau(\mathbf{p})$ form the smallest topology on X such that $p_n(x, \cdot) : X \to [0, \infty)$ is continuous for all $n \in \mathbb{N}$ and $x \in X$, i.e. $\tau(\mathbf{p}) := \tau(p_n(x, \cdot) : n \in \mathbb{N}$ and $x \in X)$.

Exercise 35.34. Using Notation 35.72, show that collection of balls,

$$\mathcal{B} := \{B_n(x, \varepsilon) : n \in \mathbb{N},\ x \in X \text{ and } \varepsilon > 0\},$$

forms a base for the topology $\tau(\mathbf{p})$. **Hint:** Use Exercise 35.24 to show $\mathcal{B}$ is a sub-base for the topology $\tau(\mathbf{p})$ and then use Exercise 35.2 to show $\mathcal{B}$ is in fact a base for the topology $\tau(\mathbf{p})$.

Exercise 35.35 (A minor variant of Exercise 13.16). Let p_n be as in Notation 35.72 and

$$d(x,y) := \sum_{n=0}^{\infty} 2^{-n} \frac{p_n(x,y)}{1+p_n(x,y)}.$$

Show d is a metric on X and $\tau_d = \tau(\mathbf{p})$. Conclude that a sequence $\{x_k\}_{k=1}^{\infty} \subset X$ converges to $x \in X$ iff

$$\lim_{k\to\infty} p_n(x_k, x) = 0 \text{ for all } n \in \mathbb{N}.$$

Exercise 35.36. Let $\{(X_n, d_n)\}_{n=1}^{\infty}$ be a sequence of metric spaces, $X := \prod_{n=1}^{\infty} X_n$, and for $x = (x(n))_{n=1}^{\infty}$ and $y = (y(n))_{n=1}^{\infty}$ in X let

$$d(x,y) = \sum_{n=1}^{\infty} 2^{-n} \frac{d_n(x(n), y(n))}{1+d_n(x(n), y(n))}.$$

(See Exercise 13.16.) Moreover, let $\pi_n : X \to X_n$ be the projection maps, show

$$\tau_d = \otimes_{n=1}^{\infty} \tau_{d_n} := \tau(\{\pi_n : n \in \mathbb{N}\}).$$

That is show the d – metric topology is the same as the product topology on X. **Suggestions:** 1) show π_n is τ_d continuous for each n and 2) show for each $x \in X$ that $d(x,\cdot)$ is $\otimes_{n=1}^{\infty} \tau_{d_n}$ – continuous. For the second assertion notice that $d(x,\cdot) = \sum_{n=1}^{\infty} f_n$ where $f_n = 2^{-n} \left(\frac{d_n(x(n),\cdot)}{1+d_n(x(n),\cdot)} \right) \circ \pi_n$.

35.7.4 Compactness Problems

Exercise 35.37 (Tychonoff's Theorem for Compact Metric Spaces). Let us continue the Notation used in Exercise 13.16. Further assume that the spaces X_n are compact for all n. Show (without using Theorem 36.16 below) that (X,d) is compact. **Hint:** Either use Cantor's method to show every sequence $\{x_m\}_{m=1}^{\infty} \subset X$ has a convergent subsequence or alternatively show (X,d) is complete and totally bounded. (Compare with Example 35.63 and see Theorem 36.16 below for the general version of this theorem.)

36

Compactness

36.1 Local and σ – Compactness

Notation 36.1 *If X is a topological space and Y is a normed space, let*

$$BC(X,Y) := \{f \in C(X,Y) : \sup_{x\in X} \|f(x)\|_Y < \infty\}$$

and

$$C_c(X,Y) := \{f \in C(X,Y) : \operatorname{supp}(f) \text{ is compact}\}.$$

If $Y = \mathbb{R}$ or $\mathbb{C}$ we will simply write $C(X)$, $BC(X)$ and $C_c(X)$ for $C(X,Y)$, $BC(X,Y)$ and $C_c(X,Y)$ respectively.

Remark 36.2. Let X be a topological space and Y be a Banach space. By combining Exercise 35.16 and Theorem 35.60 it follows that $C_c(X,Y) \subset BC(X,Y)$.

Definition 36.3 (Local and σ – compactness). *Let (X,τ) be a topological space.*

1. *(X,τ) is **locally compact** if for all $x \in X$ there exists an open neighborhood $V \subset X$ of x such that $\bar{V}$ is compact. (Alternatively, in view of Definition 35.29 (also see Definition 13.5), this is equivalent to requiring that to each $x \in X$ there exists a compact neighborhood N_x of x.)*
2. *(X,τ) is σ – **compact** if there exists compact sets $K_n \subset X$ such that $X = \cup_{n=1}^\infty K_n$. (Notice that we may assume, by replacing K_n by $K_1 \cup K_2 \cup \dots \cup K_n$ if necessary, that $K_n \uparrow X$.)*

Example 36.4. Any open subset of $U \subset \mathbb{R}^n$ is a locally compact and σ – compact metric space. The proof of local compactness is easy and is left to the reader. To see that U is σ – compact, for $k \in \mathbb{N}$, let

$$K_k := \{x \in U : |x| \le k \text{ and } d_{U^c}(x) \ge 1/k\}.$$

Then K_k is a closed and bounded subset of $\mathbb{R}^n$ and hence compact. Moreover $K_k^o \uparrow U$ as $k \to \infty$ since[1]

$$K_k^o \supset \{x \in U : |x| < k \text{ and } d_{U^c}(x) > 1/k\} \uparrow U \text{ as } k \to \infty.$$

[1] In fact this is an equality, but we will not need this here.

Exercise 36.1. If (X,τ) is locally compact and second countable, then there is a countable basis $\mathcal{B}_0$ for the topology consisting of precompact open sets. Use this to show (X,τ) is σ - compact.

Exercise 36.2. Every separable locally compact metric space is σ – compact.

Exercise 36.3. Every σ – compact metric space is second countable (or equivalently separable), see Corollary 35.61.

Exercise 36.4. Suppose that (X,d) is a metric space and $U \subset X$ is an open subset.

1. If X is locally compact then (U,d) is locally compact.
2. If X is σ – compact then (U,d) is σ – compact. **Hint:** Mimic Example 36.4, replacing $\{x \in \mathbb{R}^n : |x| \le k\}$ by compact sets $X_k \sqsubset\sqsubset X$ such that $X_k \uparrow X$.

Lemma 36.5. *Let (X,τ) be locally and σ – compact. Then there exists compact sets $K_n \uparrow X$ such that $K_n \subset K_{n+1}^o \subset K_{n+1}$ for all n.*

Proof. Suppose that $C \subset X$ is a compact set. For each $x \in C$ let $V_x \subset_o X$ be an open neighborhood of x such that $\bar{V}_x$ is compact. Then $C \subset \cup_{x\in C} V_x$ so there exists $\Lambda \subset\subset C$ such that

$$C \subset \cup_{x\in\Lambda} V_x \subset \cup_{x\in\Lambda} \bar{V}_x =: K.$$

Then K is a compact set, being a finite union of compact subsets of X, and $C \subset \cup_{x\in\Lambda} V_x \subset K^o$. Now let $C_n \subset X$ be compact sets such that $C_n \uparrow X$ as $n \to \infty$. Let $K_1 = C_1$ and then choose a compact set K_2 such that $C_2 \subset K_2^o$. Similarly, choose a compact set K_3 such that $K_2 \cup C_3 \subset K_3^o$ and continue inductively to find compact sets K_n such that $K_n \cup C_{n+1} \subset K_{n+1}^o$ for all n. Then $\{K_n\}_{n=1}^\infty$ is the desired sequence. ■

Remark 36.6. Lemma 36.5 may also be stated as saying there exists precompact open sets $\{G_n\}_{n=1}^\infty$ such that $G_n \subset \bar{G}_n \subset G_{n+1}$ for all n and $G_n \uparrow X$ as $n \to \infty$. Indeed if $\{G_n\}_{n=1}^\infty$ are as above, let $K_n := \bar{G}_n$ and if $\{K_n\}_{n=1}^\infty$ are as in Lemma 36.5, let $G_n := K_n^o$.

Proposition 36.7. *Suppose X is a locally compact metric space and $U \subset_o X$ and $K \sqsubset\sqsubset U$. Then there exists $V \subset_o X$ such that $K \subset V \subset \overline{V} \subset U \subset X$ and $\bar{V}$ is compact.*

Proof. (This is done more generally in Proposition 37.7 below.) By local compactness of X, for each $x \in K$ there exists $\varepsilon_x > 0$ such that $\overline{B_x(\varepsilon_x)}$ is compact and by shrinking ε_x if necessary we may assume,

$$\overline{B_x(\varepsilon_x)} \subset C_x(\varepsilon_x) \subset B_x(2\varepsilon_x) \subset U$$

for each $x \in K$. By compactness of K, there exists $\Lambda \subset\subset K$ such that $K \subset \cup_{x\in\Lambda} B_x(\varepsilon_x) =: V$. Notice that $\bar{V} \subset \cup_{x\in\Lambda}\overline{B_x(\varepsilon_x)} \subset U$ and $\bar{V}$ is a closed subset of the compact set $\cup_{x\in\Lambda}\overline{B_x(\varepsilon_x)}$ and hence compact as well. ∎

Definition 36.8. *Let U be an open subset of a topological space (X,τ). We will write $f \prec U$ to mean a function $f \in C_c(X,[0,1])$ such that* $\mathrm{supp}(f) := \overline{\{f \neq 0\}} \subset U$.

Lemma 36.9 (Urysohn's Lemma for Metric Spaces). *Let X be a locally compact metric space and $K \sqsubset\sqsubset U \subset_o X$. Then there exists $f \prec U$ such that $f = 1$ on K. In particular, if K is compact and C is closed in X such that $K \cap C = \emptyset$, there exists $f \in C_c(X,[0,1])$ such that $f = 1$ on K and $f = 0$ on C.*

Proof. Let V be as in Proposition 36.7 and then use Lemma 13.21 to find a function $f \in C(X,[0,1])$ such that $f = 1$ on K and $f = 0$ on V^c. Then $\mathrm{supp}(f) \subset \bar{V} \subset U$ and hence $f \prec U$. ∎

36.2 Function Space Compactness Criteria

In this section, let (X,τ) be a topological space.

Definition 36.10. *Let $\mathcal{F} \subset C(X)$.*

1. $\mathcal{F}$ is **equicontinuous at** $x \in X$ iff for all $\varepsilon > 0$ there exists $U \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon$ for all $y \in U$ and $f \in \mathcal{F}$.
2. $\mathcal{F}$ is **equicontinuous** if $\mathcal{F}$ is equicontinuous at all points $x \in X$.
3. $\mathcal{F}$ is **pointwise bounded** if $\sup\{|f(x)| : f \in \mathcal{F}\} < \infty$ for all $x \in X$.

Theorem 36.11 (Ascoli-Arzela Theorem). *Let (X,τ) be a compact topological space and $\mathcal{F} \subset C(X)$. Then $\mathcal{F}$ is precompact in $C(X)$ iff $\mathcal{F}$ is equicontinuous and point-wise bounded.*

Proof. ($\Leftarrow$) Since $C(X) \subset \ell^\infty(X)$ is a complete metric space, we must show $\mathcal{F}$ is totally bounded. Let $\varepsilon > 0$ be given. By equicontinuity, for all $x \in X$, there exists $V_x \in \tau_x$ such that $|f(y) - f(x)| < \varepsilon/2$ if $y \in V_x$ and $f \in \mathcal{F}$. Since X is compact we may choose $\Lambda \subset\subset X$ such that $X = \cup_{x\in\Lambda} V_x$. We have now decomposed X into "blocks" $\{V_x\}_{x\in\Lambda}$ such that each $f \in \mathcal{F}$ is constant to within ε on V_x. Since $\sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} < \infty$, it is now evident that

$$\begin{aligned} M &= \sup\{|f(x)| : x \in X \text{ and } f \in \mathcal{F}\} \\ &\leq \sup\{|f(x)| : x \in \Lambda \text{ and } f \in \mathcal{F}\} + \varepsilon < \infty. \end{aligned}$$

Let $\mathbb{D} := \{k\varepsilon/2 : k \in \mathbb{Z}\} \cap [-M, M]$. If $f \in \mathcal{F}$ and $\varphi \in \mathbb{D}^\Lambda$ (i.e. $\varphi : \Lambda \to \mathbb{D}$ is a function) is chosen so that $|\varphi(x) - f(x)| \leq \varepsilon/2$ for all $x \in \Lambda$, then

$$|f(y) - \varphi(x)| \leq |f(y) - f(x)| + |f(x) - \varphi(x)| < \varepsilon \; \forall \; x \in \Lambda \text{ and } y \in V_x.$$

From this it follows that $\mathcal{F} = \bigcup\left\{\mathcal{F}_\varphi : \varphi \in \mathbb{D}^\Lambda\right\}$ where, for $\varphi \in \mathbb{D}^\Lambda$,

$$\mathcal{F}_\varphi := \{f \in \mathcal{F} : |f(y) - \varphi(x)| < \varepsilon \text{ for } y \in V_x \text{ and } x \in \Lambda\}.$$

Let $\Gamma := \left\{\varphi \in \mathbb{D}^\Lambda : \mathcal{F}_\varphi \neq \emptyset\right\}$ and for each $\varphi \in \Gamma$ choose $f_\varphi \in \mathcal{F}_\varphi \cap \mathcal{F}$. For $f \in \mathcal{F}_\varphi$, $x \in \Lambda$ and $y \in V_x$ we have

$$|f(y) - f_\varphi(y)| \leq |f(y) - \varphi(x))| + |\varphi(x) - f_\varphi(y)| < 2\varepsilon.$$

So $\|f - f_\varphi\|_\infty < 2\varepsilon$ for all $f \in \mathcal{F}_\varphi$ showing that $\mathcal{F}_\varphi \subset B_{f_\varphi}(2\varepsilon)$. Therefore,

$$\mathcal{F} = \cup_{\varphi\in\Gamma}\mathcal{F}_\varphi \subset \cup_{\varphi\in\Gamma} B_{f_\varphi}(2\varepsilon)$$

and because $\varepsilon > 0$ was arbitrary we have shown that $\mathcal{F}$ is totally bounded.

($\Rightarrow$) (*The rest of this proof may safely be skipped.) Since $\|\cdot\|_\infty : C(X) \to [0,\infty)$ is a continuous function on $C(X)$ it is bounded on any compact subset $\mathcal{F} \subset C(X)$. This shows that $\sup\{\|f\|_\infty : f \in \mathcal{F}\} < \infty$ which clearly implies that $\mathcal{F}$ is pointwise bounded.[2] Suppose $\mathcal{F}$ were **not** equicontinuous at some point $x \in X$ that is to say there exists $\varepsilon > 0$ such that for all $V \in \tau_x$, $\sup\limits_{y\in V}\sup\limits_{f\in\mathcal{F}}|f(y) - f(x)| > \varepsilon$.[3] Equivalently said, to each $V \in \tau_x$ we may choose

[2] One could also prove that $\mathcal{F}$ is pointwise bounded by considering the continuous evaluation maps $e_x : C(X) \to \mathbb{R}$ given by $e_x(f) = f(x)$ for all $x \in X$.

[3] If X is first countable we could finish the proof with the following argument. Let $\{V_n\}_{n=1}^\infty$ be a neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By the assumption that $\mathcal{F}$ is not equicontinuous at x, there exist $f_n \in \mathcal{F}$ and $x_n \in V_n$ such that $|f_n(x) - f_n(x_n)| \geq \epsilon \; \forall \; n$. Since $\mathcal{F}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some $f \in \mathcal{F}$. Because $x_n \to x$ as $n \to \infty$ we learn that

$$\begin{aligned} \epsilon \leq |f_n(x) - f_n(x_n)| &\leq |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\ &\leq 2\|f_n - f\| + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty \end{aligned}$$

which is a contradiction.

$$f_V \in \mathcal{F} \text{ and } x_V \in V \ni |f_V(x) - f_V(x_V)| \geq \varepsilon. \tag{36.1}$$

Set $\mathcal{C}_V = \overline{\{f_W : W \in \tau_x \text{ and } W \subset V\}}^{\|\cdot\|_\infty} \subset \mathcal{F}$ and notice for any $\mathcal{V} \subset\subset \tau_x$ that

$$\cap_{V\in\mathcal{V}} \mathcal{C}_V \supseteq \mathcal{C}_{\cap\mathcal{V}} \neq \emptyset,$$

so that $\{\mathcal{C}_V\}_V \in \tau_x \subset \mathcal{F}$ has the finite intersection property.[4] Since $\mathcal{F}$ is compact, it follows that there exists some

$$f \in \bigcap_{V\in\tau_x} \mathcal{C}_V \neq \emptyset.$$

Since f is continuous, there exists $V \in \tau_x$ such that $|f(x) - f(y)| < \varepsilon/3$ for all $y \in V$. Because $f \in \mathcal{C}_V$, there exists $W \subset V$ such that $\|f - f_W\| < \varepsilon/3$. We now arrive at a contradiction;

$$\begin{aligned} \varepsilon &\leq |f_W(x) - f_W(x_W)| \\ &\leq |f_W(x) - f(x)| + |f(x) - f(x_W)| + |f(x_W) - f_W(x_W)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

Alternate proof of the first part. For $\varepsilon > 0$ let $\Lambda_\varepsilon \subset_f X$ and $\{V_x^\varepsilon\}_{x\in\Lambda_\varepsilon}$ be a finite open cover of X with the property; for all $x \in X$ we have

$$|f(y) - f(x)| < \varepsilon \ \forall\ y \in V_x \text{ and } f \in \mathcal{F}.$$

Let $D := \cup_{m=1}^\infty \Lambda_{1/m}$ – countable set and suppose that $\{f_n\} \subset \mathcal{F}$ is a given sequence. Since $\{f_n(x)\}_{n=1}^\infty$ is bounded in $\mathbb{R}$ for all $x \in D$, by Cantor's diagonalization argument, we may choose a subsequence, $g_k := f_{n_k}$ such that $g_0(x) := \lim_{k\to\infty} g_k(x)$ exists for all $x \in D$. To finish the proof we need only show $\{g_k\}$ is uniformly Cauchy. To this end, observe that for $y \in X$ and $m \in \mathbb{N}$ we may choose an $x \in \Lambda_{1/m}$ such that $y \in V_x^{1/m}$ and therefore,

$$\begin{aligned} |g_k(y) - g_l(y)| &\leq |g_k(y) - g_k(x)| + |g_k(x) - g_l(x)| + |g_l(x) - g_l(y)| \\ &\leq 2/m + |g_k(x) - g_l(x)| \end{aligned}$$

and therefore,

$$\|g_k - g_l\|_u \leq 2/m + \max_{x\in\Lambda_{1/m}} |g_k(x) - g_l(x)|.$$

Passing to the limit as $k, l \to \infty$ then shows

$$\limsup_{k,l\to\infty} \|g_k - g_l\|_u \leq 2/m \to 0 \text{ as } m \to \infty.$$

■

[4] If we are willing to use Net's described in Appendix **??** below we could finish the proof as follows. Since $\mathcal{F}$ is compact, the net $\{f_V\}_{V\in\tau_x} \subset \mathcal{F}$ has a cluster point $f \in \mathcal{F} \subset C(X)$. Choose a subnet $\{g_\alpha\}_{\alpha\in A}$ of $\{f_V\}_{V\in\tau_X}$ such that $g_\alpha \to f$ uniformly. Then, since $x_V \to x$ implies $x_{V_\alpha} \to x$, we may conclude from Eq. (36.1) that

$$\epsilon \leq |g_\alpha(x) - g_\alpha(x_{V_\alpha})| \to |g(x) - g(x)| = 0$$

which is a contradiction.

Exercise 36.5. Give an alternative proof of the implication, $(\Leftarrow)$, in Theorem 36.11 by showing every subsequence $\{f_n : n \in \mathbb{N}\} \subset \mathcal{F}$ has a convergence subsequence.

Exercise 36.6. Suppose $k \in C\left([0,1]^2, \mathbb{R}\right)$ and for $f \in C([0,1], \mathbb{R})$, let

$$Kf(x) := \int_0^1 k(x,y) f(y)\, dy \text{ for all } x \in [0,1].$$

Show K is a compact operator on $(C([0,1],\mathbb{R}), \|\cdot\|_\infty)$.

The following result is a corollary of Lemma 36.5 and Theorem 36.11.

Corollary 36.12 (Locally Compact Ascoli-Arzela Theorem). *Let (X,τ) be a locally compact and σ – compact topological space and $\{f_m\} \subset C(X)$ be a pointwise bounded sequence of functions such that $\{f_m|_K\}$ is equicontinuous for any compact subset $K \subset X$. Then there exists a subsequence $\{m_n\} \subset \{m\}$ such that $\{g_n := f_{m_n}\}_{n=1}^\infty \subset C(X)$ is a sequence which is uniformly convergent on compact subsets of X.*

Proof. Let $\{K_n\}_{n=1}^\infty$ be the compact subsets of X constructed in Lemma 36.5. We may now apply Theorem 36.11 repeatedly to find a nested family of subsequences

$$\{f_m\} \supset \{g_m^1\} \supset \{g_m^2\} \supset \{g_m^3\} \supset \dots$$

such that the sequence $\{g_m^n\}_{m=1}^\infty \subset C(X)$ is uniformly convergent on K_n. Using Cantor's trick, define the subsequence $\{h_n\}$ of $\{f_m\}$ by $h_n := g_n^n$. Then $\{h_n\}$ is uniformly convergent on K_l for each $l \in \mathbb{N}$. Now if $K \subset X$ is an arbitrary compact set, there exists $l < \infty$ such that $K \subset K_l^o \subset K_l$ and therefore $\{h_n\}$ is uniformly convergent on K as well. ■

Proposition 36.13. *Let $\Omega \subset_o \mathbb{R}^d$ such that $\bar{\Omega}$ is compact and $0 \leq \alpha < \beta \leq 1$. Then the inclusion map $i : C^\beta(\overline{\Omega}) \hookrightarrow C^\alpha(\overline{\Omega})$ is a compact operator. See Chapter 15 and Lemma 15.9 for the notation being used here.*

Let $\{u_n\}_{n=1}^\infty \subset C^\beta(\overline{\Omega})$ such that $\|u_n\|_{C^\beta} \le 1$, i.e. $\|u_n\|_\infty \le 1$ and

$$|u_n(x) - u_n(y)| \le |x-y|^\beta \text{ for all } x, y \in \overline{\Omega}.$$

By the Arzela-Ascoli Theorem 36.11, there exists a subsequence of $\{\tilde{u}_n\}_{n=1}^\infty$ of $\{u_n\}_{n=1}^\infty$ and $u \in C^o(\bar{\Omega})$ such that $\tilde{u}_n \to u$ in C^0. Since

$$|u(x) - u(y)| = \lim_{n\to\infty} |\tilde{u}_n(x) - \tilde{u}_n(y)| \le |x-y|^\beta,$$

$u \in C^\beta$ as well. Define $g_n := u - \tilde{u}_n \in C^\beta$, then

$$[g_n]_\beta + \|g_n\|_{C^0} = \|g_n\|_{C^\beta} \le 2$$

and $g_n \to 0$ in C^0. To finish the proof we must show that $g_n \to 0$ in C^α. Given $\delta > 0$,

$$[g_n]_\alpha = \sup_{x\ne y} \frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} \le A_n + B_n$$

where

$$\begin{aligned} A_n &= \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} : x \ne y \text{ and } |x-y| \le \delta\right\} \\ &= \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\beta} \cdot |x-y|^{\beta-\alpha} : x \ne y \text{ and } |x-y| \le \delta\right\} \\ &\le \delta^{\beta-\alpha} \cdot [g_n]_\beta \le 2\delta^{\beta-\alpha} \end{aligned}$$

and

$$B_n = \sup\left\{\frac{|g_n(x) - g_n(y)|}{|x-y|^\alpha} : |x-y| > \delta\right\} \le 2\delta^{-\alpha} \|g_n\|_{C^0} \to 0 \text{ as } n \to \infty.$$

Therefore,

$$\limsup_{n\to\infty} [g_n]_\alpha \le \limsup_{n\to\infty} A_n + \limsup_{n\to\infty} B_n \le 2\delta^{\beta-\alpha} + 0 \to 0 \text{ as } \delta \downarrow 0.$$

This proposition generalizes to the following theorem which the reader is asked to prove in Exercise 36.18 below.

Theorem 36.14. *Let Ω be a precompact open subset of $\mathbb{R}^d$, $\alpha, \beta \in [0,1]$ and $k, j \in \mathbb{N}_0$. If $j + \beta > k + \alpha$, then $C^{j,\beta}(\bar{\Omega})$ is compactly contained in $C^{k,\alpha}(\bar{\Omega})$.*

36.3 Tychonoff's Theorem

The goal of this section is to show that arbitrary products of compact spaces is still compact. Before going to the general case of an arbitrary number of factors let us start with only two factors.

Proposition 36.15. *Suppose that X and Y are non-empty compact topological spaces, then $X \times Y$ is compact in the product topology.*

Proof. Let $\mathcal{U}$ be an open cover of $X \times Y$. Then for each $(x,y) \in X \times Y$ there exist $U \in \mathcal{U}$ such that $(x,y) \in U$. By definition of the product topology, there also exist $V_x \in \tau_x^X$ and $W_y \in \tau_y^Y$ such that $V_x \times W_y \subset U$. Therefore $\mathcal{V} := \{V_x \times W_y : (x,y) \in X \times Y\}$ is also an open cover of $X \times Y$. We will now show that $\mathcal{V}$ has a finite sub-cover, say $\mathcal{V}_0 \subset\subset \mathcal{V}$. Assuming this is proved for the moment, this implies that $\mathcal{U}$ also has a finite subcover because each $V \in \mathcal{V}_0$ is contained in some $U_V \in \mathcal{U}$. So to complete the proof it suffices to show every cover $\mathcal{V}$ of the form $\mathcal{V} = \{V_\alpha \times W_\alpha : \alpha \in A\}$ where $V_\alpha \subset_o X$ and $W_\alpha \subset_o Y$ has a finite subcover. Given $x \in X$, let $f_x : Y \to X \times Y$ be the map $f_x(y) = (x,y)$ and notice that f_x is continuous since $\pi_X \circ f_x(y) = x$ and $\pi_Y \circ f_x(y) = y$ are continuous maps. From this we conclude that $\{x\} \times Y = f_x(Y)$ is compact. Similarly, it follows that $X \times \{y\}$ is compact for all $y \in Y$. Since $\mathcal{V}$ is a cover of $\{x\} \times Y$, there exist $\Gamma_x \subset\subset A$ such that $\{x\} \times Y \subset \bigcup_{\alpha\in\Gamma_x} (V_\alpha \times W_\alpha)$ without loss of generality we may assume that Γ_x is chosen so that $x \in V_\alpha$ for all $\alpha \in \Gamma_x$. Let $U_x := \bigcap_{\alpha\in\Gamma_x} V_\alpha \subset_o X$, and notice that

$$\bigcup_{\alpha\in\Gamma_x} (V_\alpha \times W_\alpha) \supset \bigcup_{\alpha\in\Gamma_x} (U_x \times W_\alpha) = U_x \times Y, \tag{36.2}$$

see Figure 36.1 below. Since $\{U_x\}_{x\in X}$ is now an open cover of X and X is

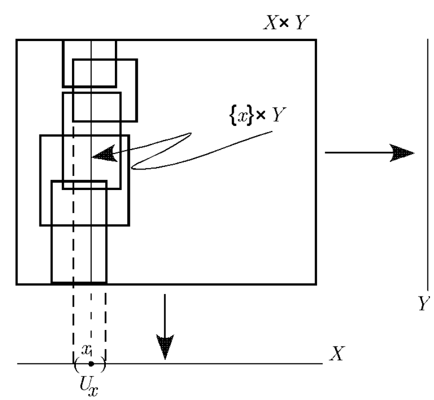


Fig. 36.1. Constructing the open set U_x.

compact, there exists $\Lambda \subset\subset X$ such that $X = \cup_{x\in\Lambda} U_x$. The finite subcollection,

$\mathcal{V}_0 := \{V_\alpha \times W_\alpha : \alpha \in \cup_{x\in\Lambda}\Gamma_x\}$, of $\mathcal{V}$ is the desired finite subcover. Indeed using Eq. (36.2),

$$\cup\mathcal{V}_0 = \cup_{x\in\Lambda}\cup_{\alpha\in\Gamma_x}(V_\alpha \times W_\alpha) \supset \cup_{x\in\Lambda}(U_x \times Y) = X \times Y.$$

■

The results of Exercises 35.37 and 35.36 prove Tychonoff's Theorem for a countable product of compact metric spaces. We now state the general version of the theorem.

Theorem 36.16 (Tychonoff's Theorem). *Let $\{X_\alpha\}_{\alpha\in A}$ be a collection of non-empty compact spaces. Then $X := X_A = \prod_{\alpha\in A} X_\alpha$ is compact in the product space topology. (Compare with Exercise 35.37 which covers the special case of a countable product of compact metric spaces.)*

Proof. (The proof is taken from Loomis [14] which followed Bourbaki. Remark 36.17 below should help the reader understand the strategy of the proof to follow.) The proof requires a form of "induction" known as Zorn's lemma which is equivalent to the axiom of choice, see Theorem 2.14 of Appendix 2.2.

For $\alpha \in A$ let π_α denote the projection map from X to X_α. Suppose that $\mathcal{F}$ is a family of closed subsets of X which has the finite intersection property, see Definition 35.56. By Proposition 35.57 the proof will be complete if we can show $\cap\mathcal{F} \neq \emptyset$.

The first step is to apply Zorn's lemma to construct a maximal collection, $\mathcal{F}_m$, of (not necessarily closed) subsets of X with the finite intersection property such that $\mathcal{F} \subset \mathcal{F}_m$. To do this, let $\Gamma := \left\{\mathcal{G} \subset 2^X : \mathcal{F} \subset \mathcal{G}\right\}$ equipped with the partial order, $\mathcal{G}_1 < \mathcal{G}_2$ if $\mathcal{G}_1 \subset \mathcal{G}_2$. If Φ is a linearly ordered subset of Γ, then $\mathcal{G} := \cup\Phi$ is an upper bound for Φ which still has the finite intersection property as the reader should check. So by Zorn's lemma, Γ has a maximal element $\mathcal{F}_m$. The maximal $\mathcal{F}_m$ has the following properties.

1. $\mathcal{F}_m$ is closed under finite intersections. Indeed, if we let $(\mathcal{F}_m)_f$ denote the collection of all finite intersections of elements from $\mathcal{F}_m$, then $(\mathcal{F}_m)_f$ has the finite intersection property and contains $\mathcal{F}_m$. Since $\mathcal{F}_m$ is maximal, this implies $(\mathcal{F}_m)_f = \mathcal{F}_m$.
2. If $B \subset X$ and $B \cap F \neq \emptyset$ for all $F \in \mathcal{F}_m$ then $B \in \mathcal{F}_m$. For if not $\mathcal{F}_m \cup \{B\}$ would still satisfy the finite intersection property and would properly contain $\mathcal{F}_m$ and this would violate the maximallity of $\mathcal{F}_m$.
3. For each $\alpha \in A$,

$$\pi_a(\mathcal{F}_m) := \{\pi_\alpha(F) \subset X_\alpha : F \in \mathcal{F}_m\}$$

has the finite intersection property. Indeed, if $\{F_i\}_{i=1}^n \subset \mathcal{F}_m$, then $\cap_{i=1}^n \pi_\alpha(F_i) \supset \pi_\alpha\left(\cap_{i=1}^n F_i\right) \neq \emptyset$.

Since X_α is compact, property 3. above along with Proposition 35.57 implies $\cap_{F\in\mathcal{F}_m}\overline{\pi_\alpha(F)} \neq \emptyset$. Since this true for each $\alpha \in A$, using the axiom of choice, there exists $p \in X$ such that $p_\alpha = \pi_\alpha(p) \in \cap_{F\in\mathcal{F}_m}\overline{\pi_\alpha(F)}$ for all $\alpha \in A$. The proof will be completed by showing $\cap\mathcal{F} \neq \emptyset$ by showing $p \in \cap\mathcal{F}$. As $\mathcal{F} \subset \left\{\bar{F} : F \in \mathcal{F}_m\right\}$, it follows that $C := \cap\left\{\bar{F} : F \in \mathcal{F}_m\right\} \subset \cap\mathcal{F}$. So to finish the proof it suffices to show $p \in C$, i.e. $p \in \bar{F}$ for all $F \in \mathcal{F}_m$.

Let U be any open neighborhood of p in X. By the definition of the product topology (or item 2. of Proposition 35.25), there exists $\Lambda \subset\subset A$ and open sets $U_\alpha \subset X_\alpha$ for all $\alpha \in \Lambda$ such that $p \in \cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(U_\alpha) \subset U$, i.e. $p_\alpha \in U_\alpha$ for $\alpha \in \Lambda$. Since $p_\alpha \in \cap_{F\in\mathcal{F}_m}\overline{\pi_\alpha(F)}$ and $p_\alpha \in U_\alpha$ for all $\alpha \in \Lambda$, it follows that $U_\alpha \cap \pi_\alpha(F) \neq \emptyset$ for all $F \in \mathcal{F}_m$ and all $\alpha \in \Lambda$. This then implies, for all $F \in \mathcal{F}_m$ and all $\alpha \in \Lambda$, that $\pi_\alpha^{-1}(U_\alpha) \cap F \neq \emptyset$.[5] By property 2.[6] above we concluded that $\pi_\alpha^{-1}(U_\alpha) \in \mathcal{F}_m$ for all $\alpha \in \Lambda$ and then by property 1. that $\cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(U_\alpha) \in \mathcal{F}_m$. In particular, for all $F \in \mathcal{F}_m$,

$$\emptyset \neq F \cap \left(\cap_{\alpha\in\Lambda}\pi_\alpha^{-1}(U_\alpha)\right) \subset F \cap U$$

which shows $p \in \bar{F}$ for each $F \in \mathcal{F}_m$, i.e. $p \in C$. ■

Remark 36.17. Consider the following simple example where $X = [-1,1] \times [-1,1]$ and $\mathcal{F} = \{F_1, F_2\}$ as in Figure 36.2. Notice that $\pi_i(F_1) \cap \pi_i(F_2) = [-1,1]$ for each i and so gives no help in trying to find the i^{th} – coordinate of one of the two points in $F_1 \cap F_2$. This is why it is necessary to introduce the collection $\mathcal{F}_0$ in the proof of Theorem 36.16. In this case one might take $\mathcal{F}_0$ to be the collection of all subsets $F \subset X$ such that $p \in F$. We then have $\cap_{F\in\mathcal{F}_0}\pi_i(F) = \{p_i\}$, so the i^{th} – coordinate of p may now be determined by observing the sets, $\{\pi_i(F) : F \in \mathcal{F}_0\}$.

Theorem 36.18 (Generalized Ascoli-Arzela Theorem). *Let (Ω, τ) be a compact topological space, X be a Banach space, and $\mathcal{F} \subset C(\Omega, X)$. Then $\mathcal{F}$ is precompact in $C(\Omega, X)$ iff $\mathcal{F}$ is equicontinuous and* ***point-wise totally bounded****, i.e. $K_\omega := \overline{\{f(\omega) : f \in \mathcal{F}\}}$ is compact in X for all $\omega \in \Omega$.*

Proof. ($\Leftarrow$) Since $C(\Omega, X) \subset \ell^\infty(\Omega, X)$ is a complete metric space, we must show $\mathcal{F}$ is totally bounded. Let $\varepsilon > 0$ be given. By equicontinuity, for all $\omega \in \Omega$, there exists $V_\omega \in \tau_\omega$ such that $|f(\gamma) - f(\omega)| < \varepsilon$ if $\gamma \in V_\omega$ and $f \in \mathcal{F}$. Since Ω is compact we may choose $\Lambda \subset_f \Omega$ such that $\Omega = \cup_{\omega\in\Lambda}V_\omega$. We have now decomposed Ω into "blocks" $\{V_\omega\}_{\omega\in\Lambda}$ such that each $f \in \mathcal{F}$ is constant to within ε on V_ω. Let $K_\Lambda := \prod_{\omega\in\Lambda} K_\omega$ which by Tychnoff's theorem is a compact subset of $\ell^\infty(\Lambda, X)$. Therefore for all $\varepsilon > 0$ there exists a finite subset $\Gamma \subset_f K_\Lambda$

[5] If $q \in F$ such that $\pi_\alpha(q) \in U_\alpha$, then $q \in \pi_\alpha^{-1}(U) \cap F$.

[6] Here is where we use that $\mathcal{F}_0$ is maximal among the collection of all, not just closed, sets having the finite intersection property and containing $\mathcal{F}$.

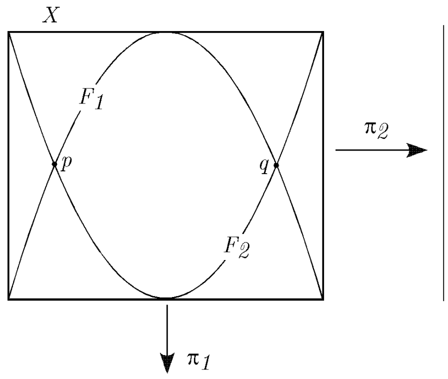


Fig. 36.2. Here $\mathcal{F} = \{F_1, F_2\}$ where F_1 and F_2 are the two parabolic arcs and $F_1 \cap F_2 = \{p, q\}$.

such that $K_\Lambda \subset \cup_{\varphi\in\Gamma} B_\varphi^{\ell^\infty(\Lambda,X)}(\varepsilon)$. By construction if $f \in \mathcal{F}$, then $f|_\Lambda \in K_\Lambda$ and therefore there exists $\varphi \in \Gamma$ such that $\|f|_\Lambda - \varphi\|_{\ell^\infty(\Lambda,X)} < \varepsilon$. This shows that $\mathcal{F} = \bigcup_{\varphi\in\Gamma} \mathcal{F}_\varphi$ where, for $\varphi \in \lneqq$,

$$\mathcal{F}_\varphi := \{f \in \mathcal{F} : \|f|_\Lambda - \varphi\|_{\ell^\infty(\Lambda,X)} < \varepsilon\}.$$

Let $\tilde{\Gamma} = \{\varphi \in \Gamma : \mathcal{F}_\varphi \neq \emptyset\}$ and for $\varphi \in \tilde{\Gamma}$ let $f_\varphi \in \mathcal{F}_\varphi$. If $f \in \mathcal{F}_\varphi$ and $\gamma \in \Omega$, choose $\omega \in \Lambda$ such that $\gamma \in V_\omega$. Then we find

$$\begin{aligned}|f(\gamma) - f_\varphi(\gamma)| &\le |f(\gamma) - f(\omega)| + |f(\omega) - \varphi(\omega)| \\ &\quad + |\varphi(\omega) - f_\varphi(\omega)| + |f_\varphi(\omega) - f_\varphi(\gamma)| < 4\varepsilon.\end{aligned}$$

As $\gamma \in \Omega$ is arbitrary we have shown $\|f - f_\varphi\|_\infty \le 4\varepsilon$ for all $f \in \mathcal{F}_\varphi$, i.e.

$$\mathcal{F} = \bigcup_{\varphi\in\tilde{\Gamma}} \mathcal{F}_\varphi \subset \bigcup_{\varphi\in\tilde{\Gamma}} B_{f_\varphi}^{C(\Omega,X)}(5\varepsilon).$$

This shows that $\mathcal{F}$ is 5ε – bounded for all $\varepsilon > 0$.

($\Rightarrow$) (*The rest of this proof may safely be skipped.) Let me first give the argument under the added restriction that $\tau = \tau_d$ for some metric, d, and Ω. Since $\|\cdot\|_\infty : C(\Omega, X) \to [0, \infty)$ is a continuous function on $C(\Omega, X)$ it is bounded on any compact subset $\mathcal{F} \subset C(\Omega, X)$. This shows that $\sup\{\|f\|_\infty : f \in \mathcal{F}\} < \infty$ which clearly implies that $\mathcal{F}$ is pointwise bounded.[7] Suppose $\mathcal{F}$ were **not** equicontinuous at some point $\omega \in \Omega$, i.e. there exists $\varepsilon > 0$ such that for all $V \in \tau_\omega$, $\sup_{\gamma\in V} \sup_{f\in\mathcal{F}} \|f(\gamma) - f(\omega)\| > \varepsilon$. Let $\{V_n = B_\omega(1/n)\}_{n=1}^\infty$. By the assumption that $\mathcal{F}$ is not equicontinuous at ω, there exist $f_n \in \mathcal{F}$ and $\omega_n \in V_n$ such that $|f_n(\omega) - f_n(\omega_n)| \ge \varepsilon$ for all n. Since $\overline{\mathcal{F}}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some $f \in \mathcal{F}$. Because $\omega_n \to \omega$ as $n \to \infty$ we learn that

$$\begin{aligned}\varepsilon &\le \|f_n(\omega) - f_n(\omega_n)\| \le \|f_n(\omega) - f(\omega)\| + \|f(\omega) - f(\omega_n)\| + \|f(\omega_n) - f_n(\omega_n)\| \\ &\le 2\|f_n - f\|_\infty + \|f(\omega) - f(\omega_n)\| \to 0 \text{ as } n \to \infty\end{aligned}$$

which is a contradiction.

($\Rightarrow$) (Here is the general argument for arbitrary topological spaces.) As in the proof above, $\mathcal{F}$ is pointwise bounded. Suppose $\mathcal{F}$ were **not** equicontinuous at some point $\omega \in \Omega$, i.e. there exists $\varepsilon > 0$ such that for all $V \in \tau_\omega$, $\sup_{\gamma\in V} \sup_{f\in\mathcal{F}} \|f(\gamma) - f(\omega)\| > \varepsilon$.[8] Equivalently said, to each $V \in \tau_\omega$ we may choose

$$f_V \in \mathcal{F} \text{ and } \omega_V \in V \ni |f_V(\omega) - f_V(\omega_V)| \ge \varepsilon.$$

Set $\mathcal{C}_V = \overline{\{f_W : W \in \tau_\omega \text{ and } W \subset V\}}^{\|\cdot\|_\infty} \subset \mathcal{F}$ and notice for any $\mathcal{V} \subset_f \tau_\omega$ that

$$\cap_{V\in\mathcal{V}} \mathcal{C}_V \supseteq \mathcal{C}_{\cap\mathcal{V}} \neq \emptyset,$$

so that $\{\mathcal{C}_V\}_V \in \tau_\omega \subset \mathcal{F}$ has the finite intersection property.[9] Since $\mathcal{F}$ is compact, it follows that there exists some

$$f \in \bigcap_{V\in\tau_\omega} \mathcal{C}_V \neq \emptyset.$$

[7] One could also prove that $\mathcal{F}$ is pointwise bounded by considering the continuous evaluation maps $e_x : C(X) \to \mathbb{R}$ given by $e_x(f) = f(x)$ for all $x \in X$.

[8] If X is first countable we could finish the proof with the following argument. Let $\{V_n\}_{n=1}^\infty$ be a neighborhood base at x such that $V_1 \supset V_2 \supset V_3 \supset \dots$. By the assumption that $\mathcal{F}$ is not equicontinuous at x, there exist $f_n \in \mathcal{F}$ and $x_n \in V_n$ such that $|f_n(x) - f_n(x_n)| \ge \epsilon \ \forall\ n$. Since $\mathcal{F}$ is a compact metric space by passing to a subsequence if necessary we may assume that f_n converges uniformly to some $f \in \mathcal{F}$. Because $x_n \to x$ as $n \to \infty$ we learn that

$$\begin{aligned}\epsilon &\le |f_n(x) - f_n(x_n)| \le |f_n(x) - f(x)| + |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \\ &\le 2\|f_n - f\| + |f(x) - f(x_n)| \to 0 \text{ as } n \to \infty\end{aligned}$$

which is a contradiction.

[9] If we are willing to use Net's described in Appendix **??** below we could finish the proof as follows. Since $\mathcal{F}$ is compact, the net $\{f_V\}_{V\in\tau_x} \subset \mathcal{F}$ has a cluster point $f \in \mathcal{F} \subset C(X)$. Choose a subnet $\{g_\alpha\}_{\alpha\in A}$ of $\{f_V\}_{V\in\tau_X}$ such that $g_\alpha \to f$ uniformly. Then, since $x_V \to x$ implies $x_{V_\alpha} \to x$, we may conclude from Eq. (36.1) that

$$\epsilon \le |g_\alpha(x) - g_\alpha(x_{V_\alpha})| \to |g(x) - g(x)| = 0$$

which is a contradiction.

Since f is continuous, there exists $V \in \tau_\omega$ such that $\|f(\omega) - f(\gamma)\| < \varepsilon/3$ for all $\gamma \in V$. Because $f \in \mathcal{C}_V$, there exists $W \subset V$ such that $\|f - f_W\| < \varepsilon/3$. We now arrive at a contradiction;

$$\begin{aligned}\varepsilon &\le \|f_W(\omega) - f_W(\omega_W)\| \\ &\le \|f_W(\omega) - f(\omega)\| + \|f(\omega) - f(\omega_W)\| + \|f(\omega_W) - f_W(\omega_W)\| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

■

36.4 Banach – Alaoglu's Theorem

36.4.1 Weak and Strong Topologies

Definition 36.19. *Let X and Y be be a normed vector spaces and $L(X,Y)$ the normed space of bounded linear transformations from X to Y.*

1. *The **weak topology** on X is the topology generated by X^*, i.e. the smallest topology on X such that every element $f \in X^*$ is continuous.*
2. *The **weak-$*$ topology** on X^* is the topology generated by X, i.e. the smallest topology on X^* such that the maps $f \in X^* \to f(x) \in \mathbb{C}$ are continuous for all $x \in X$.*
3. *The **strong operator topology** on $L(X,Y)$ is the smallest topology such that $T \in L(X,Y) \longrightarrow Tx \in Y$ is continuous for all $x \in X$.*
4. *The **weak operator topology** on $L(X,Y)$ is the smallest topology such that $T \in L(X,Y) \longrightarrow f(Tx) \in \mathbb{C}$ is continuous for all $x \in X$ and $f \in Y^*$.*

Let us be a little more precise about the topologies described in the above definitions.

1. The **weak topology** has a neighborhood base at $x_0 \in X$ consisting of sets of the form
$$N = \cap_{i=1}^n \{x \in X : |f_i(x) - f_i(x_0)| < \varepsilon\}$$
where $f_i \in X^*$ and $\varepsilon > 0$.
2. The **weak-$*$ topology** on X^* has a neighborhood base at $f \in X^*$ consisting of sets of the form
$$N := \cap_{i=1}^n \{g \in X^* : |f(x_i) - g(x_i)| < \varepsilon\}$$
where $x_i \in X$ and $\varepsilon > 0$.
3. The **strong operator topology** on $L(X,Y)$ has a neighborhood base at $T \in X^*$ consisting of sets of the form
$$N := \cap_{i=1}^n \{S \in L(X,Y) : \|Sx_i - Tx_i\| < \varepsilon\}$$
where $x_i \in X$ and $\varepsilon > 0$.
4. The **weak operator topology** on $L(X,Y)$ has a neighborhood base at $T \in X^*$ consisting of sets of the form
$$N := \cap_{i=1}^n \{S \in L(X,Y) : |f_i(Sx_i - Tx_i)| < \varepsilon\}$$
where $x_i \in X$, $f_i \in X^*$ and $\varepsilon > 0$.

Theorem 36.20 (Alaoglu's Theorem). *If X is a normed space the closed unit ball,*
$$C^* := \{f \in X^* : \|f\| \le 1\} \subset X^*,$$
is weak - $*$ compact. (Also see Theorem 36.31 and Proposition 41.16.)

Proof. For all $x \in X$ let $D_x = \{z \in \mathbb{C} : |z| \le \|x\|\}$. Then $D_x \subset \mathbb{C}$ is a compact set and so by Tychonoff's Theorem $\Omega := \prod_{x\in X} D_x$ is compact in the product topology. If $f \in C^*$, $|f(x)| \le \|f\| \|x\| \le \|x\|$ which implies that $f(x) \in D_x$ for all $x \in X$, i.e. $C^* \subset \Omega$. The topology on C^* inherited from the weak–$*$ topology on X^* is the same as that relative topology coming from the product topology on Ω. So to finish the proof it suffices to show C^* is a closed subset of the compact space Ω. To prove this let $\pi_x(f) = f(x)$ be the projection maps. Then

$$\begin{aligned}C^* &= \{f \in \Omega : f \text{ is linear}\} \\ &= \{f \in \Omega : f(x+cy) - f(x) - cf(y) = 0 \text{ for all } x, y \in X \text{ and } c \in \mathbb{C}\} \\ &= \bigcap_{x,y\in X} \bigcap_{c\in\mathbb{C}} \{f \in \Omega : f(x+cy) - f(x) - cf(y) = 0\} \\ &= \bigcap_{x,y\in X} \bigcap_{c\in\mathbb{C}} (\pi_{x+cy} - \pi_x - c\pi_y)^{-1}(\{0\})\end{aligned}$$

which is closed because $(\pi_{x+cy} - \pi_x - c\pi_y) : \Omega \to \mathbb{C}$ is continuous. ■

Example 36.21 (Compactness does not imply sequential compactness). (This example was taken from [9].) According to Theorem 14.21 $\ell^\infty([0,2\pi]) \cong \ell^1([0,2\pi])^*$. In this case the functions, $f_n(\theta) := e^{in\theta}$ are in the closed unit ball in $\ell^\infty([0,2\pi])$. We are going to show that $\{f_n\}_{n=1}^\infty$ does **not** have a weak-$*$ convergent subsequence. For if it did, there would exists $g_k := f_{n_k}$ with $n_k \uparrow \infty$ and $f \in \ell^\infty([0,2\pi])$ such that

$$\sum_\theta g_k(\theta)\psi(\theta) \to \sum_\theta f(\theta)\psi(\theta) \text{ for all } \psi \in \ell^1([0,2\pi]).$$

Taking $\psi(\theta) = \delta_{\alpha,\theta}$ with $\alpha \in [0,2\pi]$ we would infer that $\lim_{k\to\infty} g_k(\alpha) = f(\alpha)$ for all α. In particular, it would follows that $f : [0,2\pi] \to \mathbb{C}$ is a Borel measurable function such that $|f(\alpha)| = 1$ for all α. So by the dominated convergence theorem, it would follow that

$$0 = \lim_{k\to\infty} \int_0^{2\pi} g_k(\alpha) e^{-in\alpha} d\alpha = \int_0^{2\pi} f(\alpha) e^{-in\alpha} d\alpha \text{ for all } n \in \mathbb{Z},$$

wherein we have used $\int_0^{2\pi} g_k(\alpha) e^{-in\alpha} d\alpha = 0$ for all k where $n_k \neq n$ which eventually happens for k large. This then implies that $f(\alpha) = 0$ for m – a.e. α which contradicts the assertion that $|f(\alpha)| = 1$ for all α. See also Blue Rudin around p. 143.

There are a number of situations where the pathology in the above example does not happen. One is described in Theorem 36.25 below. The other is when X is a reflexive Banach space.

Theorem 36.22. *If X is a reflexive Banach space, then weak and the weak-* topologies on X^* are the same. Moreover, the closed unit ball in X^* is weak-$*$ (= weakly) sequentially compact. In particular this result holds for L^p – spaces with $1 < p < \infty$.*

Proof. Since $X^{**} = \hat{X}$, it follows that the weak-* topology on X^* is the same as the weak topology on X^*. (See Theorem 21.19 below where it is shown that X is reflexive iff X^* is reflexive.) Hence the unit ball in X^* is also now weakly compact. The Eberlein-Smulian Theorem 36.24 now guarantees that the unit ball in X^* is also weakly sequentially compact. See problem 28c on p. 86 of Rudin's functional analysis and the note on p. 376. For a short proof, see Whitley [23] and Henry B. Cohen [1]. ■

Lemma 36.23. *If X is a separable Banach space, there exists a countable subset, $\{\varphi_n\}_{n=1}^\infty$ contained in the unit ball in X^* such that if $x \in X$ and $\varphi_n(x) = 0$ for all n then $x = 0$. Moreover if $K \subset X$ is a weakly compact set, then for $x, y \in K$,*

$$d(x,y) := \sum_{n=1}^\infty \frac{1}{2^n} |\varphi_n(x-y)| \tag{36.3}$$

is a metric on K which induces the weak topology on K.

Proof. Let $\{x_n\}_{n=1}^\infty$ be a countable dense subset in the unit sphere in X. By the Hahn Banach Theorem 21.7 (or Corollary 21.8) below, there exists $\varphi_n \in X^*$ with $\|\varphi_n\| = 1$ such that $\varphi_n(x_n) = 1$. Notice that for $x \in X$, we have

$$\sup_n |\varphi_n(x)| = \|x\| \sup_n \left|\varphi_n\left(\frac{x}{\|x\|}\right)\right| \le \|x\|.$$

Moreover we may choose x_{n_k} such that $x_{n_k} \to \frac{x}{\|x\|} = y$. Since and

$$\varphi_{n_k}(y) = \varphi_{n_k}(x_{n_k}) + \varphi_{n_k}(x_{n_k} - y) = 1 + \varphi_{n_k}(x_{n_k} - y)$$

and $|\varphi_{n_k}(x_{n_k} - y)| \le \|x_{n_k} - y\| \to 0$ as $k \to \infty$, it follows that in fact $\sup_n |\varphi_n(x)| = \|x\|$ and hence $\{\varphi_n\}_{n=1}^\infty$ has the desired properties.

If $K \subset X$ is a weakly compact set, then $\varphi(K)$ is compact for all $\varphi \in X^*$ and in particular,

$$\sup_{x\in K} |\hat{x}(\varphi)| = \sup_{x\in K} |\varphi(x)| < \infty \text{ for all } \varphi \in X^*.$$

By the uniform boundedness principle,

$$\sup_{x\in K} \|x\|_X = \sup_{x\in K} \|\hat{x}\|_{X^{**}} < \infty$$

which is to say K is norm bounded.

The function d is easily seen to be a metric on X. Moreover if B is any norm bounded subset of X, $d|_{B\times B}$ is the uniform limit of continuous functions relative to the product of the weak topologies on B. Therefore $d|_{B\times B}$ is weak product topology continuous. In particular, any open d – ball in B is a weakly open subset of B.

By the previous paragraph, it follows that $id : (K, \tau_w) \to (K, \tau_d)$ is a continuous bijective map. Since (K, τ_w) is compact and closed subsets of compact sets are compact, it follows that id takes closed sets to compact subsets of (K, τ_d) which are necessarily closed since (K, τ_d) is Hausdorff. Therefore id is a homeomorphism of topological spaces as was to be proved. ■

Theorem 36.24 (Eberlein-Smulian Theorem). *For a Banach space X with the weak topology, a subset $A \subset X$ is weakly precompact iff it is weakly countably compact iff it is weakly sequentially compact.*

Proof. The direction of most interest to us is fairly easy. Namely suppose that $A \subset X$ weakly precompact and $\{a_n\}_{n=1}^\infty$ is a sequence in A, we will show that $\{a_n\}_{n=1}^\infty$ has a weakly convergent subsequence. Let $Y := \overline{\operatorname{span}(\{a_n\}_{n=1}^\infty)}^{\|\cdot\|_X}$. By the Hahn Banach Theorem 21.7 (or Corollary 21.8) below, Y is also weakly closed. Therefore $\bar{A}^{\tau_w} \cap Y = \bar{A}^{\tau_w|_Y}$ (see Lemma 35.32) is compact as well. Since Y is separable, $\bar{A}^{\tau_w} \cap Y$ is metrizable by Lemma 36.23 and therefore compactness implies completness and sequential compactness. Hence there exists an $a \in \bar{A}^{\tau_w|_Y} = \bar{A}^{\tau_w} \cap Y$ and a convergent subsequence $a'_k = a_{n_k} \to a$ in $\bar{A}^{\tau_w|_Y} = \bar{A}^{\tau_w} \cap Y$. Thus for every $\varphi \in Y^*$, we have $\varphi(a_{n_k}) \to \varphi(a)$ and in particular for every $\varphi \in X^*$, $\varphi|_Y \in Y^*$ and hence $\varphi(a_{n_k}) \to \varphi(a)$. This shows $a_{n_k} \to a$ relative to the weak topology on X. ■

Theorem 36.25 (Alaoglu's Theorem for separable spaces). *Suppose that X is a separable Banach space, $C^* := \{f \in X^* : \|f\| \leq 1\}$ is the closed unit ball in X^* and $\{x_n\}_{n=1}^\infty$ is an countable dense subset of $C := \{x \in X : \|x\| \leq 1\}$. Then*

$$\rho(f,g) := \sum_{n=1}^{\infty} \frac{1}{2^n} |f(x_n) - g(x_n)| \tag{36.4}$$

defines a metric on C^* which is compatible with the weak topology on C^*, $\tau_{C^*} := (\tau_{w^*})_{C^*} = \{V \cap C : V \in \tau_{w^*}\}$. Moreover (C^*, ρ) is a compact metric space.

Proof. The routine check that ρ is a metric is left to the reader. Let τ_ρ be the topology on C^* induced by ρ. For any $g \in X^*$ and $n \in \mathbb{N}$, the map $f \in X^* \to (f(x_n) - g(x_n)) \in \mathbb{C}$ is τ_{w^*} continuous and since the sum in Eq. (36.4) is uniformly convergent for $f \in C^*$, it follows that $f \to \rho(f,g)$ is τ_{C^*} – continuous. This implies the open balls relative to ρ are contained in τ_{C^*} and therefore $\tau_\rho \subset \tau_{C^*}$.

We now wish to prove $\tau_{C^*} \subset \tau_\rho$. Since τ_{C^*} is the topology generated by $\{\hat{x}|_{C^*} : x \in C\}$, it suffices to show $\hat{x}$ is τ_ρ – continuous for all $x \in C$. But given $x \in C$ there exists a subsequence $y_k := x_{n_k}$ of $\{x_n\}_{n=1}^\infty$ such that such that $x = \lim_{k\to\infty} y_k$. Since

$$\sup_{f \in C^*} |\hat{x}(f) - \hat{y}_k(f)| = \sup_{f \in C^*} |f(x - y_k)| \leq \|x - y_k\| \to 0 \text{ as } k \to \infty,$$

$\hat{y}_k \to \hat{x}$ uniformly on C^* and using $\hat{y}_k$ is τ_ρ – continuous for all k (as is easily checked) we learn $\hat{x}$ is also τ_ρ continuous. Hence $\tau_{C^*} = \tau(\hat{x}|_{C^*} : x \in X) \subset \tau_\rho$.

The compactness assertion follows from Theorem 36.20. The compactness assertion may also be verified directly using: 1) sequential compactness is equivalent to compactness for metric spaces and 2) a Cantor's diagonalization argument as in the proof of Theorem 36.31. (See Proposition 41.16 below.) ∎

36.5 Weak Convergence in Hilbert Spaces

Suppose H is an infinite dimensional Hilbert space and $\{x_n\}_{n=1}^\infty$ is an orthonormal subset of H. Then, by Eq. (18.1), $\|x_n - x_m\|^2 = 2$ for all $m \neq n$ and in particular, $\{x_n\}_{n=1}^\infty$ has no convergent subsequences. From this we conclude that $C := \{x \in H : \|x\| \leq 1\}$, the closed unit ball in H, is not compact. To overcome this problems it is sometimes useful to introduce a weaker topology on X having the property that C is compact.

Definition 36.26. *Let $(X, \|\cdot\|)$ be a Banach space and X^* be its continuous dual. The weak topology, τ_w, on X is the topology generated by X^*. If $\{x_n\}_{n=1}^\infty \subset X$ is a sequence we will write $x_n \overset{w}{\to} x$ as $n \to \infty$ to mean that $x_n \to x$ in the weak topology.*

Because $\tau_w = \tau(X^*) \subset \tau_{\|\cdot\|} := \tau(\{\|x - \cdot\| : x \in X\})$, it is harder for a function $f : X \to \mathbb{F}$ to be continuous in the τ_w – topology than in the norm topology, $\tau_{\|\cdot\|}$. In particular if $\varphi : X \to \mathbb{F}$ is a linear functional which is τ_w – continuous, then φ is $\tau_{\|\cdot\|}$ – continuous and hence $\varphi \in X^*$.

Exercise 36.7. Show the vector space operations of X are continuous in the weak topology, i.e. show:

1. $(x,y) \in X \times X \to x + y \in X$ is $(\tau_w \otimes \tau_w, \tau_w)$ – continuous and
2. $(\lambda, x) \in \mathbb{F} \times X \to \lambda x \in X$ is $(\tau_{\mathbb{F}} \otimes \tau_w, \tau_w)$ – continuous.

Proposition 36.27. *Let $\{x_n\}_{n=1}^\infty \subset X$ be a sequence, then $x_n \overset{w}{\to} x \in X$ as $n \to \infty$ iff $\varphi(x) = \lim_{n\to\infty} \varphi(x_n)$ for all $\varphi \in X^*$.*

Proof. By definition of τ_w, we have $x_n \overset{w}{\to} x \in X$ iff for all $\Gamma \subset\subset X^*$ and $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|\varphi(x) - \varphi(x_n)| < \varepsilon$ for all $n \geq N$ and $\varphi \in \Gamma$. This later condition is easily seen to be equivalent to $\varphi(x) = \lim_{n\to\infty} \varphi(x_n)$ for all $\varphi \in X^*$. ∎

The topological space (X, τ_w) is still Hausdorff as follows from the Hahn Banach Theorem, see Theorem 21.9 below. For the moment we will concentrate on the special case where $X = H$ is a Hilbert space in which case $H^* = \{\varphi_z := \langle \cdot | z \rangle : z \in H\}$, see Theorem 18.17. If $x, y \in H$ and $z := y - x \neq 0$, then

$$0 < \varepsilon := \|z\|^2 = \varphi_z(z) = \varphi_z(y) - \varphi_z(x).$$

Thus

$$V_x := \{w \in H : |\varphi_z(x) - \varphi_z(w)| < \varepsilon/2\} \text{ and}$$
$$V_y := \{w \in H : |\varphi_z(y) - \varphi_z(w)| < \varepsilon/2\}$$

are disjoint sets from τ_w which contain x and y respectively. This shows that (H, τ_w) is a Hausdorff space. In particular, this shows that weak limits are unique if they exist.

Remark 36.28. Suppose that H is an infinite dimensional Hilbert space $\{x_n\}_{n=1}^\infty$ is an orthonormal subset of H. Then Bessel's inequality (Proposition 18.21) implies $x_n \overset{w}{\to} 0 \in H$ as $n \to \infty$. This points out the fact that if $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, it is no longer necessarily true that $\|x\| = \lim_{n\to\infty} \|x_n\|$. However we do always have $\|x\| \leq \liminf_{n\to\infty} \|x_n\|$ because,

$$\|x\|^2 = \lim_{n\to\infty} \langle x_n | x \rangle \leq \liminf_{n\to\infty} [\|x_n\| \|x\|] = \|x\| \liminf_{n\to\infty} \|x_n\|.$$

Proposition 36.29. *Let H be a Hilbert space, $\beta \subset H$ be an orthonormal basis for H and $\{x_n\}_{n=1}^\infty \subset H$ be a bounded sequence, then the following are equivalent:*

1. $x_n \overset{w}{\to} x \in H$ *as* $n \to \infty$.
2. $\langle x|y\rangle = \lim_{n\to\infty}\langle x_n|y\rangle$ *for all* $y \in H$.
3. $\langle x|y\rangle = \lim_{n\to\infty}\langle x_n|y\rangle$ *for all* $y \in \beta$.

Moreover, if $c_y := \lim_{n\to\infty}\langle x_n|y\rangle$ *exists for all* $y \in \beta$, *then* $\sum_{y\in\beta}|c_y|^2 < \infty$ *and* $x_n \overset{w}{\to} x := \sum_{y\in\beta} c_y y \in H$ *as* $n \to \infty$.

Proof. 1. $\implies$ 2. This is a consequence of Theorem 18.17 and Proposition 36.27. 2. $\implies$ 3. is trivial. 3. $\implies$ 1. Let $M := \sup_n \|x_n\|$ and H_0 denote the algebraic span of β. Then for $y \in H$ and $z \in H_0$,

$$|\langle x - x_n|y\rangle| \le |\langle x - x_n|z\rangle| + |\langle x - x_n|y - z\rangle| \le |\langle x - x_n|z\rangle| + 2M\|y - z\|.$$

Passing to the limit in this equation implies $\limsup_{n\to\infty}|\langle x - x_n|y\rangle| \le 2M\|y - z\|$ which shows $\limsup_{n\to\infty}|\langle x - x_n|y\rangle| = 0$ since H_0 is dense in H. To prove the last assertion, let $\Gamma \subset\subset \beta$. Then by Bessel's inequality (Proposition 18.21),

$$\sum_{y\in\Gamma}|c_y|^2 = \lim_{n\to\infty}\sum_{y\in\Gamma}|\langle x_n|y\rangle|^2 \le \liminf_{n\to\infty}\|x_n\|^2 \le M^2.$$

Since $\Gamma \subset\subset \beta$ was arbitrary, we conclude that $\sum_{y\in\beta}|c_y|^2 \le M < \infty$ and hence we may define $x := \sum_{y\in\beta} c_y y$. By construction we have

$$\langle x|y\rangle = c_y = \lim_{n\to\infty}\langle x_n|y\rangle \text{ for all } y \in \beta$$

and hence $x_n \overset{w}{\to} x \in H$ as $n \to \infty$ by what we have just proved. ■

Theorem 36.30. *Suppose* $\{x_n\}_{n=1}^\infty$ *is a bounded sequence in a Hilbert space,* H. *Then there exists a subsequence* $y_k := x_{n_k}$ *of* $\{x_n\}_{n=1}^\infty$ *and* $x \in H$ *such that* $y_k \overset{w}{\to} x$ *as* $k \to \infty$.

Proof. This is a consequence of Proposition 36.29 and a Cantor's diagonalization argument which is left to the reader, see Exercise 36.12. ■

Theorem 36.31 (Alaoglu's Theorem for Hilbert Spaces). *Suppose that* H *is a separable Hilbert space,* $C := \{x \in H : \|x\| \le 1\}$ *is the closed unit ball in* H *and* $\{e_n\}_{n=1}^\infty$ *is an orthonormal basis for* H. *Then*

$$\rho(x, y) := \sum_{n=1}^\infty \frac{1}{2^n}|\langle x - y|e_n\rangle| \tag{36.5}$$

defines a metric on C *which is compatible with the weak topology on* C, $\tau_C := (\tau_w)_C = \{V \cap C : V \in \tau_w\}$. *Moreover* (C, ρ) *is a compact metric space. (This theorem will be extended to Banach spaces, see Theorems 36.20 and 36.25 below.)*

Proof. The routine check that ρ is a metric is left to the reader. Let τ_ρ be the topology on C induced by ρ. For any $y \in H$ and $n \in \mathbb{N}$, the map $x \in H \to \langle x - y|e_n\rangle = \langle x|e_n\rangle - \langle y|e_n\rangle$ is τ_w continuous and since the sum in Eq. (36.5) is uniformly convergent for $x, y \in C$, it follows that $x \to \rho(x, y)$ is τ_C − continuous. This implies the open balls relative to ρ are contained in τ_C and therefore $\tau_\rho \subset \tau_C$. For the converse inclusion, let $z \in H$, $x \to \varphi_z(x) = \langle x|z\rangle$ be an element of H^*, and for $N \in \mathbb{N}$ let $z_N := \sum_{n=1}^N \langle z|e_n\rangle e_n$. Then $\varphi_{z_N} = \sum_{n=1}^N \overline{\langle z|e_n\rangle}\varphi_{e_n}$ is ρ continuous, being a finite linear combination of the φ_{e_n} which are easily seen to be ρ − continuous. Because $z_N \to z$ as $N \to \infty$ it follows that

$$\sup_{x\in C}|\varphi_z(x) - \varphi_{z_N}(x)| = \|z - z_N\| \to 0 \text{ as } N \to \infty.$$

Therefore $\varphi_z|_C$ is ρ − continuous as well and hence $\tau_C = \tau(\varphi_z|_C : z \in H) \subset \tau_\rho$. The last assertion follows directly from Theorem 36.30 and the fact that sequential compactness is equivalent to compactness for metric spaces. ■

The next theorem give an elementary argument to show that bounded sets in a Hilbert space are always weakly sequentially compact.

Theorem 36.32 (Same as Theorem 36.30). *Suppose* $\{x_n\}_{n=1}^\infty$ *is a bounded sequence in* H *(i.e.* $C := \sup_n\|x_n\| < \infty$*), then there exists a sub-sequence,* $y_k := x_{n_k}$ *and an* $x \in H$ *such that* $\lim_{k\to\infty}\langle y_k|h\rangle = \langle x|h\rangle$ *for all* $h \in H$. *We say that* $y_k \to x$ *weakly in this case.*

Proof. Let $H_0 := \overline{\text{span}(x_k : k \in \mathbb{N})}$. Then H_0 is a closed separable Hilbert subspace of H and $\{x_k\}_{k=1}^\infty \subset H_0$. Let $\{h_n\}_{n=1}^\infty$ be a countable dense subset of H_0. Since $|\langle x_k|h_n\rangle| \le \|x_k\|\|h_n\| \le C\|h_n\| < \infty$, the sequence, $\{\langle x_k|h_n\rangle\}_{k=1}^\infty \subset \mathbb{C}$, is bounded and hence has a convergent sub-sequence for all $n \in \mathbb{N}$. By the Cantor's diagonalization argument we can find a a sub-sequence, $y_k := x_{n_k}$, of $\{x_n\}$ such that $\lim_{k\to\infty}\langle y_k|h_n\rangle$ exists for all $n \in \mathbb{N}$.

We now show $\varphi(z) := \lim_{k\to\infty}\langle y_k|z\rangle$ exists for all $z \in H_0$. Indeed, for any $k, l, n \in \mathbb{N}$, we have

$$\begin{aligned}|\langle y_k|z\rangle - \langle y_l|z\rangle| &= |\langle y_k - y_l|z\rangle| \le |\langle y_k - y_l|h_n\rangle| + |\langle y_k - y_l|z - h_n\rangle| \\ &\le |\langle y_k - y_l|h_n\rangle| + 2C\|z - h_n\|.\end{aligned}$$

Letting $k, l \to \infty$ in this estimate then shows

$$\limsup_{k,l\to\infty}|\langle y_k|z\rangle - \langle y_l|z\rangle| \le 2C\|z - h_n\|.$$

Since we may choose $n \in \mathbb{N}$ such that $\|z - h_n\|$ is as small as we please, we may conclude that $\limsup_{k,l\to\infty}|\langle y_k|z\rangle - \langle y_l|z\rangle|$, i.e. $\varphi(z) := \lim_{k\to\infty}\langle y_k|z\rangle$ exists.

The function, $\bar{\varphi}(z) = \lim_{k\to\infty}\langle z|y_k\rangle$ is a bounded linear functional on H because

$$|\bar{\varphi}(z)| = \liminf_{k\to\infty} |\langle z|y_k\rangle| \le C\,\|z\|\,.$$

Therefore by the Riesz Theorem 18.17, there exists $x \in H_0$ such that $\bar{\varphi}(z) = \langle z|x\rangle$ for all $z \in H_0$. Thus, for this $x \in H_0$ we have shown

$$\lim_{k\to\infty} \langle y_k|z\rangle = \langle x|z\rangle \text{ for all } z \in H_0. \tag{36.6}$$

To finish the proof we need only observe that Eq. (36.6) is valid for all $z \in H$. Indeed if $z \in H$, then $z = z_0 + z_1$ where $z_0 = P_{H_0}z \in H_0$ and $z_1 = z - P_{H_0}z \in H_0^\perp$. Since $y_k,\ x \in H_0$, we have

$$\lim_{k\to\infty} \langle y_k|z\rangle = \lim_{k\to\infty} \langle y_k|z_0\rangle = \langle x|z_0\rangle = \langle x|z\rangle \text{ for all } z \in H.$$

■

36.6 Exercises

Exercise 36.8. Prove Lemma 35.58 which is repeated here. Let $E \subset X$ be a subset of a metric space (X,d). Then the following are equivalent:

Lemma 36.33. *1. $x \in X$ is an accumulation point of E.*
2. $B_x'(\varepsilon) \cap E \neq \emptyset$ for all $\varepsilon > 0$.
3. $B_x(\varepsilon) \cap E$ is an infinite set for all $\varepsilon > 0$.
4. There exists $\{x_n\}_{n=1}^\infty \subset E \setminus \{x\}$ with $\lim_{n\to\infty} x_n = x$.

Exercise 36.9. Let C be a closed proper subset of $\mathbb{R}^n$ and $x \in \mathbb{R}^n \setminus C$. Show there exists a $y \in C$ such that $d(x,y) = d_C(x)$.

Exercise 36.10. Let $\mathbb{F} = \mathbb{R}$ in this problem and $A \subset \ell^2(\mathbb{N})$ be defined by

$$\begin{aligned} A &= \{x \in \ell^2(\mathbb{N}) : x(n) \ge 1 + 1/n \text{ for some } n \in \mathbb{N}\} \\ &= \cup_{n=1}^\infty \{x \in \ell^2(\mathbb{N}) : x(n) \ge 1 + 1/n\}. \end{aligned}$$

Show A is a closed subset of $\ell^2(\mathbb{N})$ with the property that $d_A(0) = 1$ while there is no $y \in A$ such that $d(0,y) = 1$. (Remember that in general an infinite union of closed sets need not be closed.)

Exercise 36.11. Let $p \in [1,\infty]$ and X be an infinite set. Show directly, without using Theorem 35.68, the closed unit ball in $\ell^p(X)$ is not compact.

36.6.1 Weak Convergence Problems

Exercise 36.12. Prove Theorem 36.30. **Hint:** Let $H_0 := \overline{\text{span}\{x_n : n \in \mathbb{N}\}}$ – a separable Hilbert subspace of H. Let $\{\lambda_m\}_{m=1}^\infty \subset H_0$ be an orthonormal basis and use Cantor's diagonalization argument to find a subsequence $y_k := x_{n_k}$ such that $c_m := \lim_{k\to\infty}\langle y_k|\lambda_m\rangle$ exists for all $m \in \mathbb{N}$. Finish the proof by appealing to Proposition 36.29.

Definition 36.34. *We say a sequence $\{x_n\}_{n=1}^\infty$ of a Hilbert space, H, converges weakly to $x \in H$ (and denote this by writing $x_n \overset{w}{\to} x \in H$ as $n \to \infty$) iff $\lim_{n\to\infty} \langle x_n|y\rangle = \langle x|y\rangle$ for all $y \in H$.*

Exercise 36.13. Suppose that $\{x_n\}_{n=1}^\infty \subset H$ and $x_n \overset{w}{\to} x \in H$ as $n \to \infty$. Show $x_n \to x$ as $n \to \infty$ (i.e. $\lim_{n\to\infty} \|x - x_n\| = 0$) iff $\lim_{n\to\infty} \|x_n\| = \|x\|$.

Exercise 36.14 (Banach-Saks). Suppose that $\{x_n\}_{n=1}^\infty \subset H$, $x_n \overset{w}{\to} x \in H$ as $n \to \infty$, and $c := \sup_n \|x_n\| < \infty$.[10] Show there exists a subsequence, $y_k = x_{n_k}$ such that

$$\lim_{N\to\infty} \left\| x - \frac{1}{N}\sum_{k=1}^N y_k \right\| = 0,$$

i.e. $\frac{1}{N}\sum_{k=1}^N y_k \to x$ as $N \to \infty$. **Hints:** 1. show it suffices to assume $x = 0$ and then choose $\{y_k\}_{k=1}^\infty$ so that $|\langle y_k|y_l\rangle| \le l^{-1}$ (or even smaller if you like) for all $k \le l$.

36.6.2 Arzela-Ascoli Theorem Problems

Exercise 36.15. Let (X,τ) be a compact topological space and $\mathcal{F} := \{f_n\}_{n=1}^\infty \subset C(X)$ is a sequence of functions which are equicontinuous and pointwise convergent. Show $f(x) := \lim_{n\to\infty} f_n(x)$ is continuous and that $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$, i.e. $f_n \to f$ uniformly as $n \to \infty$.

Exercise 36.16. Let $T \in (0,\infty)$ and $\mathcal{F} \subset C([0,T])$ be a family of functions such that:

1. $\dot{f}(t)$ exists for all $t \in (0,T)$ and $f \in \mathcal{F}$.
2. $\sup_{f\in\mathcal{F}} |f(0)| < \infty$ and
3. $M := \sup_{f\in\mathcal{F}} \sup_{t\in(0,T)} \left|\dot{f}(t)\right| < \infty$.

Show $\mathcal{F}$ is precompact in the Banach space $C([0,T])$ equipped with the norm $\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|$.

[10] The assumption that $c < \infty$ is superfluous because of the "uniform boundedness principle," see Theorem 23.9 below.

Exercise 36.17 (Peano's Existence Theorem). Suppose $Z : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is a bounded continuous function. Then for each $T < \infty$[11] there exists a solution to the differential equation

$$\dot{x}(t) = Z(t, x(t)) \text{ for } -T < t < T \text{ with } x(0) = x_0. \tag{36.7}$$

Do this by filling in the following outline for the proof.

1. Given $\varepsilon > 0$, show there exists a unique function $x_\varepsilon \in C([-\varepsilon, \infty) \to \mathbb{R}^d)$ such that $x_\varepsilon(t) := x_0$ for $-\varepsilon \le t \le 0$ and
$$x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_\varepsilon(\tau - \varepsilon))d\tau \text{ for all } t \ge 0. \tag{36.8}$$
Here
$$\int_0^t Z(\tau, x_\varepsilon(\tau-\varepsilon))d\tau = \left(\int_0^t Z_1(\tau, x_\varepsilon(\tau - \varepsilon))d\tau, \dots, \int_0^t Z_d(\tau, x_\varepsilon(\tau - \varepsilon))d\tau\right)$$
where $Z = (Z_1, \dots, Z_d)$ and the integrals are either the Lebesgue or the Riemann integral since they are equal on continuous functions. **Hint:** For $t \in [0, \varepsilon]$, it follows from Eq. (36.8) that
$$x_\varepsilon(t) = x_0 + \int_0^t Z(\tau, x_0)d\tau.$$
Now that $x_\varepsilon(t)$ is known for $t \in [-\varepsilon, \varepsilon]$ it can be found by integration for $t \in [-\varepsilon, 2\varepsilon]$. The process can be repeated.
2. Then use Exercise 36.16 to show there exists $\{\varepsilon_k\}_{k=1}^\infty \subset (0, \infty)$ such that $\lim_{k\to\infty} \varepsilon_k = 0$ and x_{ε_k} converges to some $x \in C([0, T])$ with respect to the sup-norm: $\|x\|_\infty = \sup_{t\in[0,T]} |x(t)|$). Also show for this sequence that
$$\lim_{k\to\infty} \sup_{\varepsilon_k \le \tau \le T} |x_{\varepsilon_k}(\tau - \varepsilon_k) - x(\tau)| = 0.$$
3. Pass to the limit (**with justification**) in Eq. (36.8) with ε replaced by ε_k to show x satisfies
$$x(t) = x_0 + \int_0^t Z(\tau, x(\tau))d\tau \ \forall \ t \in [0, T].$$
4. Conclude from this that $\dot{x}(t)$ exists for $t \in (0, T)$ and that x solves Eq. (36.7).
5. Apply what you have just proved to the ODE,
$$\dot{y}(t) = -Z(-t, y(t)) \text{ for } 0 \le t < T \text{ with } y(0) = x_0.$$
Then extend $x(t)$ above to $(-T, T)$ by setting $x(t) = y(-t)$ if $t \in (-T, 0]$. Show x so defined solves Eq. (36.7) for $t \in (-T, T)$.

[11] Using Corollary 36.12, we may in fact allow $T = \infty$.

Exercise 36.18. Prove Theorem 36.14. **Hint:** First prove $C^{j,\beta}(\bar{\Omega}) \sqsubset\sqsubset C^{j,\alpha}(\bar{\Omega})$ is compact if $0 \le \alpha < \beta \le 1$. Then use Lemma 19.27 repeatedly to handle all of the other cases.

37

Locally Compact Hausdorff Spaces

In this section X will always be a topological space with topology τ. We are now interested in restrictions on τ in order to insure there are "plenty" of continuous functions. One such restriction is to assume $\tau = \tau_d$ – is the topology induced from a metric on X. For example the results in Lemma 13.21 and Theorem 14.5 above shows that metric spaces have lots of continuous functions.

The main thrust of this section is to study locally compact (and σ – compact) "Hausdorff" spaces as defined in Definitions 37.2 and 36.3. We will see again that this class of topological spaces have an ample supply of continuous functions. We will start out with the notion of a Hausdorff topology. The following example shows a pathology which occurs when there are not enough open sets in a topology.

Example 37.1. As in Example 35.36, let

$$X := \{1, 2, 3\} \text{ with } \tau := \{X, \emptyset, \{1, 2\}, \{2, 3\}, \{2\}\}.$$

Example 35.36 shows limits need not be unique in this space and moreover it is easy to verify that the only continuous functions, $f : Y \to \mathbb{R}$, are necessarily constant.

Definition 37.2 (Hausdorff Topology). *A topological space, (X, τ), is **Hausdorff** if for each pair of distinct points, $x, y \in X$, there exists disjoint open neighborhoods, U and V of x and y respectively. (Metric spaces are typical examples of Hausdorff spaces.)*

Remark 37.3. When τ is Hausdorff the "pathologies" appearing in Example 37.1 do not occur. Indeed if $x_n \to x \in X$ and $y \in X \setminus \{x\}$ we may choose $V \in \tau_x$ and $W \in \tau_y$ such that $V \cap W = \emptyset$. Then $x_n \in V$ a.a. implies $x_n \notin W$ for all but a finite number of n and hence $x_n \nrightarrow y$, so limits are unique.

Proposition 37.4. *Let (X_α, τ_α) be Hausdorff topological spaces. Then the product space $X_A = \prod_{\alpha \in A} X_\alpha$ equipped with the product topology is Hausdorff.*

Proof. Let $x, y \in X_A$ be distinct points. Then there exists $\alpha \in A$ such that $\pi_\alpha(x) = x_\alpha \neq y_\alpha = \pi_\alpha(y)$. Since X_α is Hausdorff, there exists disjoint open sets $U, V \subset X_\alpha$ such $\pi_\alpha(x) \in U$ and $\pi_\alpha(y) \in V$. Then $\pi_\alpha^{-1}(U)$ and $\pi_\alpha^{-1}(V)$ are disjoint open sets in X_A containing x and y respectively. ■

Proposition 37.5. *Suppose that (X, τ) is a Hausdorff space, $K \sqsubset\sqsubset X$ and $x \in K^c$. Then there exists $U, V \in \tau$ such that $U \cap V = \emptyset$, $x \in U$ and $K \subset V$. In particular K is closed. (So compact subsets of Hausdorff topological spaces are closed.) More generally if K and F are two disjoint compact subsets of X, there exist disjoint open sets $U, V \in \tau$ such that $K \subset V$ and $F \subset U$.*

Proof. Because X is Hausdorff, for all $y \in K$ there exists $V_y \in \tau_y$ and $U_y \in \tau_x$ such that $V_y \cap U_y = \emptyset$. The cover $\{V_y\}_{y \in K}$ of K has a finite subcover, $\{V_y\}_{y \in \Lambda}$ for some $\Lambda \subset\subset K$. Let $V = \cup_{y \in \Lambda} V_y$ and $U = \cap_{y \in \Lambda} U_y$, then $U, V \in \tau$ satisfy $x \in U$, $K \subset V$ and $U \cap V = \emptyset$. This shows that K^c is open and hence that K is closed. Suppose that K and F are two disjoint compact subsets of X. For each $x \in F$ there exists disjoint open sets U_x and V_x such that $K \subset V_x$ and $x \in U_x$. Since $\{U_x\}_{x \in F}$ is an open cover of F, there exists a finite subset Λ of F such that $F \subset U := \cup_{x \in \Lambda} U_x$. The proof is completed by defining $V := \cap_{x \in \Lambda} V_x$. ■

Exercise 37.1. Show any finite set X admits exactly one Hausdorff topology τ.

Exercise 37.2. Let (X, τ) and (Y, τ_Y) be topological spaces.

1. Show τ is Hausdorff iff $\Delta := \{(x, x) : x \in X\}$ is a closed set in $X \times X$ equipped with the product topology $\tau \otimes \tau$.
2. Suppose τ is Hausdorff and $f, g : Y \to X$ are continuous maps. If $\overline{\{f = g\}}^Y = Y$ then $f = g$. **Hint:** make use of the map $f \times g : Y \to X \times X$ defined by $(f \times g)(y) = (f(y), g(y))$.

Exercise 37.3. Give an example of a topological space which has a non-closed compact subset.

Proposition 37.6. *Suppose that X is a compact topological space, Y is a Hausdorff topological space, and $f : X \to Y$ is a continuous bijection then f is a homeomorphism, i.e. $f^{-1} : Y \to X$ is continuous as well.*

Proof. Since closed subsets of compact sets are compact, continuous images of compact subsets are compact and compact subsets of Hausdorff spaces are closed, it follows that $\left(f^{-1}\right)^{-1}(C) = f(C)$ is closed in X for all closed subsets C of X. Thus f^{-1} is continuous. ■

The next two results show that locally compact Hausdorff spaces have plenty of open sets and plenty of continuous functions.

Proposition 37.7. *Suppose X is a locally compact Hausdorff space and $U \subset_o X$ and $K \sqsubset\sqsubset U$. Then there exists $V \subset_o X$ such that $K \subset V \subset \overline{V} \subset U \subset X$ and $\bar{V}$ is compact. (Compare with Proposition 36.7 above.)*

Proof. By local compactness, for all $x \in K$, there exists $U_x \in \tau_x$ such that $\bar{U}_x$ is compact. Since K is compact, there exists $\Lambda \subset\subset K$ such that $\{U_x\}_{x\in\Lambda}$ is a cover of K. The set $O = U \cap (\cup_{x\in\Lambda} U_x)$ is an open set such that $K \subset O \subset U$ and O is precompact since $\bar{O}$ is a closed subset of the compact set $\cup_{x\in\Lambda}\bar{U}_x$. ($\cup_{x\in\Lambda}\bar{U}_x$. is compact because it is a finite union of compact sets.) So by replacing U by O if necessary, we may assume that $\bar{U}$ is compact. Since $\bar{U}$ is compact and $\mathrm{bd}(U) = \bar{U} \cap U^c$ is a closed subset of $\bar{U}$, $\mathrm{bd}(U)$ is compact. Because $\mathrm{bd}(U) \subset U^c$, it follows that $\mathrm{bd}(U) \cap K = \emptyset$, so by Proposition 37.5, there exists disjoint open sets V and W such that $K \subset V$ and $\mathrm{bd}(U) \subset W$. By replacing V by $V \cap U$ if necessary we may further assume that $K \subset V \subset U$, see Figure 37.1. Because

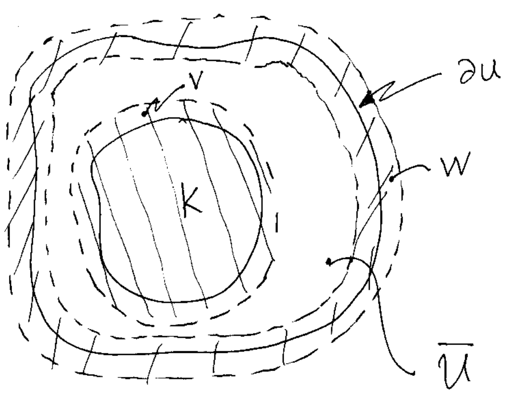


Fig. 37.1. The construction of V.

$\bar{U} \cap W^c$ is a closed set containing V and $\mathrm{bd}(U) \cap W^c = \emptyset$,

$$\bar{V} \subset \bar{U} \cap W^c = (U \cup \mathrm{bd}(U)) \cap W^c = U \cap W^c \subset U \subset \bar{U}.$$

Since $\bar{U}$ is compact it follows that $\bar{V}$ is compact and the proof is complete. ■

The following Lemma is analogous to Lemma 36.9.

Lemma 37.8 (Urysohn's Lemma for LCH Spaces). *Let X be a locally compact Hausdorff space and $K \sqsubset\sqsubset U \subset_o X$. Then there exists $f \prec U$ (see Definition 36.8) such that $f = 1$ on K. In particular, if K is compact and C is closed in X such that $K \cap C = \emptyset$, there exists $f \in C_c(X, [0,1])$ such that $f = 1$ on K and $f = 0$ on C.*

Proof. For notational ease later it is more convenient to construct $g := 1 - f$ rather than f. To motivate the proof, suppose $g \in C(X, [0,1])$ such that $g = 0$ on K and $g = 1$ on U^c. For $r > 0$, let $U_r = \{g < r\}$. Then for $0 < r < s \le 1$, $U_r \subset \{g \le r\} \subset U_s$ and since $\{g \le r\}$ is closed this implies

$$K \subset U_r \subset \bar{U}_r \subset \{g \le r\} \subset U_s \subset U.$$

Therefore associated to the function g is the collection open sets $\{U_r\}_{r>0} \subset \tau$ with the property that $K \subset U_r \subset \bar{U}_r \subset U_s \subset U$ for all $0 < r < s \le 1$ and $U_r = X$ if $r > 1$. Finally let us notice that we may recover the function g from the sequence $\{U_r\}_{r>0}$ by the formula

$$g(x) = \inf\{r > 0 : x \in U_r\}. \tag{37.1}$$

The idea of the proof to follow is to turn these remarks around and define g by Eq. (37.1).

Step 1. (Construction of the U_r.) Let

$$\mathbb{D} := \left\{k2^{-n} : k = 1, 2, \dots, 2^{-n}, n = 1, 2, \dots\right\}$$

be the dyadic rationals in $(0,1]$. Use Proposition 37.7 to find a precompact open set U_1 such that $K \subset U_1 \subset \bar{U}_1 \subset U$. Apply Proposition 37.7 again to construct an open set $U_{1/2}$ such that

$$K \subset U_{1/2} \subset \bar{U}_{1/2} \subset U_1$$

and similarly use Proposition 37.7 to find open sets $U_{1/2}, U_{3/4} \subset_o X$ such that

$$K \subset U_{1/4} \subset \bar{U}_{1/4} \subset U_{1/2} \subset \bar{U}_{1/2} \subset U_{3/4} \subset \bar{U}_{3/4} \subset U_1.$$

Likewise there exists open set $U_{1/8}, U_{3/8}, U_{5/8}, U_{7/8}$ such that

$$\begin{aligned} K &\subset U_{1/8} \subset \bar{U}_{1/8} \subset U_{1/4} \subset \bar{U}_{1/4} \subset U_{3/8} \subset \bar{U}_{3/8} \subset U_{1/2} \\ &\subset \bar{U}_{1/2} \subset U_{5/8} \subset \bar{U}_{5/8} \subset U_{3/4} \subset \bar{U}_{3/4} \subset U_{7/8} \subset \bar{U}_{7/8} \subset U_1. \end{aligned}$$

Continuing this way inductively, one shows there exists precompact open sets $\{U_r\}_{r\in\mathbb{D}} \subset \tau$ such that

$$K \subset U_r \subset \overline{U}_r \subset U_s \subset U_1 \subset \bar{U}_1 \subset U$$

for all $r, s \in \mathbb{D}$ with $0 < r < s \le 1$.

Step 2. Let $U_r := X$ if $r > 1$ and define

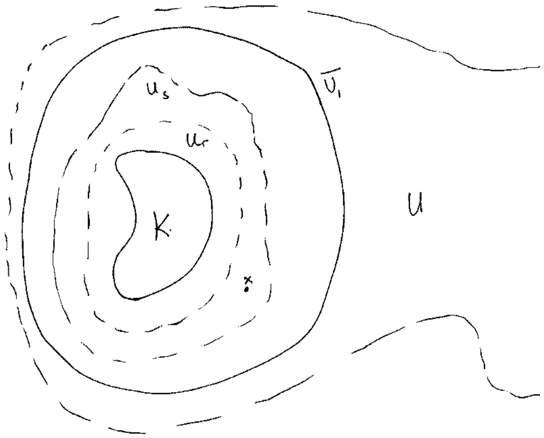


Fig. 37.2. Determining g from $\{U_r\}$.

$$g(x) = \inf\{r \in \mathbb{D} \cup (1,2) : x \in U_r\},$$

see Figure 37.2. Then $g(x) \in [0,1]$ for all $x \in X$, $g(x) = 0$ for $x \in K$ since $x \in K \subset U_r$ for all $r \in \mathbb{D}$. If $x \in U_1^c$, then $x \notin U_r$ for all $r \in \mathbb{D}$ and hence $g(x) = 1$. Therefore $f := 1 - g$ is a function such that $f = 1$ on K and $\{f \neq 0\} = \{g \neq 1\} \subset U_1 \subset \bar{U}_1 \subset U$ so that $\operatorname{supp}(f) = \overline{\{f \neq 0\}} \subset \bar{U}_1 \subset U$ is a compact subset of U. Thus it only remains to show f, or equivalently g, is continuous.

Since $\mathcal{E} = \{(\alpha, \infty), (-\infty, \alpha) : \alpha \in \mathbb{R}\}$ generates the standard topology on $\mathbb{R}$, to prove g is continuous it suffices to show $\{g < \alpha\}$ and $\{g > \alpha\}$ are open sets for all $\alpha \in \mathbb{R}$. But $g(x) < \alpha$ iff there exists $r \in \mathbb{D} \cup (1, \infty)$ with $r < \alpha$ such that $x \in U_r$. Therefore

$$\{g < \alpha\} = \bigcup \{U_r : r \in \mathbb{D} \cup (1,\infty) \ni r < \alpha\}$$

which is open in X. If $\alpha \geq 1$, $\{g > \alpha\} = \emptyset$ and if $\alpha < 0$, $\{g > \alpha\} = X$. If $\alpha \in (0,1)$, then $g(x) > \alpha$ iff there exists $r \in \mathbb{D}$ such that $r > \alpha$ and $x \notin U_r$. Now if $r > \alpha$ and $x \notin U_r$ then for $s \in \mathbb{D} \cap (\alpha, r)$, $x \notin \bar{U}_s \subset U_r$. Thus we have shown that

$$\{g > \alpha\} = \bigcup \left\{ \left(\overline{U}_s\right)^c : s \in \mathbb{D} \ni s > \alpha \right\}$$

which is again an open subset of X. ∎

Theorem 37.9 (Locally Compact Tietz Extension Theorem). *Let (X,τ) be a locally compact Hausdorff space, $K \sqsubset\sqsubset U \subset_o X$, $f \in C(K, \mathbb{R})$, $a = \min f(K)$ and $b = \max f(K)$. Then there exists $F \in C(X, [a,b])$ such that $F|_K = f$. Moreover given $c \in [a,b]$, F can be chosen so that $\operatorname{supp}(F - c) = \overline{\{F \neq c\}} \subset U$.*

The proof of this theorem is similar to Theorem 14.5 and will be left to the reader, see Exercise 37.7.

37.1 Locally compact form of Urysohn's Metrization Theorem

Definition 37.10 (Polish spaces). *A **Polish space** is a separable topological space (X,τ) which admits a complete metric, ρ, such that $\tau = \tau_\rho$.*

Notation 37.11 *Let $Q := [0,1]^{\mathbb{N}}$ denote the (infinite dimensional) **unit cube** in $\mathbb{R}^{\mathbb{N}}$. For $a, b \in Q$ let*

$$d(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n} |a_n - b_n|. \tag{37.2}$$

The metric introduced in Exercise 35.37 would be defined, in this context, as $\tilde{d}(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|a_n - b_n|}{1 + |a_n - b_n|}$. Since $1 \leq 1 + |a_n - b_n| \leq 2$, it follows that $\tilde{d} \leq d \leq 2d$. So the metrics d and $\tilde{d}$ are equivalent and in particular the topologies induced by d and $\tilde{d}$ are the same. By Exercises 35.36, the d – topology on Q is the same as the product topology and by Tychonoff's Theorem 36.16 or by Exercise 35.37, (Q, d) is a compact metric space.

Theorem 37.12. *To every separable metric space (X, ρ), there exists a continuous injective map $G : X \to Q$ such that $G : X \to G(X) \subset Q$ is a homeomorphism. Moreover if the metric, ρ, is also complete, then $G(X)$ is a G_δ –set, i.e. the $G(X)$ is the countable intersection of open subsets of (Q,d). In short, any separable metrizable space X is homeomorphic to a subset of (Q,d) and if X is a Polish space then X is homeomorphic to a G_δ – subset of (Q,d).*

Proof. (See Rogers and Williams [19], Theorem 82.5 on p. 106.) By replacing ρ by $\frac{\rho}{1+\rho}$ if necessary, we may assume that $0 \leq \rho < 1$. Let $D = \{a_n\}_{n=1}^{\infty}$ be a countable dense subset of X and define

$$G(x) = (\rho(x, a_1), \rho(x, a_2), \rho(x, a_3), \dots) \in Q$$

and

$$\gamma(x,y) = d(G(x), G(y)) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\rho(x, a_n) - \rho(y, a_n)|$$

for $x, y \in X$. To prove the first assertion, we must show G is injective and γ is a metric on X which is compatible with the topology determined by ρ.

If $G(x) = G(y)$, then $\rho(x,a) = \rho(y,a)$ for all $a \in D$. Since D is a dense subset of X, we may choose $\alpha_k \in D$ such that

$$0 = \lim_{k\to\infty} \rho(x,\alpha_k) = \lim_{k\to\infty} \rho(y,\alpha_k) = \rho(y,x)$$

and therefore $x = y$. A simple argument using the dominated convergence theorem shows $y \to \gamma(x,y)$ is ρ – continuous, i.e. $\gamma(x,y)$ is small if $\rho(x,y)$ is small. Conversely,

$$\begin{aligned}\rho(x,y) &\le \rho(x,a_n) + \rho(y,a_n) = 2\rho(x,a_n) + \rho(y,a_n) - \rho(x,a_n)\\ &\le 2\rho(x,a_n) + |\rho(x,a_n) - \rho(y,a_n)| \le 2\rho(x,a_n) + 2^n\gamma(x,y).\end{aligned}$$

Hence if $\varepsilon > 0$ is given, we may choose n so that $2\rho(x,a_n) < \varepsilon/2$ and so if $\gamma(x,y) < 2^{-(n+1)}\varepsilon$, it will follow that $\rho(x,y) < \varepsilon$. This shows $\tau_\gamma = \tau_\rho$. Since $G : (X,\gamma) \to (Q,d)$ is isometric, G is a homeomorphism.

Now suppose that (X,ρ) is a complete metric space. Let $S := G(X)$ and σ be the metric on S defined by $\sigma(G(x), G(y)) = \rho(x,y)$ for all $x,y \in X$. Then (S,σ) is a complete metric (being the isometric image of a complete metric space) and by what we have just prove, $\tau_\sigma = \tau_{d_S}$.Consequently, if $u \in S$ and $\varepsilon > 0$ is given, we may find $\delta'(\varepsilon)$ such that $B_\sigma(u,\delta'(\varepsilon)) \subset B_d(u,\varepsilon)$. Taking $\delta(\varepsilon) = \min(\delta'(\varepsilon),\varepsilon)$, we have $\operatorname{diam}_d(B_d(u,\delta(\varepsilon))) < \varepsilon$ and $\operatorname{diam}_\sigma(B_d(u,\delta(\varepsilon))) < \varepsilon$ where

$$\begin{aligned}\operatorname{diam}_\sigma(A) &:= \{\sup \sigma(u,v) : u,v \in A\} \text{ and}\\ \operatorname{diam}_d(A) &:= \{\sup d(u,v) : u,v \in A\}.\end{aligned}$$

Let $\bar{S}$ denote the closure of S inside of (Q,d) and for each $n \in \mathbb{N}$ let

$$\mathcal{N}_n := \{N \in \tau_d : \operatorname{diam}_d(N) \vee \operatorname{diam}_\sigma(N \cap S) < 1/n\}$$

and let $U_n := \cup\mathcal{N}_n \in \tau_d$. From the previous paragraph, it follows that $S \subset U_n$ and therefore $S \subset \bar{S} \cap (\cap_{n=1}^\infty U_n)$.

Conversely if $u \in \bar{S} \cap (\cap_{n=1}^\infty U_n)$ and $n \in \mathbb{N}$, there exists $N_n \in \mathcal{N}_n$ such that $u \in N_n$. Moreover, since $N_1 \cap \dots \cap N_n$ is an open neigborhood of $u \in \bar{S}$, there exists $u_n \in N_1 \cap \dots \cap N_n \cap S$ for each $n \in \mathbb{N}$. From the definition of $\mathcal{N}_n$, we have $\lim_{n\to\infty} d(u,u_n) = 0$ and $\sigma(u_n,u_m) \le \max(n^{-1}, m^{-1}) \to 0$ as $m,n \to \infty$. Since (S,σ) is complete, it follows that $\{u_n\}_{n=1}^\infty$ is convergent in (S,σ) to some element $u_0 \in S$. Since (S,d_S) has the same topology as (S,σ) it follows that $d(u_n,u_0) \to 0$ as well and thus that $u = u_0 \in S$. We have now shown, $S = \bar{S} \cap (\cap_{n=1}^\infty U_n)$. This completes the proof because we may write $\bar{S} = \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ where $S_{1/n} := \left\{u \in Q : d(u,\bar{S}) < 1/n\right\}$ and therefore, $S = \left(\bigcap_{n=1}^\infty U_n\right) \cap \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ is a G_δ set. ∎

Theorem 37.13 (Urysohn Metrization Theorem for LCH's). *Every second countable locally compact Hausdorff space, (X,τ), is metrizable, i.e. there is a metric ρ on X such that $\tau = \tau_\rho$. Moreover, ρ may be chosen so that X is isometric to a subset $Q_0 \subset Q$ equipped with the metric d in Eq. (37.2). In this metric X is totally bounded and hence the completion of X (which is isometric to $\bar{Q}_0 \subset Q$) is compact. (Also see Theorem 37.45.)*

Proof. Let $\mathcal{B}$ be a countable base for τ and set

$$\Gamma := \{(U,V) \in \mathcal{B} \times \mathcal{B} \mid \bar{U} \subset V \text{ and } \bar{U} \text{ is compact}\}.$$

To each $O \in \tau$ and $x \in O$ there exist $(U,V) \in \Gamma$ such that $x \in U \subset V \subset O$. Indeed, since $\mathcal{B}$ is a base for τ, there exists $V \in \mathcal{B}$ such that $x \in V \subset O$. Now apply Proposition 37.7 to find $U' \subset_o X$ such that $x \in U' \subset \bar{U}' \subset V$ with $\bar{U}'$ being compact. Since $\mathcal{B}$ is a base for τ, there exists $U \in \mathcal{B}$ such that $x \in U \subset U'$ and since $\bar{U} \subset \bar{U}'$, $\bar{U}$ is compact so $(U,V) \in \Gamma$. In particular this shows that $\mathcal{B}' := \{U \in \mathcal{B} : (U,V) \in \Gamma \text{ for some } V \in \mathcal{B}\}$ is still a base for τ. If Γ is a finite, then $\mathcal{B}'$ is finite and τ only has a finite number of elements as well. Since (X,τ) is Hausdorff, it follows that X is a finite set. Letting $\{x_n\}_{n=1}^N$ be an enumeration of X, define $T : X \to Q$ by $T(x_n) = e_n$ for $n = 1,2,\dots,N$ where $e_n = (0,0,\dots,0,1,0,\dots)$, with the 1 occurring in the n^{th} spot. Then $\rho(x,y) := d(T(x),T(y))$ for $x,y \in X$ is the desired metric.

So we may now assume that Γ is an infinite set and let $\{(U_n,V_n)\}_{n=1}^\infty$ be an enumeration of Γ. By Urysohn's Lemma 37.8 there exists $f_{U,V} \in C(X,[0,1])$ such that $f_{U,V} = 0$ on $\bar{U}$ and $f_{U,V} = 1$ on V^c. Let $\mathcal{F} := \{f_{U,V} \mid (U,V) \in \Gamma\}$ and set $f_n := f_{U_n,V_n}$ – an enumeration of $\mathcal{F}$. We will now show that

$$\rho(x,y) := \sum_{n=1}^\infty \frac{1}{2^n} |f_n(x) - f_n(y)|$$

is the desired metric on X. The proof will involve a number of steps.

1. (ρ is a metric on X.) It is routine to show ρ satisfies the triangle inequality and ρ is symmetric. If $x,y \in X$ are distinct points then there exists $(U_{n_0},V_{n_0}) \in \Gamma$ such that $x \in U_{n_0}$ and $V_{n_0} \subset O := \{y\}^c$. Since $f_{n_0}(x) = 0$ and $f_{n_0}(y) = 1$, it follows that $\rho(x,y) \ge 2^{-n_0} > 0$.
2. (Let $\tau_0 = \tau(f_n : n \in \mathbb{N})$, then $\tau = \tau_0 = \tau_\rho$.) As usual we have $\tau_0 \subset \tau$. Since, for each $x \in X$, $y \to \rho(x,y)$ is τ_0 – continuous (being the uniformly convergent sum of continuous functions), it follows that $B_x(\varepsilon) := \{y \in X : \rho(x,y) < \varepsilon\} \in \tau_0$ for all $x \in X$ and $\varepsilon > 0$. Thus $\tau_\rho \subset \tau_0 \subset \tau$. Suppose that $O \in \tau$ and $x \in O$. Let $(U_{n_0},V_{n_0}) \in \Gamma$ be such that $x \in U_{n_0}$ and $V_{n_0} \subset O$. Then $f_{n_0}(x) = 0$ and $f_{n_0} = 1$ on O^c. Therefore if $y \in X$ and $f_{n_0}(y) < 1$, then $y \in O$ so $x \in \{f_{n_0} < 1\} \subset O$. This shows that O may be written as a union of elements from τ_0 and therefore $O \in \tau_0$. So $\tau \subset \tau_0$ and hence $\tau = \tau_0$. Moreover, if $y \in B_x(2^{-n_0})$ then $2^{-n_0} > \rho(x,y) \ge 2^{-n_0} f_{n_0}(y)$ and therefore $x \in B_x(2^{-n_0}) \subset \{f_{n_0} < 1\} \subset O$. This shows O is ρ – open and hence $\tau_\rho \subset \tau_0 \subset \tau \subset \tau_\rho$.

3. (X is isometric to some $Q_0 \subset Q$.) Let $T : X \to Q$ be defined by $T(x) = (f_1(x), f_2(x), \dots, f_n(x), \dots)$. Then T is an isometry by the very definitions of d and ρ and therefore X is isometric to $Q_0 := T(X)$. Since Q_0 is a subset of the compact metric space (Q, d), Q_0 is totally bounded and therefore X is totally bounded.

■

BRUCE: Add Stone Chech Compactification results.

37.2 Partitions of Unity

Definition 37.14. *Let (X, τ) be a topological space and $X_0 \subset X$ be a set. A collection of sets $\{B_\alpha\}_{\alpha \in A} \subset 2^X$ is* ***locally finite*** *on X_0 if for all $x \in X_0$, there is an open neighborhood $N_x \in \tau$ of x such that $\#\{\alpha \in A : B_\alpha \cap N_x \neq \emptyset\} < \infty$.*

Definition 37.15. *Suppose that $\mathcal{U}$ is an open cover of $X_0 \subset X$. A collection $\{\varphi_\alpha\}_{\alpha \in A} \subset C(X, [0, 1])$ ($N = \infty$ is allowed here) is a* ***partition of unity*** *on X_0 subordinate to the cover $\mathcal{U}$ if:*

1. *for all α there is a $U \in \mathcal{U}$ such that $\operatorname{supp}(\varphi_\alpha) \subset U$,*
2. *the collection of sets, $\{\operatorname{supp}(\varphi_\alpha)\}_{\alpha \in A}$, is locally finite on X, and*
3. *$\sum_{\alpha \in A} \varphi_\alpha = 1$ on X_0.*

Notice by item 2. that, for each $x \in X$, there is a neighborhood N_x such that

$$\Lambda := \{\alpha \in A : \operatorname{supp}(\varphi_\alpha) \cap N_x \neq \emptyset\}$$

is a finite set. Therefore, $\sum_{\alpha \in A} \varphi_\alpha|_{N_x} = \sum_{\alpha \in \Lambda} \varphi_\alpha|_{N_x}$ which shows the sum $\sum_{\alpha \in A} \varphi_\alpha$ is well defined and defines a continuous function on N_x and therefore on X since continuity is a local property. We will summarize these last comments by saying the sum, $\sum_{\alpha \in A} \varphi_\alpha$, is ***locally finite****.*

Proposition 37.16 (Partitions of Unity: The Compact Case). *Suppose that X is a locally compact Hausdorff space, $K \subset X$ is a compact set and $\mathcal{U} = \{U_j\}_{j=1}^n$ is an open cover of K. Then there exists a partition of unity $\{h_j\}_{j=1}^n$ of K such that $h_j \prec U_j$ for all $j = 1, 2, \dots, n$.*

Proof. For all $x \in K$ choose a precompact open neighborhood, V_x, of x such that $\overline{V}_x \subset U_j$. Since K is compact, there exists a finite subset, Λ, of K such that $K \subset \bigcup_{x \in \Lambda} V_x$. Let

$$F_j = \cup \left\{\bar{V}_x : x \in \Lambda \text{ and } \overline{V}_x \subset U_j\right\}.$$

Then F_j is compact, $F_j \subset U_j$ for all j, and $K \subset \cup_{j=1}^n F_j$. By Urysohn's Lemma 37.8 there exists $f_j \prec U_j$ such that $f_j = 1$ on F_j for $j = 1, 2, \dots, n$ and by convention let $f_{n+1} \equiv 1$. We will now give two methods to finish the proof.

Method 1. Let $h_1 = f_1$, $h_2 = f_2(1 - h_1) = f_2(1 - f_1)$,

$$h_3 = f_3(1 - h_1 - h_2) = f_3(1 - f_1 - (1 - f_1)f_2) = f_3(1 - f_1)(1 - f_2)$$

and continue on inductively to define

$$h_k = (1 - h_1 - \dots - h_{k-1})f_k = f_k \cdot \prod_{j=1}^{k-1}(1 - f_j) \; \forall \, k = 2, 3, \dots, n \tag{37.3}$$

and to show

$$h_{n+1} = (1 - h_1 - \dots - h_n) \cdot 1 = 1 \cdot \prod_{j=1}^{n}(1 - f_j). \tag{37.4}$$

From these equations it clearly follows that $h_j \in C_c(X, [0, 1])$ and that $\operatorname{supp}(h_j) \subset \operatorname{supp}(f_j) \subset U_j$, i.e. $h_j \prec U_j$. Since $\prod_{j=1}^n (1 - f_j) = 0$ on K, $\sum_{j=1}^n h_j = 1$ on K and $\{h_j\}_{j=1}^n$ is the desired partition of unity.

Method 2. Let $g := \sum_{j=1}^{n} f_j \in C_c(X)$. Then $g \geq 1$ on K and hence $K \subset \{g > \frac{1}{2}\}$. Choose $\varphi \in C_c(X, [0, 1])$ such that $\varphi = 1$ on K and $\operatorname{supp}(\varphi) \subset \{g > \frac{1}{2}\}$ and define $f_0 := 1 - \varphi$. Then $f_0 = 0$ on K, $f_0 = 1$ if $g \leq \frac{1}{2}$ and therefore,

$$f_0 + f_1 + \dots + f_n = f_0 + g > 0$$

on X. The desired partition of unity may be constructed as

$$h_j(x) = \frac{f_j(x)}{f_0(x) + \dots + f_n(x)}.$$

Indeed $\operatorname{supp}(h_j) = \operatorname{supp}(f_j) \subset U_j$, $h_j \in C_c(X, [0, 1])$ and on K,

$$h_1 + \dots + h_n = \frac{f_1 + \dots + f_n}{f_0 + f_1 + \dots + f_n} = \frac{f_1 + \dots + f_n}{f_1 + \dots + f_n} = 1.$$

■

Proposition 37.17. *Let (X, τ) be a locally compact and σ – compact Hausdorff space. Suppose that $\mathcal{U} \subset \tau$ is an open cover of X. Then we may construct two locally finite open covers $\mathcal{V} = \{V_i\}_{i=1}^N$ and $\mathcal{W} = \{W_i\}_{i=1}^N$ of X ($N = \infty$ is allowed here) such that:*

1. *$W_i \subset \bar{W}_i \subset V_i \subset \bar{V}_i$ and $\bar{V}_i$ is compact for all i.*

2. *For each i there exist $U \in \mathcal{U}$ such that $\bar{V}_i \subset U$.*

Proof. By Remark 36.6, there exists an open cover of $\mathcal{G} = \{G_n\}_{n=1}^\infty$ of X such that $G_n \subset \bar{G}_n \subset G_{n+1}$. Then $X = \cup_{k=1}^\infty (\bar{G}_k \setminus \bar{G}_{k-1})$, where by convention $G_{-1} = G_0 = \emptyset$. For the moment fix $k \geq 1$. For each $x \in \bar{G}_k \setminus G_{k-1}$, let $U_x \in \mathcal{U}$ be chosen so that $x \in U_x$ and by Proposition 37.7 choose an open neighborhood N_x of x such that $\bar{N}_x \subset U_x \cap (G_{k+1} \setminus \bar{G}_{k-2})$, see Figure 37.3 below. Since

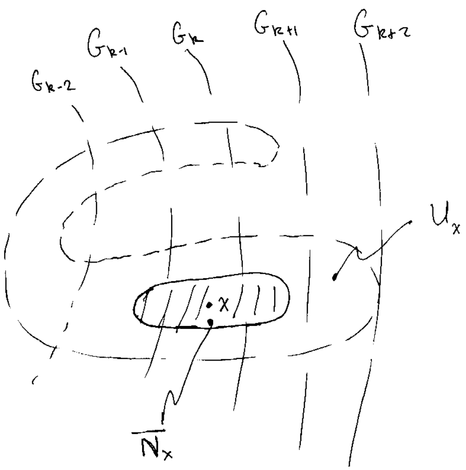


Fig. 37.3. Constructing the $\{W_i\}_{i=1}^N$.

$\{N_x\}_{x \in \bar{G}_k \setminus G_{k-1}}$ is an open cover of the compact set $\bar{G}_k \setminus G_{k-1}$, there exist a finite subset $\Gamma_k \subset \{N_x\}_{x \in \bar{G}_k \setminus G_{k-1}}$ which also covers $\bar{G}_k \setminus G_{k-1}$.

By construction, for each $W \in \Gamma_k$, there is a $U \in \mathcal{U}$ such that $\bar{W} \subset U \cap (G_{k+1} \setminus \bar{G}_{k-2})$ and by another application of Proposition 37.7, there exists an open set V_W such that $\bar{W} \subset V_W \subset \bar{V}_W \subset U \cap (G_{k+1} \setminus \bar{G}_{k-2})$. We now choose and enumeration $\{W_i\}_{i=1}^N$ of the countable open cover, $\cup_{k=1}^\infty \Gamma_k$, of X and define $V_i = V_{W_i}$. Then the collection $\{W_i\}_{i=1}^N$ and $\{V_i\}_{i=1}^N$ are easily checked to satisfy all the conclusions of the proposition. In particular notice that for each k; $V_i \cap G_k \neq \emptyset$ for only a finite number of i's. ∎

Theorem 37.18 (Partitions of Unity for σ − Compact LCH Spaces). *Let (X, τ) be locally compact, σ − compact and Hausdorff and let $\mathcal{U} \subset \tau$ be an open cover of X. Then there exists a partition of unity of $\{h_i\}_{i=1}^N$ ($N = \infty$ is allowed here) subordinate to the cover $\mathcal{U}$ such that* $\operatorname{supp}(h_i)$ *is compact for all i.*

Proof. Let $\mathcal{V} = \{V_i\}_{i=1}^N$ and $\mathcal{W} = \{W_i\}_{i=1}^N$ be open covers of X with the properties described in Proposition 37.17. By Urysohn's Lemma 37.8, there exists $f_i \prec V_i$ such that $f_i = 1$ on $\bar{W}_i$ for each i. As in the proof of Proposition 37.16 there are two methods to finish the proof.

Method 1. Define $h_1 = f_1$, h_j by Eq. (37.3) for all other j. Then as in Eq. (37.4), for all $n < N + 1$,

$$1 - \sum_{j=1}^\infty h_j = \lim_{n \to \infty} \left(f_n \prod_{j=1}^n (1 - f_j) \right) = 0$$

since for $x \in X$, $f_j(x) = 1$ for some j. As in the proof of Proposition 37.16, it is easily checked that $\{h_i\}_{i=1}^N$ is the desired partition of unity.

Method 2. Let $f := \sum_{i=1}^N f_i$, a locally finite sum, so that $f \in C(X)$. Since $\{W_i\}_{i=1}^\infty$ is a cover of X, $f \geq 1$ on X so that $1/f \in C(X))$ as well. The functions $h_i := f_i / f$ for $i = 1, 2, \ldots, N$ give the desired partition of unity. ∎

Lemma 37.19. *Let (X, τ) be a locally compact Hausdorff space.*

1. *A subset $E \subset X$ is closed iff $E \cap K$ is closed for all $K \sqsubset\sqsubset X$.*
2. *Let $\{C_\alpha\}_{\alpha \in A}$ be a locally finite collection of closed subsets of X, then $C = \cup_{\alpha \in A} C_\alpha$ is closed in X. (Recall that in general closed sets are only closed under finite unions.)*

Proof. 1. Since compact subsets of Hausdorff spaces are closed, $E \cap K$ is closed if E is closed and K is compact. Now suppose that $E \cap K$ is closed for all compact subsets $K \subset X$ and let $x \in E^c$. Since X is locally compact, there exists a precompact open neighborhood, V, of x.[1] By assumption $E \cap \bar{V}$ is closed so $x \in \left(E \cap \bar{V}\right)^c$ – an open subset of X. By Proposition 37.7 there exists an open set U such that $x \in U \subset \bar{U} \subset \left(E \cap \bar{V}\right)^c$, see Figure 37.4. Let $W := U \cap V$. Since

$$W \cap E = U \cap V \cap E \subset U \cap \bar{V} \cap E = \emptyset,$$

and W is an open neighborhood of x and $x \in E^c$ was arbitrary, we have shown E^c is open hence E is closed.

2. Let K be a compact subset of X and for each $x \in K$ let N_x be an open neighborhood of x such that $\#\{\alpha \in A : C_\alpha \cap N_x \neq \emptyset\} < \infty$. Since K is compact, there exists a finite subset $\Lambda \subset K$ such that $K \subset \cup_{x \in \Lambda} N_x$. Letting $\Lambda_0 := \{\alpha \in A : C_\alpha \cap K \neq \emptyset\}$, then

[1] If X were a metric space we could finish the proof as follows. If there does not exist an open neighborhood of x which is disjoint from E, then there would exists $x_n \in E$ such that $x_n \to x$. Since $E \cap \bar{V}$ is closed and $x_n \in E \cap \bar{V}$ for all large n, it follows (see Exercise 13.4) that $x \in E \cap \bar{V}$ and in particular that $x \in E$. But we chose $x \in E^c$.

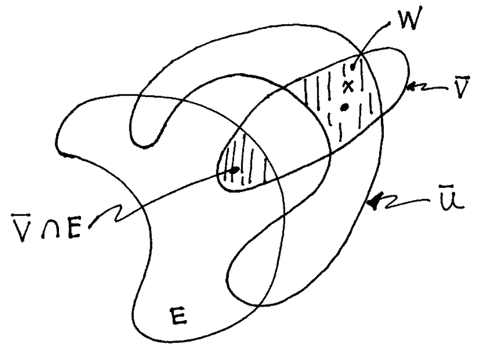


Fig. 37.4. Showing E^c is open.

$$\#\left(\Lambda_0\right) \le \sum_{x\in\Lambda} \#\{\alpha\in A : C_\alpha \cap N_x \ne \emptyset\} < \infty$$

and hence $K\cap\left(\cup_{\alpha\in A}C_\alpha\right) = K\cap\left(\cup_{\alpha\in\Lambda_0}C_\alpha\right)$. The set $\left(\cup_{\alpha\in\Lambda_0}C_\alpha\right)$ is a finite union of closed sets and hence closed. Therefore, $K\cap\left(\cup_{\alpha\in A}C_\alpha\right)$ is closed and by item 1. it follows that $\cup_{\alpha\in A}C_\alpha$ is closed as well. ■

Corollary 37.20. *Let (X,τ) be a locally compact and σ – compact Hausdorff space and $\mathcal{U}=\{U_\alpha\}_{\alpha\in A}\subset\tau$ be an open cover of X. Then there exists a partition of unity of $\{h_\alpha\}_{\alpha\in A}$ subordinate to the cover $\mathcal{U}$ such that $\operatorname{supp}(h_\alpha)\subset U_\alpha$ for all $\alpha\in A$. (Notice that we do not assert that h_α has compact support. However if $\bar{U}_\alpha$ is compact then $\operatorname{supp}(h_\alpha)$ will be compact.)*

Proof. By the σ – compactness of X, we may choose a countable subset, $\{\alpha_i\}_{i=1}^N$ ($N=\infty$ allowed here), of A such that $\{U_i := U_{\alpha_i}\}_{i=1}^N$ is still an open cover of X. Let $\{g_j\}_{j=1}^\infty$ be a partition of unity[2] subordinate to the cover $\{U_i\}_{i=1}^N$ as in Theorem 37.18. Define $\tilde{\Gamma}_k := \{j : \operatorname{supp}(g_j)\subset U_k\}$ and $\Gamma_k := \tilde{\Gamma}_k \setminus \left(\cup_{j=1}^{k-1}\tilde{\Gamma}_k\right)$, where by convention $\tilde{\Gamma}_0=\emptyset$. Then

$$\mathbb{N} = \bigcup_{k=1}^\infty \tilde{\Gamma}_k = \coprod_{k=1}^\infty \Gamma_k.$$

If $\Gamma_k=\emptyset$ let $h_k := 0$ otherwise let $h_k := \sum_{j\in\Gamma_k} g_j$, a locally finite sum. Then

$$\sum_{k=1}^N h_k = \sum_{j=1}^\infty g_j = 1.$$

[2] So as to simplify the indexing we assume there countable number of g_j's. This can always be arranged by taking $g_k\equiv 0$ for large k if necessary.

By Item 2. of Lemma 37.19, $\cup_{j\in\Gamma_k}\operatorname{supp}(g_j)$ is closed and therefore,

$$\operatorname{supp}(h_k) = \overline{\{h_k\ne 0\}} = \overline{\cup_{j\in\Gamma_k}\{g_j\ne 0\}} \subset \cup_{j\in\Gamma_k}\operatorname{supp}(g_j)\subset U_k$$

and hence $h_k \prec U_k$ and the sum $\sum_{k=1}^N h_k$ is still locally finite. (Why?) The desired partition of unity is now formed by letting $h_{\alpha_k} := h_k$ for $k<N+1$ and $h_\alpha\equiv 0$ if $\alpha\notin\{\alpha_i\}_{i=1}^N$. ■

Corollary 37.21. *Let (X,τ) be a locally compact and σ – compact Hausdorff space and A,B be disjoint closed subsets of X. Then there exists $f\in C(X,[0,1])$ such that $f=1$ on A and $f=0$ on B. In fact f can be chosen so that $\operatorname{supp}(f)\subset B^c$.*

Proof. Let $U_1=A^c$ and $U_2=B^c$, then $\{U_1,U_2\}$ is an open cover of X. By Corollary 37.20 there exists $h_1,h_2\in C(X,[0,1])$ such that $\operatorname{supp}(h_i)\subset U_i$ for $i=1,2$ and $h_1+h_2=1$ on X. The function $f=h_2$ satisfies the desired properties. ■

37.3 $C_0(X)$ and the Alexanderov Compactification

Definition 37.22. *Let (X,τ) be a topological space. A continuous function $f : X\to\mathbb{C}$ is said to* ***vanish at infinity*** *if $\{|f|\ge\varepsilon\}$ is compact in X for all $\varepsilon>0$. The functions, $f\in C(X)$, vanishing at infinity will be denoted by $C_0(X)$. (Notice that $C_0(X)=C(X)$ whenever X is compact.)*

Proposition 37.23. *Let X be a topological space, $BC(X)$ be the space of bounded continuous functions on X with the supremum norm topology. Then*

1. *$C_0(X)$ is a closed subspace of $BC(X)$.*
2. *If we further assume that X is a locally compact Hausdorff space, then $C_0(X)=\overline{C_c(X)}$.*

Proof.

1. If $f\in C_0(X)$, $K_1 := \{|f|\ge 1\}$ is a compact subset of X and therefore $f(K_1)$ is a compact and hence bounded subset of $\mathbb{C}$ and so $M := \sup_{x\in K_1}|f(x)| < \infty$. Therefore $\|f\|_\infty \le M\vee 1 < \infty$ showing $f\in BC(X)$. Now suppose $f_n\in C_0(X)$ and $f_n\to f$ in $BC(X)$. Let $\varepsilon>0$ be given and choose n sufficiently large so that $\|f-f_n\|_\infty\le\varepsilon/2$. Since
$$|f|\le|f_n|+|f-f_n|\le|f_n|+\|f-f_n\|_\infty\le|f_n|+\varepsilon/2,$$
$$\{|f|\ge\varepsilon\}\subset\{|f_n|+\varepsilon/2\ge\varepsilon\}=\{|f_n|\ge\varepsilon/2\}.$$
Because $\{|f|\ge\varepsilon\}$ is a closed subset of the compact set $\{|f_n|\ge\varepsilon/2\}$, $\{|f|\ge\varepsilon\}$ is compact and we have shown $f\in C_0(X)$.

2. Since $\overline{C_0(X)}$ is a closed subspace of $BC(X)$ and $C_c(X) \subset C_0(X)$, we always have $\overline{C_c(X)} \subset C_0(X)$. Now suppose that $f \in C_0(X)$ and let $K_n := \{|f| \geq \frac{1}{n}\} \sqsubset\sqsubset X$. By Lemma 37.8 we may choose $\varphi_n \in C_c(X, [0,1])$ such that $\varphi_n \equiv 1$ on K_n. Define $f_n := \varphi_n f \in C_c(X)$. Then

$$\|f - f_n\|_u = \|(1-\varphi_n)f\|_\infty \leq \frac{1}{n} \to 0 \text{ as } n \to \infty.$$

This shows that $f \in \overline{C_c(X)}$.

■

Proposition 37.24 (Alexanderov Compactification). *Suppose that (X,τ) is a non-compact locally compact Hausdorff space. Let $X^* = X \cup \{\infty\}$, where $\{\infty\}$ is a new symbol not in X. The collection of sets,*

$$\tau^* = \tau \cup \{X^* \setminus K : K \sqsubset\sqsubset X\} \subset 2^{X^*},$$

is a topology on X^* and (X^*, τ^*) is a compact Hausdorff space. Moreover $f \in C(X)$ extends continuously to X^* iff $f = g + c$ with $g \in C_0(X)$ and $c \in \mathbb{C}$ in which case the extension is given by $f(\infty) = c$.

Proof. 1. (τ^* is a topology.) Let $\mathcal{F} := \{F \subset X^* : X^* \setminus F \in \tau^*\}$, i.e. $F \in \mathcal{F}$ iff F is a compact subset of X or $F = F_0 \cup \{\infty\}$ with F_0 being a closed subset of X. Since the finite union of compact (closed) subsets is compact (closed), it is easily seen that $\mathcal{F}$ is closed under finite unions. Because arbitrary intersections of closed subsets of X are closed and closed subsets of compact subsets of X are compact, it is also easily checked that $\mathcal{F}$ is closed under arbitrary intersections. Therefore $\mathcal{F}$ satisfies the axioms of the closed subsets associated to a topology and hence τ^* is a topology.

2. ((X^*, τ^*) is a Hausdorff space.) It suffices to show any point $x \in X$ can be separated from ∞. To do this use Proposition 37.7 to find an open precompact neighborhood, U, of x. Then U and $V := X^* \setminus \bar{U}$ are disjoint open subsets of X^* such that $x \in U$ and $\infty \in V$.

3. ((X^*, τ^*) is compact.) Suppose that $\mathcal{U} \subset \tau^*$ is an open cover of X^*. Since $\mathcal{U}$ covers ∞, there exists a compact set $K \subset X$ such that $X^* \setminus K \in \mathcal{U}$. Clearly X is covered by $\mathcal{U}_0 := \{V \setminus \{\infty\} : V \in \mathcal{U}\}$ and by the definition of τ^* (or using (X^*, τ^*) is Hausdorff), $\mathcal{U}_0$ is an open cover of X. In particular $\mathcal{U}_0$ is an open cover of K and since K is compact there exists $\Lambda \subset\subset \mathcal{U}$ such that $K \subset \cup \{V \setminus \{\infty\} : V \in \Lambda\}$. It is now easily checked that $\Lambda \cup \{X^* \setminus K\} \subset \mathcal{U}$ is a finite subcover of X^*.

4. (Continuous functions on $C(X^*)$ statements.) Let $i : X \to X^*$ be the inclusion map. Then i is continuous and open, i.e. $i(V)$ is open in X^* for all V open in X. If $f \in C(X^*)$, then $g = f|_X - f(\infty) = f \circ i - f(\infty)$ is continuous on X. Moreover, for all $\varepsilon > 0$ there exists an open neighborhood $V \in \tau^*$ of ∞ such that

$$|g(x)| = |f(x) - f(\infty)| < \varepsilon \text{ for all } x \in V.$$

Since V is an open neighborhood of ∞, there exists a compact subset, $K \subset X$, such that $V = X^* \setminus K$. By the previous equation we see that $\{x \in X : |g(x)| \geq \varepsilon\} \subset K$, so $\{|g| \geq \varepsilon\}$ is compact and we have shown g vanishes at ∞.

Conversely if $g \in C_0(X)$, extend g to X^* by setting $g(\infty) = 0$. Given $\varepsilon > 0$, the set $K = \{|g| \geq \varepsilon\}$ is compact, hence $X^* \setminus K$ is open in X^*. Since $g(X^* \setminus K) \subset (-\varepsilon, \varepsilon)$ we have shown that g is continuous at ∞. Since g is also continuous at all points in X it follows that g is continuous on X^*. Now it $f = g + c$ with $c \in \mathbb{C}$ and $g \in C_0(X)$, it follows by what we just proved that defining $f(\infty) = c$ extends f to a continuous function on X^*. ■

Example 37.25. Let X be an uncountable set and τ be the discrete topology on X. Let $(X^* = X \cup \{\infty\}, \tau^*)$ be the one point compactification of X. The smallest dense subset of X^* is the uncountable set X. Hence X^* is a compact but non-separable and hence non-metrizable space.

Exercise 37.4. Let $X := \{0,1\}^{\mathbb{R}}$ and τ be the product topology on X where $\{0,1\}$ is equipped with the discrete topology. Show (X, τ) is separable. (Combining this with Exercise 35.10 and Tychonoff's Theorem 36.16, we see that (X, τ) is compact and separable but not first countable.)

The next proposition gathers a number of results involving countability assumptions which have appeared in the exercises.

Proposition 37.26 (Summary). *Let (X, τ) be a topological space.*

1. *If (X,τ) is second countable, then (X,τ) is separable; see Exercise 35.12.*
2. *If (X,τ) is separable and metrizable then (X,τ) is second countable; see Exercise 35.13.*
3. *If (X,τ) is locally compact and metrizable then (X,τ) is σ – compact iff (X,τ) is separable; see Exercises 36.3 and 36.4.*
4. *If (X,τ) is locally compact and second countable, then (X,τ) is σ - compact, see Exercise 36.1.*
5. *If (X,τ) is locally compact and metrizable, then (X,τ) is σ – compact iff (X,τ) is separable, see Exercises 36.2 and 36.3.*
6. *There exists spaces, (X,τ), which are both compact and separable but not first countable and in particular not metrizable, see Exercise 37.4.*

37.4 Stone-Weierstrass Theorem

We now wish to generalize Theorem 32.39 to more general topological spaces. We will first need some definitions.

Definition 37.27. *Let X be a topological space and $\mathcal{A} \subset C(X) = C(X, \mathbb{R})$ or $C(X, \mathbb{C})$ be a collection of functions. Then*

1. *$\mathcal{A}$ is said to **separate points** if for all distinct points $x, y \in X$ there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.*
2. *$\mathcal{A}$ is an **algebra** if $\mathcal{A}$ is a vector subspace of $C(X)$ which is closed under pointwise multiplication. (**Note well:** we do not assume $1 \in \mathcal{A}$.)*
3. *$\mathcal{A} \subset C(X, \mathbb{R})$ is called a **lattice** if $f \vee g := \max(f, g)$ and $f \wedge g = \min(f, g) \in \mathcal{A}$ for all $f, g \in \mathcal{A}$.*
4. *$\mathcal{A} \subset C(X, \mathbb{C})$ is closed under conjugation if $\bar{f} \in \mathcal{A}$ whenever $f \in \mathcal{A}$.*

Remark 37.28. If X is a topological space such that $C(X, \mathbb{R})$ separates points then X is Hausdorff. Indeed if $x, y \in X$ and $f \in C(X, \mathbb{R})$ such that $f(x) \neq f(y)$, then $f^{-1}(J)$ and $f^{-1}(I)$ are disjoint open sets containing x and y respectively when I and J are disjoint intervals containing $f(x)$ and $f(y)$ respectively.

Lemma 37.29. *If $\mathcal{A}$ is a closed sub-algebra of $BC(X, \mathbb{R})$ then $|f| \in \mathcal{A}$ for all $f \in \mathcal{A}$ and $\mathcal{A}$ is a lattice.*

Proof. Let $f \in \mathcal{A}$ and let $M = \sup_{x \in X} |f(x)|$. By either Exercise 4.9 or Theorems 7.36, 31.47, or 32.39, there are polynomials $p_n(t)$ such that

$$\lim_{n \to \infty} \sup_{|t| \leq M} ||t| - p_n(t)| = 0.$$

By replacing p_n by $p_n - p_n(0)$ if necessary we may assume that $p_n(0) = 0$. Since $\mathcal{A}$ is an algebra, it follows that $f_n = p_n(f) \in \mathcal{A}$ and $|f| \in \mathcal{A}$, because $|f|$ is the uniform limit of the f_n's. Since

$$f \vee g = \frac{1}{2}\,(f + g + |f - g|) \text{ and}$$
$$f \wedge g = \frac{1}{2}\,(f + g - |f - g|),$$

we have shown $\mathcal{A}$ is a lattice. ∎

Lemma 37.30. *Let $\mathcal{A} \subset C(X, \mathbb{R})$ be an algebra which separates points and suppose x and y are distinct points of X. If there exists $f, g \in \mathcal{A}$ such that*

$$f(x) \neq 0 \text{ and } g(y) \neq 0, \tag{37.5}$$

then

$$V := \{(f(x), f(y)) : f \in \mathcal{A}\} = \mathbb{R}^2. \tag{37.6}$$

Proof. It is clear that V is a non-zero subspace of $\mathbb{R}^2$. If $\dim(V) = 1$, then $V = \text{span}(a, b)$ for some $(a, b) \in \mathbb{R}^2$ which, necessarily by Eq. (37.5), satisfy $a \neq 0 \neq b$. Since $(a, b) = (f(x), f(y))$ for some $f \in \mathcal{A}$ and $f^2 \in \mathcal{A}$, it follows that $(a^2, b^2) = (f^2(x), f^2(y)) \in V$ as well. Since $\dim V = 1$, (a, b) and (a^2, b^2) are linearly dependent and therefore

$$0 = \det \begin{pmatrix} a & b \\ a^2 & b^2 \end{pmatrix} = ab^2 - a^2 b = ab(b - a)$$

which implies that $a = b$. But this the implies that $f(x) = f(y)$ for all $f \in \mathcal{A}$, violating the assumption that $\mathcal{A}$ separates points. Therefore we conclude that $\dim(V) = 2$, i.e. $V = \mathbb{R}^2$. ∎

Theorem 37.31 (Stone-Weierstrass Theorem). *Suppose X is a locally compact Hausdorff space and $\mathcal{A} \subset C_0(X, \mathbb{R})$ is a **closed** subalgebra which separates points. For $x \in X$ let*

$$\mathcal{A}_x := \{f(x) : f \in \mathcal{A}\} \text{ and}$$
$$\mathcal{I}_x = \{f \in C_0(X, \mathbb{R}) : f(x) = 0\}.$$

Then either one of the following two cases hold.

1. *$\mathcal{A} = C_0(X, \mathbb{R})$ or*
2. *there exists a unique point $x_0 \in X$ such that $\mathcal{A} = \mathcal{I}_{x_0}$.*

Moreover, case 1. holds iff $\mathcal{A}_x = \mathbb{R}$ for all $x \in X$ and case 2. holds iff there exists a point $x_0 \in X$ such that $\mathcal{A}_{x_0} = \{0\}$.

Proof. If there exists x_0 such that $\mathcal{A}_{x_0} = \{0\}$ (x_0 is unique since $\mathcal{A}$ separates points) then $\mathcal{A} \subset \mathcal{I}_{x_0}$. If such an x_0 exists let $\mathcal{C} = \mathcal{I}_{x_0}$ and if $\mathcal{A}_x = \mathbb{R}$ for all x, set $\mathcal{C} = C_0(X, \mathbb{R})$. Let $f \in \mathcal{C}$ be given. By Lemma 37.30, for all $x, y \in X$ such that $x \neq y$, there exists $g_{xy} \in \mathcal{A}$ such that $f = g_{xy}$ on $\{x, y\}$.[3] When X is compact the basic idea of the proof is contained in the following identity,

$$f(z) = \inf_{x \in X} \sup_{y \in X} g_{xy}(z) \text{ for all } z \in X. \tag{37.7}$$

To prove this identity, let $g_x := \sup_{y \in X} g_{xy}$ and notice that $g_x \geq f$ since $g_{xy}(y) = f(y)$ for all $y \in X$. Moreover, $g_x(x) = f(x)$ for all $x \in X$ since $g_{xy}(x) = f(x)$ for all x. Therefore,

$$\inf_{x \in X} \sup_{y \in X} g_{xy} = \inf_{x \in X} g_x = f.$$

[3] If $\mathcal{A}_{x_0} = \{0\}$ and $x = x_0$ or $y = x_0$, then g_{xy} exists merely by the fact that $\mathcal{A}$ separates points.

The rest of the proof is devoted to replacing the inf and the sup above by min and max over finite sets at the expense of Eq. (37.7) becoming only an approximate identity. We also have to modify Eq. (37.7) slightly to take care of the non-compact case.

Claim. Given $\varepsilon > 0$ and $x \in X$ there exists $g_x \in \mathcal{A}$ such that $g_x(x) = f(x)$ and $f < g_x + \varepsilon$ on X.

To prove this, let V_y be an open neighborhood of y such that $|f - g_{xy}| < \varepsilon$ on V_y; in particular $f < \varepsilon + g_{xy}$ on V_y. Also let $g_{x,\infty}$ be any fixed element in $\mathcal{A}$ such that $g_{x,\infty}(x) = f(x)$ and let

$$K = \left\{|f| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|g_{x,\infty}| \geq \frac{\varepsilon}{2}\right\}. \tag{37.8}$$

Since K is compact, there exists $\Lambda \subset\subset K$ such that $K \subset \bigcup_{y\in\Lambda} V_y$. Define

$$g_x(z) = \max\{g_{xy} : y \in \Lambda \cup \{\infty\}\}.$$

Since

$$f < \varepsilon + g_{xy} < \varepsilon + g_x \text{ on } V_y,$$

for any $y \in \Lambda$, and

$$f < \frac{\varepsilon}{2} < \varepsilon + g_{x,\infty} \leq g_x + \varepsilon \text{ on } K^c,$$

$f < \varepsilon + g_x$ on X and by construction $f(x) = g_x(x)$, see Figure 37.5. This completes the proof of the claim.

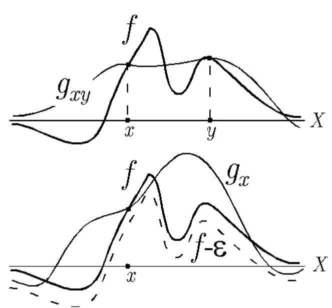


Fig. 37.5. Constructing the "dominating approximates," g_x for each $x \in X$.

To complete the proof of the theorem, let g_∞ be a fixed element of $\mathcal{A}$ such that $f < g_\infty + \varepsilon$ on X; for example let $g_\infty = g_{x_0} \in \mathcal{A}$ for some fixed $x_0 \in X$. For each $x \in X$, let U_x be a neighborhood of x such that $|f - g_x| < \varepsilon$ on U_x. Choose

$$\Gamma \subset\subset F := \left\{|f| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|g_\infty| \geq \frac{\varepsilon}{2}\right\}$$

such that $F \subset \bigcup_{x\in\Gamma} U_x$ (Γ exists since F is compact) and define

$$g = \min\{g_x : x \in \Gamma \cup \{\infty\}\} \in \mathcal{A}.$$

Then, for $x \in F$, $g_x < f + \varepsilon$ on U_x and hence $g < f + \varepsilon$ on $\bigcup_{x\in\Gamma} U_x \supset F$. Likewise,

$$g \leq g_\infty < \varepsilon/2 < f + \varepsilon \text{ on } F^c.$$

Therefore we have now shown,

$$f < g + \varepsilon \text{ and } g < f + \varepsilon \text{ on } X,$$

i.e. $|f - g| < \varepsilon$ on X. Since $\varepsilon > 0$ is arbitrary it follows that $f \in \bar{\mathcal{A}} = \mathcal{A}$ and so $\mathcal{A} = \mathcal{C}$. ■

Corollary 37.32 (Complex Stone-Weierstrass Theorem). *Let X be a locally compact Hausdorff space. Suppose $\mathcal{A} \subset C_0(X, \mathbb{C})$ is closed in the uniform topology, separates points, and is closed under complex conjugation. Then either $\mathcal{A} = C_0(X, \mathbb{C})$ or*

$$\mathcal{A} = \mathcal{I}_{x_0}^{\mathbb{C}} := \{f \in C_0(X, \mathbb{C}) : f(x_0) = 0\}$$

for some $x_0 \in X$.

Proof. Since

$$\operatorname{Re} f = \frac{f + \bar{f}}{2} \text{ and } \operatorname{Im} f = \frac{f - \bar{f}}{2i},$$

$\operatorname{Re} f$ and $\operatorname{Im} f$ are both in $\mathcal{A}$. Therefore

$$\mathcal{A}_{\mathbb{R}} = \{\operatorname{Re} f, \operatorname{Im} f : f \in \mathcal{A}\}$$

is a real sub-algebra of $C_0(X, \mathbb{R})$ which separates points. Therefore either $\mathcal{A}_{\mathbb{R}} = C_0(X, \mathbb{R})$ or $\mathcal{A}_{\mathbb{R}} = \mathcal{I}_{x_0} \cap C_0(X, \mathbb{R})$ for some x_0 and hence $\mathcal{A} = C_0(X, \mathbb{C})$ or $\mathcal{I}_{x_0}^{\mathbb{C}}$ respectively. ■

As an easy application, Theorem 37.31 and Corollary 37.32 imply Theorem 32.39 and Corollary 32.41 respectively. Here are a few more applications.

Example 37.33. Let $f \in C([a, b])$ be a positive function which is injective. Then functions of the form $\sum_{k=1}^N a_k f^k$ with $a_k \in \mathbb{C}$ and $N \in \mathbb{N}$ are dense in $C([a, b])$. For example if $a = 1$ and $b = 2$, then one may take $f(x) = x^\alpha$ for any $\alpha \neq 0$, or $f(x) = e^x$, etc.

Exercise 37.5. Let (X,d) be a separable compact metric space. Show that $C(X)$ is also separable. **Hint:** Let $E \subset X$ be a countable dense set and then consider the algebra, $\mathcal{A} \subset C(X)$, generated by $\{d(x,\cdot)\}_{x\in E}$.

Example 37.34. Let $X = [0,\infty)$, $\lambda > 0$ be fixed, $\mathcal{A}$ be the real algebra generated by $t \to e^{-\lambda t}$. So the general element $f \in \mathcal{A}$ is of the form $f(t) = p(e^{-\lambda t})$, where $p(x)$ is a polynomial function in x with real coefficients. Since $\mathcal{A} \subset C_0(X,\mathbb{R})$ separates points and $e^{-\lambda t} \in \mathcal{A}$ is pointwise positive, $\bar{\mathcal{A}} = C_0(X,\mathbb{R})$.

As an application of Example 37.34, suppose that $g \in C_c(X,\mathbb{R})$ satisfies,

$$\int_0^\infty g(t)\, e^{-\lambda t} dt = 0 \text{ for all } \lambda > 0. \tag{37.9}$$

(Note well that the integral in Eq. (37.9) is really over a finite interval since g is compactly supported.) Equation (37.9) along with linearity of the Riemann integral implies

$$\int_0^\infty g(t)\, f(t)\, dt = 0 \text{ for all } f \in \mathcal{A}.$$

We may now choose $f_n \in \mathcal{A}$ such that $f_n \to g$ uniformly and therefore, using the continuity of the Riemann integral under uniform convergence (see Proposition 32.5),

$$0 = \lim_{n\to\infty} \int_0^\infty g(t)\, f_n(t)\, dt = \int_0^\infty g^2(t)\, dt.$$

From this last equation it is easily deduced, using the continuity of g, that $g \equiv 0$. See Theorem 31.12 below, where this is done in greater generality.

Definition 37.35 (Laplace Transform). *Suppose that $f : \mathbb{R}_+ \to \mathbb{C}$ is a measurable function such that*

$$\int_0^\infty |f(t)|\, e^{-at} dt < \infty \text{ for some } a \in (0,\infty).$$

Then for $\lambda \in (a,\infty)$ we define $(\mathcal{L}f)(\lambda) := \int_0^\infty f(t)\, e^{-\lambda t} dt$ and refer to $\mathcal{L}f$ as the Laplace transform of f.

Theorem 37.36 (Injectivity of the Laplace Transform). *Continuing the notation in Definition 37.35 we have;*

1. *$\mathcal{L}f \in C^\infty((a,\infty))$ and for all $n \in \mathbb{N}_0$ we have*
$$\left(-\frac{d}{d\lambda}\right)^n (\mathcal{L}f)(\lambda) = \mathcal{L}(t \to t^n f(t))(\lambda) = \int_0^\infty t^n f(t)\, e^{-\lambda t} dt.$$
2. *If $\mathcal{L}f \equiv 0$, then $f(t) = 0$ for m -a.e. t.*

Proof. The first assertion is an easy consequence of Corollary 10.30 on differentiating past integrals. For the second, let $\lambda > 0$ be fixed, then by assumption,

$$\int_0^\infty f(t)\, e^{-at} p\left(e^{-\lambda t}\right) dt = 0 \tag{37.10}$$

for all polynomials $p(\cdot)$ without constant term. It now follows by the Stone-Wierstrass Theorem (Corollary 37.32) that

$$\mathcal{A} = \left\{p\left(e^{-\lambda t}\right) : p \text{ is a polynomial w/o constant term}\right\}$$

is dense in $C_0([0,\infty))$. Since $f(t)\, e^{-at} \in L^1([0,\infty), dt)$, we may combine the previous assertion with the dominated convergence theorem and Eq. (37.10) in order to learn,

$$\int_0^\infty f(t)\, e^{-at} g(t)\, dt = 0 \text{ for all } g \in C_0([0,\infty)).$$

An application of the multiplicative system theorem now shows this equation holds for all bounded measurable g. Taking $g(t) = \overline{\operatorname{sgn}(f(t)\, e^{-at})}$ then shows,

$$\int_0^\infty \left|f(t)\, e^{-at}\right| dt = 0 \implies f(t)\, e^{-at} = 0 \text{ for } m \text{ - a.e. } t.$$

As $e^{-at} > 0$ for all the proof is complete. ■

Exercise 37.6. Let $g \in L^1([0,\infty), m)$. Show that $g = 0$ a.e. if

$$\int_0^\infty g(x) \left(\frac{x}{1+x}\right)^n dx = 0 \text{ for } n = 1, 2, \dots$$

37.5 *More on Separation Axioms: Normal Spaces

(This section may safely be omitted on the first reading.)

Definition 37.37 ($T_0 - T_2$ Separation Axioms). *Let (X,τ) be a topological space. The topology τ is said to be:*

1. *T_0 if for $x \neq y$ in X there exists $V \in \tau$ such that $x \in V$ and $y \notin V$ or V such that $y \in V$ but $x \notin V$.*
2. *T_1 if for every $x, y \in X$ with $x \neq y$ there exists $V \in \tau$ such that $x \in V$ and $y \notin V$. Equivalently, τ is T_1 iff all one point subsets of X are closed.*[4]

[4] If one point subsets are closed and $x \neq y$ in X then $V := \{x\}^c$ is an open set containing y but not x. Conversely if τ is T_1 and $x \in X$ there exists $V_y \in \tau$ such that $y \in V_y$ and $x \notin V_y$ for all $y \neq x$. Therefore, $\{x\}^c = \cup_{y\neq x} V_y \in \tau$.

3. T_2 *if it is Hausdorff.*

Note T_2 implies T_1 which implies T_0. The topology in Example 37.1 is T_0 but not T_1. If X is a finite set and τ is a T_1 – topology on X then $\tau = 2^X$. To prove this let $x \in X$ be fixed. Then for every $y \neq x$ in X there exists $V_y \in \tau$ such that $x \in V_y$ while $y \notin V_y$. Thus $\{x\} = \cap_{y \neq x} V_y \in \tau$ showing τ contains all one point subsets of X and therefore all subsets of X. So we have to look to infinite sets for an example of T_1 topology which is not T_2.

Example 37.38. Let X be any infinite set and let $\tau = \{A \subset X : \#(A^c) < \infty\} \cup \{\emptyset\}$ – the so called **cofinite** topology. This topology is T_1 because if $x \neq y$ in X, then $V = \{x\}^c \in \tau$ with $x \notin V$ while $y \in V$. This topology however is not T_2. Indeed if $U, V \in \tau$ are open sets such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$ then $U \subset V^c$. But this implies $\#(U) < \infty$ which is impossible unless $U = \emptyset$ which is impossible since $x \in U$.

The uniqueness of limits of sequences which occurs for Hausdorff topologies (see Remark 37.3) need not occur for T_1 – spaces. For example, let $X = \mathbb{N}$ and τ be the cofinite topology on X as in Example 37.38. Then $x_n = n$ is a sequence in X such that $x_n \to x$ as $n \to \infty$ **for all** $x \in \mathbb{N}$. For the most part we will avoid these pathologies in the future by only considering Hausdorff topologies.

Definition 37.39 (Normal Spaces: T_4 – Separation Axiom). *A topological space (X, τ) is said to be **normal** or T_4 if:*

1. *X is Hausdorff and*
2. *if for any two closed disjoint subsets $A, B \subset X$ there exists disjoint open sets $V, W \subset X$ such that $A \subset V$ and $B \subset W$.*

Example 37.40. By Lemma 13.21 and Corollary 37.21 it follows that metric spaces and topological spaces which are locally compact, σ – compact and Hausdorff (in particular compact Hausdorff spaces) are normal. Indeed, in each case if A, B are disjoint closed subsets of X, there exists $f \in C(X, [0,1])$ such that $f = 1$ on A and $f = 0$ on B. Now let $U = \left\{f > \frac{1}{2}\right\}$ and $V = \{f < \frac{1}{2}\}$.

Remark 37.41. A topological space, (X, τ), is normal iff for any $C \subset W \subset X$ with C being closed and W being open there exists an open set $U \subset_o X$ such that

$$C \subset U \subset \bar{U} \subset W.$$

To prove this first suppose X is normal. Since W^c is closed and $C \cap W^c = \emptyset$, there exists disjoint open sets U and V such that $C \subset U$ and $W^c \subset V$. Therefore $C \subset U \subset V^c \subset W$ and since V^c is closed, $C \subset U \subset \bar{U} \subset V^c \subset W$.

For the converse direction suppose A and B are disjoint closed subsets of X. Then $A \subset B^c$ and B^c is open, and so by assumption there exists $U \subset_o X$ such that $A \subset U \subset \bar{U} \subset B^c$ and by the same token there exists $W \subset_o X$ such that $\bar{U} \subset W \subset \bar{W} \subset B^c$. Taking complements of the last expression implies

$$B \subset \bar{W}^c \subset W^c \subset \bar{U}^c.$$

Let $V = \bar{W}^c$. Then $A \subset U \subset_o X$, $B \subset V \subset_o X$ and $U \cap V \subset U \cap W^c = \emptyset$.

Theorem 37.42 (Urysohn's Lemma for Normal Spaces). *Let X be a normal space. Assume A, B are disjoint closed subsets of X. Then there exists $f \in C(X, [0,1])$ such that $f = 0$ on A and $f = 1$ on B.*

Proof. To make the notation match Lemma 37.8, let $U = A^c$ and $K = B$. Then $K \subset U$ and it suffices to produce a function $f \in C(X, [0,1])$ such that $f = 1$ on K and $\operatorname{supp}(f) \subset U$. The proof is now identical to that for Lemma 37.8 except we now use Remark 37.41 in place of Proposition 37.7. ∎

Theorem 37.43 (Tietze Extension Theorem). *Let (X, τ) be a normal space, D be a closed subset of X, $-\infty < a < b < \infty$ and $f \in C(D, [a,b])$. Then there exists $F \in C(X, [a,b])$ such that $F|_D = f$.*

Proof. The proof is identical to that of Theorem 14.5 except we now use Theorem 37.42 in place of Lemma 13.21. ∎

Corollary 37.44. *Suppose that X is a normal topological space, $D \subset X$ is closed, $F \in C(D, \mathbb{R})$. Then there exists $F \in C(X)$ such that $F|_D = f$.*

Proof. Let $g = \arctan(f) \in C(D, (-\frac{\pi}{2}, \frac{\pi}{2}))$. Then by the Tietze extension theorem, there exists $G \in C(X, [-\frac{\pi}{2}, \frac{\pi}{2}])$ such that $G|_D = g$. Let $B := G^{-1}(\{-\frac{\pi}{2}, \frac{\pi}{2}\}) \sqsubset X$, then $B \cap D = \emptyset$. By Urysohn's lemma (Theorem 37.42) there exists $h \in C(X, [0,1])$ such that $h \equiv 1$ on D and $h = 0$ on B and in particular $hG \in C(D, (-\frac{\pi}{2}, \frac{\pi}{2}))$ and $(hG)|_D = g$. The function $F := \tan(hG) \in C(X)$ is an extension of f. ∎

Theorem 37.45 (Urysohn Metrization Theorem for Normal Spaces). *Every second countable normal space, (X, τ), is metrizable, i.e. there is a metric ρ on X such that $\tau = \tau_\rho$. Moreover, ρ may be chosen so that X is isometric to a subset $Q_0 \subset Q$ (Q is as in Notation 37.11) equipped with the metric d in Eq. (37.2). In this metric X is totally bounded and hence the completion of X (which is isometric to $\bar{Q}_0 \subset Q$) is compact.*

Proof. (The proof here will be very similar to the proof of Theorem 37.13.) Let $\mathcal{B}$ be a countable base for τ and set

$$\Gamma := \{(U, V) \in \mathcal{B} \times \mathcal{B} \mid \bar{U} \subset V\}.$$

To each $O \in \tau$ and $x \in O$ there exist $(U, V) \in \Gamma$ such that $x \in U \subset V \subset O$. Indeed, since $\mathcal{B}$ is a base for τ, there exists $V \in \mathcal{B}$ such that $x \in V \subset O$. Because

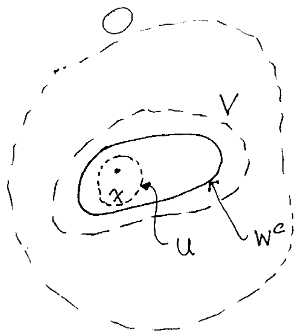


Fig. 37.6. Constructing $(U,V) \in \Gamma$.

$\{x\} \cap V^c = \emptyset$, there exists disjoint open sets $\widetilde{U}$ and W such that $x \in \widetilde{U}$, $V^c \subset W$ and $\widetilde{U} \cap W = \emptyset$. Choose $U \in \mathcal{B}$ such that $x \in U \subset \widetilde{U}$. Since $U \subset \widetilde{U} \subset W^c$, $\overline{U} \subset W^c \subset V$ and hence $(U,V) \in \Gamma$. See Figure 37.6 below. In particular this shows that

$$\mathcal{B}_0 := \{U \in \mathcal{B} : (U,V) \in \Gamma \text{ for some } V \in \mathcal{B}\}$$

is still a base for τ.

If Γ is a finite set, the previous comment shows that τ only has a finite number of elements as well. Since (X,τ) is Hausdorff, it follows that X is a finite set. Letting $\{x_n\}_{n=1}^N$ be an enumeration of X, define $T : X \to Q$ by $T(x_n) = e_n$ for $n = 1, 2, \ldots, N$ where $e_n = (0, 0, \ldots, 0, 1, 0, \ldots)$, with the 1 occurring in the n^{th} spot. Then $\rho(x,y) := d(T(x), T(y))$ for $x, y \in X$ is the desired metric.

So we may now assume that Γ is an infinite set and let $\{(U_n, V_n)\}_{n=1}^\infty$ be an enumeration of Γ. By Urysohn's Lemma for normal spaces (Theorem 37.42) there exists $f_{U,V} \in C(X,[0,1])$ such that $f_{U,V} = 0$ on $\bar{U}$ and $f_{U,V} = 1$ on V^c. Let $\mathcal{F} := \{f_{U,V} \mid (U,V) \in \Gamma\}$ and set $f_n := f_{U_n,V_n}$ – an enumeration of $\mathcal{F}$. The proof that

$$\rho(x,y) := \sum_{n=1}^\infty \frac{1}{2^n} |f_n(x) - f_n(y)|$$

is the desired metric on X now follows exactly as the corresponding argument in the proof of Theorem 37.13. ■

37.6 Exercises

Exercise 37.7. Prove Theorem 37.9. **Hints:**

1. By Proposition 37.7, there exists a precompact open set V such that $K \subset V \subset \bar{V} \subset U$. Now suppose that $f : K \to [0,\alpha]$ is continuous with $\alpha \in (0,1]$ and let $A := f^{-1}([0, \frac{1}{3}\alpha])$ and $B := f^{-1}([\frac{2}{3}\alpha, 1])$. Appeal to Lemma 37.8 to find a function $g \in C(X, [0, \alpha/3])$ such that $g = \alpha/3$ on B and $\text{supp}(g) \subset V \setminus A$.
2. Now follow the argument in the proof of Theorem 14.5 to construct $F \in C(X,[a,b])$ such that $F|_K = f$.
3. For $c \in [a,b]$, choose $\varphi \prec U$ such that $\varphi = 1$ on K and replace F by $F_c := \varphi F + (1-\varphi)c$.

Exercise 37.8 (Sterographic Projection). Let $X = \mathbb{R}^n$, $X^* := X \cup \{\infty\}$ be the one point compactification of X, $S^n := \{y \in \mathbb{R}^{n+1} : |y| = 1\}$ be the unit sphere in $\mathbb{R}^{n+1}$ and $N = (0, \ldots, 0, 1) \in \mathbb{R}^{n+1}$. Define $f : S^n \to X^*$ by $f(N) = \infty$, and for $y \in S^n \setminus \{N\}$ let $f(y) = b \in \mathbb{R}^n$ be the unique point such that $(b, 0)$ is on the line containing N and y, see Figure 37.7 below. Find a formula for f and show $f : S^n \to X^*$ is a homeomorphism. (So the one point compactification of $\mathbb{R}^n$ is homeomorphic to the n sphere.)

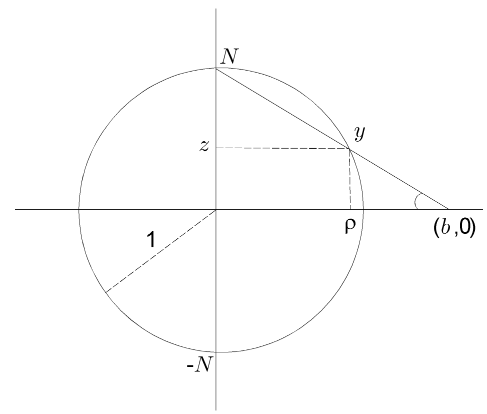


Fig. 37.7. Sterographic projection and the one point compactification of $\mathbb{R}^n$.

Exercise 37.9. Let (X,τ) be a locally compact Hausdorff space. Show (X,τ) is separable iff (X^*, τ^*) is separable.

Exercise 37.10. Show by example that there exists a locally compact metric space (X,d) such that the one point compactification, $(X^* := X \cup \{\infty\}, \tau^*)$, is **not** metrizable. **Hint**: use exercise 37.9.

Exercise 37.11. Suppose (X, d) is a locally compact and σ – compact metric space. Show the one point compactification, $(X^* := X \cup \{\infty\}, \tau^*)$, is metrizable.

Exercise 37.12. In this problem, suppose Theorem 37.31 has only been proved when X is compact. Show that it is possible to prove Theorem 37.31 by using Proposition 37.24 to reduce the non-compact case to the compact case.

Hints:

1. If $\mathcal{A}_x = \mathbb{R}$ for all $x \in X$ let $X^* = X \cup \{\infty\}$ be the one point compactification of X.
2. If $\mathcal{A}_{x_0} = \{0\}$ for some $x_0 \in X$, let $Y := X \setminus \{x_0\}$ and $Y^* = Y \cup \{\infty\}$ be the one point compactification of Y.
3. For $f \in \mathcal{A}$ define $f(\infty) = 0$. In this way $\mathcal{A}$ may be considered to be a sub-algebra of $C(X^*, \mathbb{R})$ in case 1. or a sub-algebra of $C(Y^*, \mathbb{R})$ in case 2.

Exercise 37.13. Given a continuous function $f : \mathbb{R} \to \mathbb{C}$ which is 2π - periodic and $\varepsilon > 0$. Show there exists a trigonometric polynomial, $p(\theta) = \sum_{n=-N}^{n} \alpha_n e^{in\theta}$, such that $|f(\theta) - P(\theta)| < \varepsilon$ for all $\theta \in \mathbb{R}$. **Hint:** show that there exists a unique function $F \in C(S^1)$ such that $f(\theta) = F(e^{i\theta})$ for all $\theta \in \mathbb{R}$.

Remark 37.46. Exercise 37.13 generalizes to 2π – periodic functions on $\mathbb{R}^d$, i.e. functions such that $f(\theta + 2\pi e_i) = f(\theta)$ for all $i = 1, 2, \ldots, d$ where $\{e_i\}_{i=1}^d$ is the standard basis for $\mathbb{R}^d$. A trigonometric polynomial $p(\theta)$ is a function of $\theta \in \mathbb{R}^d$ of the form

$$p(\theta) = \sum_{n \in \Gamma} \alpha_n e^{in \cdot \theta}$$

where Γ is a finite subset of $\mathbb{Z}^d$. The assertion is again that these trigonometric polynomials are dense in the 2π – periodic functions relative to the supremum norm.

Examples of Measures

In this chapter we are going to state a couple of construction theorems for measures. The proofs of these theorems will be deferred until the next chapter, also see Chapter **??**. Our goal in this chapter is to apply these construction theorems to produce a fairly broad class of examples of measures.

38.1 The Riesz-Markov Theorem

Now suppose that X is a locally compact Hausdorff space and $\mathcal{B} = \mathcal{B}_X$ is the Borel σ − algebra on X. Open subsets of $\mathbb{R}^d$ and locally compact separable metric spaces are examples of such spaces, see Section 36.1.

Definition 38.1. *A linear functional I on $C_c(X)$ is **positive** if $I(f) \geq 0$ for all $f \in C_c(X, [0, \infty))$.*

Proposition 38.2. *If I is a positive linear functional on $C_c(X)$ and K is a compact subset of X, then there exists $C_K < \infty$ such that $|I(f)| \leq C_K \|f\|_\infty$ for all $f \in C_c(X)$ with $\operatorname{supp}(f) \subset K$.*

Proof. By Urysohn's Lemma 37.8, there exists $\varphi \in C_c(X, [0, 1])$ such that $\varphi = 1$ on K. Then for all $f \in C_c(X, \mathbb{R})$ such that $\operatorname{supp}(f) \subset K$, $|f| \leq \|f\|_\infty \varphi$ or equivalently $\|f\|_\infty \varphi \pm f \geq 0$. Hence $\|f\|_\infty I(\varphi) \pm I(f) \geq 0$ or equivalently which is to say $|I(f)| \leq \|f\|_\infty I(\varphi)$. Letting $C_K := I(\varphi)$, we have shown that $|I(f)| \leq C_K \|f\|_\infty$ for all $f \in C_c(X, \mathbb{R})$ with $\operatorname{supp}(f) \subset K$. For general $f \in C_c(X, \mathbb{C})$ with $\operatorname{supp}(f) \subset K$, choose $|\alpha| = 1$ such that $\alpha I(f) \geq 0$. Then

$$|I(f)| = \alpha I(f) = I(\alpha\ f) = I(\operatorname{Re}(\alpha f)) \leq C_K \| \operatorname{Re}(\alpha f) \|_\infty \leq C_K \|f\|_\infty.$$

■

Example 38.3. If μ is a K-finite measure on X, then

$$I_\mu(f) = \int_X f d\mu \ \forall f \in C_c(X)$$

defines a positive linear functional on $C_c(X)$. In the future, we will often simply write $\mu(f)$ for $I_\mu(f)$.

The Riesz-Markov Theorem 38.9 below asserts that every positive linear functional on $C_c(X)$ comes from a K-finite measure μ.

Example 38.4. Let $X = \mathbb{R}$ and $\tau = \tau_d = 2^X$ be the discrete topology on X. Now let $\mu(A) = 0$ if A is countable and $\mu(A) = \infty$ otherwise. Since $K \subset X$ is compact iff $\#(K) < \infty$, μ is a K-finite measure on X and

$$I_\mu(f) = \int_X f d\mu = 0 \text{ for all } f \in C_c(X).$$

This shows that the correspondence $\mu \to I_\mu$ from K-finite measures to positive linear functionals on $C_c(X)$ is not injective without further restriction.

Definition 38.5. *Suppose that μ is a Borel measure on X and $B \in \mathcal{B}_X$. We say μ is **inner regular on** B if*

$$\mu(B) = \sup\{\mu(K) : K \sqsubset\sqsubset B\} \tag{38.1}$$

*and μ **is outer regular on** B if*

$$\mu(B) = \inf\{\mu(U) : B \subset U \subset_o X\}. \tag{38.2}$$

*The measure μ is said to be a **regular Borel measure** on X, if it is both inner and outer regular on all Borel measurable subsets of X.*

Definition 38.6. *A measure $\mu : \mathcal{B}_X \to [0, \infty]$ is a **Radon measure** on X if μ is a K-finite measure which is inner regular on all open subsets of X and outer regular on all Borel subsets of X.*

The measure in Example 38.4 is an example of a K-finite measure on X which is not a Radon measure on X. BRUCE: Add exercise stating the sum of two radon measures is still a radon measure. It is not true for countable sums since this does not even preserve the K - finite condition.

Example 38.7. If the topology on a set, X, is the discrete topology, then a measure μ on $\mathcal{B}_X$ is a Radon measure iff μ is of the form

$$\mu = \sum_{x \in X} \mu_x \delta_x \tag{38.3}$$

where $\mu_x \in [0,\infty)$ for all $x \in X$. To verify this first notice that $\mathcal{B}_X = \tau_X = 2^X$ and hence every measure on $\mathcal{B}_X$ is necessarily outer regular on all subsets of X. The measure μ is K-finite iff $\mu_x := \mu(\{x\}) < \infty$ for all $x \in X$. If μ is a Radon measure, then for $A \subset X$ we have, by inner regularity,

$$\mu(A) = \sup\{\mu(\Lambda) : \Lambda \subset\subset A\} = \sup\left\{\sum_{x\in\Lambda} \mu_x : \Lambda \subset\subset A\right\} = \sum_{x\in A} \mu_x.$$

On the other hand if μ is given by Eq. (38.3) and $A \subset X$, then

$$\mu(A) = \sum_{x\in A} \mu_x = \sup\left\{\mu(\Lambda) = \sum_{x\in\Lambda} \mu_x : \Lambda \subset\subset A\right\}$$

showing μ is inner regular on all (open) subsets of X.

Recall from Definition 36.8 that if U is an open subset of X, we write $f \prec U$ to mean that $f \in C_c(X,[0,1])$ with $\operatorname{supp}(f) := \overline{\{f \neq 0\}} \subset U$.

Notation 38.8 *Given a positive linear functional, I, on $C_c(X)$ define $\mu = \mu_I$ on $\mathcal{B}_X$ by*

$$\mu(U) = \sup\{I(f) : f \prec U\} \tag{38.4}$$

for all $U \subset_o X$ and then define

$$\mu(B) = \inf\{\mu(U) : B \subset U \text{ and } U \text{ is open}\}. \tag{38.5}$$

Theorem 38.9 (Riesz-Markov Theorem). *The map $\mu \to I_\mu$ taking Radon measures on X to positive linear functionals on $C_c(X)$ is bijective. Moreover if I is a positive linear functional on $C_c(X)$, the function $\mu := \mu_I$ defined in Notation 38.8 has the following properties.*

1. *μ is a Radon measure on X and the map $I \to \mu_I$ is the inverse to the map $\mu \to I_\mu$.*
2. *For all compact subsets $K \subset X$,*

$$\mu(K) = \inf\{I(f) : 1_K \le f \prec X\}. \tag{38.6}$$

3. *If $\|I_\mu\|$ denotes the dual norm of $I = I_\mu$ on $C_c(X,\mathbb{R})^*$, then $\|I\| = \mu(X)$. In particular, the linear functional, I_μ, is bounded iff $\mu(X) < \infty$.*

Proof. (Also see Theorem **??** and related material about the Daniel integral.) The proof of the surjectivity of the map $\mu \to I_\mu$ and the assertion in item 1. is the content of Theorem 38.11 below.

Injectivity of $\mu \to I_\mu$. Suppose that μ is a is a Radon measure on X. To each open subset $U \subset X$ let

$$\mu_0(U) := \sup\{I_\mu(f) : f \prec U\}. \tag{38.7}$$

It is evident that $\mu_0(U) \le \mu(U)$ because $f \prec U$ implies $f \le 1_U$. Given a compact subset $K \subset U$, Urysohn's Lemma 37.8 implies there exists $f \prec U$ such that $f = 1$ on K. Therefore,

$$\mu(K) \le \int_X f d\mu \le \mu_0(U) \le \mu(U) \tag{38.8}$$

By assumption μ is inner regular on open sets, and therefore taking the supremum of Eq. (38.8) over compact subsets, K, of U shows

$$\mu(U) = \mu_0(U) = \sup\{I_\mu(f) : f \prec U\}. \tag{38.9}$$

If μ and ν are two Radon measures such that $I_\mu = I_\nu$. Then by Eq. (38.9) it follows that $\mu = \nu$ on all open sets. Then by outer regularity, $\mu = \nu$ on $\mathcal{B}_X$ and this shows the map $\mu \to I_\mu$ is injective.

Item 2. Let $K \subset X$ be a compact set, then by monotonicity of the integral,

$$\mu(K) \le \inf\{I_\mu(f) : f \in C_c(X) \text{ with } f \ge 1_K\}. \tag{38.10}$$

To prove the reverse inequality, choose, by outer regularity, $U \subset_o X$ such that $K \subset U$ and $\mu(U \setminus K) < \varepsilon$. By Urysohn's Lemma 37.8 there exists $f \prec U$ such that $f = 1$ on K and hence,

$$I_\mu(f) = \int_X f\, d\mu = \mu(K) + \int_{U\setminus K} f\, d\mu \le \mu(K) + \mu(U\setminus K) < \mu(K) + \varepsilon.$$

Consequently,

$$\inf\{I_\mu(f) : f \in C_c(X) \text{ with } f \ge 1_K\} < \mu(K) + \varepsilon$$

and because $\varepsilon > 0$ was arbitrary, the reverse inequality in Eq. (38.10) holds and Eq. (38.6) is verified.

Item 3. If $f \in C_c(X)$, then

$$|I_\mu(f)| \le \int_X |f|\, d\mu = \int_{\operatorname{supp}(f)} |f|\, d\mu \le \|f\|_\infty \mu(\operatorname{supp}(f)) \le \|f\|_\infty \mu(X) \tag{38.11}$$

and thus $\|I_\mu\| \le \mu(X)$. For the reverse inequality let K be a compact subset of X and use Urysohn's Lemma 37.8 again to find a function $f \prec X$ such that $f = 1$ on K. By Eq. (38.8) we have

$$\mu(K) \le \int_X f d\mu = I_\mu(f) \le \|I_\mu\| \|f\|_\infty = \|I_\mu\|,$$

which by the inner regularity of μ on open sets implies

$$\mu(X) = \sup\{\mu(K) : K \sqsubset\sqsubset X\} \leq \|I_\mu\| .$$

■

Example 38.10 (Discrete Version of Theorem 38.9). Suppose X is a set, $\tau = 2^X$ is the discrete topology on X and for $x \in X$, let $e_x \in C_c(X)$ be defined by $e_x(y) = 1_{\{x\}}(y)$. Let I be positive linear functional on $C_c(X)$ and define a Radon measure, μ, on X by

$$\mu(A) := \sum_{x \in A} I(e_x) \text{ for all } A \subset X.$$

Then for $f \in C_c(X)$ (so f is a complex valued function on X supported on a finite set),

$$\int_X f d\mu = \sum_{x \in X} f(x) I(e_x) = I\left(\sum_{x \in X} f(x) e_x\right) = I(f),$$

so that $I = I_\mu$. It is easy to see in this example that μ defined above is the unique regular radon measure on X such that $I = I_\mu$ while example Example 38.4 shows the uniqueness is lost if the regularity assumption is dropped.

38.2 Proof of the Riesz-Markov Theorem 38.9

This section is devoted to completing the proof of the Riesz-Markov Theorem 38.9.

Theorem 38.11. *Suppose (X, τ) is a locally compact Hausdorff space, I is a positive linear functional on $C_c(X)$ and $\mu := \mu_I$ be as in Notation 38.8. Then μ is a Radon measure on X such that $I = I_\mu$, i.e.*

$$I(f) = \int_X f d\mu \text{ for all } f \in C_c(X) .$$

Proof. Let $\mu : \tau \to [0, \infty]$ be as in Eq. (38.4) and $\mu^* : 2^X \to [0, \infty]$ be the associate outer measure as in Proposition **??**. As we have seen in Lemma **??**, μ is sub-additive on τ and

$$\mu^*(E) = \inf\{\mu(U) : E \subset U \subset_o X\} .$$

By Theorem **??**, $\mathcal{M} := \mathcal{M}(\mu^*)$ is a σ-algebra and $\mu^*|_{\mathcal{M}}$ is a measure on $\mathcal{M}$.

To show $\mathcal{B}_X \subset \mathcal{M}$ it suffices to show $U \in \mathcal{M}$ for all $U \in \tau$, i.e. we must show;

$$\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \setminus U) \tag{38.12}$$

for every $E \subset X$ such that $\mu^*(E) < \infty$. First suppose E is open, in which case $E \cap U$ is open as well. Let $f \prec E \cap U$ and $K := \mathrm{supp}(f)$. Then $E \setminus U \subset E \setminus K$ and if $g \prec E \setminus K \in \tau$ then $f + g \prec E$ (see Figure 38.1) and hence

$$\mu^*(E) \geq I(f + g) = I(f) + I(g) .$$

Taking the supremum of this inequality over $g \prec E \setminus K$ shows

$$\mu^*(E) \geq I(f) + \mu^*(E \setminus K) \geq I(f) + \mu^*(E \setminus U) .$$

Taking the supremum of this inequality over $f \prec U$ shows Eq. (38.12) is valid for $E \in \tau$.

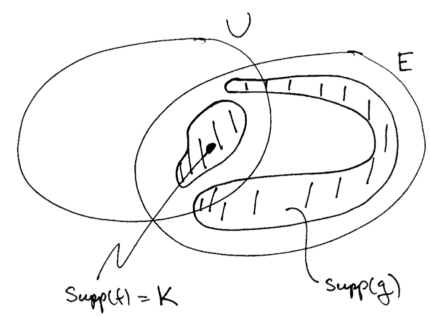


Fig. 38.1. Constructing a function g which approximates $1_{E \setminus U}$.

For general $E \subset X$, let $V \in \tau$ with $E \subset V$, then

$$\mu^*(V) \geq \mu^*(V \cap U) + \mu^*(V \setminus U) \geq \mu^*(E \cap U) + \mu^*(E \setminus U)$$

and taking the infimum of this inequality over such V shows Eq. (38.12) is valid for general $E \subset X$. Thus $U \in \mathcal{M}$ for all $U \in \tau$ and therefore $\mathcal{B}_X \subset \mathcal{M}$.

Up to this point it has been shown that $\mu = \mu^*|_{\mathcal{B}_X}$ is a measure which, by very construction, is outer regular. We now verify that μ satisfies Eq. (38.6), namely that $\mu(K) = \nu(K)$ for all compact sets $K \subset X$ where

$$\nu(K) := \inf\{I(f) : f \in C_c(X, [0,1]) \ni f \geq 1_K\} .$$

To do this let $f \in C_c(X, [0,1])$ with $f \geq 1_K$ and $\varepsilon > 0$ be given. Let $U_\varepsilon := \{f > 1 - \varepsilon\} \in \tau$ and $g \prec U_\varepsilon$, then $g \leq (1 - \varepsilon)^{-1} f$ and hence

$I(g) \le (1-\varepsilon)^{-1} I(f)$. Taking the supremum of this inequality over all $g \prec U_\varepsilon$ then gives,

$$\mu(K) \le \mu(U_\varepsilon) \le (1-\varepsilon)^{-1} I(f).$$

Since $\varepsilon > 0$ was arbitrary, we learn $\mu(K) \le I(f)$ for all $1_K \le f \prec X$ and therefore, $\mu(K) \le \nu(K)$. Now suppose that $U \in \tau$ and $K \subset U$. By Urysohn's Lemma 37.8 (also see Lemma 36.9), there exists $f \prec U$ such that $f \ge 1_K$ and therefore

$$\mu(K) \le \nu(K) \le I(f) \le \mu(U).$$

By the outer regularity of μ, we have

$$\mu(K) \le \nu(K) \le \inf\{\mu(U) : K \subset U \subset_o X\} = \mu(K),$$

i.e.

$$\mu(K) = \nu(K) = \inf\{I(f) : f \in C_c(X,[0,1]) \ni f \ge 1_K\}. \tag{38.13}$$

This inequality clearly establishes that μ is K-finite and therefore $C_c(X,[0,\infty)) \subset L^1(\mu)$.

Next we will establish,

$$I(f) = I_\mu(f) := \int_X f d\mu \tag{38.14}$$

for all $f \in C_c(X)$. By the linearity, it suffices to verify Eq. (38.14) holds for $f \in C_c(X,[0,\infty))$. To do this we will use the "layer cake method" to slice f into thin pieces. Explicitly, fix an $N \in \mathbb{N}$ and for $n \in \mathbb{N}$ let

$$f_n := \min\left(\max\left(f - \frac{n-1}{N}, 0\right), \frac{1}{N}\right), \tag{38.15}$$

see Figure 38.2. It should be clear from Figure 38.2 that $f = \sum_{n=1}^\infty f_n$ with the sum actually being a finite sum since $f_n \equiv 0$ for all n sufficiently large. Let $K_0 := \text{supp}(f)$ and $K_n := \left\{f \ge \frac{n}{N}\right\}$. Then (again see Figure 38.2) for all $n \in \mathbb{N}$,

$$1_{K_n} \le N f_n \le 1_{K_{n-1}}$$

which upon integrating on μ gives

$$\mu(K_n) \le N I_\mu(f_n) \le \mu(K_{n-1}). \tag{38.16}$$

Moreover, if U is any open set containing K_{n-1}, then $N f_n \prec U$ and so by Eq. (38.13) and the definition of μ, we have

$$\mu(K_n) \le N I(f_n) \le \mu(U). \tag{38.17}$$

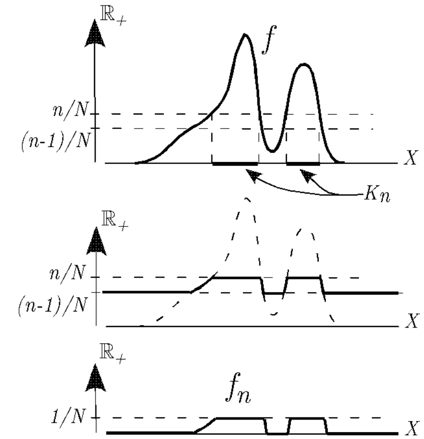


Fig. 38.2. This sequence of figures shows how the function f_n is constructed. The idea is to think of f as describing a "cake" set on a "table," X. We then slice the cake into slabs, each of which is placed back on the table. Each of these slabs is described by one of the functions, f_n, as in Eq. (38.15).

From the outer regularity of μ, it follows from Eq. (38.17) that

$$\mu(K_n) \le N I(f_n) \le \mu(K_{n-1}). \tag{38.18}$$

As a consequence of Eqs. (38.16) and (38.18), we have

$$N|I_\mu(f_n) - I(f_n)| \le \mu(K_{n-1}) - \mu(K_n) = \mu(K_{n-1} \setminus K_n).$$

Therefore

$$\begin{aligned} |I_\mu(f) - I(f)| &= \left|\sum_{n=1}^\infty I_\mu(f_n) - I(f_n)\right| \le \sum_{n=1}^\infty |I_\mu(f_n) - I(f_n)| \\ &\le \frac{1}{N}\sum_{n=1}^\infty \mu(K_{n-1} \setminus K_n) = \frac{1}{N}\mu(K_0) \to 0 \text{ as } N \to \infty \end{aligned}$$

which establishes Eq. (38.14).

It now only remains to show μ is inner regular on open sets to complete the proof. If $U \in \tau$ and $\mu(U) < \infty$, then for any $\varepsilon > 0$ there exists $f \prec U$ such that

$$\mu(U) \le I(f) + \varepsilon = \int_X f d\mu + \varepsilon \le \mu(\text{supp}(f)) + \varepsilon.$$

Hence if $K = \text{supp}(f)$, we have $K \subset U$ and $\mu(U \setminus K) < \varepsilon$ and this shows μ is inner regular on open sets with finite measure. Finally if $U \in \tau$ and $\mu(U) = \infty$, there exists $f_n \prec U$ such that $I(f_n) \uparrow \infty$ as $n \to \infty$. Then, letting $K_n = \text{supp}(f_n)$, we have $K_n \subset U$ and $\mu(K_n) \geq I(f_n)$ and therefore $\mu(K_n) \uparrow \mu(U) = \infty$. ∎

38.2.1 Rudin's Proof of the Riesz-Markov Theorem 38.9

Proof. As usual we let $\mu : \tau \to [0, \infty]$ be as in Eq. (38.4) and by Lemma **??** we know that μ is sub-additive on τ. We now define $\mu : 2^X \to [0, \infty]$ by setting

$$\mu(A) := \inf\{\mu(V) : A \subset V \subset_o X\}.$$

I claim that μ is the outer measure associated to $\mu|_\tau$. Indeed, if $A \subset V := \cup V_i$ with $V_i \in \tau$,

$$\mu(A) \leq \mu(V) \leq \sum_i \mu(V_i)$$

from which it follows that $\mu(A) \leq \mu|_\tau^*(A)$. The reverse inequality is trivial. It now follows by Proposition **??** that μ is subadditive on 2^X as well. This is also easily proved directly since if $A = \cup A_i$ and $A_i \subset V_i \subset_o X$, then $A \subset V := \cup_i V_i$ so that

$$\mu(A) \leq \mu(V) \leq \sum_i \mu(V_i).$$

Since the $V_i \in \tau$ is arbitrary subject to the restriction that $A_i \subset V_i$, it follows that

$$\mu(A) \leq \sum_i \mu(A_i).$$

Now let

$$\mathcal{M}_F := \left\{A \in 2^X : \infty > \mu(A) = \sup\{\mu(K) : K \subset A\}\right\}$$

be those sets A of X which are μ -finite and and are μ – inner regular and let

$$\mathcal{M} := \left\{A \in 2^X : A \cap K \in \mathcal{M}_F \text{ for all } K \sqsubset\sqsubset X\right\}$$

be those sets which are **locally** μ – inner regular.

1. Suppose $K \sqsubset\sqsubset X$ and choose $f \prec X$ such that $f = 1$ on K. For $\alpha \in (0,1)$, let $V_\alpha := \{f > \alpha\}$, then $V_\alpha \in \tau$ and $K \subset V_\alpha$. Hence if $g \prec V_\alpha$, then we have $\alpha g \leq f$ so that $\alpha I(g) \leq I(f)$ which shows

$$\mu(K) \leq \mu(V_\alpha) \leq \alpha^{-1} I(f).$$

Letting $\alpha \uparrow 1$ in this last inequality shows that $\mu(K) \leq I(f) < \infty$ which shows $K \in \mathcal{M}_F$ and that

$$\mu(K) \leq \inf\{I(f) : 1_K \leq f \prec X\}.$$

Given $\varepsilon > 0$, let $V \in \tau$ be chosen so that $K \subset V$ and $\mu(V) < \mu(K) + \varepsilon$ and then choose f such that $1_K \leq f \prec V$. Then

$$I(f) \leq \mu(V) < \mu(K) + \varepsilon$$

from which it follows that

$$\inf\{I(f) : 1_K \leq f \prec X\} \leq \mu(K)$$

and we have shown that

$$\mu(K) = \inf\{I(f) : 1_K \leq f \prec X\} < \infty.$$

2. Now suppose that $V \in \tau$ with $\mu(V) < \infty$ and let $\alpha \in (0, \mu(V))$. Choose $f \prec V$ such that $\alpha \leq I(f) \leq \mu(V)$. Letting $K = \text{supp}(f)$ and $W \in \tau$ such that $K \subset W$, we have $f \prec W$ and therefore that $I(f) \leq \mu(W)$. Since $W \in \tau$ such that $K \subset W$ was arbitrary, it follows that

$$\alpha \leq I(f) \leq \inf\{\mu(W) : K \subset W \subset_o X\} = \mu(K).$$

Since $\alpha < \mu(V)$ was arbitrary, it follows that $\mu(V) = \sup\{\mu(K) : K \subset V\}$ and therefore that $V \in \mathcal{M}_F$.
3. Suppose that K and F are pairwise disjoint compact subsets of X and choose $1_K + 1_F \leq f \prec X$ such $I(f) \leq \mu(K \cup F) + \varepsilon$. Let $\alpha \in C_c(X, [0,1])$ be chosen so that $\alpha = 1$ on K and $\alpha = 0$ on F. Then $1_K \leq \alpha f$ and $1_F \leq (1 - \alpha) f$ so that

$$\mu(K) + \mu(F) \leq I(\alpha f) + I((1-\alpha) f) = I(f) \leq \mu(K \cup F) + \varepsilon \leq \mu(K) + \mu(F) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$\mu(K) + \mu(F) = \mu(K \cup F).$$

4. Now suppose that $\{A_i\}_{i=1}^\infty$ are pairwise disjoint members $\mathcal{M}_F$ and let $A := \cup_{i=1}^\infty A_i$. As we have already seen

$$\mu(A) \leq \sum_{i=1}^\infty \mu(A_i)$$

with equality if $\mu(A) = \infty$. We now suppose that $\mu(A) < \infty$. There exists $K_i \subset A_i$ such that $\mu(A_i) \leq \mu(K_i) + \varepsilon_i$ for any $\varepsilon_i > 0$. Thus

$$\sum_{i=1}^N \mu(A_i) \leq \sum_{i=1}^N [\mu(K_i) + \varepsilon_i] = \mu\left(\bigcup_{i=1}^N K_i\right) + \sum_{i=1}^N \varepsilon_i \leq \mu(A) + \sum_{i=1}^\infty \varepsilon_i$$

and hence
$$\sum_{i=1}^{\infty}\mu(A_i)=\lim_{N\uparrow\infty}\sum_{i=1}^{N}\mu(A_i)\le\mu(A)+\sum_{i=1}^{\infty}\varepsilon_i.$$
Since that $\varepsilon_i>0$ were arbitrary, it follows that
$$\sum_{i=1}^{\infty}\mu(A_i)\le\mu(A)\le\sum_{i=1}^{\infty}\mu(A_i)$$
and hence that
$$\sum_{i=1}^{\infty}\mu(A_i)=\mu(A).$$
In particular if $\mu(A)<\infty$, then $A\in\mathcal{M}_F$ as well.

5. Suppose that $A\in\mathcal{M}_F$ and K is compact and V is open so that $K\subset A\subset V$ and $\mu(V)-\mu(K)<\varepsilon$. We have already shown that K and $V\setminus K\in\tau$ are in $\mathcal{M}_F$. Since $V=K\cup(V\setminus K)$, it follows that $\mu(V)=\mu(K)+\mu(V\setminus K)$, i.e. that
$$\mu(V\setminus K)=\mu(V)-\mu(K).$$
6. We now show that $\mathcal{M}_F$ is closed under finite unions, intersections and differences. Indeed if $A_i\in\mathcal{M}_F$ we may choose $K_i\subset A_i\subset V_i$ such that $\mu(V_i\setminus K_i)<\varepsilon$ for $i=1,2$. Then
$$V_1\setminus K_2\subset(V_1\setminus K_1)\cup(K_1\setminus K_2),$$
$$K_1\setminus K_2\subset(K_1\setminus V_2)\cup(V_2\setminus K_2),$$
and hence
$$K_1\setminus V_2\subset A_1\setminus A_2\subset V_1\setminus K_2\subset(V_1\setminus K_1)\cup(K_1\setminus V_2)\cup(V_2\setminus K_2).$$
and hence it follows that
$$\mu(V_1\setminus K_2)\le 2\varepsilon+\mu((K_1\setminus V_2))$$
and since $K_1\setminus V_2$ is compact we learn that $A_1\setminus A_2\in\mathcal{M}_F$. Furthermore,
$$A_1\cup A_2=(A_1\setminus A_2)\cup A_2\in\mathcal{M}_F$$
and
$$A_1\cap A_2=A_1\setminus(A_1\setminus A_2)\in\mathcal{M}_F.$$
Alternatively,
$$K_1\cup K_2\subset A_1\cup A_2\subset V_1\cup V_2$$
so that
$$\mu(V_1\cup V_2\setminus(K_1\cup K_2))\le\mu((V_1\setminus K_1)\cup(V_2\setminus K_2))\le\mu((V_1\setminus K_1))+\mu((V_2\setminus K_2))\le 2\varepsilon$$
from which it follows that $A_1\cup A_2\in\mathcal{M}_F$ etc.

7. $\mathcal{M}$ is a σ - algebra which contains $\mathcal{B}_X$. If $A\in\mathcal{M}$ and K is compact then $A\cap K\in\mathcal{M}_F$ and hence
$$A^c\cap K=K\setminus A=K\setminus(A\cap K)\in\mathcal{M}_F.$$
Since K was arbitrary it follows that $A\in\mathcal{M}$ and we have shown $\mathcal{M}$ is stable under complementation. Now suppose that $A=\cup A_i$ with $A_i\in\mathcal{M}$ and K is compact. Then
$$A\cap K=\bigcup_{i=1}^{\infty}(A_i\cap K)=\bigcup_{i=1}^{\infty}B_i$$
where
$$B_i:=(A_i\cap K)\setminus\left[\cup_{j=1}^{i}(A_j\cap K)\right]\in\mathcal{M}_F.$$
Since that $B_i\in\mathcal{M}_F$ are pairwise disjoint, it follows that $A\cap K\in\mathcal{M}_F$ as well and hence that $A\in\mathcal{M}$.
Moreover if C is a closed set then $C\cap K$ is compact and hence in $\mathcal{M}_F$ for all compact sets K. Thus $\mathcal{M}$ is a σ algebra which contains all closed sets and therefore the contains the Borel $\sigma-$ algebra.
8. We have $\mathcal{M}_F=\{A\in\mathcal{M}:\mu(A)<\infty\}$. As we have seen if $A\in\mathcal{M}_F$ then $A\cap K\in\mathcal{M}_F$ for all K compact and hence it easily follows that $\mathcal{M}_F\subset\mathcal{M}$. Conversely if $A\in\mathcal{M}$ with $\mu(A)<\infty$, then choose $V\in\tau$ such that $A\subset V$ and $\mu(A)\cong\mu(V)<\infty$. Then choose $K\sqsubset\sqsubset V$ such that $\mu(V\setminus K)\cong 0$. Since $K\cap A\in\mathcal{M}_F$ there exists a compact set H such that $H\subset A\cap K\subset A$ such that $\mu(A\cap K)\cong\mu(H)$. Since
$$A\subset(A\cap K)\cup(A\setminus K)\subset(A\cap K)\cup(V\setminus K)$$
it follows that
$$\mu(A)\le\mu(A\cap K)+\mu(V\setminus K)\cong\mu(H)$$
which shows that A is inner regular and hence that $A\in\mathcal{M}_F$.
9. μ is a measure on $\mathcal{M}$. To see this suppose that A is the disjoint union of $A_i\in\mathcal{M}$. If $\mu(A)<\infty$, then $A_i\in\mathcal{M}_F$ for all i and we know that $\mu(A)=\sum_i\mu(A_i)$. Conversely, if $\mu(A)=\infty$ then $\infty=\mu(A)\le\sum_i\mu(A_i)$ as desired.
10. The fact that $I(f)=\int_X fd\mu$ for all $f\in C_c(X)$ follows as in the previous proof.

■

38.2.2 Regularity Results For Radon Measures

Proposition 38.12. *If μ is a Radon measure on X then μ is inner regular on all σ-finite Borel sets.*

Proof. Suppose $A \in \mathcal{B}_X$ and $\mu(A) < \infty$ and $\varepsilon > 0$ is given. By outer regularity of μ, there exist an open set $U \subset_o X$ such that $A \subset U$ and $\mu(U \setminus A) < \varepsilon$. By inner regularity on open sets, there exists a compact set $F \sqsubset\sqsubset U$ such that $\mu(U \setminus F) < \varepsilon$. Again by outer regularity of μ, there exist $V \subset_o X$ such that $(U \setminus A) \subset V$ and $\mu(V) < \varepsilon$. Then $K := F \setminus V$ is compact set and

$$K \subset F \setminus (U \setminus A) = F \cap (U \cap A^c)^c = F \cap (U^c \cup A) = F \cap A,$$

see Figure 38.3. Since,

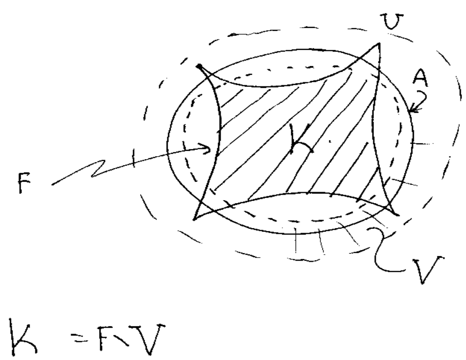


Fig. 38.3. Constructing the compact set K.

$$\mu(K) = \mu(F) - \mu(F \cap V) \approx \mu(U) \approx \mu(A),$$

or more formally,

$$\begin{aligned}\mu(K) &= \mu(F) - \mu(F \cap V) \geq \mu(U) - \varepsilon - \mu(F \cap V) \\ &\geq \mu(U) - 2\varepsilon \geq \mu(A) - 3\varepsilon,\end{aligned}$$

we see that $\mu(A \setminus K) \leq 3\varepsilon$. This proves the proposition when $\mu(A) < \infty$.

If $\mu(A) = \infty$ and there exists $A_n \uparrow A$ as $n \to \infty$ with $\mu(A_n) < \infty$. Then by the first part, there exist compact set K_n such that $K_n \subset A_n$ and $\mu(A_n \setminus K_n) < 1/n$ or equivalently $\mu(K_n) > \mu(A_n) - 1/n \to \infty$ as $n \to \infty$. ■

Corollary 38.13. *Every σ-finite Radon measure, μ, is a regular Borel measure, i.e. μ is both outer and inner regular on all Borel subsets.*

Notation 38.14 *If (X, τ) is a topological space, let F_σ denote the collection of sets formed by taking countable unions of closed sets and $G_\delta = \tau_\delta$ denote the collection of sets formed by taking countable intersections of open sets.*

Proposition 38.15. *Suppose that μ is a σ -finite Radon measure and $B \in \mathcal{B}$. Then*

1. *For all $\varepsilon > 0$ there exists sets $F \subset B \subset U$ with F closed, U open and $\mu(U \setminus F) < \varepsilon$.*
2. *There exists $A \in F_\sigma$ and $C \in G_\delta$ such that $A \subset B \subset C$ such that and $\mu(C \setminus A) = 0$.*

Proof. 1. Let $X_n \in \mathcal{B}$ such that $X_n \uparrow X$ and $\mu(X_n) < \infty$ and choose open set U_n such that $B \cap X_n \subset U_n$ and $\mu(U_n \setminus (B \cap X_n)) < \varepsilon 2^{-(n+1)}$. Then $U := \bigcup_{n=1}^\infty U_n$ is an open set such that

$$\mu(U \setminus B) \leq \sum_{n=1}^\infty \mu(U_n \setminus B) \leq \sum_{n=1}^\infty \mu(U_n \setminus (B \cap X_n)) < \frac{\varepsilon}{2}.$$

Applying this same result to B^c allows us to find a closed set F such that $B^c \subset F^c$ and

$$\mu(B \setminus F) = \mu(F^c \setminus B^c) < \frac{\varepsilon}{2}.$$

Thus $F \subset B \subset U$ and $\mu(U \setminus F) < \varepsilon$ as desired.

2. This a simple consequence of item 1. ■

Theorem 38.16. *Let X be a locally compact Hausdorff space such that every open set $V \subset_o X$ is σ – compact, i.e. there exists $K_n \sqsubset\sqsubset V$ such that $V = \cup_n K_n$. Then any K-finite measure ν on X is a Radon measure and in fact is a regular Borel measure. (The reader should check that if X is second countable, then open sets are σ compact, see Exercise 36.1. In particular this condition holds for $\mathbb{R}^n$ with the standard topology.)*

Proof. By the Riesz-Markov Theorem 38.9, the positive linear functional,

$$I(f) := \int_X f d\nu \text{ for all } f \in C_c(X),$$

may be represented by a Radon measure μ on $(X, \mathcal{B})$, i.e. such that $I(f) = \int_X f d\mu$ for all $f \in C_c(X)$. By Corollary 38.13, μ is also a regular Borel measure on $(X, \mathcal{B})$. So to finish the proof it suffices to show $\nu = \mu$. We will give two proofs of this statement.

First Proof. The same arguments used in the proof of Lemma 11.34 shows $\sigma(C_c(X)) = \mathcal{B}_X$. Let K be a compact subset of X and use Urysohn's Lemma

37.8 to find $\varphi \prec X$ such that $\varphi \geq 1_K$. By a simple application of the multiplicative system Theorem 11.28 one shows

$$\int_X \varphi f d\nu = \int_X \varphi f d\mu$$

for all bounded $\mathcal{B}_X = \sigma\left(C_c\left(X\right)\right)$ – measurable functions on X. Taking $f = 1_K$ then shows that $\nu\left(K\right) = \mu\left(K\right)$ with $K \sqsubset\sqsubset X$. An application of Theorem **??** implies $\mu = \nu$ on σ – algebra generated by the compact sets. This completes the proof, since, by assumption, this σ – algebra contains all of the open sets and hence is the Borel σ – algebra.

Second Proof. Since μ is a Radon measure on X, it follows from Eq. (38.9), that

$$\mu(U) = \sup\left\{\int_X f d\mu : f \prec U\right\} = \sup\left\{\int_X f d\nu : f \prec U\right\} \leq \nu\left(U\right) \qquad (38.19)$$

for all open subsets U of X. For each compact subset $K \subset U$, there exists, by Uryshon's Lemma 37.8, a function $f \prec U$ such that $f \geq 1_K$. Thus

$$\nu(K) \leq \int_X f d\nu = \int_X f d\mu \leq \mu\left(U\right). \qquad (38.20)$$

Combining Eqs. (38.19) and (38.20) implies $\nu(K) \leq \mu(U) \leq \nu(U)$. By assumption there exists compact sets, $K_n \subset U$, such that $K_n \uparrow U$ as $n \to \infty$ and therefore by continuity of ν,

$$\nu(U) = \lim_{n\to\infty} \nu(K_n) \leq \mu(U) \leq \nu(U).$$

Hence we have shown, $\nu\left(U\right) = \mu\left(U\right)$ for all $U \in \tau$.

If $B \in \mathcal{B} = \mathcal{B}_X$ and $\varepsilon > 0$, by Proposition 38.15, there exists $F \subset B \subset U$ such that F is closed, U is open and $\mu(U \setminus F) < \varepsilon$. Since $U \setminus F$ is open, $\nu(U \setminus F) = \mu(U \setminus F) < \varepsilon$ and therefore

$$\nu\left(U\right) - \varepsilon \leq \nu\left(B\right) \leq \nu\left(U\right) \text{ and}$$
$$\mu\left(U\right) - \varepsilon \leq \mu\left(B\right) \leq \mu\left(U\right).$$

Since $\nu\left(U\right) = \mu\left(U\right)$, $\nu\left(B\right) = \infty$ iff $\mu\left(B\right) = \infty$ and if $\nu\left(B\right) < \infty$ then $\left|\nu\left(B\right) - \mu\left(B\right)\right| < \varepsilon$. Because $\varepsilon > 0$ is arbitrary, we may conclude that $\nu\left(B\right) = \mu\left(B\right)$ for all $B \in \mathcal{B}$. ■

Proposition 38.17 (Density of $C_c\left(X\right)$ in $L^p\left(\mu\right)$). *If μ is a Radon measure on X, then $C_c(X)$ is dense in $L^p(\mu)$ for all $1 \leq p < \infty$.*

Proof. Let $\varepsilon > 0$ and $B \in \mathcal{B}_X$ with $\mu(B) < \infty$. By Proposition 38.12, there exists $K \sqsubset\sqsubset B \subset U \subset_o X$ such that $\mu(U \setminus K) < \varepsilon^p$ and by Urysohn's Lemma 37.8, there exists $f \prec U$ such that $f = 1$ on K. This function f satisfies

$$\|f - 1_B\|_p^p = \int_X |f - 1_B|^p \, d\mu \leq \int_{U\setminus K} |f - 1_B|^p \, d\mu \leq \mu(U \setminus K) < \varepsilon^p.$$

From this it easy to conclude that $C_c\left(X\right)$ is dense in $\mathbb{S}_f(\mathcal{B}, \mu)$ – the simple functions on X which are in $L^1\left(\mu\right)$. Combining this with Lemma 31.3 which asserts that $\mathbb{S}_f(\mathcal{B}, \mu)$ is dense in $L^p\left(\mu\right)$ completes the proof of the theorem. ■

Theorem 38.18 (Lusin's Theorem). *Suppose (X, τ) is a locally compact Hausdorff space, $\mathcal{B}_X$ is the Borel σ – algebra on X, and μ is a Radon measure on $(X, \mathcal{B}_X)$. Also let $\varepsilon > 0$ be given. If $f : X \to \mathbb{C}$ is a measurable function such that $\mu(f \neq 0) < \infty$, there exists a compact set $K \subset \{f \neq 0\}$ such that $f|_K$ is continuous and $\mu(\{f \neq 0\} \setminus K) < \varepsilon$. Moreover there exists $\varphi \in C_c(X)$ such that $\mu(f \neq \varphi) < \varepsilon$ and if f is bounded the function φ may be chosen so that*

$$\|\varphi\|_u \leq \|f\|_u := \sup_{x\in X} |f(x)|.$$

Proof. Suppose first that f is bounded, in which case

$$\int_X |f| \, d\mu \leq \|f\|_u \, \mu(f \neq 0) < \infty.$$

By Proposition 38.17, there exists $f_n \in C_c(X)$ such that $f_n \to f$ in $L^1(\mu)$ as $n \to \infty$. By passing to a subsequence if necessary, we may assume $\|f - f_n\|_1 < \varepsilon n^{-1} 2^{-n}$ and hence by Chebyshev's inequality (Lemma **??**),

$$\mu\left(|f - f_n| > n^{-1}\right) < \varepsilon 2^{-n} \text{ for all } n.$$

Let $E := \cup_{n=1}^\infty \left\{|f - f_n| > n^{-1}\right\}$, so that $\mu(E) < \varepsilon$. On E^c, $|f - f_n| \leq 1/n$, i.e. $f_n \to f$ uniformly on E^c and hence $f|_{E^c}$ is continuous. By Proposition 38.12, there exists a compact set K and open set V such that

$$K \subset \{f \neq 0\} \setminus E \subset V$$

such that $\mu(V \setminus K) < \varepsilon$. Notice that

$$\begin{aligned}\mu(\{f \neq 0\} \setminus K) &= \mu\left((\{f \neq 0\} \setminus K) \setminus E\right) + \mu\left((\{f \neq 0\} \setminus K) \cap E\right) \\ &\leq \mu(V \setminus K) + \mu(E) < 2\varepsilon.\end{aligned}$$

By the Tietze extension Theorem 37.9, there exists $F \in C\left(X\right)$ such that $f = F|_K$. By Urysohn's Lemma 37.8 there exists $\psi \prec V$ such that $\psi = 1$ on K. So letting $\varphi = \psi F \in C_c(X)$, we have $\varphi = f$ on K, $\|\varphi\|_\infty \leq \|f\|_\infty$ and since $\{\varphi \neq f\} \subset E \cup (V \setminus K)$, $\mu(\varphi \neq f) < 3\varepsilon$. This proves the assertions in the theorem when f is bounded.

Suppose that $f : X \to \mathbb{C}$ is (possibly) unbounded and $\varepsilon > 0$ is given. Then $B_N := \{0 < |f| \leq N\} \uparrow \{f \neq 0\}$ as $N \to \infty$ and therefore for all N sufficiently large,

$$\mu(\{f \neq 0\} \setminus B_N) < \varepsilon/3.$$

Since μ is a Radon measure, Proposition 38.12, guarantee's there is a compact set $C \subset \{f \neq 0\}$ such that $\mu(\{f \neq 0\} \setminus C) < \varepsilon/3$. Therefore,

$$\mu(\{f \neq 0\} \setminus (B_N \cap C)) < 2\varepsilon/3.$$

We may now apply the bounded case already proved to the function $1_{B_N \cap C} f$ to find a compact subset K and an open set V such that $K \subset V$,

$$K \subset \{1_{B_N \cap C} f \neq 0\} = B_N \cap C \cap \{f \neq 0\}$$

such that $\mu((B_N \cap C \cap \{f \neq 0\}) \setminus K) < \varepsilon/3$ and $\varphi \in C_c(X)$ such that $\varphi = 1_{B_N \cap C} f = f$ on K. This completes the proof, since

$$\mu(\{f \neq 0\} \setminus K) \leq \mu((B_N \cap C \cap \{f \neq 0\}) \setminus K) + \mu(\{f \neq 0\} \setminus (B_N \cap C)) < \varepsilon$$

which implies $\mu(f \neq \varphi) < \varepsilon$. ■

Example 38.19. To illustrate Theorem 38.18, suppose that $X = (0,1)$, $\mu = m$ is Lebesgue measure and $f = 1_{(0,1)\cap\mathbb{Q}}$. Then Lusin's theorem asserts for any $\varepsilon > 0$ there exists a compact set $K \subset (0,1)$ such that $m((0,1) \setminus K) < \varepsilon$ and $f|_K$ is continuous. To see this directly, let $\{r_n\}_{n=1}^\infty$ be an enumeration of the rationals in $(0,1)$,

$$J_n = (r_n - \varepsilon 2^{-n}, r_n + \varepsilon 2^{-n}) \cap (0,1) \text{ and } W = \cup_{n=1}^\infty J_n.$$

Then W is an open subset of X and $\mu(W) < \varepsilon$. Therefore $K_n := [1/n, 1-1/n] \setminus W$ is a compact subset of X and $m(X \setminus K_n) \leq \frac{2}{n} + \mu(W)$. Taking n sufficiently large we have $m(X \setminus K_n) < \varepsilon$ and $f|_{K_n} \equiv 0$ which is of course continuous.

The following result is a slight generalization of Lemma 31.11.

Corollary 38.20. *Let (X,τ) be a locally compact Hausdorff space, $\mu : \mathcal{B}_X \to [0,\infty]$ be a Radon measure on X and $h \in L^1_{loc}(\mu)$. If*

$$\int_X fhd\mu = 0 \text{ for all } f \in C_c(X) \tag{38.21}$$

then $1_K h = 0$ for μ – a.e. for every compact subset $K \subset X$. (BRUCE: either show $h = 0$ a.e. or give a counterexample. Also, either prove or give a counterexample to the question to the statement the $d\nu = \rho d\mu$ is a Radon measure if $\rho \geq 0$ and in $L^1_{loc}(\mu)$.)

Proof. By considering the real and imaginary parts of h we may assume with out loss of generality that h is real valued. Let K be a compact subset of X. Then $1_K \text{sgn}(\bar{h}) \in L^1(\mu)$ and by Proposition 38.17, there exists $f_n \in C_c(X)$ such that $\lim_{n\to\infty} \|f_n - 1_K \text{sgn}(h)\|_{L^1(\mu)} = 0$. Let $\varphi \in C_c(X, [0,1])$ such that $\varphi = 1$ on K and $g_n = \varphi \min(-1, \max(1, f_n))$. Since

$$|g_n - 1_K \text{sgn}(h)| \leq |f_n - 1_K \text{sgn}(h)|$$

we have found $g_n \in C_c(X, \mathbb{R})$ such that $|g_n| \leq 1_{\text{supp}(\varphi)}$ and $g_n \to 1_K \text{sgn}(h)$ in $L^1(\mu)$. By passing to a sub-sequence if necessary we may assume the convergence happens μ – almost everywhere. Using Eq. (38.21) and the dominated convergence theorem (the dominating function is $|h| 1_{\text{supp}(\varphi)}$) we conclude that

$$0 = \lim_{n\to\infty} \int_X g_n h d\mu = \int_X 1_K \text{sgn}(\bar{h}) h d\mu = \int_K |h| \, d\mu$$

which shows $h(x) = 0$ for μ-a.e. $x \in K$. ■

38.2.3 The dual of $C_0(X)$

Definition 38.21. *Let (X,τ) be a locally compact Hausdorff space and $\mathcal{B} = \sigma(\tau)$ be the Borel σ – algebra. A* ***signed Radon measure*** *is a signed measure μ on $\mathcal{B}$ such that the measures, $\mu_\pm$, in the Jordan decomposition of μ are both Radon measures. A* ***complex Radon measure*** *is a complex measure μ on $\mathcal{B}$ such that $\operatorname{Re}\mu$ and $\operatorname{Im}\mu$ are signed radon measures.*

Example 38.22. Every complex measure μ on $\mathcal{B}_{\mathbb{R}^d}$ is a Radon measure. BRUCE: add some more examples and perhaps some exercises here.

BRUCE: Compare and combine with results from Section **??**.

Proposition 38.23. *Suppose (X,τ) is a topological space and $I \in C_0(X,\mathbb{R})^*$. Then we may write $I = I_+ - I_-$ where $I_\pm \in C_0(X,\mathbb{R})^*$ are positive linear functionals.*

Proof. For $f \in C_0(X, [0,\infty))$, let

$$I_+(f) := \sup\{I(g) : g \in C_0(X, [0,\infty)) \text{ and } g \leq f\}$$

and notice that $|I_+(f)| \leq \|I\| \|f\|$. If $c > 0$, then $I_+(cf) = cI_+(f)$. Suppose that $f_1, f_2 \in C_0(X, [0,\infty))$ and $g_i \in C_0(X, [0,\infty))$ such that $g_i \leq f_i$, then $g_1 + g_2 \leq f_1 + f_2$ so that

$$I(g_1) + I(g_2) = I(g_1 + g_2) \leq I_+(f_1 + f_2)$$

and therefore

$$I_+(f_1) + I_+(f_2) \le I_+(f_1 + f_2). \tag{38.22}$$

Moreover, if $g \in C_0(X, [0, \infty))$ and $g \le f_1 + f_2$, let $g_1 = \min(f_1, g)$, so that

$$0 \le g_2 := g - g_1 \le f_1 - g_1 + f_2 \le f_2.$$

Hence $I(g) = I(g_1) + I(g_2) \le I_+(f_1) + I_+(f_2)$ for all such g and therefore,

$$I_+(f_1 + f_2) \le I_+(f_1) + I_+(f_2). \tag{38.23}$$

Combining Eqs. (38.22) and (38.23) shows that $I_+(f_1 + f_2) = I_+(f_1) + I_+(f_2)$. For general $f \in C_0(X, \mathbb{R})$, let $I_+(f) = I_+(f_+) - I_+(f_-)$ where $f_+ = \max(f, 0)$ and $f_- = -\min(f, 0)$. (Notice that $f = f_+ - f_-$.) If $f = h - g$ with $h, g \in C_0(X, \mathbb{R})$, then $g + f_+ = h + f_-$ and therefore,

$$I_+(g) + I_+(f_+) = I_+(h) + I_+(f_-)$$

and hence $I_+(f) = I_+(h) - I_+(g)$. In particular,

$$I_+(-f) = I_+(f_- - f_+) = I_+(f_-) - I_+(f_+) = -I_+(f)$$

so that $I_+(cf) = cI_+(f)$ for all $c \in \mathbb{R}$. Also,

$$\begin{aligned} I_+(f + g) &= I_+(f_+ + g_+ - (f_- + g_-)) = I_+(f_+ + g_+) - I_+(f_- + g_-) \\ &= I_+(f_+) + I_+(g_+) - I_+(f_-) - I_+(g_-) \\ &= I_+(f) + I_+(g). \end{aligned}$$

Therefore I_+ is linear. Moreover,

$$|I_+(f)| \le \max(|I_+(f_+)|, |I_+(f_-)|) \le \|I\| \max(\|f_+\|, \|f_-\|) = \|I\| \|f\|$$

which shows that $\|I_+\| \le \|I\|$. Let $I_- = I_+ - I \in C_0(X, \mathbb{R})^*$, then for $f \ge 0$,

$$I_-(f) = I_+(f) - I(f) \ge 0$$

by definition of I_+, so $I_- \ge 0$ as well. ∎

Remark 38.24. The above proof works for functionals on linear spaces of bounded functions which are closed under taking $f \wedge g$ and $f \vee g$. As an example, let $\lambda(f) = \int_0^1 f(x)dx$ for all bounded measurable functions $f : [0, 1] \to \mathbb{R}$. By the Hahn Banach Theorem 21.7 (or Corollary 21.8) below, we may extend λ to a linear functional Λ on all bounded functions on $[0, 1]$ in such a way that $\|\Lambda\| = 1$. Let Λ_+ be as above, then $\Lambda_+ = \lambda$ on bounded measurable functions and $\|\Lambda_+\| = 1$. Define $\mu(A) := \Lambda(1_A)$ for all $A \subset [0, 1]$ and notice that if A is measurable, the $\mu(A) = m(A)$. So μ is a finitely additive extension of m to **all** subsets of $[0, 1]$.

Exercise 38.1. Suppose that μ is a signed Radon measure and $I = I_\mu$. Let μ_+ and μ_- be the Radon measures associated to $I_\pm$ with $I_\pm$ being constructed as in the proof of Proposition 38.23. Show that $\mu = \mu_+ - \mu_-$ is the Jordan decomposition of μ.

Theorem 38.25. *Let X be a locally compact Hausdorff space, $M(X)$ be the space of complex Radon measures on X and for $\mu \in M(X)$ let $\|\mu\| = |\mu|(X)$. Then the map*

$$\mu \in M(X) \to I_\mu \in C_0(X)^*$$

is an isometric isomorphism. Here again $I_\mu(f) := \int_X f \, d\mu$.

Proof. To show that the map $M(X) \to C_0(X)^*$ is surjective, let $I \in C_0(X)^*$ and then write $I = I^{re} + iI^{im}$ be the decomposition into real and imaginary parts. Then further decompose these into there plus and minus parts so

$$I = I_+^{re} - I_-^{re} + i\left(I_+^{im} - I_-^{im}\right)$$

and let $\mu_\pm^{re}$ and $\mu_\pm^{im}$ be the corresponding positive Radon measures associated to $I_\pm^{re}$ and $I_\pm^{im}$. Then $I = I_\mu$ where

$$\mu = \mu_+^{re} - \mu_-^{re} + i\left(\mu_+^{im} - \mu_-^{im}\right).$$

To finish the proof it suffices to show $\|I_\mu\|_{C_0(X)^*} = \|\mu\| = |\mu|(X)$. We have

$$\begin{aligned} \|I_\mu\|_{C_0(X)^*} &= \sup\left\{\left|\int_X f d\mu\right| : f \in C_0(X) \ni \|f\|_\infty \le 1\right\} \\ &\le \sup\left\{\left|\int_X f d\mu\right| : f \text{ measurable and } \|f\|_\infty \le 1\right\} = \|\mu\|. \end{aligned}$$

To prove the opposite inequality, write $d\mu = gd|\mu|$ with g a complex measurable function such that $|g| = 1$. By Proposition 38.17, there exist $f_n \in C_c(X)$ such that $f_n \to g$ in $L^1(|\mu|)$ as $n \to \infty$. Let $g_n = \varphi(f_n)$ where $\varphi : \mathbb{C} \to \mathbb{C}$ is the continuous function defined by $\varphi(z) = z$ if $|z| \le 1$ and $\varphi(z) = z/|z|$ if $|z| \ge 1$. Then $|g_n| \le 1$ and $g_n \to g$ in $L^1(\mu)$. Thus

$$\|\mu\| = |\mu|(X) = \int_X d|\mu| = \int_X \bar{g} d\mu = \lim_{n\to\infty} \int_X \bar{g}_n d\mu \le \|I_\mu\|_{C_0(X)^*}.$$

∎

Exercise 38.2. Let (X, τ) be a compact Hausdorff space which supports a positive measure ν on $\mathcal{B} = \sigma(\tau)$ such that $\nu(X) \neq \sum_{x \in X} \nu(\{x\})$, i.e. ν is a not a counting type measure. (Example $X = [0, 1]$.) Then $C(X)$ is not reflexive.

Hint: recall that $C(X)^*$ is isomorphic to the space of complex Radon measures on $(X, \mathcal{B})$. Using this isomorphism, define $\lambda \in C(X)^{**}$ by

$$\lambda(\mu) = \sum_{x \in X} \mu(\{x\})$$

and then show $\lambda \neq \hat{f}$ for any $f \in C(X)$.

38.3 Classifying Radon Measures on $\mathbb{R}$

Throughout this section, let $X = \mathbb{R}$, $\mathcal{E}$ be the elementary class

$$\mathcal{E} = \{(a, b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\}, \tag{38.24}$$

and $\mathcal{A} = \mathcal{A}(\mathcal{E})$ be the algebra formed by taking finite disjoint unions of elements from $\mathcal{E}$, see Proposition **??**. The aim of this section is to prove Theorem **??** which we restate here for convenience.

Theorem 38.26. *The collection of K-finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ are in one to one correspondence with a right continuous non-decreasing functions, $F : \mathbb{R} \to \mathbb{R}$, with $F(0) = 0$. The correspondence is as follows. If F is a right continuous non-decreasing function $F : \mathbb{R} \to \mathbb{R}$, then there exists a unique measure, μ_F, on $\mathcal{B}_{\mathbb{R}}$ such that*

$$\mu_F((a, b]) = F(b) - F(a) \ \forall \ -\infty < a \leq b < \infty$$

and this measure may be defined by

$$\begin{aligned} \mu_F(A) &= \inf\left\{\sum_{i=1}^{\infty}(F(b_i) - F(a_i)) : A \subset \cup_{i=1}^{\infty}(a_i, b_i]\right\} \\ &= \inf\left\{\sum_{i=1}^{\infty}(F(b_i) - F(a_i)) : A \subset \coprod_{i=1}^{\infty}(a_i, b_i]\right\} \end{aligned} \tag{38.25}$$

for all$A \in \mathcal{B}_{\mathbb{R}}$. Conversely if μ is K-finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, then

$$F(x) := \begin{cases} -\mu((x, 0]) & \text{if } x \leq 0 \\ \mu((0, x]) & \text{if } x \geq 0 \end{cases} \tag{38.26}$$

is a right continuous non-decreasing function and this map is the inverse to the map, $F \to \mu_F$.

There are three aspects to this theorem; namely the existence of the map $F \to \mu_F$, the surjectivity of the map and the injectivity of this map. Assuming the map $F \to \mu_F$ exists, the surjectivity follows from Eq. (38.26) and the injectivity is an easy consequence of Theorem **??**. The rest of this section is devoted to giving two proofs for the existence of the map $F \to \mu_F$.

Exercise 38.3. Show by direct means any measure $\mu = \mu_F$ satisfying Eq. (38.25) is outer regular on all Borel sets. **Hint:** it suffices to show if $B := \coprod_{i=1}^{\infty}(a_i, b_i]$, then there exists $V \subset_o \mathbb{R}$ such that $\mu(V \setminus B)$ is as small as you please.

38.3.1 Classifying Radon Measures on $\mathbb{R}$ using Theorem ??

Corollary 38.27. *The map $F \to \mu_F$ in Theorem 38.26 exists.*

Proof. This is simply a combination of Proposition **??** and Theorem **??**. ∎

38.3.2 Classifying Radon Measures on $\mathbb{R}$ using the Riesz-Markov Theorem 38.9

38.3.3 The Lebesgue-Stieljtes Integral

Notation 38.28 *Given an increasing function $F : \mathbb{R} \to \mathbb{R}$, let $F(x-) = \lim_{y \uparrow x} F(y)$, $F(x+) = \lim_{y \downarrow x} F(y)$ and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x) \in \bar{\mathbb{R}}$. Since F is increasing all of theses limits exists.*

Theorem 38.29. *If $F : \mathbb{R} \to \mathbb{R}$ is an increasing function (**not** necessarily right continuous), there exists a unique measure $\mu = \mu_F$ on $\mathcal{B}_{\mathbb{R}}$ such that*

$$\int_{-\infty}^{\infty} f dF = \int_{\mathbb{R}} f d\mu \text{ for all } f \in C_c(\mathbb{R}, \mathbb{R}), \tag{38.27}$$

*where $\int_{-\infty}^{\infty} f dF$ is as in Lemma **??** above. This measure may also be characterized as the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that*

$$\mu((a, b]) = F(b+) - F(a+) \text{ for all } -\infty < a < b < \infty. \tag{38.28}$$

Moreover, if $A \in \mathcal{B}_{\mathbb{R}}$ then

$$\begin{aligned} \mu_F(A) &= \inf\left\{\sum_{i=1}^{\infty}(F(b_i+) - F(a_i+)) : A \subset \cup_{i=1}^{\infty}(a_i, b_i]\right\} \\ &= \inf\left\{\sum_{i=1}^{\infty}(F(b_i+) - F(a_i+)) : A \subset \coprod_{i=1}^{\infty}(a_i, b_i]\right\}. \end{aligned} \tag{38.29}$$

Proof. An application of the Riesz-Markov Theorem 38.9 implies there exists a unique measure μ on $\mathcal{B}_{\mathbb{R}}$ such Eq. (38.27) is valid. Let $-\infty < a < b < \infty$, $\varepsilon > 0$ be small and $\varphi_\varepsilon(x)$ be the function defined in Figure 38.4, i.e. φ_ε is one on $[a+2\varepsilon, b+\varepsilon]$, linearly interpolates to zero on $[b+\varepsilon, b+2\varepsilon]$ and on $[a+\varepsilon, a+2\varepsilon]$ and is zero on $(a, b+2\varepsilon)^c$. Since $\varphi_\varepsilon \to 1_{(a,b]}$ it follows by the dominated convergence theorem that

$$\mu((a, b]) = \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \varphi_\varepsilon d\mu = \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \varphi_\varepsilon dF. \tag{38.30}$$

On the other hand we have

$$1_{(a+2\varepsilon, b+\varepsilon]} \leq \varphi_\varepsilon \leq 1_{(a+\varepsilon, b+2\varepsilon]}, \tag{38.31}$$

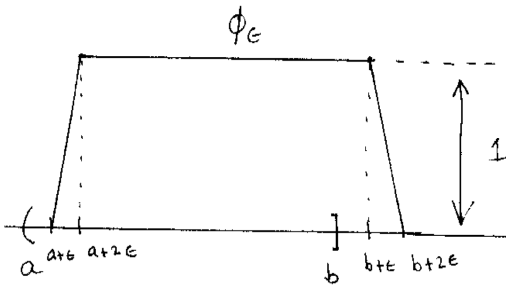


Fig. 38.4. .

and therefore applying I_F to both sides of Eq. (38.31) shows;

$$\begin{aligned} F(b+\varepsilon) - F(a+2\varepsilon) &= \int_{\mathbb{R}} 1_{(a+2\varepsilon, b+\varepsilon]} dF \\ &\leq \int_{\mathbb{R}} \varphi_\varepsilon dF \\ &\leq \int_{\mathbb{R}} 1_{(a+\varepsilon, b+2\varepsilon]} dF = F(b+2\varepsilon) - F(a+\varepsilon). \end{aligned}$$

Letting $\varepsilon \downarrow 0$ in this equation and using Eq. (38.30) shows

$$F(b+) - F(a+) \leq \mu((a,b]) \leq F(b+) - F(a+).$$

For the last assertion let

$$\begin{aligned} \mu^0(A) &= \inf \left\{ \sum_{i=1}^{\infty} (F(b_i) - F(a_i)) : A \subset \coprod_{i=1}^{\infty} (a_i, b_i] \right\} \\ &= \inf \left\{ \mu(B) : A \subset B \in \mathcal{A}_\sigma \right\}, \end{aligned}$$

where $\mathcal{A}$ is the algebra generated by the half open intervals on $\mathbb{R}$. By monotonicity of μ, it follows that

$$\mu^0(A) \geq \mu(A) \text{ for all } A \in \mathcal{B}. \tag{38.32}$$

For the reverse inequality, let $A \subset V \subset_o \mathbb{R}$ and notice by Exercise 35.22 that $V = \coprod_{i=1}^{\infty} (a_i, b_i)$ for some collection of disjoint open intervals in $\mathbb{R}$. Since $(a_i, b_i) \in \mathcal{A}_\sigma$ (as the reader should verify!), it follows that $V \in \mathcal{A}_\sigma$ and therefore,

$$\mu^0(A) \leq \inf \{\mu(V) : A \subset V \subset_o \mathbb{R}\} = \mu(A).$$

Combining this with Eq. (38.32) shows $\mu^0(A) = \mu(A)$ which is precisely Eq. (38.29). ∎

Corollary 38.30. *The map $F \to \mu_F$ is a one to one correspondence between right continuous non-decreasing functions F such that $F(0) = 0$ and Radon (same as K - finite) measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.*

38.4 Kolmogorov's Existence of Measure on Products Spaces

Throughout this section, let $\{(X_\alpha, \tau_\alpha)\}_{\alpha \in A}$ be second countable locally compact Hausdorff spaces and let $X := \prod_{\alpha \in A} X_\alpha$ be equipped with the product topology, $\tau := \otimes_{\alpha \in A} \tau_\alpha$. More generally for $\Lambda \subset A$, let $X_\Lambda := \prod_{\alpha \in \Lambda} X_\alpha$ and $\tau_\Lambda := \otimes_{\alpha \in \Lambda} \tau_\alpha$ and $\Lambda \subset \Gamma \subset A$, let $\pi_{\Lambda,\Gamma} : X_\Gamma \to X_\Lambda$ be the projection map; $\pi_{\Lambda,\Gamma}(x) = x|_\Lambda$ for $x \in X_\Gamma$. We will simply write π_Λ for $\pi_{\Lambda,A} : X \to X_\Lambda$. (Notice that if Λ is a finite subset of A then $(X_\Lambda, \tau_\Lambda)$ is still second countable as the reader should verify.) Let $\mathcal{M} = \otimes_{\alpha \in A} \mathcal{B}_\alpha$ be the product σ – algebra on $X = X_A$ and $\mathcal{B}_\Lambda = \sigma(\tau_\Lambda)$ be the Borel σ – algebra on X_Λ.

Theorem 38.31 (Kolmogorov's Existence Theorem). *Suppose $\{\mu_\Lambda : \Lambda \subset\subset A\}$ are probability measures on $(X_\Lambda, \mathcal{B}_\Lambda)$ satisfying the following compatibility condition:*

- $(\pi_{\Lambda,\Gamma})_* \mu_\Gamma = \mu_\Lambda$ *whenever* $\Lambda \subset \Gamma \subset\subset A$.

Then there exists a unique probability measure, μ, on $(X, \mathcal{M})$ such that $(\pi_\Lambda)_* \mu = \mu_\Lambda$ whenever $\Lambda \subset\subset A$. Recall, see Exercise **??**, *that the condition $(\pi_\Lambda)_* \mu = \mu_\Lambda$ is equivalent to the statement;*

$$\int_X F(\pi_\Lambda(x)) d\mu(x) = \int_{X_\Lambda} F(y) d\mu_\Lambda(y) \tag{38.33}$$

for all $\Lambda \subset\subset A$ and $F : X_\Lambda \to \mathbb{R}$ bounded a measurable.

We will first prove the theorem in the following special case. The full proof will be given after Exercise 38.4 below.

Theorem 38.32. *Theorem 38.31 holds under the additional assumption that each of the spaces, $\{(X_\alpha, \tau_\alpha)\}_{\alpha \in A}$, are* ***compact*** *second countable and Hausdorff and A is countable.*

Proof. Recall from Theorem **??** that the Borel σ – algebra, $\mathcal{B}_\Lambda = \sigma(\tau_\Lambda)$, and the product σ – algebra, $\otimes_{\alpha \in \Lambda} \mathcal{B}_\alpha$, are the same for any $\Lambda \subset A$. By Tychonoff's Theorem 36.16 and Proposition 37.4, X and X_Λ for any $\Lambda \subset A$ are still compact Hausdorff spaces which are second countable if Λ is finite. By the Stone Weierstrass Theorem 37.31,

$$\mathcal{D} := \{f \in C(X) : f = F \circ \pi_\Lambda \text{ with } F \in C(X_\Lambda) \text{ and } \Lambda \subset\subset A\}$$

is a dense subspace of $C(X)$. For $f = F \circ \pi_\Lambda \in \mathcal{D}$, let

$$I(f) = \int_{X_\Lambda} F \circ \pi_\Lambda(x) d\mu_\Lambda(x). \tag{38.34}$$

Let us verify that I is well defined. Suppose that f may also be expressed as $f = F' \circ \pi_{\Lambda'}$ with $\Lambda' \subset\subset A$ and $F' \in C(X_{\Lambda'})$. Let $\Gamma := \Lambda \cup \Lambda'$ and define $G \in C(X_\Gamma)$ by $G := F \circ \pi_{\Lambda,\Gamma}$. Hence, using Exercise **??**,

$$\int_{X_\Gamma} G d\mu_\Gamma = \int_{X_\Gamma} F \circ \pi_{\Lambda,\Gamma}\, d\mu_\Gamma = \int_{X_\Lambda} F\, d\left[(\pi_{\Lambda,\Gamma})_* \mu_\Gamma\right] = \int_{X_\Lambda} F\, d\mu_\Lambda$$

wherein we have used the compatibility condition in the last equality. Similarly, using $G = F' \circ \pi_{\Lambda',\Gamma}$ (as the reader should verify), one shows

$$\int_{X_\Gamma} G d\mu_\Gamma = \int_{X_{\Lambda'}} F'\, d\mu_{\Lambda'}.$$

Therefore

$$\int_{X_{\Lambda'}} F'\, d\mu_{\Lambda'} = \int_{X_\Gamma} G d\mu_\Gamma = \int_{X_\Lambda} F\, d\mu_\Lambda,$$

which shows I in Eq. (38.34) is well defined.

Since $|I(f)| \le \|f\|_\infty$, the B.L.T. Theorem 32.4 allows us to extend I from the dense subspace, $\mathcal{D}$, to a continuous linear functional, $\bar{I}$, on $C(X)$. Because I was positive on $\mathcal{D}$, it is easy to check that $\bar{I}$ is still positive on $C(X)$. So by the Riesz-Markov Theorem 38.9, there exists a Radon measure on $\mathcal{B} = \mathcal{M}$ such that $\bar{I}(f) = \int_X f d\mu$ for all $f \in C(X)$. By the definition of $\bar{I}$ in now follows that

$$\int_{X^\Lambda} F d(\pi_\Lambda)_* \mu = \int_{X^\Lambda} F \circ \pi_\Lambda d\mu = \bar{I}(F \circ \pi_\Lambda) = \int_{X^\Lambda} F d\mu_\Lambda$$

for all $F \in C(X^\Lambda)$ and $\Lambda \subset\subset A$. Since X_Λ is a second countable locally compact Hausdorff space, this identity implies, see Theorem 31.8[1], that $(\pi_\Lambda)_* \mu = \mu_\Lambda$. The uniqueness assertion of the theorem follows from the fact that the measure μ is determined uniquely by its values on the algebra $\mathcal{A} := \cup_{\Lambda \subset\subset A} \pi_\Lambda^{-1}(\mathcal{B}_{X_\Lambda})$ which generates $\mathcal{B} = \mathcal{M}$, see Theorem **??**. ■

Exercise 38.4. Let (Y, τ) be a locally compact Hausdorff space and $(Y^* = Y \cup \{\infty\}, \tau^*)$ be the one point compactification of Y. Then

[1] Alternatively, use Theorems 38.16 and the uniquness assertion in Markov-Riesz Theorem 38.9 to conclude $(\pi_\Lambda)_* \mu = \mu_\Lambda$.

$$\mathcal{B}_{Y^*} := \sigma(\tau^*) = \{A \subset Y^* : A \cap Y \in \mathcal{B}_Y = \sigma(\tau)\}$$

or equivalently put

$$\mathcal{B}_{Y^*} = \mathcal{B}_Y \cup \{A \cup \{\infty\} : A \in \mathcal{B}_Y\}.$$

Also shows that $(Y^* = Y \cup \{\infty\}, \tau^*)$ is second countable if (Y, τ) was second countable.

Proof. Proof of Theorem 38.31.

Case 1; A is a countable. Let $(X_\alpha^* = X_\alpha \cup \{\infty_\alpha\}, \tau_\alpha^*)$ be the one point compactification of (X_α, τ_α). For $\Lambda \subset A$, let $X_\Lambda^* := \prod_{\alpha \in \Lambda} X_\alpha^*$ equipped with the product topology and Borel $\sigma -$ algebra, $\mathcal{B}_\Lambda^*$. Since Λ is at most countable, the set,

$$X_\Lambda := \bigcap_{\alpha \in A} \{\pi_\alpha = \infty_\alpha\},$$

is a measurable subset of X_Λ^*. Therefore for each $\Lambda \subset\subset A$, we may extend μ_Λ to a measure, $\bar{\mu}_\Lambda$, on $(X_\Lambda^*, \mathcal{B}_\Lambda^*)$ using the formula,

$$\bar{\mu}_\Lambda(B) = \mu_\Lambda(B \cap X_\Lambda) \text{ for all } B \in X_\Lambda^*.$$

An application of Theorem 38.32 shows there exists a unique probability measure, $\bar{\mu}$, on $X^* := X_A^*$ such that $(\pi_\Lambda)_* \bar{\mu} = \bar{\mu}_\Lambda$ for all $\Lambda \subset\subset A$. Since

$$X^* \setminus X = \bigcup_{\alpha \in A} \{\pi_\alpha = \infty_\alpha\}$$

and $\bar{\mu}(\{\pi_\alpha = \infty\}) = \bar{\mu}_{\{\alpha\}}(\{\infty_\alpha\}) = 0$, it follows that $\bar{\mu}(X^* \setminus X) = 0$. Hence $\mu := \bar{\mu}|_{\mathcal{B}_X}$ is a probability measure on $(X, \mathcal{B}_X)$. Finally if $B \in \mathcal{B}_X \subset \mathcal{B}_{X^*}$,

$$\begin{aligned}\mu_\Lambda(B) &= \bar{\mu}_\Lambda(B) = (\pi_\Lambda)_* \bar{\mu}(B) = \bar{\mu}\left(\pi_\Lambda^{-1}(B)\right) \\ &= \bar{\mu}\left(\pi_\Lambda^{-1}(B) \cap X\right) = \mu\left(\pi_\Lambda|_X^{-1}(B)\right)\end{aligned}$$

which shows μ is the required probability measure on $\mathcal{B}_X$.

Case 2; A is uncountable. By case 1. for each countable or finite subset $\Gamma \subset A$ there is a measure μ_Γ on $(X_\Gamma, \mathcal{B}_\Gamma)$ such that $(\pi_{\Lambda,\Gamma})_* \mu_\Gamma = \mu_\Lambda$ for all $\Lambda \subset\subset \Gamma$. By Exercise 11.9,

$$\mathcal{M} = \bigcup \left\{\pi_\Gamma^{-1}(\mathcal{B}_\Gamma) : \Gamma \text{ is a countable subset of } A\right\},$$

i.e. every $B \in \mathcal{M}$ may be written in the form $B = \pi_\Gamma^{-1}(C)$ for some countable subset, $\Gamma \subset A$, and $C \in \mathcal{B}_\Gamma$. For such a B we define $\mu(B) := \mu_\Gamma(C)$. It is left to the reader to check that μ is well defined and that μ is a measure on $\mathcal{M}$.

(Keep in mind the countable union of countable sets is countable.) If $\Lambda \subset\subset A$ and $C \in \mathcal{B}_\Lambda$, then

$$[(\pi_\Lambda)_* \mu](C) = \mu\left(\pi_\Lambda^{-1}(C)\right) := \mu_\Lambda(C),$$

i.e. $(\pi_\Lambda)_* \mu = \mu_\Lambda$ as desired. ∎

Corollary 38.33. *Suppose that $\{\mu_\alpha\}_{\alpha\in A}$ are probability measure on $(X_\alpha, \mathcal{B}_\alpha)$ for all $\alpha \in A$ and if $\Lambda \subset\subset A$ let $\mu_\Lambda := \otimes_{\alpha\in\Lambda}\mu_\alpha$ be the product measure on $(X_\Lambda, \mathcal{B}_\Lambda = \otimes_{\alpha\in\Lambda}\mathcal{B}_\alpha)$. Then there exists a unique probability measure, μ, on $(X, \mathcal{M})$ such that $(\pi_\Lambda)_* \mu = \mu_\Lambda$ for all $\Lambda \subset\subset A$. (It is possible remove the topology from this corollary, see Theorem* **??** *below.)*

Exercise 38.5. Prove Corollary 38.33 by showing the measures $\mu_\Lambda := \otimes_{\alpha\in\Lambda}\mu_\alpha$ satisfy the compatibility condition in Theorem 38.31.

*** Beginning of WORK material. ***

Lemma 38.34 (Is this true? I am not so sure.). *Suppose $\Lambda \subset A$ and μ is Radon measure on (X, τ). Then $(\pi_\Lambda)_* \mu := \mu \circ \pi_\Lambda^{-1}$ is a Radon measure on $(X_\Lambda, \tau_\Lambda)$.*

Proof. Let $Y := X_\Lambda$, $Z := X_{\Lambda^c}$ and $\pi : Y \times Z \to Y$ be the canonical projection map. We equip $Y \times Z$ with the product topology. The mapping $\varphi = (\pi_\Lambda, \pi_{\Lambda^c}) : X \to Y \times Z$ is easily seen to be continuous and bijective and therefore a homeomorphism by Proposition 37.6. Because of this observation it suffices to prove; if $\nu := \varphi_*\mu$ is a Radon probability measure on $Y \times Z$ then $\pi_*\nu$ is a Radon probability measure on Y.

Outer regularity. Suppose that $B \in \mathcal{B}_Y$ and U is an open subset in $X \times Y$ such that $B \times Z = \pi^{-1}(B) \subset U$. Letting then

$$\inf\{\pi_*\nu(U) : B \subset U \in \tau_Y\} = \inf\left\{\nu\left(\pi^{-1}(U)\right) : B \subset U \in \tau_Y\right\}$$

and U is an open subset of Y is outer regular on all Borel sets and inner regular on all open subsets of Y. ∎

*** End of WORK material. ***

Probability Measures on Lusin Spaces

Definition 39.1 (Lusin spaces). *A **Lusin space** is a topological space (X, τ) which is homeomorphic to a Borel subset of a compact metric space.*

Example 39.2. By Theorem 37.12, every Polish (i.e. complete separable metric space) is a Lusin space. Moreover, any Borel subset of Lusin space is again a Lusin space.

Definition 39.3. *Two measurable spaces, $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are said to be **isomorphic** if there exists a bijective map, $f : X \to Y$ such that $f(\mathcal{M}) = \mathcal{N}$ and $f^{-1}(\mathcal{N}) = \mathcal{M}$, i.e. both f and f^{-1} are measurable.*

39.1 Weak Convergence Results

The following is an application of theorem 35.60 characterizing compact sets in metric spaces. (BRUCE: add Helly's selection principle here.)

Proposition 39.4. *Suppose that (X, ρ) is a complete separable metric space and μ is a probability measure on $\mathcal{B} = \sigma(\tau_\rho)$. Then for all $\varepsilon > 0$, there exists $K_\varepsilon \sqsubset\sqsubset X$ such that $\mu(K_\varepsilon) \geq 1 - \varepsilon$.*

Proof. Let $\{x_k\}_{k=1}^\infty$ be a countable dense subset of X. Then $X = \cup_k C_{x_k}(1/n)$ for all $n \in \mathbb{N}$. Hence by continuity of μ, there exists, for all $n \in \mathbb{N}$, $N_n < \infty$ such that $\mu(F_n) \geq 1 - \varepsilon 2^{-n}$ where $F_n := \cup_{k=1}^{N_n} C_{x_k}(1/n)$. Let $K := \cap_{n=1}^\infty F_n$ then

$$\begin{aligned}\mu(X \setminus K) &= \mu(\cup_{n=1}^\infty F_n^c) \\ &\leq \sum_{n=1}^\infty \mu(F_n^c) = \sum_{n=1}^\infty (1 - \mu(F_n)) \leq \sum_{n=1}^\infty \varepsilon 2^{-n} = \varepsilon\end{aligned}$$

so that $\mu(K) \geq 1 - \varepsilon$. Moreover K is compact since K is closed and totally bounded; $K \subset F_n$ for all n and each F_n is $1/n$ – bounded. ∎

Definition 39.5. *A sequence of probability measures $\{P_n\}_{n=1}^\infty$ is said to converge to a probability P if for every $f \in BC(X)$, $P_n(f) \to P(f)$. This is actually weak-* convergence when viewing $P_n \in BC(X)^*$.*

Proposition 39.6. *The following are equivalent:*

1. *$P_n \overset{w}{\to} P$ as $n \to \infty$*
2. *$P_n(f) \to P(f)$ for every $f \in BC(X)$ which is uniformly continuous.*
3. *$\limsup_{n\to\infty} P_n(F) \leq P(F)$ for all $F \sqsubset X$.*
4. *$\liminf_{n\to\infty} P_n(G) \geq P(G)$ for all $G \subset_o X$.*
5. *$\lim_{n\to\infty} P_n(A) = P(A)$ for all $A \in \mathcal{B}$ such that $P(\mathrm{bd}(A)) = 0$.*

Proof. 1. $\Longrightarrow$ 2. is obvious. For 2. $\Longrightarrow$ 3.,

$$\varphi(t) := \begin{cases} 1 & \text{if } t \leq 0 \\ 1 - t & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if } t \geq 1 \end{cases} \tag{39.1}$$

and let $f_n(x) := \varphi(nd(x, F))$. Then $f_n \in BC(X, [0,1])$ is uniformly continuous, $0 \leq 1_F \leq f_n$ for all n and $f_n \downarrow 1_F$ as $n \to \infty$. Passing to the limit $n \to \infty$ in the equation

$$0 \leq P_n(F) \leq P_n(f_m)$$

gives

$$0 \leq \limsup_{n\to\infty} P_n(F) \leq P(f_m)$$

and then letting $m \to \infty$ in this inequality implies item 3. 3. $\Longleftrightarrow$ 4. Assuming item 3., let $F = G^c$, then

$$\begin{aligned}1 - \liminf_{n\to\infty} P_n(G) &= \limsup_{n\to\infty} (1 - P_n(G)) = \limsup_{n\to\infty} P_n(G^c) \\ &\leq P(G^c) = 1 - P(G)\end{aligned}$$

which implies 4. Similarly 4. $\Longrightarrow$ 3. 3. $\Longleftrightarrow$ 5. Recall that $\mathrm{bd}(A) = \bar{A} \setminus A^o$, so if $P(\mathrm{bd}(A)) = 0$ and 3. (and hence also 4. holds) we have

$$\begin{aligned}&\limsup_{n\to\infty} P_n(A) \leq \limsup_{n\to\infty} P_n(\bar{A}) \leq P(\bar{A}) = P(A) \text{ and} \\ &\liminf_{n\to\infty} P_n(A) \geq \liminf_{n\to\infty} P_n(A^o) \geq P(A^o) = P(A)\end{aligned}$$

from which it follows that $\lim_{n\to\infty} P_n(A) = P(A)$. Conversely, let $F \sqsubset X$ and set $F_\delta := \{x \in X : \rho(x, F) \leq \delta\}$. Then

$$\mathrm{bd}(F_\delta) \subset F_\delta \setminus \{x \in X : \rho(x, F) < \delta\} = A_\delta$$

where $A_\delta := \{x \in X : \rho(x, F) = \delta\}$. Since $\{A_\delta\}_{\delta>0}$ are all disjoint, we must have

$$\sum_{\delta>0} P(A_\delta) \le P(X) \le 1$$

and in particular the set $\Lambda := \{\delta > 0 : P(A_\delta) > 0\}$ is at most countable. Let $\delta_n \notin \Lambda$ be chosen so that $\delta_n \downarrow 0$ as $n \to \infty$, then

$$P(F_{\delta_m}) = \lim_{n\to\infty} P_n(F_{\delta_n}) \ge \limsup_{n\to\infty} P_n(F).$$

Let $m \to \infty$ this equation to conclude $P(F) \ge \limsup_{n\to\infty} P_n(F)$ as desired. To finish the proof we will now show 3. $\Longrightarrow$ 1. By an affine change of variables it suffices to consider $f \in C(X, (0,1))$ in which case we have

$$\sum_{i=1}^{k} \frac{(i-1)}{k} 1_{\left\{\frac{(i-1)}{k} \le f < \frac{i}{k}\right\}} \le f \le \sum_{i=1}^{k} \frac{i}{k} 1_{\left\{\frac{(i-1)}{k} \le f < \frac{i}{k}\right\}}. \tag{39.2}$$

Let $F_i := \left\{\frac{i}{k} \le f\right\}$ and notice that $F_k = \emptyset$, then we for any probability P that

$$\sum_{i=1}^{k} \frac{(i-1)}{k} \left[P(F_{i-1}) - P(F_i)\right] \le P(f) \le \sum_{i=1}^{k} \frac{i}{k} \left[P(F_{i-1}) - P(F_i)\right]. \tag{39.3}$$

Now

$$\begin{aligned}
&\sum_{i=1}^{k} \frac{(i-1)}{k} \left[P(F_{i-1}) - P(F_i)\right] \\
&\quad = \sum_{i=1}^{k} \frac{(i-1)}{k} P(F_{i-1}) - \sum_{i=1}^{k} \frac{(i-1)}{k} P(F_i) \\
&\quad = \sum_{i=1}^{k-1} \frac{i}{k} P(F_i) - \sum_{i=1}^{k} \frac{i-1}{k} P(F_i) = \frac{1}{k} \sum_{i=1}^{k-1} P(F_i)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{k} \frac{i}{k} \left[P(F_{i-1}) - P(F_i)\right] \\
&\quad = \sum_{i=1}^{k} \frac{i-1}{k} \left[P(F_{i-1}) - P(F_i)\right] + \sum_{i=1}^{k} \frac{1}{k} \left[P(F_{i-1}) - P(F_i)\right] \\
&\quad = \sum_{i=1}^{k-1} P(F_i) + \frac{1}{k}
\end{aligned}$$

so that Eq. (39.3) becomes,

$$\frac{1}{k} \sum_{i=1}^{k-1} P(F_i) \le P(f) \le \frac{1}{k} \sum_{i=1}^{k-1} P(F_i) + 1/k.$$

Using this equation with $P = P_n$ and then with $P = P$ we find

$$\begin{aligned}
\limsup_{n\to\infty} P_n(f) &\le \limsup_{n\to\infty} \left[\frac{1}{k} \sum_{i=1}^{k-1} P_n(F_i) + 1/k\right] \\
&\le \frac{1}{k} \sum_{i=1}^{k-1} P(F_i) + 1/k \le P(f) + 1/k. \\
&\le
\end{aligned}$$

Since k is arbitrary,

$$\limsup_{n\to\infty} P_n(f) \le P(f).$$

This inequality also hold for $1 - f$ and this implies $\liminf_{n\to\infty} P_n(f) \ge P(f)$ and hence $\lim_{n\to\infty} P_n(f) = P(f)$ as claimed. ■

Definition 39.7. *Let X be a topological space. A collection of probability measures Λ on $(X, \mathcal{B}_X)$ is said to be **tight** if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \in \mathcal{B}_X$ such that $P(K_\varepsilon) \ge 1 - \varepsilon$ for all $P \in \Lambda$.*

Theorem 39.8. *Suppose X is a separable metrizable space and $\Lambda = \{P_n\}_{n=1}^\infty$ is a tight sequence of probability measures on $\mathcal{B}_X$. Then there exists a subsequence $\{P_{n_k}\}_{k=1}^\infty$ which is weakly convergent to a probability measure P on $\mathcal{B}_X$.*

Proof. First suppose that X is compact. In this case $C(X)$ is a Banach space which is separable by the Stone – Weirstrass theorem, see Exercise 37.5. By the Riesz theorem, Corollary **??**, we know that $C(X)^*$ is in one to one correspondence with complex measure on $(X, \mathcal{B}_X)$. We have also seen that $C(X)^*$ is metrizable and the unit ball in $C(X)^*$ is weak - * compact, see Theorem 36.25. Hence there exists a subsequence $\{P_{n_k}\}_{k=1}^\infty$ which is weak -* convergent to a probability measure P on X. Alternatively, use the Cantor's diagonalization procedure on a countable dense set $\Gamma \subset C(X)$ so find $\{P_{n_k}\}_{k=1}^\infty$ such that $\Lambda(f) := \lim_{k\to\infty} P_{n_k}(f)$ exists for all $f \in \Gamma$. Then for $g \in C(X)$ and $f \in \Gamma$, we have

$$\begin{aligned}
|P_{n_k}(g) - P_{n_l}(g)| &\le |P_{n_k}(g) - P_{n_k}(f)| + |P_{n_k}(f) - P_{n_l}(f)| \\
&\quad + |P_{n_l}(f) - P_{n_l}(g)| \\
&\le 2\,\|g - f\|_\infty + |P_{n_k}(f) - P_{n_l}(f)|
\end{aligned}$$

which shows

$$\limsup_{k,l\to\infty} |P_{n_k}(g) - P_{n_l}(g)| \le 2\|g-f\|_\infty .$$

Letting $f \in \Lambda$ tend to g in $C(X)$ shows $\limsup_{k,l\to\infty} |P_{n_k}(g) - P_{n_l}(g)| = 0$ and hence $\Lambda(g) := \lim_{k\to\infty} P_{n_k}(g)$ for all $g \in C(X)$. It is now clear that $\Lambda(g) \ge 0$ for all $g \ge 0$ so that Λ is a positive linear functional on X and thus there is a probability measure P such that $\Lambda(g) = P(g)$.

General case. By Theorem 37.12 we may assume that X is a subset of a compact metric space which we will denote by $\bar{X}$. We now extend P_n to $\bar{X}$ by setting $\bar{P}_n(A) := \bar{P}_n(A \cap \bar{X})$ for all $A \in \mathcal{B}_{\bar{X}}$. By what we have just proved, there is a subsequence $\left\{\bar{P}'_k := \bar{P}_{n_k}\right\}_{k=1}^\infty$ such that $\bar{P}'_k$ converges weakly to a probability measure $\bar{P}$ on $\bar{X}$. The main thing we now have to prove is that "$\bar{P}(X) = 1$," this is where the tightness assumption is going to be used. Given $\varepsilon > 0$, let $K_\varepsilon \subset X$ be a compact set such that $\bar{P}_n(K_\varepsilon) \ge 1-\varepsilon$ for all n. Since K_ε is compact in X it is compact in $\bar{X}$ as well and in particular a closed subset of $\bar{X}$. Therefore by Proposition 39.6

$$\bar{P}(K_\varepsilon) \ge \limsup_{k\to\infty} \bar{P}_k^{'}(K_\varepsilon) = 1-\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows with $X_0 := \cup_{n=1}^\infty K_{1/n}$ satisfies $\bar{P}(X_0) = 1$. Because $X_0 \in \mathcal{B}_X \cap \mathcal{B}_{\bar{X}}$, we may view $\bar{P}$ as a measure on $\mathcal{B}_X$ by letting $P(A) := \bar{P}(A \cap X_0)$ for all $A \in \mathcal{B}_X$. Given a closed subset $F \subset X$, choose $\tilde{F} \sqsubset \bar{X}$ such that $F = \tilde{F} \cap X$. Then

$$\limsup_{k\to\infty} P'_k(F) = \limsup_{k\to\infty} \bar{P}'_k(\tilde{F}) \le \bar{P}(\tilde{F}) = \bar{P}(\tilde{F} \cap X_0) = P(F),$$

which shows $P'_k \overset{w}{\to} P$. ■

39.2 Haar Measures

To be written.

39.3 Hausdorff Measure

To be written.

39.4 Exercises

Exercise 39.1. Let $E \in \mathcal{B}_\mathbb{R}$ with $m(E) > 0$. Then for any $\alpha \in (0,1)$ there exists a bounded open interval $J \subset \mathbb{R}$ such that $m(E \cap J) \ge \alpha m(J)$.[1] **Hints:** 1. Reduce to the case where $m(E) \in (0,\infty)$. 2) Approximate E from the outside by an open set $V \subset \mathbb{R}$. 3. Make use of Exercise 35.22, which states that V may be written as a disjoint union of open intervals.

Exercise 39.2. Let $F : \mathbb{R} \to \mathbb{R}$ be a right continuous increasing function and $\mu = \mu_F$ be as in Theorem 38.26. For $a < b$, find the values of $\mu(\{a\})$, $\mu([a,b))$, $\mu([a,b])$ and $\mu((a,b))$ in terms of the function F.

Exercise 39.3. Suppose that $F \in C^1(\mathbb{R})$ is an increasing function and μ_F is the unique Borel measure on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a \le b$. Show that $d\mu_F = \rho dm$ for some function $\rho \ge 0$. Find ρ explicitly in terms of F.

Exercise 39.4. Suppose that $F(x) = e1_{x\ge 3} + \pi 1_{x \ge 7}$ and μ_F is the is the unique Borel measure on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a \le b$. Give an explicit description of the measure μ_F.

Exercise 39.5. Let (X,τ) be a locally compact Hausdorff space and $I : C_0(X,\mathbb{R}) \to \mathbb{R}$ be a positive linear functional. Show I is necessarily bounded, i.e. there exists a $C < \infty$ such that $|I(f)| \le C\|f\|_\infty$ for all $f \in C_0(X,\mathbb{R})$.

Outline. (BRUCE: see Nate's proof below and then rewrite this outline to make the problem much easier and to handle more general circumstances.)

1. By the Riesz-Markov Theorem 38.9, there exists a unique Radon measure μ on $(X,\mathcal{B}_X)$ such that $\mu(f) := \int_X f d\mu = I(f)$ for all $f \in C_c(X,\mathbb{R})$. Show $\mu(f) \le I(f)$ for all $f \in C_0(X,[0,\infty))$.
2. Show that if $\mu(X) = \infty$, then there exists a function $f \in C_0(X,[0,\infty))$ such that $\infty = \mu(f) \le I(f)$ contradicting the assumption that I is real valued.
3. By item 2., $\mu(X) < \infty$ and therefore $C_0(X,\mathbb{R}) \subset L^1(\mu)$ and $\mu : C_0(X,\mathbb{R}) \to \mathbb{R}$ is a well defined positive linear functional. Let $J(f) := I(f) - \mu(f)$, which by item 1. is a positive linear functional such that $J|_{C_c(X,\mathbb{R})} \equiv 0$. Show that any positive linear functional, J, on $C_0(X,\mathbb{R})$ satisfying these properties must necessarily be zero. Thus $I(f) = \mu(f)$ and $\|I\| = \mu(X) < \infty$ as claimed.

Exercise 39.6. BRUCE (Drop this exercise or move somewhere else, it is already proved in the notes in more general terms.) Suppose that $I : C_c^\infty(\mathbb{R},\mathbb{R}) \to \mathbb{R}$ is a positive linear functional. Show

[1] See also the Lebesgue differentiation Theorem 25.14 from which one may prove the much stronger form of this theorem, namely for m -a.e. $x \in E$ there exits $r_\alpha(x) > 0$ such that $m(E \cap (x-r, x+r)) \ge \alpha m((x-r,x+r))$ for all $r \le r_\alpha(x)$.

1. For each compact subset $K \sqsubset\sqsubset \mathbb{R}$ there exists a constant $C_K < \infty$ such that
$$|I(f)| \leq C_K \|f\|_\infty$$
whenever $\text{supp}(f) \subset K$.
2. Show there exists a unique Radon measure μ on $\mathcal{B}_\mathbb{R}$ (the Borel σ – algebra on $\mathbb{R}$) such that $I(f) = \int_\mathbb{R} f d\mu$ for all $f \in C_c^\infty(\mathbb{R}, \mathbb{R})$.

39.4.1 The Laws of Large Number Exercises

For the rest of the problems of this section, let ν be a probability measure on $\mathcal{B}_\mathbb{R}$ such that
$$\int_\mathbb{R} |x| \, d\nu(x) < \infty.$$
By Corollary 38.33, there exists a unique measure μ on $(X := \mathbb{R}^\mathbb{N}, \mathcal{B} := \mathcal{B}_{\mathbb{R}^\mathbb{N}} = \otimes_{n=1}^\infty \mathcal{B}_\mathbb{R})$ such that
$$\int_X f(x_1, x_2, \dots, x_N) d\mu(x) = \int_{\mathbb{R}^N} f(x_1, x_2, \dots, x_N) d\nu(x_1) \dots d\nu(x_N) \quad (39.4)$$
for all $N \in \mathbb{N}$ and bounded measurable functions $f : \mathbb{R}^N \to \mathbb{R}$, i.e. $\mu = \otimes_{n=1}^\infty \mu_n$ with $\mu_n = \nu$ for every n. We will also use the following notation:
$$\begin{aligned} S_n(x) &:= \frac{1}{n} \sum_{k=1}^n x_k \text{ for } x \in X, \\ m &:= \int_\mathbb{R} x d\nu(x) \\ \sigma^2 &:= \int_\mathbb{R} (x-m)^2 d\nu(x) = \int_\mathbb{R} x^2 d\nu(x) - m^2, \text{ and} \\ \gamma &:= \int_\mathbb{R} (x-m)^4 d\nu(x). \end{aligned}$$
As is customary, m is said to be the mean or average of ν and σ^2 is the variance of ν.

Exercise 39.7 (Weak Law of Large Numbers). Assume $\sigma^2 < \infty$. Show $\int_X S_n d\mu = m$.
$$\|S_n - m\|_2^2 = \int_X (S_n - m)^2 \, d\mu = \frac{\sigma^2}{n},$$
and $\mu(|S_n - m| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$ for all $\varepsilon > 0$ and $n \in \mathbb{N}$.

Exercise 39.8 (A simple form of the Strong Law of Large Numbers). Suppose now that $\gamma := \int_\mathbb{R} (x-m)^4 d\nu(x) < \infty$. Show for all $\varepsilon > 0$ and $n \in \mathbb{N}$ that
$$\begin{aligned} \|S_n - m\|_4^4 = \int_X (S_n - m)^4 \, d\mu &= \frac{1}{n^4}\left(n\gamma + 3n(n-1)\sigma^4\right) \\ &= \frac{1}{n^2}\left[n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4\right] \text{ and} \\ \mu(|S_n - m| > \varepsilon) &\leq \frac{n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4}{\varepsilon^4 n^2}. \end{aligned}$$
Conclude from the last estimate and the first Borel Cantelli Lemma **??** that $\lim_{n\to\infty} S_n(x) = m$ for μ – a.e. $x \in X$.

Exercise 39.9. Suppose $\gamma := \int_\mathbb{R} (x-m)^4 d\nu(x) < \infty$ and $m = \int_\mathbb{R} (x-m) d\nu(x) \neq 0$. For $\lambda > 0$ let $T_\lambda : \mathbb{R}^\mathbb{N} \to \mathbb{R}^\mathbb{N}$ be defined by $T_\lambda(x) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n, \dots)$, $\mu_\lambda = \mu \circ T_\lambda^{-1}$ and
$$X_\lambda := \left\{ x \in \mathbb{R}^\mathbb{N} : \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n x_j = \lambda \right\}.$$
Show
$$\mu_\lambda(X_{\lambda'}) = \delta_{\lambda,\lambda'} = \begin{cases} 1 \text{ if } \lambda = \lambda' \\ 0 \text{ if } \lambda \neq \lambda' \end{cases}$$
and use this to show if $\lambda \neq 1$, then $d\mu_\lambda \neq \rho d\mu$ for any measurable function $\rho : \mathbb{R}^\mathbb{N} \to [0, \infty]$.

Part IX

Further Hilbert and Banach Space Techniques

40

$\mathbf{L^2}$ - Hilbert Spaces Techniques and Fourier Series

This section is concerned with Hilbert spaces presented as in the following example.

Example 40.1. Let $(X, \mathcal{M}, \mu)$ be a measure space. Then $H := L^2(X, \mathcal{M}, \mu)$ with inner product

$$\langle f|g\rangle = \int_X f \cdot \bar{g} d\mu$$

is a Hilbert space.

It will be convenient to define

$$\langle f, g\rangle := \int_X f(x)\, g(x)\, d\mu(x) \tag{40.1}$$

for all measurable functions f, g on X such that $fg \in L^1(\mu)$. So with this notation we have $\langle f|g\rangle = \langle f, \bar{g}\rangle$ for all $f, g \in H$.

40.1 Fourier Series Considerations

Throughout this section we will let $d\theta$, dx, $d\alpha$, etc. denote Lebesgue measure on $\mathbb{R}^d$ normalized so that the cube, $Q := (-\pi, \pi]^d$, has measure one, i.e. $d\theta = (2\pi)^{-d} dm(\theta)$ where m is standard Lebesgue measure on $\mathbb{R}^d$. As usual, for $\alpha \in \mathbb{N}_0^d$, let

$$D_\theta^\alpha = \left(\frac{1}{i}\right)^{|\alpha|} \frac{\partial^{|\alpha|}}{\partial\theta_1^{\alpha_1} \dots \partial\theta_d^{\alpha_d}}.$$

Notation 40.2 *Let $C_{per}^k(\mathbb{R}^d)$ denote the 2π – periodic functions in $C^k(\mathbb{R}^d)$, that is $f \in C_{per}^k(\mathbb{R}^d)$ iff $f \in C^k(\mathbb{R}^d)$ and $f(\theta + 2\pi e_i) = f(\theta)$ for all $\theta \in \mathbb{R}^d$ and $i = 1, 2, \dots, d$. Further let $\langle\cdot|\cdot\rangle$ denote the inner product on the Hilbert space, $H := L^2([-\pi, \pi]^d)$, given by*

$$\langle f|g\rangle := \int_Q f(\theta)\bar{g}(\theta) d\theta = \left(\frac{1}{2\pi}\right)^d \int_Q f(\theta)\bar{g}(\theta) dm(\theta)$$

*and define $\varphi_k(\theta) := e^{ik\cdot\theta}$ for all $k \in \mathbb{Z}^d$. For $f \in L^1(Q)$, we will write $\tilde{f}(k)$ for the **Fourier coefficient**,*

$$\tilde{f}(k) := \langle f|\varphi_k\rangle = \int_Q f(\theta) e^{-ik\cdot\theta} d\theta. \tag{40.2}$$

Since any 2π – periodic functions on $\mathbb{R}^d$ may be identified with function on the d - dimensional torus, $\mathbb{T}^d \cong \mathbb{R}^d/(2\pi\mathbb{Z})^d \cong \left(S^1\right)^d$, I may also write $C^k(\mathbb{T}^d)$ for $C_{per}^k(\mathbb{R}^d)$ and $L^p(\mathbb{T}^d)$ for $L^p(Q)$ where elements in $f \in L^p(Q)$ are to be thought of as there extensions to 2π – periodic functions on $\mathbb{R}^d$. The following theorem is a repeat of Theorem 18.38 above.

Theorem 40.3 (Fourier Series). *The functions $\beta := \left\{\varphi_k : k \in \mathbb{Z}^d\right\}$ form an orthonormal basis for H, i.e. if $f \in H$ then*

$$f = \sum_{k \in \mathbb{Z}^d} \langle f|\varphi_k\rangle \varphi_k = \sum_{k \in \mathbb{Z}^d} \tilde{f}(k)\varphi_k \tag{40.3}$$

where the convergence takes place in $L^2([-\pi, \pi]^d)$.

Proof. Simple computations show $\beta := \left\{\varphi_k : k \in \mathbb{Z}^d\right\}$ is an orthonormal set. We now claim that β is an orthonormal basis. To see this recall that $C_c((-\pi, \pi)^d)$ is dense in $L^2((-\pi, \pi)^d, dm)$. Any $f \in C_c((-\pi, \pi)^d)$ may be extended to be a continuous 2π – periodic function on $\mathbb{R}$ and hence by Exercise 7.15 (see also Theorem 7.42, Exercise 37.13 and Remark 37.46), f may uniformly (and hence in L^2) be approximated by a trigonometric polynomial. Therefore β is a total orthonormal set, i.e. β is an orthonormal basis. ■

Exercise 40.1. Let A be the operator defined in Lemma 20.7 and for $g \in L^2(\mathbb{T})$, let $Ug(k) := \tilde{g}(k)$ so that $U : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ is unitary. Show $U^{-1}AU = M_a$ where $a \in C_{per}^\infty(\mathbb{R})$ is a function to be found. Use this representation and the results in Exercise 20.2 to give a simple proof of the results in Lemma 20.7.

40.1.1 Dirichlet, Fejér and Kernels

Although the sum in Eq. (40.3) is guaranteed to converge relative to the Hilbertian norm on H it certainly need not converge pointwise even if $f \in C_{per}(\mathbb{R}^d)$ as will be proved in Section 23.3.1 below. Nevertheless, if f is sufficiently regular, then the sum in Eq. (40.3) will converge pointwise as we will now show. In the process we will give a direct and constructive proof of the result in Exercise 37.13, see Theorem 40.5 below.

Let us restrict our attention to $d = 1$ here. Consider

$$\begin{aligned} f_n(\theta) &= \sum_{|k|\le n} \tilde{f}(k)\varphi_k(\theta) = \sum_{|k|\le n} \frac{1}{2\pi}\left[\int_{[-\pi,\pi]} f(x)e^{-ik\cdot x}dx\right]\varphi_k(\theta) \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)\sum_{|k|\le n} e^{ik\cdot(\theta-x)}dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)D_n(\theta-x)dx \end{aligned} \tag{40.4}$$

where

$$D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta}$$

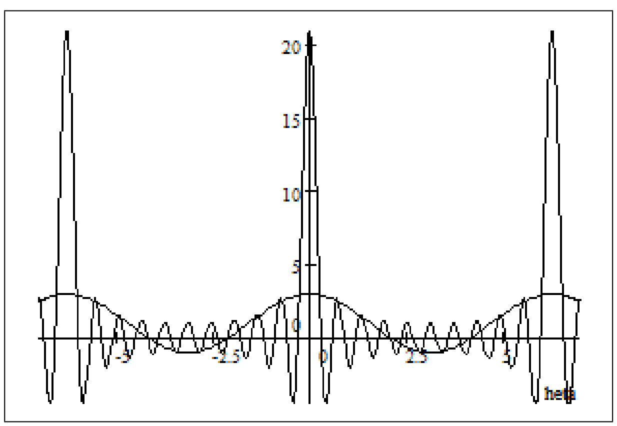
is called the **Dirichlet kernel**. Letting $\alpha = e^{i\theta/2}$, we have

$$\begin{aligned} D_n(\theta) &= \sum_{k=-n}^{n} \alpha^{2k} = \frac{\alpha^{2(n+1)} - \alpha^{-2n}}{\alpha^2 - 1} = \frac{\alpha^{2n+1} - \alpha^{-(2n+1)}}{\alpha - \alpha^{-1}} \\ &= \frac{2i\sin(n+\frac{1}{2})\theta}{2i\sin\frac{1}{2}\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}. \end{aligned}$$

and therefore

$$D_n(\theta) := \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}, \tag{40.5}$$

see Figure 40.1.1.



This is a plot D_1 and D_{10}.

with the understanding that the right side of this equation is $2n + 1$ whenever $\theta \in 2\pi\mathbb{Z}$.

Theorem 40.4. *Suppose $f \in L^1([-\pi,\pi], dm)$ and f is differentiable at some $\theta \in [-\pi,\pi]$, then $\lim_{n\to\infty} f_n(\theta) = f(\theta)$ where f_n is as in Eq. (40.4).*

Proof. Observe that

$$\frac{1}{2\pi}\int_{[-\pi,\pi]} D_n(\theta-x)dx = \frac{1}{2\pi}\int_{[-\pi,\pi]}\sum_{|k|\le n} e^{ik\cdot(\theta-x)}dx = 1$$

and therefore,

$$\begin{aligned} f_n(\theta) - f(\theta) &= \frac{1}{2\pi}\int_{[-\pi,\pi]} [f(x) - f(\theta)]\, D_n(\theta-x)dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} [f(x) - f(\theta-x)]\, D_n(x)dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} \left[\frac{f(\theta-x) - f(\theta)}{\sin\frac{1}{2}x}\right]\sin(n+\frac{1}{2})x\, dx. \end{aligned} \tag{40.6}$$

If f is differentiable at θ, the last expression in Eq. (40.6) tends to 0 as $n \to \infty$ by the Riemann Lebesgue Lemma (Corollary 31.17 or Lemma 31.40) and the fact that $1_{[-\pi,\pi]}(x)\frac{f(\theta-x)-f(\theta)}{\sin\frac{1}{2}x} \in L^1(dx)$. ■

Despite the Dirichlet kernel not being positive, it still satisfies the approximate δ – sequence property, $\frac{1}{2\pi}D_n \to \delta_0$ as $n \to \infty$, when acting on C^1 – periodic functions in θ. In order to improve the convergence properties it is reasonable to try to replace $\{f_n : n \in \mathbb{N}_0\}$ by the sequence of averages (see Exercise 14.16),

$$\begin{aligned} F_N(\theta) &= \frac{1}{N+1}\sum_{n=0}^{N} f_n(\theta) = \frac{1}{N+1}\sum_{n=0}^{N}\frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)\sum_{|k|\le n} e^{ik\cdot(\theta-x)}dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} K_N(\theta-x)f(x)dx \end{aligned}$$

where

$$K_N(\theta) := \frac{1}{N+1}\sum_{n=0}^{N}\sum_{|k|\le n} e^{ik\cdot\theta} \tag{40.7}$$

is the **Fejér kernel.**

Theorem 40.5. *The Fejér kernel K_N in Eq. (40.7) satisfies:*

1.

$$K_N(\theta) = \sum_{n=-N}^{N} \left[1 - \frac{|n|}{N+1}\right] e^{in\theta} \tag{40.8}$$
$$= \frac{1}{N+1} \frac{\sin^2\left(\frac{N+1}{2}\theta\right)}{\sin^2\left(\frac{\theta}{2}\right)}. \tag{40.9}$$

2. $K_N(\theta) \geq 0$.
3. $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(\theta)d\theta = 1$
4. $\sup_{\varepsilon\leq|\theta|\leq\pi} K_N(\theta) \to 0$ *as* $N \to \infty$ *for all* $\varepsilon > 0$*, see Figure 40.1.*
5. For any continuous 2π *– periodic function* f *on* $\mathbb{R}$, $K_N * f(\theta) \to f(\theta)$ *uniformly in* θ *as* $N \to \infty$*, where*

$$K_N * f(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(\theta - \alpha) f(\alpha) d\alpha$$
$$= \sum_{n=-N}^{N} \left[1 - \frac{|n|}{N+1}\right] \tilde{f}(n) e^{in\theta}. \tag{40.10}$$

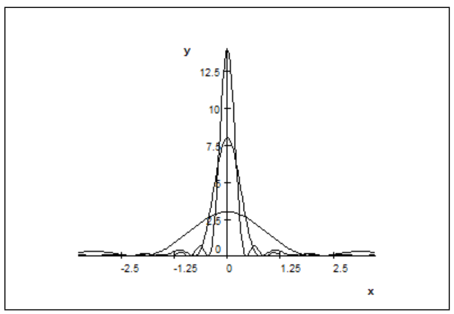


Fig. 40.1. Plots of $K_N(\theta)$ for $N = 2$, 7 and 13.

Proof. 1. Equation (40.8) is a consequence of the identity,

$$\sum_{n=0}^{N}\sum_{|k|\leq n} e^{ik\cdot\theta} = \sum_{|k|\leq n\leq N} e^{ik\cdot\theta} = \sum_{|k|\leq N}(N+1-|k|)\, e^{ik\cdot\theta}.$$

Moreover, letting $\alpha = e^{i\theta/2}$ and using Eq. (3.3) shows

$$\begin{aligned}
K_N(\theta) &= \frac{1}{N+1}\sum_{n=0}^{N}\sum_{|k|\leq n}\alpha^{2k} = \frac{1}{N+1}\sum_{n=0}^{N}\frac{\alpha^{2n+2}-\alpha^{-2n}}{\alpha^2-1} \\
&= \frac{1}{(N+1)(\alpha-\alpha^{-1})}\sum_{n=0}^{N}\left[\alpha^{2n+1}-\alpha^{-2n-1}\right] \\
&= \frac{1}{(N+1)(\alpha-\alpha^{-1})}\sum_{n=0}^{N}\left[\alpha\alpha^{2n}-\alpha^{-1}\alpha^{-2n}\right] \\
&= \frac{1}{(N+1)(\alpha-\alpha^{-1})}\left[\alpha\frac{\alpha^{2N+2}-1}{\alpha^2-1}-\alpha^{-1}\frac{\alpha^{-2N-2}-1}{\alpha^{-2}-1}\right] \\
&= \frac{1}{(N+1)(\alpha-\alpha^{-1})^2}\left[\alpha^{2(N+1)}-1+\alpha^{-2(N+1)}-1\right] \\
&= \frac{1}{(N+1)(\alpha-\alpha^{-1})^2}\left[\alpha^{(N+1)}-\alpha^{-(N+1)}\right]^2 \\
&= \frac{1}{N+1}\frac{\sin^2((N+1)\theta/2)}{\sin^2(\theta/2)}.
\end{aligned}$$

Items 2. and 3. follow easily from Eqs. (40.9) and (40.8) respectively. Item 4. is a consequence of the elementary estimate;

$$\sup_{\varepsilon\leq|\theta|\leq\pi} K_N(\theta) \leq \frac{1}{N+1}\frac{1}{\sin^2\left(\frac{\varepsilon}{2}\right)}$$

and is clearly indicated in Figure 40.1. Item 5. now follows by the standard approximate δ – function arguments, namely,

$$\begin{aligned}
|K_N * f(\theta) - f(\theta)| &= \frac{1}{2\pi}\left|\int_{-\pi}^{\pi} K_N(\theta-\alpha)\left[f(\alpha)-f(\theta)\right]d\alpha\right| \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(\alpha)\left|f(\theta-\alpha)-f(\theta)\right|d\alpha \\
&\leq \frac{1}{\pi}\frac{1}{N+1}\frac{1}{\sin^2\left(\frac{\varepsilon}{2}\right)}\|f\|_\infty + \frac{1}{2\pi}\int_{|\alpha|\leq\varepsilon} K_N(\alpha)\left|f(\theta-\alpha)-f(\theta)\right|d\alpha \\
&\leq \frac{1}{\pi}\frac{1}{N+1}\frac{1}{\sin^2\left(\frac{\varepsilon}{2}\right)}\|f\|_\infty + \sup_{|\alpha|\leq\varepsilon}\left|f(\theta-\alpha)-f(\theta)\right|.
\end{aligned}$$

Therefore,

$$\limsup_{N\to\infty}\|K_N * f - f\|_\infty \leq \sup_\theta \sup_{|\alpha|\leq\varepsilon}\left|f(\theta-\alpha)-f(\theta)\right| \to 0 \text{ as } \varepsilon \downarrow 0.$$

■

40.1.2 The Dirichlet Problems on D and the Poisson Kernel

Let $D := \{z \in \mathbb{C} : |z| < 1\}$ be the open unit disk in $\mathbb{C} \cong \mathbb{R}^2$, write $z \in \mathbb{C}$ as $z = x + iy$ or $z = re^{i\theta}$, and let $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ be the **Laplacian** acting on $C^2(D)$.

Theorem 40.6 (Dirichlet problem for D). *To every continuous function $g \in C(\mathrm{bd}(D))$ there exists a unique function $u \in C(\bar{D}) \cap C^2(D)$ solving*

$$\Delta u(z) = 0 \text{ for } z \in D \text{ and } u|_{\partial D} = g. \tag{40.11}$$

Moreover for $r < 1$, u is given by,

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta - \alpha)u(e^{i\alpha})d\alpha =: P_r * u(e^{i\theta}) \tag{40.12}$$

$$= \frac{1}{2\pi}\operatorname{Re}\int_{-\pi}^{\pi} \frac{1 + re^{i(\theta-\alpha)}}{1 - re^{i(\theta-\alpha)}}u(e^{i\alpha})d\alpha \tag{40.13}$$

where P_r is the ***Poisson kernel*** *defined by*

$$P_r(\delta) := \frac{1 - r^2}{1 - 2r\cos\delta + r^2}.$$

(The problem posed in Eq. (40.11) is called the ***Dirichlet problem for*** *D.)*

Proof. In this proof, we are going to be identifying $S^1 = \mathrm{bd}(D) := \{z \in \bar{D} : |z| = 1\}$ with $[-\pi, \pi]/(\pi \sim -\pi)$ by the map $\theta \in [-\pi, \pi] \to e^{i\theta} \in S^1$. Also recall that the Laplacian Δ may be expressed in polar coordinates as,

$$\Delta u = r^{-1}\partial_r\left(r^{-1}\partial_r u\right) + \frac{1}{r^2}\partial_\theta^2 u,$$

where

$$(\partial_r u)\left(re^{i\theta}\right) = \frac{\partial}{\partial r}u\left(re^{i\theta}\right) \text{ and } (\partial_\theta u)\left(re^{i\theta}\right) = \frac{\partial}{\partial \theta}u\left(re^{i\theta}\right).$$

Uniqueness. Suppose u is a solution to Eq. (40.11) and let

$$\tilde{g}(k) := \frac{1}{2\pi}\int_{-\pi}^{\pi} g(e^{ik\theta})e^{-ik\theta}d\theta$$

and

$$\tilde{u}(r,k) := \frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})e^{-ik\theta}d\theta \tag{40.14}$$

be the Fourier coefficients of $g(\theta)$ and $\theta \to u\left(re^{i\theta}\right)$ respectively. Then for $r \in (0,1)$,

$$\begin{aligned} r^{-1}\partial_r\left(r\partial_r\tilde{u}(r,k)\right) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} r^{-1}\partial_r\left(r^{-1}\partial_r u\right)(re^{i\theta})e^{-ik\theta}d\theta \\ &= -\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{r^2}\partial_\theta^2 u(re^{i\theta})e^{-ik\theta}d\theta \\ &= -\frac{1}{r^2}\frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})\partial_\theta^2 e^{-ik\theta}d\theta \\ &= \frac{1}{r^2}k^2\tilde{u}(r,k) \end{aligned}$$

or equivalently

$$r\partial_r\left(r\partial_r\tilde{u}(r,k)\right) = k^2\tilde{u}(r,k). \tag{40.15}$$

Recall the general solution to

$$r\partial_r\left(r\partial_r y(r)\right) = k^2 y(r) \tag{40.16}$$

may be found by trying solutions of the form $y(r) = r^\alpha$ which then implies $\alpha^2 = k^2$ or $\alpha = \pm k$. From this one sees that $\tilde{u}(r,k)$ solving Eq. (40.15) may be written as $\tilde{u}(r,k) = A_k r^{|k|} + B_k r^{-|k|}$ for some constants A_k and B_k when $k \neq 0$. If $k = 0$, the solution to Eq. (40.16) is gotten by simple integration and the result is $\tilde{u}(r,0) = A_0 + B_0 \ln r$. Since $\tilde{u}(r,k)$ is bounded near the origin for each k it must be that $B_k = 0$ for all $k \in \mathbb{Z}$. Hence we have shown there exists $A_k \in \mathbb{C}$ such that, for all $r \in (0,1)$,

$$A_k r^{|k|} = \tilde{u}(r,k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(re^{i\theta})e^{-ik\theta}d\theta. \tag{40.17}$$

Since all terms of this equation are continuous for $r \in [0,1]$, Eq. (40.17) remains valid for all $r \in [0,1]$ and in particular we have, at $r = 1$, that

$$A_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(e^{i\theta})e^{-ik\theta}d\theta = \tilde{g}(k).$$

Hence if u is a solution to Eq. (40.11) then u must be given by

$$u(re^{i\theta}) = \sum_{k\in\mathbb{Z}}\tilde{g}(k)r^{|k|}e^{ik\theta} \text{ for } r < 1. \tag{40.18}$$

or equivalently,

$$u(z) = \sum_{k\in\mathbb{N}_0}\tilde{g}(k)z^k + \sum_{k\in\mathbb{N}}\tilde{g}(-k)\bar{z}^k.$$

Notice that the theory of the Fourier series implies Eq. (40.18) is valid in the $L^2(d\theta)$ - sense. However more is true, since for $r < 1$, the series in Eq. (40.18) is absolutely convergent and in fact defines a C^∞ – function (see Exercise 4.12 or

Corollary 10.30) which must agree with the continuous function, $\theta \to u\left(re^{i\theta}\right)$, for almost every θ and hence for all θ. This completes the proof of uniqueness.

Existence. Given $g \in C(\mathrm{bd}(D))$, let u be defined as in Eq. (40.18). Then, again by Exercise 4.12 or Corollary 10.30, $u \in C^\infty(D)$. So to finish the proof it suffices to show $\lim_{x\to y} u(x) = g(y)$ for all $y \in \mathrm{bd}(D)$. Inserting the formula for $\tilde{g}(k)$ into Eq. (40.18) gives

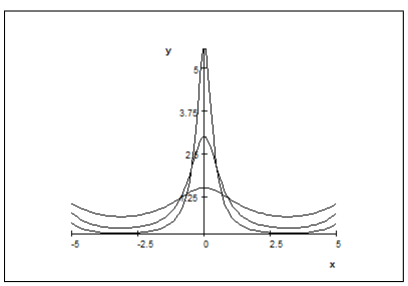
$$u(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta-\alpha)\, u(e^{i\alpha})d\alpha \text{ for all } r<1$$

where

$$\begin{aligned} P_r(\delta) &= \sum_{k\in\mathbb{Z}} r^{|k|}e^{ik\delta} = \sum_{k=0}^{\infty} r^k e^{ik\delta} + \sum_{k=0}^{\infty} r^k e^{-ik\delta} - 1 = \\ &= \mathrm{Re}\left[2\frac{1}{1-re^{i\delta}} - 1\right] = \mathrm{Re}\left[\frac{1+re^{i\delta}}{1-re^{i\delta}}\right] \\ &= \mathrm{Re}\left[\frac{\left(1+re^{i\delta}\right)\left(1-re^{-i\delta}\right)}{|1-re^{i\delta}|^2}\right] = \mathrm{Re}\left[\frac{1-r^2+2ir\sin\delta}{1-2r\cos\delta+r^2}\right] \qquad (40.19) \\ &= \frac{1-r^2}{1-2r\cos\delta+r^2}. \end{aligned}$$

The Poisson kernel again solves the usual approximate δ – function properties (see Figure 2), namely:

1. $P_r(\delta) > 0$ and
$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta-\alpha)\, d\alpha &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k\in\mathbb{Z}} r^{|k|}e^{ik(\theta-\alpha)}d\alpha \\ &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} r^{|k|}\int_{-\pi}^{\pi} e^{ik(\theta-\alpha)}d\alpha = 1 \end{aligned}$$
and
2.
$$\sup_{\varepsilon\le|\theta|\le\pi} P_r(\theta) \le \frac{1-r^2}{1-2r\cos\varepsilon+r^2} \to 0 \text{ as } r\uparrow 1.$$



A plot of $P_r(\delta)$ for $r = 0.2$, 0.5 and 0.7.

Therefore by the same argument used in the proof of Theorem 40.5,

$$\lim_{r\uparrow 1}\sup_\theta\left|u\left(re^{i\theta}\right) - g\left(e^{i\theta}\right)\right| = \lim_{r\uparrow 1}\sup_\theta\left|\left(P_r * g\right)\left(e^{i\theta}\right) - g\left(e^{i\theta}\right)\right| = 0$$

which certainly implies $\lim_{x\to y} u(x) = g(y)$ for all $y \in \mathrm{bd}(D)$. ■

Remark 40.7 (Harmonic Conjugate). Writing $z = re^{i\theta}$, Eq. (40.13) may be rewritten as

$$u(z) = \frac{1}{2\pi}\mathrm{Re}\int_{-\pi}^{\pi}\frac{1+ze^{-i\alpha}}{1-ze^{-i\alpha}}u(e^{i\alpha})d\alpha$$

which shows $u = \mathrm{Re}\, F$ where

$$F(z) := \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1+ze^{-i\alpha}}{1-ze^{-i\alpha}}u(e^{i\alpha})d\alpha.$$

Moreover it follows from Eq. (40.19) that

$$\begin{aligned} \mathrm{Im}\, F(re^{i\theta}) &= \frac{1}{\pi}\mathrm{Im}\int_{-\pi}^{\pi}\frac{r\sin(\theta-\alpha)}{1-2r\cos(\theta-\alpha)+r^2}g(e^{i\alpha})d\alpha \\ &=: \left(Q_r * u\right)\left(e^{i\theta}\right) \end{aligned}$$

where

$$Q_r(\delta) := \frac{r\sin(\delta)}{1-2r\cos(\delta)+r^2}.$$

From these remarks it follows that $v =: \left(Q_r * g\right)\left(e^{i\theta}\right)$ is the harmonic conjugate of u and $\tilde{P}_r = Q_r$. For more on this point see Section 26.7 below.

40.2 Weak L^2-Derivatives

Theorem 40.8 (Weak and Strong Differentiability). *Suppose that $f \in L^2(\mathbb{R}^n)$ and $v \in \mathbb{R}^n \setminus \{0\}$. Then the following are equivalent:*

1. *There exists $\{t_n\}_{n=1}^\infty \subset \mathbb{R}\setminus\{0\}$ such that $\lim_{n\to\infty} t_n = 0$ and*
$$\sup_n\left\|\frac{f(\cdot+t_nv)-f(\cdot)}{t_n}\right\|_2 < \infty.$$
2. *There exists $g \in L^2(\mathbb{R}^n)$ such that $\langle f, \partial_v\varphi\rangle = -\langle g, \varphi\rangle$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$.*
3. *There exists $g \in L^2(\mathbb{R}^n)$ and $f_n \in C_c^\infty(\mathbb{R}^n)$ such that $f_n \overset{L^2}{\to} f$ and $\partial_v f_n \overset{L^2}{\to} g$ as $n \to \infty$.*

4. There exists $g \in L^2$ such that

$$\frac{f(\cdot + tv) - f(\cdot)}{t} \overset{L^2}{\to} g \text{ as } t \to 0.$$

(See Theorem 41.18 for the L^p generalization of this theorem.)

Proof. 1. $\Longrightarrow$ 2. We may assume, using Theorem 36.30 and passing to a subsequence if necessary, that $\frac{f(\cdot+t_n v)-f(\cdot)}{t_n} \overset{w}{\to} g$ for some $g \in L^2(\mathbb{R}^n)$. Now for $\varphi \in C_c^\infty(\mathbb{R}^n)$,

$$\begin{aligned}\langle g|\varphi\rangle &= \lim_{n\to\infty}\left\langle \frac{f(\cdot + t_n v) - f(\cdot)}{t_n}, \varphi\right\rangle = \lim_{n\to\infty}\left\langle f, \frac{\varphi(\cdot - t_n v) - \varphi(\cdot)}{t_n}\right\rangle \\ &= \left\langle f, \lim_{n\to\infty}\frac{\varphi(\cdot - t_n v) - \varphi(\cdot)}{t_n}\right\rangle = -\langle f, \partial_v \varphi\rangle,\end{aligned}$$

wherein we have used the translation invariance of Lebesgue measure and the dominated convergence theorem. 2. $\Longrightarrow$ 3. Let $\varphi \in C_c^\infty(\mathbb{R}^n, \mathbb{R})$ such that $\int_{\mathbb{R}^n} \varphi(x)dx = 1$ and let $\varphi_m(x) = m^n \varphi(mx)$, then by Proposition 31.37, $h_m := \varphi_m * f \in C^\infty(\mathbb{R}^n)$ for all m and

$$\begin{aligned}\partial_v h_m(x) = \partial_v \varphi_m * f(x) &= \int_{\mathbb{R}^n} \partial_v \varphi_m(x-y) f(y)dy = \langle f, -\partial_v [\varphi_m(x - \cdot)]\rangle \\ &= \langle g, \varphi_m(x-\cdot)\rangle = \varphi_m * g(x).\end{aligned}$$

By Theorem 31.33, $h_m \to f \in L^2(\mathbb{R}^n)$ and $\partial_v h_m = \varphi_m * g \to g$ in $L^2(\mathbb{R}^n)$ as $m \to \infty$. This shows 3. holds except for the fact that h_m need not have compact support. To fix this let $\psi \in C_c^\infty(\mathbb{R}^n, [0,1])$ such that $\psi = 1$ in a neighborhood of 0 and let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ and $(\partial_v \psi)_\varepsilon(x) := (\partial_v \psi)(\varepsilon x)$. Then

$$\partial_v(\psi_\varepsilon h_m) = \partial_v \psi_\varepsilon h_m + \psi_\varepsilon \partial_v h_m = \varepsilon (\partial_v \psi)_\varepsilon h_m + \psi_\varepsilon \partial_v h_m$$

so that $\psi_\varepsilon h_m \to h_m$ in L^2 and $\partial_v(\psi_\varepsilon h_m) \to \partial_v h_m$ in L^2 as $\varepsilon \downarrow 0$. Let $f_m = \psi_{\varepsilon_m} h_m$ where ε_m is chosen to be greater than zero but small enough so that

$$\|\psi_{\varepsilon_m} h_m - h_m\|_2 + \|\partial_v(\psi_{\varepsilon_m} h_m) \to \partial_v h_m\|_2 < 1/m.$$

Then $f_m \in C_c^\infty(\mathbb{R}^n)$, $f_m \to f$ and $\partial_v f_m \to g$ in L^2 as $m \to \infty$. 3. $\Longrightarrow$ 4. By the fundamental theorem of calculus

$$\begin{aligned}\frac{\tau_{-tv} f_m(x) - f_m(x)}{t} &= \frac{f_m(x+tv) - f_m(x)}{t} \\ &= \frac{1}{t}\int_0^1 \frac{d}{ds} f_m(x+stv)ds = \int_0^1 (\partial_v f_m)(x+stv)ds.\end{aligned} \tag{40.20}$$

Let

$$G_t(x) := \int_0^1 \tau_{-stv} g(x)ds = \int_0^1 g(x+stv)ds$$

which is defined for almost every x and is in $L^2(\mathbb{R}^n)$ by Minkowski's inequality for integrals, Theorem 29.2. Therefore

$$\frac{\tau_{-tv} f_m(x) - f_m(x)}{t} - G_t(x) = \int_0^1 [(\partial_v f_m)(x+stv) - g(x+stv)]\, ds$$

and hence again by Minkowski's inequality for integrals,

$$\begin{aligned}\left\|\frac{\tau_{-tv} f_m - f_m}{t} - G_t\right\|_2 &\le \int_0^1 \|\tau_{-stv}(\partial_v f_m) - \tau_{-stv} g\|_2\, ds \\ &= \int_0^1 \|\partial_v f_m - g\|_2\, ds.\end{aligned}$$

Letting $m \to \infty$ in this equation implies $(\tau_{-tv} f - f)/t = G_t$ a.e. Finally one more application of Minkowski's inequality for integrals implies,

$$\begin{aligned}\left\|\frac{\tau_{-tv} f - f}{t} - g\right\|_2 = \|G_t - g\|_2 &= \left\|\int_0^1 (\tau_{-stv} g - g)\, ds\right\|_2 \\ &\le \int_0^1 \|\tau_{-stv} g - g\|_2\, ds.\end{aligned}$$

By the dominated convergence theorem and Proposition 31.25, the latter term tends to 0 as $t \to 0$ and this proves 4. The proof is now complete since 4. $\Longrightarrow$ 1. is trivial. ∎

40.3 *Conditional Expectation

In this section let $(\Omega, \mathcal{F}, P)$ be a probability space, i.e. $(\Omega, \mathcal{F}, P)$ is a measure space and $P(\Omega) = 1$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub – sigma algebra of $\mathcal{F}$ and write $f \in \mathcal{G}_b$ if $f : \Omega \to \mathbb{C}$ is bounded and f is $(\mathcal{G}, \mathcal{B}_\mathbb{C})$ – measurable. In this section we will write

$$Ef := \int_\Omega f dP.$$

Definition 40.9 (Conditional Expectation). *Let $E_\mathcal{G} : L^2(\Omega, \mathcal{F}, P) \to L^2(\Omega, \mathcal{G}, P)$ denote orthogonal projection of $L^2(\Omega, \mathcal{F}, P)$ onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$. For $f \in L^2(\Omega, \mathcal{G}, P)$, we say that $E_\mathcal{G} f \in L^2(\Omega, \mathcal{F}, P)$ is the* ***conditional expectation*** *of f.*

Theorem 40.10. *Let $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ be as above and $f, g \in L^2(\Omega, \mathcal{F}, P)$.*

1. *If* $f\ge 0$, P – *a.e. then* $E_{\mathcal{G}}f\ge 0$, P – *a.e.*
2. *If* $f\ge g$, P – *a.e. there* $E_{\mathcal{G}}f\ge E_{\mathcal{G}}g$, P – *a.e.*
3. $|E_{\mathcal{G}}f|\le E_{\mathcal{G}}|f|$, P – *a.e.*
4. $\|E_{\mathcal{G}}f\|_{L^1}\le\|f\|_{L^1}$ *for all* $f\in L^2$. *So by the B.L.T. Theorem 32.4,* $E_{\mathcal{G}}$ *extends uniquely to a bounded linear map from* $L^1(\Omega,\mathcal{F},P)$ *to* $L^1(\Omega,\mathcal{G},P)$ *which we will still denote by* $E_{\mathcal{G}}$.
5. *If* $f\in L^1(\Omega,\mathcal{F},P)$ *then* $F=E_{\mathcal{G}}f\in L^1(\Omega,\mathcal{G},P)$ *iff*
$$E(Fh)=E(fh)\text{ for all }h\in\mathcal{G}_b.$$
6. *If* $g\in\mathcal{G}_b$ *and* $f\in L^1(\Omega,\mathcal{F},P)$, *then* $E_{\mathcal{G}}(gf)=g\cdot E_{\mathcal{G}}f$, P – *a.e.*

Proof. By the definition of orthogonal projection for $h\in\mathcal{G}_b$,

$$E(fh)=E(f\cdot E_{\mathcal{G}}h)=E(E_{\mathcal{G}}f\cdot h).$$

So if $f,h\ge 0$ then $0\le E(fh)\le E(E_{\mathcal{G}}f\cdot h)$ and since this holds for all $h\ge 0$ in $\mathcal{G}_b$, $E_{\mathcal{G}}f\ge 0$, P – a.e. This proves (1). Item (2) follows by applying item (1). to $f-g$. If f is real, $\pm f\le|f|$ and so by Item (2), $\pm E_{\mathcal{G}}f\le E_{\mathcal{G}}|f|$, i.e. $|E_{\mathcal{G}}f|\le E_{\mathcal{G}}|f|$, P – a.e. For complex f, let $h\ge 0$ be a bounded and $\mathcal{G}$ – measurable function. Then

$$\begin{aligned}E\left[|E_{\mathcal{G}}f|\,h\right]&=E\left[E_{\mathcal{G}}f\cdot\overline{\operatorname{sgn}(E_{\mathcal{G}}f)}h\right]=E\left[f\cdot\overline{\operatorname{sgn}(E_{\mathcal{G}}f)}h\right]\\&\le E\left[|f|\,h\right]=E\left[E_{\mathcal{G}}|f|\cdot h\right].\end{aligned}$$

Since h is arbitrary, it follows that $|E_{\mathcal{G}}f|\le E_{\mathcal{G}}|f|$, P – a.e. Integrating this inequality implies

$$\|E_{\mathcal{G}}f\|_{L^1}\le E\,|E_{\mathcal{G}}f|\le E\left[E_{\mathcal{G}}|f|\cdot 1\right]=E\left[|f|\right]=\|f\|_{L^1}.$$

Item (5). Suppose $f\in L^1(\Omega,\mathcal{F},P)$ and $h\in\mathcal{G}_b$. Let $f_n\in L^2(\Omega,\mathcal{F},P)$ be a sequence of functions such that $f_n\to f$ in $L^1(\Omega,\mathcal{F},P)$. Then

$$\begin{aligned}E(E_{\mathcal{G}}f\cdot h)&=E(\lim_{n\to\infty}E_{\mathcal{G}}f_n\cdot h)=\lim_{n\to\infty}E(E_{\mathcal{G}}f_n\cdot h)\\&=\lim_{n\to\infty}E(f_n\cdot h)=E(f\cdot h).\end{aligned}\tag{40.21}$$

This equation uniquely determines $E_{\mathcal{G}}$, for if $F\in L^1(\Omega,\mathcal{G},P)$ also satisfies $E(F\cdot h)=E(f\cdot h)$ for all $h\in\mathcal{G}_b$, then taking $h=\overline{\operatorname{sgn}(F-E_{\mathcal{G}}f)}$ in Eq. (40.21) gives

$$0=E((F-E_{\mathcal{G}}f)\,h)=E(|F-E_{\mathcal{G}}f|).$$

This shows $F=E_{\mathcal{G}}f$, P – a.e. Item (6) is now an easy consequence of this characterization, since if $h\in\mathcal{G}_b$,

$$E\left[(gE_{\mathcal{G}}f)\,h\right]=E\left[E_{\mathcal{G}}f\cdot hg\right]=E\left[f\cdot hg\right]=E\left[gf\cdot h\right]=E\left[E_{\mathcal{G}}(gf)\cdot h\right].$$

Thus $E_{\mathcal{G}}(gf)=g\cdot E_{\mathcal{G}}f$, P – a.e. ∎

Proposition 40.11. *If* $\mathcal{G}_0\subset\mathcal{G}_1\subset\mathcal{F}$. *Then*

$$E_{\mathcal{G}_0}E_{\mathcal{G}_1}=E_{\mathcal{G}_1}E_{\mathcal{G}_0}=E_{\mathcal{G}_0}.\tag{40.22}$$

Proof. Equation (40.22) holds on $L^2(\Omega,\mathcal{F},P)$ by the basic properties of orthogonal projections. It then hold on $L^1(\Omega,\mathcal{F},P)$ by continuity and the density of $L^2(\Omega,\mathcal{F},P)$ in $L^1(\Omega,\mathcal{F},P)$. ∎

Example 40.12. Suppose that $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are two σ – finite measure spaces. Let $\Omega=X\times Y$, $\mathcal{F}=\mathcal{M}\otimes\mathcal{N}$ and $P(dx,dy)=\rho(x,y)\mu(dx)\nu(dy)$ where $\rho\in L^1(\Omega,\mathcal{F},\mu\otimes\nu)$ is a positive function such that $\int_{X\times Y}\rho d(\mu\otimes\nu)=1$. Let $\pi_X:\Omega\to X$ be the projection map, $\pi_X(x,y)=x$, and

$$\mathcal{G}:=\sigma(\pi_X)=\pi_X^{-1}(\mathcal{M})=\{A\times Y:A\in\mathcal{M}\}.$$

Then $f:\Omega\to\mathbb{R}$ is $\mathcal{G}$ – measurable iff $f=F\circ\pi_X$ for some function $F:X\to\mathbb{R}$ which is $\mathcal{N}$ – measurable, see Lemma **??**. For $f\in L^1(\Omega,\mathcal{F},P)$, we will now show $E_{\mathcal{G}}f=F\circ\pi_X$ where

$$F(x)=\frac{1}{\bar\rho(x)}1_{(0,\infty)}(\bar\rho(x))\cdot\int_Y f(x,y)\rho(x,y)\nu(dy),$$

$\bar\rho(x):=\int_Y\rho(x,y)\nu(dy)$. (By convention, $\int_Y f(x,y)\rho(x,y)\nu(dy):=0$ if $\int_Y|f(x,y)|\,\rho(x,y)\nu(dy)=\infty$.)

By Tonelli's theorem, the set

$$E:=\{x\in X:\bar\rho(x)=\infty\}\cup\left\{x\in X:\int_Y|f(x,y)|\,\rho(x,y)\nu(dy)=\infty\right\}$$

is a μ – null set. Since

$$\begin{aligned}E\left[|F\circ\pi_X|\right]&=\int_X d\mu(x)\int_Y d\nu(y)\,|F(x)|\,\rho(x,y)=\int_X d\mu(x)\,|F(x)|\,\bar\rho(x)\\&=\int_X d\mu(x)\left|\int_Y\nu(dy)f(x,y)\rho(x,y)\right|\\&\le\int_X d\mu(x)\int_Y\nu(dy)\,|f(x,y)|\,\rho(x,y)<\infty,\end{aligned}$$

$F\circ\pi_X\in L^1(\Omega,\mathcal{G},P)$. Let $h=H\circ\pi_X$ be a bounded $\mathcal{G}$ – measurable function, then

$$\begin{aligned}E\left[F\circ\pi_X\cdot h\right]&=\int_X d\mu(x)\int_Y d\nu(y)F(x)H(x)\rho(x,y)\\&=\int_X d\mu(x)F(x)H(x)\bar\rho(x)\\&=\int_X d\mu(x)H(x)\int_Y\nu(dy)f(x,y)\rho(x,y)\\&=E\left[hf\right]\end{aligned}$$

and hence $E_{\mathcal{G}}f = F \circ \pi_X$ as claimed.

This example shows that conditional expectation is a generalization of the notion of performing integration over a partial subset of the variables in the integrand. Whereas to compute the expectation, one should integrate over all of the variables. See also Exercise 40.8 to gain more intuition about conditional expectations.

Theorem 40.13 (Jensen's inequality). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Assume $f \in L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ is a function such that (for simplicity) $\varphi(f) \in L^1(\Omega, \mathcal{F}, P; \mathbb{R})$, then $\varphi(E_{\mathcal{G}}f) \leq E_{\mathcal{G}}[\varphi(f)]$, P – a.e.*

Proof. Let us first assume that φ is C^1 and f is bounded. In this case

$$\varphi(x) - \varphi(x_0) \geq \varphi'(x_0)(x - x_0) \text{ for all } x_0, x \in \mathbb{R}. \tag{40.23}$$

Taking $x_0 = E_{\mathcal{G}}f$ and $x = f$ in this inequality implies

$$\varphi(f) - \varphi(E_{\mathcal{G}}f) \geq \varphi'(E_{\mathcal{G}}f)(f - E_{\mathcal{G}}f)$$

and then applying $E_{\mathcal{G}}$ to this inequality gives

$$\begin{aligned} E_{\mathcal{G}}[\varphi(f)] - \varphi(E_{\mathcal{G}}f) &= E_{\mathcal{G}}[\varphi(f) - \varphi(E_{\mathcal{G}}f)] \\ &\geq \varphi'(E_{\mathcal{G}}f)(E_{\mathcal{G}}f - E_{\mathcal{G}}E_{\mathcal{G}}f) = 0 \end{aligned}$$

The same proof works for general φ, one need only use Proposition **??** to replace Eq. (40.23) by

$$\varphi(x) - \varphi(x_0) \geq \varphi'_-(x_0)(x - x_0) \text{ for all } x_0, x \in \mathbb{R}$$

where $\varphi'_-(x_0)$ is the left hand derivative of φ at x_0. If f is not bounded, apply what we have just proved to $f^M = f 1_{|f| \leq M}$, to find

$$E_{\mathcal{G}}\left[\varphi(f^M)\right] \geq \varphi(E_{\mathcal{G}}f^M). \tag{40.24}$$

Since $E_{\mathcal{G}} : L^1(\Omega, \mathcal{F}, P; \mathbb{R}) \to L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ is a bounded operator and $f^M \to f$ and $\varphi(f^M) \to \varphi(f)$ in $L^1(\Omega, \mathcal{F}, P; \mathbb{R})$ as $M \to \infty$, there exists $\{M_k\}_{k=1}^\infty$ such that $M_k \uparrow \infty$ and $f^{M_k} \to f$ and $\varphi(f^{M_k}) \to \varphi(f)$, P – a.e. So passing to the limit in Eq. (40.24) shows $E_{\mathcal{G}}[\varphi(f)] \geq \varphi(E_{\mathcal{G}}f)$, P – a.e. ■

40.4 Exercises

Exercise 40.2. Let $(X, \mathcal{M}, \mu)$ be a measure space and $H := L^2(X, \mathcal{M}, \mu)$. Given $f \in L^\infty(\mu)$ let $M_f : H \to H$ be the multiplication operator defined by $M_f g = fg$. Show $M_f^2 = M_f$ iff there exists $A \in \mathcal{M}$ such that $f = 1_A$ a.e.

Exercise 40.3. Let $O(n)$ be the orthogonal groups consisting of $n \times n$ real orthogonal matrices O, i.e. $O^{tr}O = I$. For $O \in O(n)$ and $f \in L^2(\mathbb{R}^n)$ let $U_O f(x) = f(O^{-1}x)$. Show

1. $U_O f$ is well defined, namely if $f = g$ a.e. then $U_O f = U_O g$ a.e.
2. $U_O : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is unitary and satisfies $U_{O_1}U_{O_2} = U_{O_1 O_2}$ for all $O_1, O_2 \in O(n)$. That is to say the map $O \in O(n) \to \mathcal{U}(L^2(\mathbb{R}^n))$ – the unitary operators on $L^2(\mathbb{R}^n)$ is a group homomorphism, i.e. a "unitary representation" of $O(n)$.
3. For each $f \in L^2(\mathbb{R}^n)$, the map $O \in O(n) \to U_O f \in L^2(\mathbb{R}^n)$ is continuous. Take the topology on $O(n)$ to be that inherited from the Euclidean topology on the vector space of all $n \times n$ matrices. **Hint**: see the proof of Proposition 31.25.

Exercise 40.4. Euclidean group representation and its infinitesimal generators including momentum and angular momentum operators.

Exercise 40.5. Spherical Harmonics.

Exercise 40.6. The gradient and the Laplacian in spherical coordinates.

Exercise 40.7. Legendre polynomials.

40.5 Conditional Expectation Exercises

Exercise 40.8. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{A} := \{A_i\}_{i=1}^\infty \subset \mathcal{F}$ is a partition of Ω. (Recall this means $\Omega = \coprod_{i=1}^\infty A_i$.) Let $\mathcal{G}$ be the σ – algebra generated by $\mathcal{A}$. Show:

1. $B \in \mathcal{G}$ iff $B = \cup_{i \in \Lambda} A_i$ for some $\Lambda \subset \mathbb{N}$.
2. $g : \Omega \to \mathbb{R}$ is $\mathcal{G}$ – measurable iff $g = \sum_{i=1}^\infty \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$.
3. For $f \in L^1(\Omega, \mathcal{F}, P)$, let $E(f|A_i) := E[1_{A_i} f] / P(A_i)$ if $P(A_i) \neq 0$ and $E(f|A_i) = 0$ otherwise. Show

$$E_{\mathcal{G}}f = \sum_{i=1}^\infty E(f|A_i) 1_{A_i}.$$

41

Weak and Strong Derivatives

For this section, let Ω be an open subset of $\mathbb{R}^d$, $p,q,r \in [1,\infty]$, $L^p(\Omega) = L^p(\Omega, \mathcal{B}_\Omega, m)$ and $L^p_{loc}(\Omega) = L^p_{loc}(\Omega, \mathcal{B}_\Omega, m)$, where m is Lebesgue measure on $\mathcal{B}_{\mathbb{R}^d}$ and $\mathcal{B}_\Omega$ is the Borel σ − algebra on Ω. If $\Omega = \mathbb{R}^d$, we will simply write L^p and L^p_{loc} for $L^p(\mathbb{R}^d)$ and $L^p_{loc}(\mathbb{R}^d)$ respectively. Also let

$$\langle f, g\rangle := \int_\Omega fg dm$$

for any pair of measurable functions $f, g : \Omega \to \mathbb{C}$ such that $fg \in L^1(\Omega)$. For example, by Hölder's inequality, if $\langle f, g\rangle$ is defined for $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$ when $q = \frac{p}{p-1}$.

Definition 41.1. *A sequence $\{u_n\}_{n=1}^\infty \subset L^p_{loc}(\Omega)$ is said to converge to $u \in L^p_{loc}(\Omega)$ if $\lim_{n\to\infty} \|u - u_n\|_{L^q(K)} = 0$ for all compact subsets $K \subset \Omega$.*

The following simple but useful remark will be used (typically without further comment) in the sequel.

Remark 41.2. Suppose $r,p,q \in [1,\infty]$ are such that $r^{-1} = p^{-1} + q^{-1}$ and $f_t \to f$ in $L^p(\Omega)$ and $g_t \to g$ in $L^q(\Omega)$ as $t \to 0$, then $f_t g_t \to fg$ in $L^r(\Omega)$. Indeed,

$$\begin{aligned}\|f_t g_t - fg\|_r &= \|(f_t - f)\, g_t + f\,(g_t - g)\|_r \\ &\le \|f_t - f\|_p \|g_t\|_q + \|f\|_p \|g_t - g\|_q \to 0 \text{ as } t \to 0\end{aligned}$$

41.1 Basic Definitions and Properties

Definition 41.3 (Weak Differentiability). *Let $v \in \mathbb{R}^d$ and $u \in L^p(\Omega)$ ($u \in L^p_{loc}(\Omega)$) then $\partial_v u$ is said to **exist weakly** in $L^p(\Omega)$ ($L^p_{loc}(\Omega)$) if there exists a function $g \in L^p(\Omega)$ ($g \in L^p_{loc}(\Omega)$) such that*

$$\langle u, \partial_v \varphi\rangle = -\langle g, \varphi\rangle \text{ for all } \varphi \in C_c^\infty(\Omega). \tag{41.1}$$

*The function g if it exists will be denoted by $\partial_v^{(w)} u$. Similarly if $\alpha \in \mathbb{N}_0^d$ and ∂^α is as in Notation 31.21, we say $\partial^\alpha u$ **exists weakly** in $L^p(\Omega)$ ($L^p_{loc}(\Omega)$) iff there exists $g \in L^p(\Omega)$ ($L^p_{loc}(\Omega)$) such that*

$$\langle u, \partial^\alpha \varphi\rangle = (-1)^{|\alpha|} \langle g, \varphi\rangle \text{ for all } \varphi \in C_c^\infty(\Omega).$$

*More generally if $p(\xi) = \sum_{|\alpha| \le N} a_\alpha \xi^\alpha$ is a polynomial in $\xi \in \mathbb{R}^n$, then $p(\partial)u$ **exists weakly** in $L^p(\Omega)$ ($L^p_{loc}(\Omega)$) iff there exists $g \in L^p(\Omega)$ ($L^p_{loc}(\Omega)$) such that*

$$\langle u, p(-\partial)\varphi\rangle = \langle g, \varphi\rangle \text{ for all } \varphi \in C_c^\infty(\Omega) \tag{41.2}$$

and we denote g by w−$p(\partial)u$.

By Corollary 31.41, there is at most one $g \in L^1_{loc}(\Omega)$ such that Eq. (41.2) holds, so w−$p(\partial)u$ is well defined.

Lemma 41.4. *Let $p(\xi)$ be a polynomial on $\mathbb{R}^d$, $k = \deg(p) \in \mathbb{N}$, and $u \in L^1_{loc}(\Omega)$ such that $p(\partial)u$ exists weakly in $L^1_{loc}(\Omega)$. Then*

1. *$\operatorname{supp}_m(\text{w}-p(\partial)u) \subset \operatorname{supp}_m(u)$, where $\operatorname{supp}_m(u)$ is the essential support of u relative to Lebesgue measure, see Definition 31.26.*
2. *If $\deg p = k$ and $u|_U \in C^k(U, \mathbb{C})$ for some open set $U \subset \Omega$, then w−$p(\partial)u = p(\partial)\, u$ a.e. on U.*

Proof.

1. Since
$$\langle \text{w}-p(\partial)u, \varphi\rangle = -\langle u, p(-\partial)\varphi\rangle = 0 \text{ for all } \varphi \in C_c^\infty(\Omega \setminus \operatorname{supp}_m(u)),$$
an application of Corollary 31.41 shows w−$p(\partial)u = 0$ a.e. on $\Omega \setminus \operatorname{supp}_m(u)$. So by Lemma 31.27, $\Omega \setminus \operatorname{supp}_m(u) \subset \Omega \setminus \operatorname{supp}_m(\text{w}-p(\partial)u)$, i.e. $\operatorname{supp}_m(\text{w}-p(\partial)u) \subset \operatorname{supp}_m(u)$.
2. Suppose that $u|_U$ is C^k and let $\psi \in C_c^\infty(U)$. (We view ψ as a function in $C_c^\infty(\mathbb{R}^d)$ by setting $\psi \equiv 0$ on $\mathbb{R}^d \setminus U$.) By Corollary 31.38, there exists $\gamma \in C_c^\infty(\Omega)$ such that $0 \le \gamma \le 1$ and $\gamma = 1$ in a neighborhood of $\operatorname{supp}(\psi)$. Then by setting $\gamma u = 0$ on $\mathbb{R}^d \setminus \operatorname{supp}(\gamma)$ we may view $\gamma u \in C_c^k(\mathbb{R}^d)$ and so by standard integration by parts (see Lemma 31.39) and the ordinary product rule,
$$\begin{aligned}\langle \text{w}-p(\partial)u, \psi\rangle &= \langle u, p(-\partial)\psi\rangle = -\langle \gamma u, p(-\partial)\psi\rangle \\ &= \langle p(\partial)(\gamma u), \psi\rangle = \langle p(\partial)u, \psi\rangle\end{aligned} \tag{41.3}$$

wherein the last equality we have γ is constant on $\mathrm{supp}(\psi)$. Since Eq. (41.3) is true for all $\psi \in C_c^\infty(U)$, an application of Corollary 31.41 with $h = \text{w}-p(\partial)u - p(\partial)u$ and $\mu = m$ shows $\text{w}-p(\partial)u = p(\partial)u$ a.e. on U.

■

Notation 41.5 *In light of Lemma 41.4 there is no danger in simply writing $p(\partial)u$ for $\text{w}-p(\partial)u$. So in the sequel we will always interpret $p(\partial)u$ in the weak or "distributional" sense.*

Example 41.6. Suppose $u(x) = |x|$ for $x \in \mathbb{R}$, then $\partial u(x) = \mathrm{sgn}(x)$ in $L^1_{loc}(\mathbb{R})$ while $\partial^2 u(x) = 2\delta(x)$ so $\partial^2 u(x)$ does not exist weakly in $L^1_{loc}(\mathbb{R})$.

Example 41.7. Suppose $d = 2$ and $u(x,y) = 1_{y>x}$. Then $u \in L^1_{loc}(\mathbb{R}^2)$, while $\partial_x 1_{y>x} = -\delta(y-x)$ and $\partial_y 1_{y>x} = \delta(y-x)$ and so that neither $\partial_x u$ or $\partial_y u$ exists weakly. On the other hand $(\partial_x + \partial_y)u = 0$ weakly. To prove these assertions, notice $u \in C^\infty(\mathbb{R}^2 \setminus \Delta)$ where $\Delta = \{(x,x) : x \in \mathbb{R}^2\}$. So by Lemma 41.4, for any polynomial $p(\xi)$ without constant term, if $p(\partial)u$ exists weakly then $p(\partial)u = 0$. However,

$$\begin{aligned}
\langle u, -\partial_x \varphi\rangle &= -\int_{y>x} \varphi_x(x,y)dxdy = -\int_{\mathbb{R}} \varphi(y,y)dy, \\
\langle u, -\partial_y \varphi\rangle &= -\int_{y>x} \varphi_y(x,y)dxdy = \int_{\mathbb{R}} \varphi(x,x)dx \text{ and} \\
\langle u, -(\partial_x + \partial_y)\varphi\rangle &= 0
\end{aligned}$$

from which it follows that $\partial_x u$ and $\partial_y u$ can not be zero while $(\partial_x + \partial_y)u = 0$.

On the other hand if $p(\xi)$ and $q(\xi)$ are two polynomials and $u \in L^1_{loc}(\Omega)$ is a function such that $p(\partial)u$ exists weakly in $L^1_{loc}(\Omega)$ and $q(\partial)[p(\partial)u]$ exists weakly in $L^1_{loc}(\Omega)$ then $(qp)(\partial)u$ exists weakly in $L^1_{loc}(\Omega)$. This is because

$$\begin{aligned}
\langle u, (qp)(-\partial)\varphi\rangle &= \langle u, p(-\partial)q(-\partial)\varphi\rangle \\
&= \langle p(\partial)u, q(-\partial)\varphi\rangle = \langle q(\partial)p(\partial)u, \varphi\rangle \text{ for all } \varphi \in C_c^\infty(\Omega).
\end{aligned}$$

Example 41.8. Let $u(x,y) = 1_{x>0} + 1_{y>0}$ in $L^1_{loc}(\mathbb{R}^2)$. Then $\partial_x u(x,y) = \delta(x)$ and $\partial_y u(x,y) = \delta(y)$ so $\partial_x u(x,y)$ and $\partial_y u(x,y)$ do **not** exist weakly in $L^1_{loc}(\mathbb{R}^2)$. However $\partial_y \partial_x u$ does exists weakly and is the zero function. This shows $\partial_y \partial_x u$ may exists weakly despite the fact both $\partial_x u$ and $\partial_y u$ do not exists weakly in $L^1_{loc}(\mathbb{R}^2)$.

Lemma 41.9. *Suppose $u \in L^1_{loc}(\Omega)$ and $p(\xi)$ is a polynomial of degree k such that $p(\partial)u$ exists weakly in $L^1_{loc}(\Omega)$ then*

$$\langle p(\partial)u, \varphi\rangle = \langle u, p(-\partial)\varphi\rangle \text{ for all } \varphi \in C_c^k(\Omega). \tag{41.4}$$

Note: *The point here is that Eq. (41.4) holds for all $\varphi \in C_c^k(\Omega)$ not just $\varphi \in C_c^\infty(\Omega)$.*

Proof. Let $\varphi \in C_c^k(\Omega)$ and choose $\eta \in C_c^\infty(B(0,1))$ such that $\int_{\mathbb{R}^d} \eta(x)dx = 1$ and let $\eta_\varepsilon(x) := \varepsilon^{-d}\eta(x/\varepsilon)$. Then $\eta_\varepsilon * \varphi \in C_c^\infty(\Omega)$ for ε sufficiently small and $p(-\partial)[\eta_\varepsilon * \varphi] = \eta_\varepsilon * p(-\partial)\varphi \to p(-\partial)\varphi$ and $\eta_\varepsilon * \varphi \to \varphi$ uniformly on compact sets as $\varepsilon \downarrow 0$. Therefore by the dominated convergence theorem,

$$\langle p(\partial)u, \varphi\rangle = \lim_{\varepsilon\downarrow 0}\langle p(\partial)u, \eta_\varepsilon * \varphi\rangle = \lim_{\varepsilon\downarrow 0}\langle u, p(-\partial)(\eta_\varepsilon * \varphi)\rangle = \langle u, p(-\partial)\varphi\rangle.$$

■

Lemma 41.10 (Product Rule). *Let $u \in L^1_{loc}(\Omega)$, $v \in \mathbb{R}^d$ and $\varphi \in C^1(\Omega)$. If $\partial_v^{(w)} u$ exists in $L^1_{loc}(\Omega)$, then $\partial_v^{(w)}(\varphi u)$ exists in $L^1_{loc}(\Omega)$ and*

$$\partial_v^{(w)}(\varphi u) = \partial_v \varphi \cdot u + \varphi \partial_v^{(w)} u \text{ a.e.}$$

Moreover if $\varphi \in C_c^1(\Omega)$ and $F := \varphi u \in L^1$ (here we define F on $\mathbb{R}^d$ by setting $F = 0$ on $\mathbb{R}^d \setminus \Omega$), then $\partial^{(w)} F = \partial_v \varphi \cdot u + \varphi \partial_v^{(w)} u$ exists weakly in $L^1(\mathbb{R}^d)$.

Proof. Let $\psi \in C_c^\infty(\Omega)$, then using Lemma 41.9,

$$\begin{aligned}
-\langle \varphi u, \partial_v \psi\rangle &= -\langle u, \varphi \partial_v \psi\rangle = -\langle u, \partial_v(\varphi\psi) - \partial_v \varphi \cdot \psi\rangle \\
&= \langle \partial_v^{(w)} u, \varphi\psi\rangle + \langle \partial_v \varphi \cdot u, \psi\rangle \\
&= \langle \varphi \partial_v^{(w)} u, \psi\rangle + \langle \partial_v \varphi \cdot u, \psi\rangle.
\end{aligned}$$

This proves the first assertion. To prove the second assertion let $\gamma \in C_c^\infty(\Omega)$ such that $0 \le \gamma \le 1$ and $\gamma = 1$ on a neighborhood of $\mathrm{supp}(\varphi)$. So for $\psi \in C_c^\infty(\mathbb{R}^d)$, using $\partial_v \gamma = 0$ on $\mathrm{supp}(\varphi)$ and $\gamma\psi \in C_c^\infty(\Omega)$, we find

$$\begin{aligned}
\langle F, \partial_v \psi\rangle &= \langle \gamma F, \partial_v \psi\rangle = \langle F, \gamma \partial_v \psi\rangle = \langle (\varphi u), \partial_v(\gamma\psi) - \partial_v \gamma \cdot \psi\rangle \\
&= \langle (\varphi u), \partial_v(\gamma\psi)\rangle = -\langle \partial_v^{(w)}(\varphi u), (\gamma\psi)\rangle \\
&= -\langle \partial_v \varphi \cdot u + \varphi \partial_v^{(w)} u, \gamma\psi\rangle = -\langle \partial_v \varphi \cdot u + \varphi \partial_v^{(w)} u, \psi\rangle.
\end{aligned}$$

This show $\partial_v^{(w)} F = \partial_v \varphi \cdot u + \varphi \partial_v^{(w)} u$ as desired. ■

Lemma 41.11. *Suppose $q \in [1,\infty)$, $p(\xi)$ is a polynomial in $\xi \in \mathbb{R}^d$ and $u \in L^q_{loc}(\Omega)$. If there exists $\{u_m\}_{m=1}^\infty \subset L^q_{loc}(\Omega)$ such that $p(\partial)u_m$ exists in $L^q_{loc}(\Omega)$ for all m and there exists $g \in L^q_{loc}(\Omega)$ such that for all $\varphi \in C_c^\infty(\Omega)$,*

$$\lim_{m\to\infty}\langle u_m, \varphi\rangle = \langle u, \varphi\rangle \text{ and } \lim_{m\to\infty}\langle p(\partial)u_m, \varphi\rangle = \langle g, \varphi\rangle$$

then $p(\partial)u$ exists in $L^q_{loc}(\Omega)$ and $p(\partial)u = g$.

Proof. Since

$$\langle u, p(-\partial)\varphi\rangle = \lim_{m\to\infty}\langle u_m, p(-\partial)\varphi\rangle = \lim_{m\to\infty}\langle p(\partial)u_m, \varphi\rangle = \langle g, \varphi\rangle$$

for all $\varphi \in C_c^\infty(\Omega)$, $p(\partial)u$ exists and is equal to $g \in L^q_{loc}(\Omega)$. ■

Conversely we have the following proposition.

Proposition 41.12 (Mollification). *Suppose $q \in [1,\infty)$, $p_1(\xi),\dots,p_N(\xi)$ is a collection of polynomials in $\xi \in \mathbb{R}^d$ and $u \in L^q_{loc}(\Omega)$ such that $p_l(\partial)u$ exists weakly in $L^q_{loc}(\Omega)$ for $l = 1,2,\dots,N$. Then there exists $u_n \in C_c^\infty(\Omega)$ such that $u_n \to u$ in $L^q_{loc}(\Omega)$ and $p_l(\partial)\, u_n \to p_l(\partial)\, u$ in $L^q_{loc}(\Omega)$ for $l = 1,2,\dots,N$.*

Proof. Let $\eta \in C_c^\infty(B(0,1))$ such that $\int_{\mathbb{R}^d} \eta dm = 1$ and $\eta_\varepsilon(x) := \varepsilon^{-d}\eta(x/\varepsilon)$ be as in the proof of Lemma 41.9. For any function $f \in L^1_{loc}(\Omega)$, $\varepsilon > 0$ and $x \in \Omega_\varepsilon := \{y \in \Omega : \operatorname{dist}(y,\Omega^c) > \varepsilon\}$, let

$$f_\varepsilon(x) := f * \eta_\varepsilon(x) := 1_\Omega f * \eta_\varepsilon(x) = \int_\Omega f(y)\eta_\varepsilon(x-y)dy.$$

Notice that $f_\varepsilon \in C^\infty(\Omega_\varepsilon)$ and $\Omega_\varepsilon \uparrow \Omega$ as $\varepsilon \downarrow 0$. Given a compact set $K \subset \Omega$ let $K_\varepsilon := \{x \in \Omega : \operatorname{dist}(x,K) \le \varepsilon\}$. Then $K_\varepsilon \downarrow K$ as $\varepsilon \downarrow 0$, there exists $\varepsilon_0 > 0$ such that $K_0 := K_{\varepsilon_0}$ is a compact subset of $\Omega_0 := \Omega_{\varepsilon_0} \subset \Omega$ (see Figure 41.1) and for $x \in K$,

$$f * \eta_\varepsilon(x) := \int_\Omega f(y)\eta_\varepsilon(x-y)dy = \int_{K_\varepsilon} f(y)\eta_\varepsilon(x-y)dy.$$

Therefore, using Theorem 31.33,

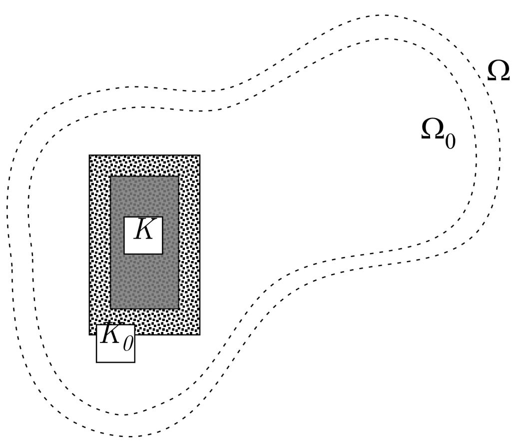


Fig. 41.1. The geomentry of $K \subset K_0 \subset \Omega_0 \subset \Omega$.

$$\begin{aligned}\|f * \eta_\varepsilon - f\|_{L^p(K)} &= \|(1_{K_0}f) * \eta_\varepsilon - 1_{K_0}f\|_{L^p(K)} \\ &\le \|(1_{K_0}f) * \eta_\varepsilon - 1_{K_0}f\|_{L^p(\mathbb{R}^d)} \to 0 \text{ as } \varepsilon \downarrow 0.\end{aligned}$$

Hence, for all $f \in L^q_{loc}(\Omega)$, $f * \eta_\varepsilon \in C^\infty(\Omega_\varepsilon)$ and

$$\lim_{\varepsilon\downarrow 0} \|f * \eta_\varepsilon - f\|_{L^p(K)} = 0. \tag{41.5}$$

Now let $p(\xi)$ be a polynomial on $\mathbb{R}^d$, $u \in L^q_{loc}(\Omega)$ such that $p(\partial)\, u \in L^q_{loc}(\Omega)$ and $v_\varepsilon := \eta_\varepsilon * u \in C^\infty(\Omega_\varepsilon)$ as above. Then for $x \in K$ and $\varepsilon < \varepsilon_0$,

$$\begin{aligned}p(\partial)v_\varepsilon(x) &= \int_\Omega u(y)p(\partial_x)\eta_\varepsilon(x-y)dy = \int_\Omega u(y)p(-\partial_y)\eta_\varepsilon(x-y)dy \\ &= \int_\Omega u(y)p(-\partial_y)\eta_\varepsilon(x-y)dy = \langle u, p(\partial)\eta_\varepsilon(x-\cdot)\rangle \\ &= \langle p(\partial)u, \eta_\varepsilon(x-\cdot)\rangle = (p(\partial)u)_\varepsilon(x). \end{aligned} \tag{41.6}$$

From Eq. (41.6) we may now apply Eq. (41.5) with $f = u$ and $f = p_l(\partial)u$ for $1 \le l \le N$ to find

$$\|v_\varepsilon - u\|_{L^p(K)} + \sum_{l=1}^N \|p_l(\partial)v_\varepsilon - p_l(\partial)u\|_{L^p(K)} \to 0 \text{ as } \varepsilon \downarrow 0.$$

For $n \in \mathbb{N}$, let

$$K_n := \{x \in \Omega : |x| \le n \text{ and } d(x,\Omega^c) \ge 1/n\}$$

(so $K_n \subset K^o_{n+1} \subset K_{n+1}$ for all n and $K_n \uparrow \Omega$ as $n \to \infty$ or see Lemma 36.5) and choose $\psi_n \in C_c^\infty(K^o_{n+1},[0,1])$, using Corollary 31.38, so that $\psi_n = 1$ on a neighborhood of K_n. Choose $\varepsilon_n \downarrow 0$ such that $K_{n+1} \subset \Omega_{\varepsilon_n}$ and

$$\|v_{\varepsilon_n} - u\|_{L^p(K_n)} + \sum_{l=1}^N \|p_l(\partial)v_{\varepsilon_n} - p_l(\partial)u\|_{L^p(K_n)} \le 1/n.$$

Then $u_n := \psi_n \cdot v_{\varepsilon_n} \in C_c^\infty(\Omega)$ and since $u_n = v_{\varepsilon_n}$ on K_n we still have

$$\|u_n - u\|_{L^p(K_n)} + \sum_{l=1}^N \|p_l(\partial)u_n - p_l(\partial)u\|_{L^p(K_n)} \le 1/n. \tag{41.7}$$

Since any compact set $K \subset \Omega$ is contained in K^o_n for all n sufficiently large, Eq. (41.7) implies

$$\lim_{n\to\infty}\left[\|u_n - u\|_{L^p(K)} + \sum_{l=1}^N \|p_l(\partial)u_n - p_l(\partial)u\|_{L^p(K)}\right] = 0.$$

■

The following proposition is another variant of Proposition 41.12 which the reader is asked to prove in Exercise 41.2 below.

Proposition 41.13. *Suppose $q \in [1,\infty)$, $p_1(\xi),\dots,p_N(\xi)$ is a collection of polynomials in $\xi \in \mathbb{R}^d$ and $u \in L^q = L^q(\mathbb{R}^d)$ such that $p_l(\partial)u \in L^q$ for $l = 1,2,\dots,N$. Then there exists $u_n \in C_c^\infty(\mathbb{R}^d)$ such that*

$$\lim_{n\to\infty}\left[\|u_n - u\|_{L^q} + \sum_{l=1}^N \|p_l(\partial)u_n - p_l(\partial)u\|_{L^q}\right] = 0.$$

Notation 41.14 (Difference quotients) *For $v \in \mathbb{R}^d$ and $h \in \mathbb{R}\setminus\{0\}$ and a function $u : \Omega \to \mathbb{C}$, let*

$$\partial_v^h u(x) := \frac{u(x+hv) - u(x)}{h}$$

for those $x \in \Omega$ such that $x + hv \in \Omega$. When v is one of the standard basis elements, e_i for $1 \le i \le d$, we will write $\partial_i^h u(x)$ rather than $\partial_{e_i}^h u(x)$. Also let

$$\nabla^h u(x) := \left(\partial_1^h u(x),\dots,\partial_n^h u(x)\right)$$

be the difference quotient approximation to the gradient.

Definition 41.15 (Strong Differentiability). *Let $v \in \mathbb{R}^d$ and $u \in L^p$, then $\partial_v u$ is said to exist* ***strongly*** *in L^p if the $\lim_{h\to 0}\partial_v^h u$ exists in L^p. We will denote the limit by $\partial_v^{(s)}u$.*

It is easily verified that if $u \in L^p$, $v \in \mathbb{R}^d$ and $\partial_v^{(s)}u \in L^p$ exists then $\partial_v^{(w)}u$ exists and $\partial_v^{(w)}u = \partial_v^{(s)}u$. The key to checking this assertion is the identity,

$$\begin{aligned}\langle\partial_v^h u,\varphi\rangle &= \int_{\mathbb{R}^d}\frac{u(x+hv)-u(x)}{h}\varphi(x)dx \\ &= \int_{\mathbb{R}^d} u(x)\frac{\varphi(x-hv)-\varphi(x)}{h}dx = \langle u,\partial_{-v}^h\varphi\rangle. \qquad (41.8)\end{aligned}$$

Hence if $\partial_v^{(s)}u = \lim_{h\to 0}\partial_v^h u$ exists in L^p and $\varphi \in C_c^\infty(\mathbb{R}^d)$, then

$$\langle\partial_v^{(s)}u,\varphi\rangle = \lim_{h\to 0}\langle\partial_v^h u,\varphi\rangle = \lim_{h\to 0}\langle u,\partial_{-v}^h\varphi\rangle = \frac{d}{dh}|_0\langle u,\varphi(\cdot - hv)\rangle = -\langle u,\partial_v\varphi\rangle$$

wherein Corollary 10.30 has been used in the last equality to bring the derivative past the integral. This shows $\partial_v^{(w)}u$ exists and is equal to $\partial_v^{(s)}u$. What is somewhat more surprising is that the converse assertion that if $\partial_v^{(w)}u$ exists then so does $\partial_v^{(s)}u$. Theorem 41.18 is a generalization of Theorem 40.8 from L^2 to L^p. For the reader's convenience, let us give a self-contained proof of the version of the Banach - Alaoglu's Theorem which will be used in the proof of Theorem 41.18. (This is the same as Theorem 36.25 above.)

Proposition 41.16 (Weak-$*$ Compactness: Banach - Alaoglu's Theorem). *Let X be a separable Banach space and $\{f_n\} \subset X^*$ be a bounded sequence, then there exist a subsequence $\{\tilde f_n\} \subset \{f_n\}$ and $f \in X^*$ such that $\lim_{n\to\infty}\tilde f_n(x) = f(x)$ for all $x \in X$.*

Proof. Let $D \subset X$ be a countable dense subset of X and let $M := \sup_n\|f_n\|_{X^*} < \infty$. Using Cantor's diagonal trick, choose $\{\tilde f_n\} \subset \{f_n\}$ such that $\lim_{n\to\infty}\tilde f_n(x) =: f(x)$ exists for all $x \in D$. For $x \in X$ and $y \in D$ we have,

$$\begin{aligned}\left|\tilde f_n(x) - \tilde f_m(x)\right| &\le \left|\tilde f_n(x) - \tilde f_n(y)\right| + \left|\tilde f_n(y) - \tilde f_m(y)\right| + \left|\tilde f_m(y) - \tilde f_m(x)\right| \\ &\le 2M\|x-y\| + \left|\tilde f_n(y) - \tilde f_m(y)\right|\end{aligned}$$

and therefore,

$$\begin{aligned}\limsup_{m,n\to\infty}\left|\tilde f_n(x) - \tilde f_m(x)\right| &\le 2M\|x-y\| + \limsup_{m,n\to\infty}\left|\tilde f_n(y) - \tilde f_m(y)\right| \\ &= 2M\|x-y\|.\end{aligned}$$

As the right side may be made as small as we please it follows that $f(x) := \lim_{n\to\infty}\tilde f_n(x)$ exists for all $x \in X$. The resulting function f is easily seen to be linear and bounded by M so that $f \in X^*$. ■

Corollary 41.17. *Let $p \in (1,\infty]$ and $q = \frac{p}{p-1}$. Then to every bounded sequence $\{u_n\}_{n=1}^\infty \subset L^p(\Omega)$ there is a subsequence $\{\tilde u_n\}_{n=1}^\infty$ and an element $u \in L^p(\Omega)$ such that*

$$\lim_{n\to\infty}\langle\tilde u_n, g\rangle = \langle u, g\rangle \text{ for all } g \in L^q(\Omega).$$

Proof. By Theorem 29.6, the map

$$v \in L^p(\Omega) \to \langle v,\cdot\rangle \in (L^q(\Omega))^*$$

is an isometric isomorphism of Banach spaces. By Theorem 31.15, $L^q(\Omega)$ is separable for all $q \in [1,\infty)$ and hence the result now follows from Proposition 41.16. ■

Theorem 41.18 (Weak and Strong Differentiability). *Suppose $p \in [1,\infty)$, $u \in L^p(\mathbb{R}^d)$ and $v \in \mathbb{R}^d\setminus\{0\}$. Then the following are equivalent:*

1. *There exists $g \in L^p(\mathbb{R}^d)$ and $\{h_n\}_{n=1}^\infty \subset \mathbb{R}\setminus\{0\}$ such that $\lim_{n\to\infty}h_n = 0$ and*
$$\lim_{n\to\infty}\langle\partial_v^{h_n}u,\varphi\rangle = \langle g,\varphi\rangle \text{ for all } \varphi \in C_c^\infty(\mathbb{R}^d).$$

2. *$\partial_v^{(w)} u$ exists and is equal to $g \in L^p(\mathbb{R}^d)$, i.e. $\langle u, \partial_v \varphi\rangle = -\langle g, \varphi\rangle$ for all $\varphi \in C_c^\infty(\mathbb{R}^d)$.*
3. *There exists $g \in L^p(\mathbb{R}^d)$ and $u_n \in C_c^\infty(\mathbb{R}^d)$ such that $u_n \overset{L^p}{\to} u$ and $\partial_v u_n \overset{L^p}{\to} g$ as $n \to \infty$.*
4. *$\partial_v^{(s)} u$ exists and is is equal to $g \in L^p(\mathbb{R}^d)$, i.e. $\partial_v^h u \to g$ in L^p as $h \to 0$.*

Moreover if $p \in (1, \infty)$ any one of the equivalent conditions 1. – 4. above are implied by the following condition.

1′. There exists $\{h_n\}_{n=1}^\infty \subset \mathbb{R}\backslash\{0\}$ such that $\lim_{n\to\infty} h_n = 0$ and $\sup_n \left\|\partial_v^{h_n} u\right\|_p < \infty$.

Proof. 4. $\implies$ 1. is simply the assertion that strong convergence implies weak convergence. 1. $\implies$ 2. For $\varphi \in C_c^\infty(\mathbb{R}^d)$, Eq. (41.8) and the dominated convergence theorem implies

$$\langle g, \varphi\rangle = \lim_{n\to\infty} \langle \partial_v^{h_n} u, \varphi\rangle = \lim_{n\to\infty} \langle u, \partial_{-v}^{h_n} \varphi\rangle = -\langle u, \partial_v \varphi\rangle.$$

2. $\implies$ 3. Let $\eta \in C_c^\infty(\mathbb{R}^d, \mathbb{R})$ such that $\int_{\mathbb{R}^d} \eta(x)dx = 1$ and let $\eta_m(x) = m^d \eta(mx)$, then by Proposition 31.37, $h_m := \eta_m * u \in C^\infty(\mathbb{R}^d)$ for all m and

$$\begin{aligned}\partial_v h_m(x) = \partial_v \eta_m * u(x) &= \int_{\mathbb{R}^d} \partial_v \eta_m(x-y) u(y) dy \\ &= \langle u, -\partial_v [\eta_m(x - \cdot)]\rangle = \langle g, \eta_m(x-\cdot)\rangle = \eta_m * g(x).\end{aligned}$$

By Theorem 31.33, $h_m \to u \in L^p(\mathbb{R}^d)$ and $\partial_v h_m = \eta_m * g \to g$ in $L^p(\mathbb{R}^d)$ as $m \to \infty$. This shows 3. holds except for the fact that h_m need not have compact support. To fix this let $\psi \in C_c^\infty(\mathbb{R}^d, [0,1])$ such that $\psi = 1$ in a neighborhood of 0 and let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ and $(\partial_v \psi)_\varepsilon(x) := (\partial_v \psi)(\varepsilon x)$. Then

$$\partial_v(\psi_\varepsilon h_m) = \partial_v \psi_\varepsilon h_m + \psi_\varepsilon \partial_v h_m = \varepsilon (\partial_v \psi)_\varepsilon h_m + \psi_\varepsilon \partial_v h_m$$

so that $\psi_\varepsilon h_m \to h_m$ in L^p and $\partial_v(\psi_\varepsilon h_m) \to \partial_v h_m$ in L^p as $\varepsilon \downarrow 0$. Let $u_m = \psi_{\varepsilon_m} h_m$ where ε_m is chosen to be greater than zero but small enough so that

$$\|\psi_{\varepsilon_m} h_m - h_m\|_p + \|\partial_v(\psi_{\varepsilon_m} h_m) \to \partial_v h_m\|_p < 1/m.$$

Then $u_m \in C_c^\infty(\mathbb{R}^d)$, $u_m \to u$ and $\partial_v u_m \to g$ in L^p as $m \to \infty$. 3. $\implies$ 4. By the fundamental theorem of calculus

$$\begin{aligned}\partial_v^h u_m(x) &= \frac{u_m(x+hv) - u_m(x)}{h} \\ &= \frac{1}{h}\int_0^1 \frac{d}{ds} u_m(x+shv) ds = \int_0^1 (\partial_v u_m)(x+shv) ds. \qquad (41.9)\end{aligned}$$

and therefore,

$$\partial_v^h u_m(x) - \partial_v u_m(x) = \int_0^1 [(\partial_v u_m)(x+shv) - \partial_v u_m(x)]\, ds.$$

So by Minkowski's inequality for integrals, Theorem 29.2,

$$\left\|\partial_v^h u_m(x) - \partial_v u_m\right\|_p \le \int_0^1 \|(\partial_v u_m)(\cdot + shv) - \partial_v u_m\|_p \, ds$$

and letting $m \to \infty$ in this equation then implies

$$\left\|\partial_v^h u - g\right\|_p \le \int_0^1 \|g(\cdot + shv) - g\|_p \, ds.$$

By the dominated convergence theorem and Proposition 31.25, the right member of this equation tends to zero as $h \to 0$ and this shows item 4. holds. (1′. $\implies$ 1. when $p > 1$) This is a consequence of Corollary 41.17 (or see Theorem 36.25 above) which asserts, by passing to a subsequence if necessary, that $\partial_v^{h_n} u \overset{w}{\to} g$ for some $g \in L^p(\mathbb{R}^d)$. ■

Example 41.19. The fact that (1′) does not imply the equivalent conditions 1 – 4 in Theorem 41.18 when $p = 1$ is demonstrated by the following example. Let $u := 1_{[0,1]}$, then

$$\int_{\mathbb{R}} \left|\frac{u(x+h) - u(x)}{h}\right| dx = \frac{1}{|h|}\int_{\mathbb{R}} \left|1_{[-h,1-h]}(x) - 1_{[0,1]}(x)\right| dx = 2$$

for $|h| < 1$. On the other hand the distributional derivative of u is $\partial u(x) = \delta(x) - \delta(x-1)$ which is not in L^1.

Alternatively, if there exists $g \in L^1(\mathbb{R}, dm)$ such that

$$\lim_{n\to\infty} \frac{u(x+h_n) - u(x)}{h_n} = g(x) \text{ in } L^1$$

for some sequence $\{h_n\}_{n=1}^\infty$ as above. Then for $\varphi \in C_c^\infty(\mathbb{R})$ we would have on one hand,

$$\begin{aligned}\int_{\mathbb{R}} \frac{u(x+h_n) - u(x)}{h_n} \varphi(x) dx &= \int_{\mathbb{R}} \frac{\varphi(x-h_n) - \varphi(x)}{h_n} u(x) dx \\ &\to -\int_0^1 \varphi'(x) dx = (\varphi(0) - \varphi(1)) \text{ as } n \to \infty,\end{aligned}$$

while on the other hand,

$$\int_{\mathbb{R}} \frac{u(x+h_n) - u(x)}{h_n} \varphi(x) dx \to \int_{\mathbb{R}} g(x)\varphi(x) dx.$$

These two equations imply

$$\int_{\mathbb{R}} g(x)\varphi(x)dx = \varphi(0) - \varphi(1) \text{ for all } \varphi \in C_c^\infty(\mathbb{R}) \tag{41.10}$$

and in particular that $\int_{\mathbb{R}} g(x)\varphi(x)dx = 0$ for all $\varphi \in C_c(\mathbb{R}\setminus\{0,1\})$. By Corollary 31.41, $g(x) = 0$ for $m-$a.e. $x \in \mathbb{R}\setminus\{0,1\}$ and hence $g(x) = 0$ for $m-$a.e. $x \in \mathbb{R}$. But this clearly contradicts Eq. (41.10). This example also shows that the unit ball in $L^1(\mathbb{R}, dm)$ is not weakly sequentially compact. Compare with Lemma 21.21 below.

Corollary 41.20. *If $1 \le p < \infty$, $u \in L^p$ such that $\partial_v u \in L^p$, then $\left\|\partial_v^h u\right\|_{L^p} \le \|\partial_v u\|_{L^p}$ for all $h \neq 0$ and $v \in \mathbb{R}^d$.*

Proof. By Minkowski's inequality for integrals, Theorem 29.2, we may let $m \to \infty$ in Eq. (41.9) to find

$$\partial_v^h u(x) = \int_0^1 (\partial_v u)(x + shv)ds \text{ for a.e. } x \in \mathbb{R}^d$$

and

$$\left\|\partial_v^h u\right\|_{L^p} \le \int_0^1 \|(\partial_v u)(\cdot + shv)\|_{L^p} ds = \|\partial_v u\|_{L^p}.$$

■

Proposition 41.21 (A weak form of Weyls Lemma). *If $u \in L^2(\mathbb{R}^d)$ such that $f := \triangle u \in L^2(\mathbb{R}^d)$ then $\partial^\alpha u \in L^2(\mathbb{R}^d)$ for $|\alpha| \le 2$. Furthermore if $k \in \mathbb{N}_0$ and $\partial^\beta f \in L^2(\mathbb{R}^d)$ for all $|\beta| \le k$, then $\partial^\alpha u \in L^2(\mathbb{R}^d)$ for $|\alpha| \le k+2$.*

Proof. By Proposition 41.13, there exists $u_n \in C_c^\infty(\mathbb{R}^d)$ such that $u_n \to u$ and $\Delta u_n \to \Delta u = f$ in $L^2(\mathbb{R}^d)$. By integration by parts we find

$$\begin{aligned}\int_{\mathbb{R}^d} |\nabla(u_n - u_m)|^2 dm &= (-\Delta(u_n - u_m), (u_n - u_m))_{L^2} \\ &\to -(f - f, u - u) = 0 \text{ as } m, n \to \infty\end{aligned}$$

and hence by item 3. of Theorem 41.18, $\partial_i u \in L^2$ for each i. Since

$$\|\nabla u\|_{L^2}^2 = \lim_{n\to\infty} \int_{\mathbb{R}^d} |\nabla u_n|^2 dm = (-\Delta u_n, u_n)_{L^2} \to -(f, u) \text{ as } n \to \infty$$

we also learn that

$$\|\nabla u\|_{L^2}^2 = -(f, u) \le \|f\|_{L^2} \cdot \|u\|_{L^2}. \tag{41.11}$$

Let us now consider

$$\begin{aligned}\sum_{i,j=1}^d \int_{\mathbb{R}^d} |\partial_i\partial_j u_n|^2 dm &= -\sum_{i,j=1}^d \int_{\mathbb{R}^d} \partial_j u_n \partial_i^2 \partial_j u_n dm \\ &= -\sum_{j=1}^d \int_{\mathbb{R}^d} \partial_j u_n \partial_j \Delta u_n dm = \sum_{j=1}^d \int_{\mathbb{R}^d} \partial_j^2 u_n \Delta u_n dm \\ &= \int_{\mathbb{R}^d} |\Delta u_n|^2 dm = \|\Delta u_n\|_{L^2}^2.\end{aligned}$$

Replacing u_n by $u_n - u_m$ in this calculation shows

$$\sum_{i,j=1}^d \int_{\mathbb{R}^d} |\partial_i\partial_j(u_n - u_m)|^2 dm = \|\Delta(u_n - u_m)\|_{L^2}^2 \to 0 \text{ as } m, n \to \infty$$

and therefore by Lemma 41.4 (also see Exercise 41.4), $\partial_i\partial_j u \in L^2(\mathbb{R}^d)$ for all i, j and

$$\sum_{i,j=1}^d \int_{\mathbb{R}^d} |\partial_i\partial_j u|^2 dm = \|\Delta u\|_{L^2}^2 = \|f\|_{L^2}^2. \tag{41.12}$$

Combining Eqs. (41.11) and (41.12) gives the estimate

$$\begin{aligned}\sum_{|\alpha|\le 2} \|\partial^\alpha u\|_{L^2}^2 &\le \|u\|_{L^2}^2 + \|f\|_{L^2} \cdot \|u\|_{L^2} + \|f\|_{L^2}^2 \\ &= \|u\|_{L^2}^2 + \|\Delta u\|_{L^2} \cdot \|u\|_{L^2} + \|\Delta u\|_{L^2}^2. \end{aligned} \tag{41.13}$$

Let us now further assume $\partial_i f = \partial_i \Delta u \in L^2(\mathbb{R}^d)$. Then for $h \in \mathbb{R}\setminus\{0\}$, $\partial_i^h u \in L^2(\mathbb{R}^d)$ and $\Delta\partial_i^h u = \partial_i^h \Delta u = \partial_i^h f \in L^2(\mathbb{R}^d)$ and hence by Eq. (41.13) and what we have just proved, $\partial^\alpha \partial_i^h u = \partial_i^h \partial^\alpha u \in L^2$ and

$$\begin{aligned}\sum_{|\alpha|\le 2} \left\|\partial_i^h \partial^\alpha u\right\|_{L^2(\mathbb{R}^d)}^2 &\le \left\|\partial_i^h u\right\|_{L^2}^2 + \left\|\partial_i^h f\right\|_{L^2} \cdot \left\|\partial_i^h u\right\|_{L^2} + \left\|\partial_i^h f\right\|_{L^2}^2 \\ &\le \|\partial_i u\|_{L^2}^2 + \|\partial_i f\|_{L^2} \cdot \|\partial_i u\|_{L^2} + \|\partial_i f\|_{L^2}^2\end{aligned}$$

where the last inequality follows from Corollary 41.20. Therefore applying Theorem 41.18 again we learn that $\partial_i\partial^\alpha u \in L^2(\mathbb{R}^d)$ for all $|\alpha| \le 2$ and

$$\begin{aligned}\sum_{|\alpha|\le 2} \|\partial_i \partial^\alpha u\|_{L^2(\mathbb{R}^d)}^2 &\le \|\partial_i u\|_{L^2}^2 + \|\partial_i f\|_{L^2} \cdot \|\partial_i u\|_{L^2} + \|\partial_i f\|_{L^2}^2 \\ &\le \|\nabla u\|_{L^2}^2 + \|\partial_i f\|_{L^2} \cdot \|\nabla u\|_{L^2} + \|\partial_i f\|_{L^2}^2 \\ &\le \|f\|_{L^2} \cdot \|u\|_{L^2} \\ &\quad + \|\partial_i f\|_{L^2} \cdot \sqrt{\|f\|_{L^2} \cdot \|u\|_{L^2}} + \|\partial_i f\|_{L^2}^2.\end{aligned}$$

The remainder of the proof, which is now an induction argument using the above ideas, is left as an exercise to the reader. ■

Theorem 41.22. *Suppose that Ω is an open subset of $\mathbb{R}^d$ and V is an open precompact subset of Ω.*

1. *If $1 \le p < \infty$, $u \in L^p(\Omega)$ and $\partial_i u \in L^p(\Omega)$, then $\|\partial_i^h u\|_{L^p(V)} \le \|\partial_i u\|_{L^p(\Omega)}$ for all $0 < |h| < \frac{1}{2}\text{dist}(V, \Omega^c)$.*
2. *Suppose that $1 < p \le \infty$, $u \in L^p(\Omega)$ and assume there exists a constants $C_V < \infty$ and $\varepsilon_V \in (0, \frac{1}{2}\text{dist}(V, \Omega^c))$ such that*
$$\|\partial_i^h u\|_{L^p(V)} \le C_V \text{ for all } 0 < |h| < \varepsilon_V.$$
Then $\partial_i u \in L^p(V)$ and $\|\partial_i u\|_{L^p(V)} \le C_V$. Moreover if $C := \sup_{V \subset\subset \Omega} C_V < \infty$ then in fact $\partial_i u \in L^p(\Omega)$ and $\|\partial_i u\|_{L^p(\Omega)} \le C$.

Proof. 1. Let $U \subset_o \Omega$ such that $\bar{V} \subset U$ and $\bar{U}$ is a compact subset of Ω. For $u \in C^1(\Omega) \cap L^p(\Omega)$, $x \in B$ and $0 < |h| < \frac{1}{2}\text{dist}(V, U^c)$,
$$\partial_i^h u(x) = \frac{u(x + he_i) - u(x)}{h} = \int_0^1 \partial_i u(x + the_i)\ dt$$
and in particular,
$$|\partial_i^h u(x)| \le \int_0^1 |\partial u(x + the_i)| dt.$$
Therefore by Minikowski's inequality for integrals,
$$\|\partial_i^h u\|_{L^p(V)} \le \int_0^1 \|\partial u(\cdot + the_i)\|_{L^p(V)} dt \le \|\partial_i u\|_{L^p(U)}. \tag{41.14}$$
For general $u \in L^p(\Omega)$ with $\partial_i u \in L^p(\Omega)$, by Proposition 41.12, there exists $u_n \in C_c^\infty(\Omega)$ such that $u_n \to u$ and $\partial_i u_n \to \partial_i u$ in $L^p_{loc}(\Omega)$. Therefore we may replace u by u_n in Eq. (41.14) and then pass to the limit to find
$$\|\partial_i^h u\|_{L^p(V)} \le \|\partial_i u\|_{L^p(U)} \le \|\partial_i u\|_{L^p(\Omega)}.$$
2. If $\|\partial_i^h u\|_{L^p(V)} \le C_V$ for all h sufficiently small then by Corollary 41.17 there exists $h_n \to 0$ such that $\partial_i^{h_n} u \overset{w}{\to} v \in L^p(V)$. Hence if $\varphi \in C_c^\infty(V)$,
$$\int_V v\varphi dm = \lim_{n\to\infty} \int_\Omega \partial_i^{h_n} u\varphi dm = \lim_{n\to\infty} \int_\Omega u\partial_i^{-h_n}\varphi dm = -\int_\Omega u\partial_i\varphi\ dm = -\int_V u\partial_i\varphi\ dm.$$
Therefore $\partial_i u = v \in L^p(V)$ and $\|\partial_i u\|_{L^p(V)} \le \|v\|_{L^p(V)} \le C_V$.[1] Finally if $C := \sup_{V\subset\subset\Omega} C_V < \infty$, then by the dominated convergence theorem,
$$\|\partial_i u\|_{L^p(\Omega)} = \lim_{V\uparrow\Omega} \|\partial_i u\|_{L^p(V)} \le C.$$
■

We will now give a couple of applications of Theorem 41.18.

Lemma 41.23. *Let $v \in \mathbb{R}^d$.*

1. *If $h \in L^1$ and $\partial_v h$ exists in L^1, then $\int_{\mathbb{R}^d} \partial_v h(x) dx = 0$.*
2. *If $p, q, r \in [1, \infty)$ satisfy $r^{-1} = p^{-1} + q^{-1}$, $f \in L^p$ and $g \in L^q$ are functions such that $\partial_v f$ and $\partial_v g$ exists in L^p and L^q respectively, then $\partial_v(fg)$ exists in L^r and $\partial_v(fg) = \partial_v f \cdot g + f \cdot \partial_v g$. Moreover if $r = 1$ we have the integration by parts formula,*
$$\langle \partial_v f, g\rangle = -\langle f, \partial_v g\rangle. \tag{41.15}$$
3. *If $p = 1$, $\partial_v f$ exists in L^1 and $g \in BC^1(\mathbb{R}^d)$ (i.e. $g \in C^1(\mathbb{R}^d)$ with g and its first derivatives being bounded) then $\partial_v(gf)$ exists in L^1 and $\partial_v(fg) = \partial_v f \cdot g + f \cdot \partial_v g$ and again Eq. (41.15) holds.*

Proof. 1) By item 3. of Theorem 41.18 there exists $h_n \in C_c^\infty(\mathbb{R}^d)$ such that $h_n \to h$ and $\partial_v h_n \to \partial_v h$ in L^1. Then
$$\int_{\mathbb{R}^d} \partial_v h_n(x) dx = \frac{d}{dt}|_0 \int_{\mathbb{R}^d} h_n(x + hv) dx = \frac{d}{dt}|_0 \int_{\mathbb{R}^d} h_n(x) dx = 0$$
and letting $n \to \infty$ proves the first assertion.

Alternatively, using the strong derivative notation we find,
$$\int_{\mathbb{R}^d} \partial_v h(x) dx = \int_{\mathbb{R}^d} L^1\text{-}\lim_{t\to 0} \frac{h(x + tv) - h(x)}{t} dx = \lim_{t\to 0} \int_{\mathbb{R}^d} \frac{h(x + tv) - h(x)}{t} dx = 0.$$

2) Similarly there exists $f_n, g_n \in C_c^\infty(\mathbb{R}^d)$ such that $f_n \to f$ and $\partial_v f_n \to \partial_v f$ in L^p and $g_n \to g$ and $\partial_v g_n \to \partial_v g$ in L^q as $n \to \infty$. So by the standard product rule and Remark 41.2, $f_n g_n \to fg \in L^r$ as $n \to \infty$ and

[1] Here we have used the result that if $f \in L^p$ and $f_n \in L^p$ such that $\langle f_n, \phi\rangle \to \langle f, \phi\rangle$ for all $\phi \in C_c^\infty(V)$, then $\|f\|_{L^p(V)} \le \liminf_{n\to\infty} \|f_n\|_{L^p(V)}$. To prove this, we have with $q = \frac{p}{p-1}$ that
$$|\langle f, \phi\rangle| = \lim_{n\to\infty} |\langle f_n, \phi\rangle| \le \liminf_{n\to\infty} \|f_n\|_{L^p(V)} \cdot \|\phi\|_{L^q(V)}$$
and therefore,
$$\|f\|_{L^p(V)} = \sup_{\phi\neq 0} \frac{|\langle f, \phi\rangle|}{\|\phi\|_{L^q(V)}} \le \liminf_{n\to\infty} \|f_n\|_{L^p(V)}.$$

$$\partial_v(f_n g_n) = \partial_v f_n \cdot g_n + f_n \cdot \partial_v g_n \to \partial_v f \cdot g + f \cdot \partial_v g \text{ in } L^r \text{ as } n \to \infty.$$

It now follows from another application of Theorem 41.18 that $\partial_v(fg)$ exists in L^r and $\partial_v(fg) = \partial_v f \cdot g + f \cdot \partial_v g$. Eq. (41.15) follows from this product rule and item 1. when $r = 1$.

3) Let $f_n \in C_c^\infty(\mathbb{R}^d)$ such that $f_n \to f$ and $\partial_v f_n \to \partial_v f$ in L^1 as $n \to \infty$. Then as above, $gf_n \to gf$ in L^1 and $\partial_v(gf_n) \to \partial_v g \cdot f + g\partial_v f$ in L^1 as $n \to \infty$. In particular if $\varphi \in C_c^\infty(\mathbb{R}^d)$, then

$$\begin{aligned}\langle gf, \partial_v \varphi\rangle &= \lim_{n\to\infty} \langle gf_n, \partial_v \varphi\rangle = -\lim_{n\to\infty} \langle \partial_v (gf_n), \varphi\rangle \\ &= -\lim_{n\to\infty} \langle \partial_v g \cdot f_n + g\partial_v f_n, \varphi\rangle = -\langle \partial_v g \cdot f + g\partial_v f, \varphi\rangle.\end{aligned}$$

This shows $\partial_v(fg)$ exists (weakly) and $\partial_v(fg) = \partial_v f \cdot g + f \cdot \partial_v g$. Again Eq. (41.15) holds in this case by item 1. already proved. ■

Lemma 41.24. *Let $p, q, r \in [1,\infty]$ satisfy $p^{-1} + q^{-1} = 1 + r^{-1}$, $f \in L^p$, $g \in L^q$ and $v \in \mathbb{R}^d$.*

1. *If $\partial_v f$ exists strongly in L^r, then $\partial_v(f * g)$ exists strongly in L^p and*
$$\partial_v(f * g) = (\partial_v f) * g.$$
2. *If $\partial_v g$ exists strongly in L^q, then $\partial_v(f * g)$ exists strongly in L^r and*
$$\partial_v(f * g) = f * \partial_v g.$$
3. *If $\partial_v f$ exists weakly in L^p and $g \in C_c^\infty(\mathbb{R}^d)$, then $f * g \in C^\infty(\mathbb{R}^d)$, $\partial_v(f * g)$ exists strongly in L^r and*
$$\partial_v(f * g) = f * \partial_v g = (\partial_v f) * g.$$

Proof. Items 1 and 2. By Young's inequality (Theorem 31.31) and simple computations:

$$\begin{aligned}&\left\|\frac{\tau_{-hv}(f * g) - f * g}{h} - (\partial_v f) * g\right\|_r \\ &= \left\|\frac{\tau_{-hv} f * g - f * g}{h} - (\partial_v f) * g\right\|_r \\ &= \left\|\left[\frac{\tau_{-hv} f - f}{h} - (\partial_v f)\right] * g\right\|_r \\ &\le \left\|\frac{\tau_{-hv} f - f}{h} - (\partial_v f)\right\|_p \|g\|_q\end{aligned}$$

which tends to zero as $h \to 0$. The second item is proved analogously, or just make use of the fact that $f * g = g * f$ and apply Item 1. Using the fact that $g(x - \cdot) \in C_c^\infty(\mathbb{R}^d)$ and the definition of the weak derivative,

$$\begin{aligned}f * \partial_v g(x) &= \int_{\mathbb{R}^d} f(y) \left(\partial_v g\right)(x - y) dy = -\int_{\mathbb{R}^d} f(y) \left(\partial_v g(x - \cdot)\right)(y) dy \\ &= \int_{\mathbb{R}^d} \partial_v f(y) g(x - y) dy = \partial_v f * g(x).\end{aligned}$$

Item 3. is a consequence of this equality and items 1. and 2. ■

Proposition 41.25. *Let $\Omega = (\alpha, \beta) \subset \mathbb{R}$ be an open interval and $f \in L^1_{loc}(\Omega)$ such that $\partial^{(w)} f = 0$ in $L^1_{loc}(\Omega)$. Then there exists $c \in \mathbb{C}$ such that $f = c$ a.e. More generally, suppose $F : C_c^\infty(\Omega) \to \mathbb{C}$ is a linear functional such that $F(\varphi') = 0$ for all $\varphi \in C_c^\infty(\Omega)$, where $\varphi'(x) = \frac{d}{dx}\varphi(x)$, then there exists $c \in \mathbb{C}$ such that*

$$F(\varphi) = \langle c, \varphi\rangle = \int_\Omega c\varphi(x) dx \text{ for all } \varphi \in C_c^\infty(\Omega). \quad (41.16)$$

Proof. Before giving a proof of the second assertion, let us show it includes the first. Indeed, if $F(\varphi) := \int_\Omega \varphi f dm$ and $\partial^{(w)} f = 0$, then $F(\varphi') = 0$ for all $\varphi \in C_c^\infty(\Omega)$ and therefore there exists $c \in \mathbb{C}$ such that

$$\int_\Omega \varphi f dm = F(\varphi) = c\langle \varphi, 1\rangle = c\int_\Omega \varphi f dm.$$

But this implies $f = c$ a.e. So it only remains to prove the second assertion. Let $\eta \in C_c^\infty(\Omega)$ such that $\int_\Omega \eta dm = 1$. Given $\varphi \in C_c^\infty(\Omega) \subset C_c^\infty(\mathbb{R})$, let $\psi(x) = \int_{-\infty}^x (\varphi(y) - \eta(y)\langle\varphi, 1\rangle) dy$. Then $\psi'(x) = \varphi(x) - \eta(x)\langle \varphi, 1\rangle$ and $\psi \in C_c^\infty(\Omega)$ as the reader should check. Therefore,

$$0 = F(\psi) = F(\varphi - \langle \varphi, \eta\rangle \eta) = F(\varphi) - \langle \varphi, 1\rangle F(\eta)$$

which shows Eq. (41.16) holds with $c = F(\eta)$. This concludes the proof, however it will be instructive to give another proof of the first assertion.

Alternative proof of first assertion. Suppose $f \in L^1_{loc}(\Omega)$ and $\partial^{(w)} f = 0$ and $f_m := f * \eta_m$ as is in the proof of Lemma 41.9. Then $f_m' = \partial^{(w)} f * \eta_m = 0$, so $f_m = c_m$ for some constant $c_m \in \mathbb{C}$. By Theorem 31.33, $f_m \to f$ in $L^1_{loc}(\Omega)$ and therefore if $J = [a, b]$ is a compact subinterval of Ω,

$$|c_m - c_k| = \frac{1}{b - a}\int_J |f_m - f_k| \, dm \to 0 \text{ as } m, k \to \infty.$$

So $\{c_m\}_{m=1}^\infty$ is a Cauchy sequence and therefore $c := \lim_{m\to\infty} c_m$ exists and $f = \lim_{m\to\infty} f_m = c$ a.e. ■

We will say more about the connection of weak derivatives to pointwise derivatives in Section 25.5 below.

41.2 Exercises

Exercise 41.1. Give another proof of Lemma 41.10 base on Proposition 41.12.

Exercise 41.2. Prove Proposition 41.13. **Hints:** 1. Use u_ε as defined in the proof of Proposition 41.12 to show it suffices to consider the case where $u \in C^\infty(\mathbb{R}^d) \cap L^q(\mathbb{R}^d)$ with $\partial^\alpha u \in L^q(\mathbb{R}^d)$ for all $\alpha \in \mathbb{N}_0^d$. 2. Then let $\psi \in C_c^\infty(B(0,1),[0,1])$ such that $\psi = 1$ on a neighborhood of 0 and let $u_n(x) := u(x)\psi(x/n)$.

Exercise 41.3. Suppose $p(\xi)$ is a polynomial in $\xi \in \mathbb{R}^d$, $p \in (1,\infty)$, $q := \frac{p}{p-1}$, $u \in L^p$ such that $p(\partial)u \in L^p$ and $v \in L^q$ such that $p(-\partial)v \in L^q$. Show $\langle p(\partial)u, v\rangle = \langle u, p(-\partial)v\rangle$.

Exercise 41.4. Let $p \in [1,\infty)$, α be a multi index (if $\alpha = 0$ let ∂^0 be the identity operator on L^p),

$$D(\partial^\alpha) := \{f \in L^p(\mathbb{R}^n) : \partial^\alpha f \text{ exists weakly in } L^p(\mathbb{R}^n)\}$$

and for $f \in D(\partial^\alpha)$ (the domain of ∂^α) let $\partial^\alpha f$ denote the α – weak derivative of f. (See Definition 41.3.)

1. Show ∂^α is a densely defined operator on L^p, i.e. $D(\partial^\alpha)$ is a dense linear subspace of L^p and $\partial^\alpha : D(\partial^\alpha) \to L^p$ is a linear transformation.
2. Show $\partial^\alpha : D(\partial^\alpha) \to L^p$ is a closed operator, i.e. the graph,
$$\Gamma(\partial^\alpha) := \{(f, \partial^\alpha f) \in L^p \times L^p : f \in D(\partial^\alpha)\},$$
is a closed subspace of $L^p \times L^p$.
3. Show $\partial^\alpha : D(\partial^\alpha) \subset L^p \to L^p$ is not bounded unless $\alpha = 0$. (The norm on $D(\partial^\alpha)$ is taken to be the L^p – norm.)

Exercise 41.5. Let $p \in [1,\infty)$, $f \in L^p$ and α be a multi index. Show $\partial^\alpha f$ exists weakly (see Definition 41.3) in L^p iff there exists $f_n \in C_c^\infty(\mathbb{R}^n)$ and $g \in L^p$ such that $f_n \to f$ and $\partial^\alpha f_n \to g$ in L^p as $n \to \infty$. **Hints:** See exercises 41.2 and 41.4.

Exercise 41.6. 8.8 on p. 246.

Exercise 41.7. Assume $n = 1$ and let $\partial = \partial_{e_1}$ where $e_1 = (1) \in \mathbb{R}^1 = \mathbb{R}$.

1. Let $f(x) = |x|$, show ∂f exists weakly in $L^1_{loc}(\mathbb{R})$ and $\partial f(x) = \operatorname{sgn}(x)$ for m – a.e. x.
2. Show $\partial(\partial f)$ does **not** exists weakly in $L^1_{loc}(\mathbb{R})$.
3. Generalize item 1. as follows. Suppose $f \in C(\mathbb{R},\mathbb{R})$ and there exists a finite set $\Lambda := \{t_1 < t_2 < \cdots < t_N\} \subset \mathbb{R}$ such that $f \in C^1(\mathbb{R}\setminus\Lambda,\mathbb{R})$. Assuming $\partial f \in L^1_{loc}(\mathbb{R})$, show ∂f exists weakly and $\partial^{(w)} f(x) = \partial f(x)$ for m – a.e. x.

Exercise 41.8. Suppose that $f \in L^1_{loc}(\Omega)$ and $v \in \mathbb{R}^d$ and $\{e_j\}_{j=1}^n$ is the standard basis for $\mathbb{R}^d$. If $\partial_j f := \partial_{e_j} f$ exists weakly in $L^1_{loc}(\Omega)$ for all $j = 1,2,\ldots,n$ then $\partial_v f$ exists weakly in $L^1_{loc}(\Omega)$ and $\partial_v f = \sum_{j=1}^n v_j \partial_j f$.

Exercise 41.9. Suppose, $f \in L^1_{loc}(\mathbb{R}^d)$ and $\partial_v f$ exists weakly and $\partial_v f = 0$ in $L^1_{loc}(\mathbb{R}^d)$ for all $v \in \mathbb{R}^d$. Then there exists $\lambda \in \mathbb{C}$ such that $f(x) = \lambda$ for m – a.e. $x \in \mathbb{R}^d$. **Hint:** See steps 1. and 2. in the outline given in Exercise 41.10 below.

Exercise 41.10 (A generalization of Exercise 41.9). Suppose Ω is a connected open subset of $\mathbb{R}^d$ and $f \in L^1_{loc}(\Omega)$. If $\partial^\alpha f = 0$ weakly for $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| = N+1$, then $f(x) = p(x)$ for m – a.e. x where $p(x)$ is a polynomial of degree at most N. Here is an outline.

1. Suppose $x_0 \in \Omega$ and $\varepsilon > 0$ such that $C := C_{x_0}(\varepsilon) \subset \Omega$ and let η_n be a sequence of approximate δ – functions such $\operatorname{supp}(\eta_n) \subset B_0(1/n)$ for all n. Then for n large enough, $\partial^\alpha(f * \eta_n) = (\partial^\alpha f) * \eta_n$ on C for $|\alpha| = N+1$. Now use Taylor's theorem to conclude there exists a polynomial p_n of degree at most N such that $f_n = p_n$ on C.
2. Show $p := \lim_{n\to\infty} p_n$ exists on C and then let $n \to \infty$ in step 1. to show there exists a polynomial p of degree at most N such that $f = p$ a.e. on C.
3. Use Taylor's theorem to show if p and q are two polynomials on $\mathbb{R}^d$ which agree on an open set then $p = q$.
4. Finish the proof with a connectedness argument using the results of steps 2. and 3. above.

Exercise 41.11. Suppose $\Omega \subset_o \mathbb{R}^d$ and $v, w \in \mathbb{R}^d$. Assume $f \in L^1_{loc}(\Omega)$ and that $\partial_v\partial_w f$ exists weakly in $L^1_{loc}(\Omega)$, show $\partial_w\partial_v f$ also exists weakly and $\partial_w\partial_v f = \partial_v\partial_w f$.

Exercise 41.12. Let $d = 2$ and $f(x,y) = 1_{x\geq 0}$. Show $\partial^{(1,1)} f = 0$ weakly in L^1_{loc} despite the fact that $\partial_1 f$ does not exist weakly in L^1_{loc}!

Exercise 41.13. Let $f \in L^1_{loc}(\mathbb{R})$ such that $f' = 0$ weakly in $L^1_{loc}(\mathbb{R})$, i.e. $\langle f, \varphi'\rangle = 0$ for all $\varphi \in C_c^\infty(\mathbb{R})$. Show there exists $c \in \mathbb{C}$ such that $f = c$ a.e. [**Hint:** you might use convolution to reduce the problem to the case where f is smooth.]

Exercise 41.14. Let $f \in L^1_{loc}(\mathbb{R})$. Then $\partial^w f = g$ exists in $L^1_{loc}(\mathbb{R})$ iff f has a continuous version $\tilde{f}$ which is absolutely continuous on all compact subintervals of $\mathbb{R}$. Moreover, $\partial^w f = \tilde{f}'$ a.e., where $\tilde{f}'(x)$ is the usual pointwise derivative. [**Hint:** when $\partial^w f = g$ exists in $L^1_{loc}(\mathbb{R})$ consider $F := f(x) - \int_0^x g(y)\,dy$ and make use of Exercise 41.13.]

Conditional Expectation

Now that we have defined the Bochner integral it is time to move on to conditional expectations. For now let us now suppose that μ is a probability measure on $(\Omega, \mathcal{F})$ and denote the Bochner integral, $\mu(F) = \int_\Omega F d\mu$ by $\mathbb{E}F$. Given a sub-σ − algebra $\mathcal{G}$ of $\mathcal{F}$ we would like to define $\mathbb{E}[F|\mathcal{G}]$ as an element of $L^1(\Omega, \mathcal{G}, \mu; X)$ for each $F \in L^1(\Omega, \mathcal{F}, \mu; X)$. As with the Bochner integral we will start with simple functions. For any $F \in \mathcal{S}(\mu; X)$ we have the identity,

$$F = \sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} x 1_{\{F=x\}}$$

and hence it is reasonable to require,

$$\mathbb{E}_{\mathcal{G}} F = \mathbb{E}[F|\mathcal{G}] = \mathbb{E}\left[\sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} x 1_{\{F=x\}} |\mathcal{G}\right] = \sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} \mathbb{E}\left[x 1_{\{F=x\}} |\mathcal{G}\right].$$

Furthermore, since $x \in X$ is constant in the previous formula we should further require,

$$\mathbb{E}_{\mathcal{G}} F = \sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} x \cdot \mathbb{E}\left[1_{\{F=x\}} |\mathcal{G}\right]. \tag{42.1}$$

Proposition 42.1. *If $\mathbb{E}_{\mathcal{G}} : \mathcal{S}(\mu; X) \to L^1(\Omega, \mathcal{G}, \mu; X)$ is the map defined by Eq. (42.1), then;*

1. *$\mathbb{E}_{\mathcal{G}}$ is linear*
2. *$\|\mathbb{E}_{\mathcal{G}} F\|_{L^1(\mu)} \le \|F\|_{L^1(\mu)}$ for all $F \in \mathcal{S}(\mu; X)$, i.e. $\mathbb{E}_{\mathcal{G}}$ is a contraction,*
3. *$\mathbb{E}_{\mathcal{G}} F$ satisfies,*
$$\mathbb{E}[\mathbb{E}_{\mathcal{G}} F : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}$$
and
$$\varphi \circ \mathbb{E}_{\mathcal{G}} F = \mathbb{E}_{\mathcal{G}}[\varphi \circ F] \text{ a.s. for all } \varphi \in X^*.$$

Proof. 1) If $0 \neq c \in \mathbb{R}$ and $F \in \mathcal{S}(\mu; X)$, then

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(cF) &= \sum_{x \in X} x \mathbb{E}\left[1_{\{cF=x\}} |\mathcal{G}\right] = \sum_{x \in X} x \mathbb{E}\left[1_{\{F=c^{-1}x\}} |\mathcal{G}\right] \\ &= \sum_{y \in X} cy\, \mathbb{E}\left[1_{\{F=y\}} |\mathcal{G}\right] = c\mathbb{E}_{\mathcal{G}}(F)\end{aligned}$$

and if $c = 0$, $\mathbb{E}_{\mathcal{G}}(0F) = 0 = 0\mathbb{E}_{\mathcal{G}}(F)$.

If $F, G \in \mathcal{S}(\mu; X)$,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(F+G) &= \sum_x x \mathbb{E}_{\mathcal{G}}\left[1_{F+G=x}\right] \\ &= \sum_x x \sum_{y+z=x} \mathbb{E}_{\mathcal{G}}\left[1_{F=y\ \&\ G=z}\right] \\ &= \sum_{y,z} (y+z) \mathbb{E}_{\mathcal{G}}\left[1_{F=y\ \&\ G=z}\right] \\ &= \sum_y y \mathbb{E}_{\mathcal{G}}\left[1_{F=y}\right] + \sum_z z \mathbb{E}_{\mathcal{G}}\left[1_{G=z}\right] \\ &= \mathbb{E}_{\mathcal{G}}(F) + \mathbb{E}_{\mathcal{G}}(G).\end{aligned}$$

2) Integrating the pointwise inequality,

$$\|\mathbb{E}_{\mathcal{G}} F\| = \sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} \|x\| \cdot \mathbb{E}\left[1_{\{F=x\}} |\mathcal{G}\right]$$

implies,

$$\begin{aligned}\|\mathbb{E}_{\mathcal{G}} F\|_{L^1(\mu)} &\le \sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} \|x\| \cdot \mathbb{E}\left(\mathbb{E}\left[1_{\{F=x\}} |\mathcal{G}\right]\right) \\ &= \sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} \|x\| \cdot \mathbb{E}\left(1_{\{F=x\}}\right) \\ &= \mathbb{E}\left(\sum_{x \in \operatorname{Ran}(F)\setminus\{0\}} \|x\|\, 1_{\{F=x\}}\right) \\ &= \mathbb{E}\|F\| = \|F\|_{L^1(\mu)}.\end{aligned}$$

3. If $A \in \mathcal{G}$ we have,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}F:A\right] &= \mathbb{E}\left[\sum_{x\neq 0} x\cdot\mathbb{E}\left[1_{\{F=x\}}|\mathcal{G}\right]\cdot 1_A\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} x\cdot\mathbb{E}\left[1_A 1_{\{F=x\}}|\mathcal{G}\right]\right] \\ &= \sum_{x\neq 0}\mathbb{E}\left[x\cdot\mathbb{E}\left[1_A 1_{\{F=x\}}|\mathcal{G}\right]\right] \\ &= \sum_{x\neq 0} x\cdot\mathbb{E}\left[\mathbb{E}\left[1_A 1_{\{F=x\}}|\mathcal{G}\right]\right] \\ &= \sum_{x\neq 0} x\cdot\mathbb{E}\left[1_A 1_{\{F=x\}}\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} x\cdot 1_A 1_{\{F=x\}}\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0} x\cdot 1_{\{1_A F=x\}}\right] \\ &= \mathbb{E}\left[F:A\right].\end{aligned}$$

where we have used Example **??** two times in the above computation and all of the above sums are really over $x\in\mathrm{Ran}\,(F)\setminus\{0\}$. Finally if $\varphi\in X^*$ we have,

$$\begin{aligned}\varphi\circ\mathbb{E}_{\mathcal{G}}F &= \sum_{x\neq 0}\varphi(x)\cdot\mathbb{E}\left[1_{\{F=x\}}|\mathcal{G}\right] \\ &= \mathbb{E}\left[\sum_{x\neq 0}\varphi(x)\cdot 1_{\{F=x\}}|\mathcal{G}\right] = \mathbb{E}\left[\varphi\circ F|\mathcal{G}\right].\end{aligned}$$

■

We may now apply the bounded linear transformation Theorem 32.4 in order to extend $\mathbb{E}_{\mathcal{G}}$ to all of $L^1(\Omega,\mathcal{F},\mu;X)$.

Theorem 42.2 (Conditional expectation). *Let $F\in L^1(\Omega,\mathcal{F},\mu;X)$. There is a linear map, $\mathbb{E}_{\mathcal{G}}:L^1(\Omega,\mathcal{F},\mu;X)\to L^1(\Omega,\mathcal{G},\mu;X)$, such that $\mathbb{E}_{\mathcal{G}}F$ is uniquely determined by either;*

1. *$\mathbb{E}_{\mathcal{G}}F$ is the unique element in $L^1(\Omega,\mathcal{G},\mu;X)$ such that*
$$\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}F:A\right]=\mathbb{E}\left[F:A\right] \text{ for all } A\in\mathcal{G}$$
or
2. *$\mathbb{E}_{\mathcal{G}}F$ is the unique element in $L^1(\Omega,\mathcal{G},\mu;X)$ such that $\varphi\circ\mathbb{E}_{\mathcal{G}}F=\mathbb{E}_{\mathcal{G}}[\varphi\circ F]$ a.s. for all $\varphi\in X^*$.*

Moreover, $\mathbb{E}_{\mathcal{G}}:L^1(\Omega,\mathcal{F},\mu;X)\to L^1(\Omega,\mathcal{G},\mu;X)$ is a contraction.

Proof. The existence of contraction $\mathbb{E}_{\mathcal{G}}:L^1(\Omega,\mathcal{F},\mu;X)\to L^1(\Omega,\mathcal{G},\mu;X)$ with the desired properties easily follows from Propositions **??** and 42.1 along with the bounded linear transformation Theorem 32.4. So it only remains to verify that $\mathbb{E}_{\mathcal{G}}F$ is uniquely determined by either of the two conditions above.

1) If $G\in L^1(\Omega,\mathcal{G},\mu;X)$ satisfies $\mathbb{E}[G:A]=\mathbb{E}[F:A]=\mathbb{E}[\mathbb{E}_{\mathcal{G}}F:A]$ for all $A\in\mathcal{G}$ then $\mathbb{E}[G-\mathbb{E}_{\mathcal{G}}F:A]=0$ for all $A\in\mathcal{G}$. If $G\neq\mathbb{E}_{\mathcal{G}}F$ a.s., we may use Lemma **??** in order to find $A\in\mathcal{G}$ with $\mu(A)>0$ and $\varphi\in X^*$ such that $\varphi\circ(G-\mathbb{E}_{\mathcal{G}}F)>0$ on A. We then may conclude,

$$0=\varphi(0)=\varphi\left(\mathbb{E}\left[G-\mathbb{E}_{\mathcal{G}}F:A\right]\right)=\mathbb{E}\left[\varphi\left(G-\mathbb{E}_{\mathcal{G}}F\right):A\right]>0$$

which is absurd and hence we must have $G\neq\mathbb{E}_{\mathcal{G}}F$ a.s.

2) If $G\in L^1(\Omega,\mathcal{G},\mu;X)$ satisfies $\varphi\circ G=\mathbb{E}_{\mathcal{G}}[\varphi\circ F]=\varphi\circ\mathbb{E}_{\mathcal{G}}F$ a.s. for all $\varphi\in X^*$ then $\varphi(G-\mathbb{E}_{\mathcal{G}}F)=0$ a.s. for all $\varphi\in X^*$ and the result follows from Corollary **??**. Let us recall the proof here. As in the proof of Proposition **??**, there exists a countable subset, $\mathbb{D}\subset X^*$, such that $\|x\|=\sup_{\varphi\in\mathbb{D}}\varphi(x)$ for all $x\in X$. Therefore we may conclude,

$$\|G-\mathbb{E}_{\mathcal{G}}F\|=\sup_{\varphi\in\mathbb{D}}\varphi\left(G-\mathbb{E}_{\mathcal{G}}F\right)=0 \text{ a.s.},$$

i.e. $G=\mathbb{E}_{\mathcal{G}}F$ a.s.

Alternate proof. If $G\in L^1(\Omega,\mathcal{G},\mu;X)$ satisfies $\varphi\circ G=\mathbb{E}_{\mathcal{G}}[\varphi\circ F]=\varphi\circ\mathbb{E}_{\mathcal{G}}F$ a.s. for all $\varphi\in X^*$ then for any $A\in\mathcal{G}$ we have

$$\varphi\left(\mathbb{E}\left[G:A\right]\right)=\mathbb{E}\left[\varphi\circ G:A\right]=\mathbb{E}\left[\varphi\circ\mathbb{E}_{\mathcal{G}}F:A\right]=\varphi\left(\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}F:A\right]\right).$$

Therefore by the Hahn - Banach theorem X^* separates points and we may conclude that $\mathbb{E}[G:A]=\mathbb{E}[\mathbb{E}_{\mathcal{G}}F:A]$ and so by item 1., $G=\mathbb{E}_{\mathcal{G}}F$ a.s. ■

Proposition 42.3. *Let $\mathcal{G}\subset\mathcal{F}$ and $F\in L^1(\Omega,\mathcal{F},\mu;X)$. The conditional expectation operator $(\mathbb{E}_{\mathcal{G}})$ satisfies the following additional properties;*

1. *$\|\mathbb{E}_{\mathcal{G}}F\|\leq\mathbb{E}_{\mathcal{G}}\|F\|$ a.s..*
2. *$\|\mathbb{E}_{\mathcal{G}}F\|_{L^p(\mu)}\leq\|F\|_{L^p(\mu)}$ for all $F\in L^p(\Omega,\mathcal{F},\mu;X)$ where $1\leq p<\infty$.*
3. *If $h\in L^\infty(\Omega,\mathcal{G},\mu)$, then $\mathbb{E}_{\mathcal{G}}[hF]=h\cdot\mathbb{E}_{\mathcal{G}}[F]$ a.s.*
4. *If $\mathcal{G}_0\subset\mathcal{G}\subset\mathcal{F}$ then $\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}}=\mathbb{E}_{\mathcal{G}_0}=\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\mathcal{G}_0}$.*

Proof. We prove each item in turn.

1. If $F \in \mathcal{S}(\mu; X)$, then

$$\|\mathbb{E}_{\mathcal{G}} F\| = \left\| \sum_{x \neq 0} x \mathbb{E}_{\mathcal{G}} 1_{F=x} \right\| \leq \sum_{x \neq 0} \|x\| \, \mathbb{E}_{\mathcal{G}} 1_{F=x}$$
$$= \mathbb{E}_{\mathcal{G}} \left(\sum_{x \neq 0} \|x\| \, 1_{F=x} \right) = \mathbb{E}_{\mathcal{G}} \|F\|.$$

For general $F \in L^1(\Omega, \mathcal{F}, \mu; X)$ we may choose $F_n \in \mathcal{S}(\mu; X)$ such that $F_n \to F$ in L^1 and therefore, $\|F_n\| \to \|F\|$, $\|\mathbb{E}_{\mathcal{G}} F_n\| \to \|\mathbb{E}_{\mathcal{G}} F\|$, and $\mathbb{E}_{\mathcal{G}} \|F_n\| \to \mathbb{E}_{\mathcal{G}} \|F\|$ in $L^1(\mu)$ as $n \to \infty$. Thus we may pass to the limit in the inequality $\|\mathbb{E}_{\mathcal{G}} F_n\| \leq \mathbb{E}_{\mathcal{G}} \|F_n\|$ in order to show $\|\mathbb{E}_{\mathcal{G}} F\| \leq \mathbb{E}_{\mathcal{G}} \|F\|$ a.s..
Alternative proof. If $\varphi \in X^*$ with $\|\varphi\|_{X^*} = 1$, we have

$$|\varphi \circ \mathbb{E}_{\mathcal{G}} F| = |\mathbb{E}_{\mathcal{G}} [\varphi \circ F]| \leq \mathbb{E}_{\mathcal{G}} |[\varphi \circ F]| \leq \mathbb{E}_{\mathcal{G}} \|F\| \text{ a.s.}$$

Therefore if we let $\{\varphi_n\}$ be as defined in the proof of Proposition **??**, then

$$\|\mathbb{E}_{\mathcal{G}} F\| = \sup_n |\varphi_n \circ \mathbb{E}_{\mathcal{G}} F| \leq \mathbb{E}_{\mathcal{G}} \|F\| \text{ a.s.}$$

2. For the second item we have,

$$\|\mathbb{E}_{\mathcal{G}} F\|^p_{L^p(\mu)} = \mathbb{E}\left[\|\mathbb{E}_{\mathcal{G}} F\|^p\right] \leq \mathbb{E}\left[(\mathbb{E}_{\mathcal{G}} \|F\|)^p\right] \leq \mathbb{E}\left[\mathbb{E}_{\mathcal{G}}\left[\|F\|^p\right]\right] = \mathbb{E}\|F\|^p.$$

3. If $\varphi \in X^*$ then

$$\varphi(h \cdot \mathbb{E}_{\mathcal{G}}[F]) = h \cdot \varphi \circ \mathbb{E}_{\mathcal{G}}[F] = h \cdot \mathbb{E}_{\mathcal{G}}[\varphi \circ F]$$
$$= \mathbb{E}_{\mathcal{G}}[h \cdot \varphi \circ F] = \mathbb{E}_{\mathcal{G}}[\varphi(h \cdot F)]$$
$$= \varphi(\mathbb{E}_{\mathcal{G}}[h \cdot F]).$$

Since $\varphi \in X^*$ was arbitrary it follows that $\mathbb{E}_{\mathcal{G}}[h \cdot F] = h \cdot \mathbb{E}_{\mathcal{G}}[F]$ a.s.
4. If $\varphi \in X^*$ then

$$\varphi(\mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}} F) = \mathbb{E}_{\mathcal{G}_0}[\varphi(\mathbb{E}_{\mathcal{G}} F)] = \mathbb{E}_{\mathcal{G}_0}[\mathbb{E}_{\mathcal{G}}(\varphi \circ F)]$$
$$= \mathbb{E}_{\mathcal{G}_0}(\varphi \circ F) = \varphi \circ \mathbb{E}_{\mathcal{G}_0} F$$

and similarly,

$$\varphi(\mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathcal{G}_0} F) = \varphi \circ \mathbb{E}_{\mathcal{G}_0} F.$$

Since $\varphi \in X^*$ was arbitrary the result follows.

■

42.1 Basic Martingale Results

In this section we will write μ for P. Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \leq p < \infty$, and let $\mathcal{F}_\infty := \vee_{n=1}^\infty \mathcal{F}_n := \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$. We will say that $M_n : \Omega \to X$ is a **martingale** if $\{M_n\}_{n=1}^\infty$ is an adapted integrable process such that $\mathbb{E}_{\mathcal{B}_n} M_{n+1} = M_n$ a.s. Notice that if $\{M_n\}$ is a martingale then $X_n := \|M_n\|$ is a positive submartingale.

Lemma 42.4. *The space $\cup_{n=1}^\infty L^p(\Omega, \mathcal{F}_n, P)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P)$.*

Proof. The spaces $L^p(\Omega, \mathcal{F}_n, P)$ form an increasing sequence of closed subspaces of $L^p(\Omega, \mathcal{F}_\infty, P)$. Further let $\mathbb{A}$ be the algebra of functions consisting of those $f \in \cup_{n=1}^\infty L^p(\Omega, \mathcal{F}_n, P)$ such that f is bounded. As a consequence of the density Theorem 16.10 (from the probability notes), we know that $\mathbb{A}$ and hence $\cup_{n=1}^\infty L^p(\Omega, \mathcal{F}_n, P)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P)$. This completes the proof. However for the readers convenience let us quickly review the proof of Theorem 16.10 (from the probability notes) in this context.

Let $\mathbb{H}$ denote those bounded $\mathcal{F}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{A}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded $\mathcal{F}_\infty$ –measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{A} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated (or bounded) convergence theorem, $\lim_{n\to\infty} \|(f - f_n)\|_{L^p(P)} = 0.$[1] We may now choose $\varphi_n \in \mathbb{A}$ such that $\|\varphi_n - f_n\|_{L^p(P)} \leq \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} \leq \limsup_{n\to\infty} \|(f - f_n)\|_{L^p(P)} + \limsup_{n\to\infty} \|f_n - \varphi_n\|_{L^p(P)} = 0,$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem, now shows $\mathbb{H}$ contains all bounded $\sigma(\mathbb{A}) = \mathcal{F}_\infty$ – measurable functions on Ω. Since for any $f \in L^p(\Omega, \mathcal{F}, P)$, $f 1_{|f| \leq n} \in \mathbb{H}$ there exists $\varphi_n \in \mathbb{A}$ such that $\|f_n - \varphi_n\|_p \leq n^{-1}$. Using the DCT we know that $f_n \to f$ in L^p and therefore by Minikowski's inequality it follows that $\varphi_n \to f$ in L^p. ■

Corollary 42.5. *The space $\cup_{n=1}^\infty L^p(\Omega, \mathcal{F}_n, P; X)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P; X)$.*

Proof. Since $\mathcal{S}(\Omega, \mathcal{F}_\infty, P; X)$ is dense in $L^p(\Omega, \mathcal{F}_\infty, P; X)$ it suffices to show that every element of $\mathcal{S}(\Omega, \mathcal{F}_\infty, P; X)$ is well approximated by some $G \in \cup_{n=1}^\infty L^p(\Omega, \mathcal{F}_n, P; X)$ and for this it suffices to show $1_A \cdot x$ is well approximated by by some $G \in \cup_{n=1}^\infty L^p(\Omega, \mathcal{F}_n, P; X)$ for all $x \in X$ and $A \in \mathcal{F}_\infty$.

[1] It is at this point that the proof would break down if $p = \infty$.

But as a consequence Lemma 42.4 we may find $h \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P)$ such that $\|h - 1_A\|_{L^p(P)}$ is as small as we please and therefore

$$\|1_A \cdot x - h \cdot x\|_{L^p(P)} \le \|x\| \cdot \|h - 1_A\|_{L^p(P)}$$

can be made as small as we please as well. ■

Theorem 42.6. *For every $F \in L^p(\Omega, \mathcal{F}, P)$, $M_n = \mathbb{E}[F|\mathcal{F}_n]$ is a martingale and $M_n \to M_\infty := \mathbb{E}[F|\mathcal{F}_\infty]$ in $L^p(\Omega, \mathcal{F}_\infty, P; X)$ as $n \to \infty$.*

Proof. The tower property immediately shows $M_n = \mathbb{E}[F|\mathcal{F}_n]$ is a martingale. Since conditional expectation is a contraction on L^p it follows that $\mathbb{E}\|M_n\|^p \le \mathbb{E}\|F\|^p < \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$. So to finish the proof we need to show $M_n \to M_\infty$ in $L^p(\Omega, \mathcal{F}, P; X)$ as $n \to \infty$.

If $F \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$, then $M_n = F$ for all sufficiently large n and for $n = \infty$ and the result holds. Now suppose that $F \in L^p(\Omega, \mathcal{F}_\infty, P; X)$ and $G \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$. Then

$$\begin{aligned}\|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_n} F\|_p &\le \|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_\infty} G\|_p + \|\mathbb{E}_{\mathcal{F}_\infty} G - \mathbb{E}_{\mathcal{F}_n} G\|_p + \|\mathbb{E}_{\mathcal{F}_n} G - \mathbb{E}_{\mathcal{F}_n} F\|_p \\ &\le 2\|F - G\|_p + \|\mathbb{E}_{\mathcal{F}_\infty} G - \mathbb{E}_{\mathcal{F}_n} G\|_p\end{aligned}$$

and hence

$$\limsup_{n\to\infty} \|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_n} F\|_p \le 2\|F - G\|_p.$$

Using the density Corollary 42.5 we may choose $G \in \cup_{n=1}^{\infty} L^p(\Omega, \mathcal{F}_n, P; X)$ as close to $F \in L^p(\Omega, \mathcal{F}_\infty, P; X)$ as we please and therefore it follows that $\limsup_{n\to\infty} \|\mathbb{E}_{\mathcal{F}_\infty} F - \mathbb{E}_{\mathcal{F}_n} F\|_p = 0$.

For general $F \in L^p(\Omega, \mathcal{F}, P)$ it suffices to observe that $M_\infty := \mathbb{E}[F|\mathcal{F}_\infty] \in L^p(\Omega, \mathcal{F}_\infty, P)$ and by the tower property of conditional expectations,

$$\mathbb{E}[M_\infty|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[F|\mathcal{F}_\infty]|\mathcal{F}_n] = \mathbb{E}[F|\mathcal{F}_n] = M_n.$$

So again $M_n \to M_\infty$ in L^p as desired. ■

The converse of Theorem 42.6 holds as well but is not really needed for our purposes. It use compactness results from the probability notes which need to be transferred here.

Theorem 42.7 (Probably should skip). *Suppose $1 \le p < \infty$ and $\{M_n\}_{n=1}^{\infty} \subset L^p(\Omega, \mathcal{F}, P; X)$ is a martingale. Further assume that $\sup_n \|M_n\|_p < \infty$ and that $\{M_n\}_{n=1}^{\infty}$ is uniformly integrable if $p = 1$. Then there exists $M_\infty \in L^p(\Omega, \mathcal{F}_\infty, P; X)$ such that $M_n := \mathbb{E}[M_\infty|\mathcal{F}_\infty]$. Moreover by Theorem 42.6 we know that $M_n \to M_\infty$ in $L^p(\Omega, \mathcal{F}_\infty, P)$ as $n \to \infty$ and hence M_∞ is uniquely determined by $\{M_n\}_{n=1}^{\infty}$.*

Proof. By Theorems **??** and **??** exists $M_\infty \in L^p(\Omega, \mathcal{F}_\infty, P)$ and a subsequence, $Y_k = M_{n_k}$ such that

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[M_\infty h] \text{ for all } h \in L^q(\Omega, \mathcal{F}_\infty, P)$$

where $q := p(p-1)^{-1}$. Using the martingale property, if $h \in (\mathcal{F}_n)_b$ for some n, it follows that $\mathbb{E}[Y_k h] = \mathbb{E}[M_n h]$ for all large k and therefore that

$$\mathbb{E}[M_\infty h] = \mathbb{E}[M_n h] \text{ for all } h \in (\mathcal{F}_n)_b.$$

This implies that $M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$ as desired. ■

Theorem 42.8 (Almost sure convergence). *Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{F}_\infty := \vee_{n=1}^{\infty} \mathcal{F}_n := \sigma(\cup_{n=1}^{\infty} \mathcal{F}_n)$. Then for every $F \in L^1(\Omega, \mathcal{F}, P; X)$, the martingale, $M_n = \mathbb{E}[F|\mathcal{F}_n]$, converges almost surely to $M_\infty := \mathbb{E}[F|\mathcal{F}_\infty]$.*

Proof. We follow the proof in Stroock [21, Corollary 5.2.7]. Let $\mathbb{H}$ denote those $F \in L^1(\Omega, \mathcal{F}_\infty, P; X)$ such that $M_n := \mathbb{E}[F|\mathcal{F}_n] \to M_\infty = F$ a.s. As we saw above $\mathbb{H}$ contains the dense subspace $\cup_{n=1}^{\infty} L^1(\Omega, \mathcal{F}_n, P; X)$. It is also easy to see that $\mathbb{H}$ is a linear space. Thus it suffices to show that $\mathbb{H}$ is closed in $L^1(P; X)$. To prove this let $F^{(k)} \in \mathbb{H}$ with $F^{(k)} \to F$ in $L^1(P)$ and let $M_n^{(k)} := \mathbb{E}\left[F^{(k)}|\mathcal{F}_n\right]$. Then by the Doob's maximal inequality applied to the sub-martingale $\left\{\left\|M_n - M_n^{(k)}\right\|\right\}_{n=1}^{\infty}$ we have

$$P\left(\sup_n \left\|M_n - M_n^{(k)}\right\| \ge a\right) \le \frac{1}{a} \sup_n \mathbb{E}\left\|M_n - M_n^{(k)}\right\| \le \frac{1}{a}\mathbb{E}\left\|F - F^{(k)}\right\|$$

for all $a > 0$ and $k \in \mathbb{N}$. Therefore,

$$\begin{aligned}&P\left(\sup_{n\ge N} \|F - M_n\| \ge 3a\right) \\ &\le P\left(\left\|F - F^{(k)}\right\| \ge a\right) + P\left(\sup_{n\ge N}\left\|F^{(k)} - M_n^{(k)}\right\| \ge a\right) \\ &\quad + P\left(\sup_{n\ge N}\left\|M_n^{(k)} - M_n\right\| \ge a\right) \\ &\le \frac{2}{a}\mathbb{E}\left\|F - F^{(k)}\right\| + P\left(\sup_{n\ge N}\left\|F^{(k)} - M_n^{(k)}\right\| \ge a\right)\end{aligned}$$

and hence

$$\limsup_{N\to\infty} P\left(\sup_{n\ge N} \|F - M_n\| \ge 3a\right) \le \frac{2}{a}\mathbb{E}\left\|F - F^{(k)}\right\| \to 0 \text{ as } k \to \infty.$$

Thus we have shown

$$\limsup_{N\to\infty} P\left(\sup_{n\geq N} \|F - M_n\| \geq 3a\right) = 0 \text{ for all } a > 0.$$

Since

$$\left\{\limsup_{n\to\infty} \|F - M_n\| \geq 3a\right\} \subset \left\{\sup_{n\geq N} \|F - M_n\| \geq 3a\right\} \text{ for all } N,$$

it follows that

$$P\left(\limsup_{n\to\infty} \|F - M_n\| \geq 3a\right) = 0 \text{ for all } a > 0$$

and therefore $\limsup_{n\to\infty} \|F - M_n\| = 0$ (P a.s.) which shows that $F \in \mathbb{H}$. ■

Part X

The Fourier Transform and Distributions

43

Fourier Transform

43.1 Motivation

We begin with a little motivation which will be fleshed out more in Exercise 43.10 below. Suppose for simplicity that $f \in C_c^1(\mathbb{R})$. For each $L > 0$ we know that $\left\{\chi_k^L(x) := e^{ikx/L} : k \in \mathbb{Z}\right\}$ is an orthonormal basis for $H_L := L^2([-\pi L, \pi L])$ equipped with the inner product

$$\langle f|g\rangle_L := \frac{1}{2\pi L}\int_{[-\pi L,\pi L]} f(x)\bar{g}(x)dx.$$

Therefore for L sufficiently large we have for $|x| \le \pi L$ that

$$\begin{aligned} f(x) &= \sum_{k\in\mathbb{Z}} \left\langle f|\chi_k^L\right\rangle_L \chi_k^L(x) \\ &= \frac{1}{\sqrt{2\pi}}\sum_{k\in\mathbb{Z}}\frac{1}{L}\left(\frac{1}{\sqrt{2\pi}}\int_{[-\pi L,\pi L]} f(y)e^{-iky/L}dy\right)\chi_k^L(x) \\ &= \frac{1}{\sqrt{2\pi}}\sum_{k\in\mathbb{Z}}\frac{1}{L}\hat{f}\left(\frac{k}{L}\right)e^{ikx/L} \end{aligned} \tag{43.1}$$

where

$$\hat{f}(\xi) := \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(y)e^{-i\xi y}dy$$

Moreover we have,and

$$\begin{aligned} \|f\|^2_{L^2(m)} &= 2\pi L\,\langle f|f\rangle_L = 2\pi L\sum_{k\in\mathbb{Z}}\left|\left\langle f|\chi_k^L\right\rangle_L\right|^2 \\ &= \frac{1}{2\pi L}\sum_{k\in\mathbb{Z}}\left|\int_{[-\pi L,\pi L]} f(y)e^{-iky/L}dy\right|^2 \\ &= \sum_{k\in\mathbb{Z}}\left|\hat{f}\left(\frac{k}{L}\right)\right|^2\frac{1}{L}. \end{aligned} \tag{43.2}$$

Formally passing to the limit in Eqs. (43.1) and (43.2) suffest that

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\hat{f}(\xi)\,e^{i\xi x}d\xi \text{ and } \|f\|^2_{L^2(m)} = \left\|\hat{f}\right\|^2_{L^2(m)}.$$

In short this predicts the Fourier transform, $f \to \hat{f}$, is a unitary operator on $L^2(\mathbb{R})$. We will eventually show this is the case after first showing how to interpert $\hat{f}$ for $f \in L^2(\mathbb{R})$.

We now generalize to the the n – dimensionsal case. The underlying space in this section is $\mathbb{R}^n$ with Lebesgue measure. As suggested above, the Fourier inversion formula is going to state that

$$f(x) = \left(\frac{1}{2\pi}\right)^n\int_{\mathbb{R}^n} d\xi e^{i\xi\cdot x}\left[\int_{\mathbb{R}^n} dy f(y)e^{-iy\cdot\xi}\right]. \tag{43.3}$$

If we let $\xi = 2\pi\eta$, this may be written as

$$f(x) = \int_{\mathbb{R}^n} d\eta e^{i2\pi\eta\cdot x}\int_{\mathbb{R}^n} dy f(y)e^{-i2\pi y\cdot\eta}$$

and we have removed the multiplicative factor of $\left(\frac{1}{2\pi}\right)^n$ in Eq. (43.3) at the expense of placing factors of 2π in the arguments of the exponentials. [This is what Folland does.] Another way to avoid writing the 2π's altogether is to redefine dx and $d\xi$ and this is what we will do here.

Notation 43.1 *Let m be Lebesgue measure on $\mathbb{R}^n$ and define:*

$$d\lambda(x) := \mathbf{d}x := \left(\frac{1}{\sqrt{2\pi}}\right)^n dm(x) \text{ and } \mathbf{d}\xi := \left(\frac{1}{\sqrt{2\pi}}\right)^n dm(\xi).$$

To be consistent with this new normalization of Lebesgue measure we will redefine $\|f\|_p$ and $\langle f, g\rangle$ as

$$\|f\|_p = \left(\int_{\mathbb{R}^n}|f(x)|^p\,\mathbf{d}x\right)^{1/p} = \left(\left(\frac{1}{2\pi}\right)^{n/2}\int_{\mathbb{R}^n}|f(x)|^p\,dm(x)\right)^{1/p}$$

and

$$\langle f, g\rangle := \int_{\mathbb{R}^n} f(x)g(x)\mathbf{d}x \text{ when } fg \in L^1.$$

Similarly we will define the convolution relative to these normalizations by $f\bigstar \mathbf{g} := \left(\frac{1}{2\pi}\right)^{n/2} f * g$, i.e.

$$f\bigstar g(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\mathbf{d}y = \int_{\mathbb{R}^n} f(x-y)g(y)\left(\frac{1}{2\pi}\right)^{n/2} dm(y).$$

The following notation will also be convenient; given a multi-index $\alpha \in \mathbb{Z}_+^n$, let $|\alpha| = \alpha_1 + \cdots + \alpha_n$,

$$x^\alpha := \prod_{j=1}^n x_j^{\alpha_j}, \ \partial_x^\alpha = \left(\frac{\partial}{\partial x}\right)^\alpha := \prod_{j=1}^n \left(\frac{\partial}{\partial x_j}\right)^{\alpha_j} \text{ and}$$
$$D_x^\alpha = \left(\frac{1}{i}\right)^{|\alpha|} \left(\frac{\partial}{\partial x}\right)^\alpha = \left(\frac{1}{i}\frac{\partial}{\partial x}\right)^\alpha.$$

Also let

$$\langle x \rangle := (1+|x|^2)^{1/2}$$

and for $s \in \mathbb{R}$ let

$$\nu_s(x) = (1+|x|)^s.$$

43.2 Fourier Transform formal development

Definition 43.2 (Fourier Transform). *For $f \in L^1$, let*

$$\hat{f}(\xi) = \mathcal{F}f(\xi) := \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\mathbf{d}x \tag{43.4}$$
$$g^\vee(x) = \mathcal{F}^{-1}g(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi} g(\xi)\mathbf{d}\xi = \mathcal{F}g(-x) \tag{43.5}$$

The next theorem summarizes some more basic properties of the Fourier transform.

Theorem 43.3. *Suppose that $f, g \in L^1$. Then*

1. *$\hat{f} \in C_0(\mathbb{R}^n)$ and $\left\|\hat{f}\right\|_\infty \le \|f\|_1$.*
2. *For $y \in \mathbb{R}^n$, $(\tau_y f)\hat{}(\xi) = e^{-iy\cdot\xi}\hat{f}(\xi)$ where, as usual, $\tau_y f(x) := f(x-y)$.*
3. *The Fourier transform takes convolution to products, i.e. $(f \bigstar g)\hat{} = \hat{f}\hat{g}$.*
4. *For $f, g \in L^1$, $\langle \hat{f}, g\rangle = \langle f, \hat{g}\rangle$.*
5. *If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, then*
$$(f \circ T)^\wedge(\xi) = |\det T|^{-1} \hat{f}\left(\left(T^{-1}\right)^* \xi\right) \text{ and}$$
$$(f \circ T)^\vee(\xi) = |\det T|^{-1} f^\vee\left(\left(T^{-1}\right)^* \xi\right)$$
6. *If $(1+|x|)^k f(x) \in L^1$, then $\hat{f} \in C^k$ and $\partial^\alpha \hat{f} \in C_0$ for all $|\alpha| \le k$. Moreover,*
$$\partial_\xi^\alpha \hat{f}(\xi) = \mathcal{F}\left[(-ix)^\alpha f(x)\right](\xi) \tag{43.6}$$
for all $|\alpha| \le k$.
7. *If $f \in C^k$ and $\partial^\alpha f \in L^1$ for all $|\alpha| \le k$, then $(1+|\xi|)^k \hat{f}(\xi) \in C_0$ and*
$$(\partial^\alpha f)\hat{}(\xi) = (i\xi)^\alpha \hat{f}(\xi) \tag{43.7}$$
for all $|\alpha| \le k$.
8. *Suppose $g \in L^1(\mathbb{R}^k)$ and $h \in L^1(\mathbb{R}^{n-k})$ and $f = g \otimes h$, i.e.*
$$f(x) = g(x_1, \dots, x_k)h(x_{k+1}, \dots, x_n),$$
then $\hat{f} = \hat{g} \otimes \hat{h}$.

Proof. Item 1. is the Riemann Lebesgue Lemma 31.40. Items 2. – 5. are proved by the following straight forward computations:

$$(\tau_y f)\hat{}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x-y)\mathbf{d}x = \int_{\mathbb{R}^n} e^{-i(x+y)\cdot\xi} f(x)\mathbf{d}x = e^{-iy\cdot\xi}\hat{f}(\xi),$$
$$\langle \hat{f}, g\rangle = \int_{\mathbb{R}^n} \hat{f}(\xi)g(\xi)\mathbf{d}\xi = \int_{\mathbb{R}^n} \mathbf{d}\xi g(\xi) \int_{\mathbb{R}^n} \mathbf{d}x e^{-ix\cdot\xi} f(x)$$
$$= \int_{\mathbb{R}^n\times\mathbb{R}^n} \mathbf{d}x\mathbf{d}\xi e^{-ix\cdot\xi} g(\xi)f(x) = \int_{\mathbb{R}^n\times\mathbb{R}^n} \mathbf{d}x\hat{g}(x)f(x) = \langle f, \hat{g}\rangle,$$
$$(f \bigstar g)\hat{}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f \bigstar g(x)\mathbf{d}x = \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\left(\int_{\mathbb{R}^n} f(x-y)g(y)\mathbf{d}y\right)\mathbf{d}x$$
$$= \int_{\mathbb{R}^n} \mathbf{d}y \int_{\mathbb{R}^n} \mathbf{d}x e^{-ix\cdot\xi} f(x-y)g(y)$$
$$= \int_{\mathbb{R}^n} \mathbf{d}y \int_{\mathbb{R}^n} \mathbf{d}x e^{-i(x+y)\cdot\xi} f(x)g(y)$$
$$= \int_{\mathbb{R}^n} \mathbf{d}y e^{-iy\cdot\xi} g(y) \int_{\mathbb{R}^n} \mathbf{d}x e^{-ix\cdot\xi} f(x) = \hat{f}(\xi)\hat{g}(\xi)$$

and letting $y = Tx$ so that $\mathbf{d}x = |\det T|^{-1}\mathbf{d}y$

$$(f \circ T)\hat{}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(Tx)\mathbf{d}x = \int_{\mathbb{R}^n} e^{-iT^{-1}y\cdot\xi} f(y)|\det T|^{-1}\mathbf{d}y$$
$$= |\det T|^{-1}\hat{f}\left(\left(T^{-1}\right)^*\xi\right).$$

Item 6. is simply a matter of differentiating under the integral sign which is easily justified because $(1+|x|)^k f(x) \in L^1$. Item 7. follows by using Lemma 31.39 repeatedly (i.e. integration by parts) to find

$$(\partial^\alpha f)\hat{}(\xi) = \int_{\mathbb{R}^n} \partial_x^\alpha f(x) e^{-ix\cdot\xi}\mathbf{d}x = (-1)^{|\alpha|}\int_{\mathbb{R}^n} f(x)\partial_x^\alpha e^{-ix\cdot\xi}\mathbf{d}x$$
$$= (-1)^{|\alpha|}\int_{\mathbb{R}^n} f(x)(-i\xi)^\alpha e^{-ix\cdot\xi}\mathbf{d}x = (i\xi)^\alpha\hat{f}(\xi).$$

Since $\partial^\alpha f \in L^1$ for all $|\alpha| \le k$, it follows that $(i\xi)^\alpha \hat{f}(\xi) = (\partial^\alpha f)\hat{}\,(\xi) \in C_0$ for all $|\alpha| \le k$. Since

$$(1+|\xi|)^k \le \left(1+\sum_{i=1}^n |\xi_i|\right)^k = \sum_{|\alpha|\le k} c_\alpha |\xi^\alpha|$$

where $0 < c_\alpha < \infty$,

$$\left|(1+|\xi|)^k \hat{f}(\xi)\right| \le \sum_{|\alpha|\le k} c_\alpha \left|\xi^\alpha \hat{f}(\xi)\right| \to 0 \text{ as } \xi \to \infty.$$

Item 8. is a simple application of Fubini's theorem. ■

Example 43.4. If $f(x) = e^{-|x|^2/2}$ then $\hat{f}(\xi) = e^{-|\xi|^2/2}$, in short

$$\mathcal{F}e^{-|x|^2/2} = e^{-|\xi|^2/2} \text{ and } \mathcal{F}^{-1}e^{-|\xi|^2/2} = e^{-|x|^2/2}. \tag{43.8}$$

More generally, for an $s > 0$,

$$\left(\mathcal{F}e^{-s|x|^2}\right)(\xi) = \left(\frac{1}{2s}\right)^{n/2} e^{-\frac{1}{4s}|\xi|^2}. \tag{43.9}$$

In particular for $t > 0$ let

$$p_t(x) := t^{-n/2} e^{-\frac{1}{2t}|x|^2} \tag{43.10}$$

then

$$\widehat{p_t}(\xi) = e^{-\frac{t}{2}|\xi|^2} \text{ and } (\widehat{p_t})^\vee(x) = p_t(x). \tag{43.11}$$

Let us now verify these assertions. By Item 8. of Theorem 43.3, to prove Eq. (43.8) it suffices to consider the 1 − dimensional case because $e^{-|x|^2/2} = \prod_{i=1}^n e^{-x_i^2/2}$. Let $g(\xi) := \left(\mathcal{F}e^{-x^2/2}\right)(\xi)$, then by Eq. (43.6) and Eq. (43.7),

$$\begin{aligned} g'(\xi) &= \mathcal{F}\left[(-ix)\,e^{-x^2/2}\right](\xi) = i\mathcal{F}\left[\frac{d}{dx}e^{-x^2/2}\right](\xi) \\ &= i(i\xi)\mathcal{F}\left[e^{-x^2/2}\right](\xi) = -\xi g(\xi). \end{aligned} \tag{43.12}$$

Lemma **??** implies

$$g(0) = \int_{\mathbb{R}} e^{-x^2/2}\mathbf{d}x = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-x^2/2}dm(x) = 1,$$

and so solving Eq. (43.12) with $g(0) = 1$ gives $\mathcal{F}\left[e^{-x^2/2}\right](\xi) = g(\xi) = e^{-\xi^2/2}$ as desired. The assertion that $\mathcal{F}^{-1}e^{-|\xi|^2/2} = e^{-|x|^2/2}$ follows similarly or by using Eq. (43.5) to conclude,

$$\mathcal{F}^{-1}\left[e^{-|\xi|^2/2}\right](x) = \mathcal{F}\left[e^{-|-\xi|^2/2}\right](x) = \mathcal{F}\left[e^{-|\xi|^2/2}\right](x) = e^{-|x|^2/2}.$$

The remaining results are now a matter of scaling. For example making the change of variables $x = y/\sqrt{2s}$ shows

$$\begin{aligned} \left(\mathcal{F}e^{-s|x|^2}\right)(\xi) &= \int_{\mathbb{R}^n} e^{-s|x|^2} e^{-i\xi\cdot x}\mathbf{d}x = \left(\frac{1}{2s}\right)^{n/2}\int_{\mathbb{R}^n} e^{-\frac{1}{2}|y|^2} e^{-i\xi\cdot y/\sqrt{2s}}\mathbf{d}y \\ &= \left(\frac{1}{2s}\right)^{n/2}\left[\mathcal{F}e^{-\frac{1}{2}|x|^2}\right]\left(\frac{\xi}{\sqrt{2s}}\right) = \left(\frac{1}{2s}\right)^{n/2} e^{-\frac{1}{4s}|\xi|^2}. \end{aligned}$$

Similarly the results in Eq. (43.11) now follow from this one or from Eq. (43.8) and item 5 of Theorem 43.3. For example, since $p_t(x) = t^{-n/2}p_1(x/\sqrt{t})$,

$$(\widehat{p_t})(\xi) = t^{-n/2}\left(\sqrt{t}\right)^n \hat{p}_1(\sqrt{t}\xi) = e^{-\frac{t}{2}|\xi|^2}.$$

This may also be written as $(\widehat{p_t})(\xi) = t^{-n/2}p_{\frac{1}{t}}(\xi)$. Using this and the fact that p_t is an even function,

$$(\widehat{p_t})^\vee(x) = \mathcal{F}\widehat{p_t}(-x) = t^{-n/2}\mathcal{F}p_{\frac{1}{t}}(-x) = t^{-n/2}t^{n/2}p_t(-x) = p_t(x).$$

43.3 Schwartz Test Functions

Definition 43.5. *A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have* ***rapid decay*** *or* ***rapid decrease*** *if*

$$\sup_{x\in\mathbb{R}^n} (1+|x|)^N |f(x)| < \infty \text{ for } N = 1, 2, \ldots.$$

Equivalently, for each $N \in \mathbb{N}$ there exists constants $C_N < \infty$ such that $|f(x)| \le C_N(1+|x|)^{-N}$ for all $x \in \mathbb{R}^n$. A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have (at most) ***polynomial growth*** *if there exists $N < \infty$ such*

$$\sup\,(1+|x|)^{-N}|f(x)| < \infty,$$

i.e. there exists $N \in \mathbb{N}$ and $C < \infty$ such that $|f(x)| \le C(1+|x|)^N$ for all $x \in \mathbb{R}^n$.

Definition 43.6 (Schwartz Test Functions). *Let $\mathcal{S}$ denote the space of functions $f \in C^\infty(\mathbb{R}^n)$ such that f and all of its partial derivatives have rapid decay and let*

$$\|f\|_{N,\alpha} = \sup_{x\in\mathbb{R}^n}\left|(1+|x|)^N\partial^\alpha f(x)\right|$$

so that

$$\mathcal{S} = \left\{ f \in C^\infty(\mathbb{R}^n) : \|f\|_{N,\alpha} < \infty \text{ for all } N \text{ and } \alpha \right\}.$$

Also let $\mathcal{P}$ denote those functions $g \in C^\infty(\mathbb{R}^n)$ such that g and all of its derivatives have at most polynomial growth, i.e. $g \in C^\infty(\mathbb{R}^n)$ is in $\mathcal{P}$ iff for all multi-indices α, there exists $N_\alpha < \infty$ such

$$\sup (1 + |x|)^{-N_\alpha} |\partial^\alpha g(x)| < \infty.$$

(Notice that any polynomial function on $\mathbb{R}^n$ is in $\mathcal{P}$.)

Remark 43.7. Since $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S} \subset L^2(\mathbb{R}^n)$, it follows that $\mathcal{S}$ is dense in $L^2(\mathbb{R}^n)$.

Exercise 43.1. Let

$$L = \sum_{|\alpha| \le k} a_\alpha(x) \partial^\alpha \tag{43.13}$$

with $a_\alpha \in \mathcal{P}$. Show $L(\mathcal{S}) \subset \mathcal{S}$ and in particular $\partial^\alpha f$ and $x^\alpha f$ are back in $\mathcal{S}$ for all multi-indices α.

Notation 43.8 *Suppose that $p(x,\xi) = \Sigma_{|\alpha| \le N} a_\alpha(x) \xi^\alpha$ where each function $a_\alpha(x)$ is a smooth function. We then set*

$$p(x, D_x) := \Sigma_{|\alpha| \le N} a_\alpha(x) D_x^\alpha$$

and if each $a_\alpha(x)$ is also a polynomial in x we will let

$$p(-D_\xi, \xi) := \Sigma_{|\alpha| \le N} a_\alpha(-D_\xi) M_{\xi^\alpha}$$

where M_{ξ^α} is the operation of multiplication by ξ^α.

Proposition 43.9. *Let $p(x,\xi)$ be as above and assume each $a_\alpha(x)$ is a polynomial in x. Then for $f \in \mathcal{S}$,*

$$\left(p(x, D_x) f\right)^\wedge(\xi) = p(-D_\xi, \xi) \hat{f}(\xi) \tag{43.14}$$

and

$$p(\xi, D_\xi) \hat{f}(\xi) = [p(D_x, -x) f(x)]^\wedge(\xi). \tag{43.15}$$

Proof. The identities $(-D_\xi)^\alpha e^{-ix\cdot\xi} = x^\alpha e^{-ix\cdot\xi}$ and $D_x^\alpha e^{ix\cdot\xi} = \xi^\alpha e^{ix\cdot\xi}$ imply, for any polynomial function q on $\mathbb{R}^n$,

$$q(-D_\xi) e^{-ix\cdot\xi} = q(x) e^{-ix\cdot\xi} \text{ and } q(D_x) e^{ix\cdot\xi} = q(\xi) e^{ix\cdot\xi}. \tag{43.16}$$

Therefore using Eq. (43.16) repeatedly,

$$\begin{aligned}
\left(p(x, D_x) f\right)^\wedge(\xi) &= \int_{\mathbb{R}^n} \sum_{|\alpha| \le N} a_\alpha(x) D_x^\alpha f(x) \cdot e^{-ix\cdot\xi} \mathbf{d}\xi \\
&= \int_{\mathbb{R}^n} \sum_{|\alpha| \le N} D_x^\alpha f(x) \cdot a_\alpha(-D_\xi) e^{-ix\cdot\xi} \mathbf{d}\xi \\
&= \int_{\mathbb{R}^n} f(x) \sum_{|\alpha| \le N} (-D_x)^\alpha \left[a_\alpha(-D_\xi) e^{-ix\cdot\xi} \right] \mathbf{d}\xi \\
&= \int_{\mathbb{R}^n} f(x) \sum_{|\alpha| \le N} a_\alpha(-D_\xi) \left[\xi^\alpha e^{-ix\cdot\xi} \right] \mathbf{d}\xi = p(-D_\xi, \xi) \hat{f}(\xi)
\end{aligned}$$

wherein the third inequality we have used Lemma 31.39 to do repeated integration by parts, the fact that mixed partial derivatives commute in the fourth, and in the last we have repeatedly used Corollary 10.30 to differentiate under the integral. The proof of Eq. (43.15) is similar:

$$\begin{aligned}
p(\xi, D_\xi) \hat{f}(\xi) &= p(\xi, D_\xi) \int_{\mathbb{R}^n} f(x) e^{-ix\cdot\xi} \mathbf{d}x = \int_{\mathbb{R}^n} f(x) p(\xi, -x) e^{-ix\cdot\xi} \mathbf{d}x \\
&= \sum_{|\alpha| \le N} \int_{\mathbb{R}^n} f(x) (-x)^\alpha a_\alpha(\xi) e^{-ix\cdot\xi} \mathbf{d}x \\
&= \sum_{|\alpha| \le N} \int_{\mathbb{R}^n} f(x) (-x)^\alpha a_\alpha(-D_x) e^{-ix\cdot\xi} \mathbf{d}x \\
&= \sum_{|\alpha| \le N} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} a_\alpha(D_x) \left[(-x)^\alpha f(x) \right] \mathbf{d}x \\
&= [p(D_x, -x) f(x)]^\wedge(\xi).
\end{aligned}$$

■

Corollary 43.10. *The Fourier transform preserves the space $\mathcal{S}$, i.e. $\mathcal{F}(\mathcal{S}) \subset \mathcal{S}$.*

Proof. Let $p(x,\xi) = \Sigma_{|\alpha| \le N} a_\alpha(x) \xi^\alpha$ with each $a_\alpha(x)$ being a polynomial function in x. If $f \in \mathcal{S}$ then $p(D_x, -x) f \in \mathcal{S} \subset L^1$ and so by Eq. (43.15), $p(\xi, D_\xi) \hat{f}(\xi)$ is bounded in ξ, i.e.

$$\sup_{\xi \in \mathbb{R}^n} |p(\xi, D_\xi) \hat{f}(\xi)| \le C(p, f) < \infty.$$

Taking $p(x,\xi) = (1 + |x|^2)^N \xi^\alpha$ with $N \in \mathbb{Z}_+$ in this estimate shows $\hat{f}(\xi)$ and all of its derivatives have rapid decay, i.e. $\hat{f}$ is in $\mathcal{S}$. ■

43.4 Fourier Inversion Formula

Theorem 43.11 (Fourier Inversion Theorem). *Suppose that $f \in L^1$ and $\hat{f} \in L^1$ (for example suppose $f \in \mathcal{S}$), then*

1. *there exists $f_0 \in C_0(\mathbb{R}^n)$ such that $f = f_0$ a.e.,*
2. *$f_0 = \mathcal{F}^{-1}\mathcal{F} f$ and $f_0 = \mathcal{F}\mathcal{F}^{-1} f$,*
3. *f and $\hat{f}$ are in $L^1 \cap L^\infty$ and*
4. *$\|f\|_2 = \left\|\hat{f}\right\|_2$.*

In particular, $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is a linear isomorphism of vector spaces.

Proof. First notice that $\hat{f} \in C_0(\mathbb{R}^n) \subset L^\infty$ and $\hat{f} \in L^1$ by assumption, so that $\hat{f} \in L^1 \cap L^\infty$. Let $p_t(x) := t^{-n/2} e^{-\frac{1}{2t}|x|^2}$ be as in Example 43.4 so that $\widehat{p_t}(\xi) = e^{-\frac{t}{2}|\xi|^2}$ and $\widehat{p}_t^\vee = p_t$. Define $f_0 := \hat{f}^\vee \in C_0$ then

$$\begin{aligned} f_0(x) = (\hat{f})^\vee(x) &= \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\xi\cdot x} \mathbf{d}\xi = \lim_{t\downarrow 0} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\xi\cdot x} \widehat{p_t}(\xi) \mathbf{d}\xi \\ &= \lim_{t\downarrow 0} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(y) e^{i\xi\cdot(x-y)} \widehat{p_t}(\xi) \mathbf{d}\xi \, \mathbf{d}y \\ &= \lim_{t\downarrow 0} \int_{\mathbb{R}^n} f(y) p_t(x-y) \mathbf{d}y = f(x) \text{ a.e.} \end{aligned}$$

wherein we have used Theorem 31.33 in the last equality along with the observations that $p_t(y) = p_1(y/\sqrt{t})$ and $\int_{\mathbb{R}^n} p_1(y)\mathbf{d}y = 1$ so that

$$L^1\text{–}\lim_{t\downarrow 0} \int_{\mathbb{R}^n} f(y) p_t(x-y) \mathbf{d}y = f(x).$$

In particular this shows that $f_0 \in L^1 \cap L^\infty$. A similar argument shows that $\mathcal{F}^{-1}\mathcal{F} f = f_0$ as well. Let us now compute the L^2 – norm of $\hat{f}$,

$$\begin{aligned} \|\hat{f}\|_2^2 &= \int_{\mathbb{R}^n} \hat{f}(\xi)\overline{\hat{f}(\xi)} \mathbf{d}\xi = \int_{\mathbb{R}^n} \mathbf{d}\xi \; \hat{f}(\xi) \int_{\mathbb{R}^n} \mathbf{d}x \overline{f(x)} e^{ix\cdot\xi} \\ &= \int_{\mathbb{R}^n} \mathbf{d}x \, \overline{f(x)} \int_{\mathbb{R}^n} \mathbf{d}\xi \; \hat{f}(\xi) e^{ix\cdot\xi} \text{ (by Fubini)} \\ &= \int_{\mathbb{R}^n} \mathbf{d}x \; \overline{f(x)} f(x) = \|f\|_2^2 \end{aligned}$$

because $\int_{\mathbb{R}^n} \mathbf{d}\xi \hat{f}(\xi) e^{ix\cdot\xi} = \mathcal{F}^{-1}\hat{f}(x) = f(x)$ a.e. ∎

In the next few exercises you are asked to compute the Fourier transform of a number of functions.

Exercise 43.2. For any $m > 0$, show

$$\mathcal{F}\left[e^{-m|x|}\right](\xi) = \frac{2m}{\sqrt{2\pi}} \frac{1}{m^2 + \xi^2}$$

and

$$\mathcal{F}\left(\frac{1}{m^2 + \xi^2}\right)(x) = \frac{\sqrt{2\pi}}{2m} e^{-m|x|}.$$

Exercise 43.3. Using the identity

$$\frac{1}{\xi^2 + 1} = \int_0^\infty e^{-s(\xi^2+1)} ds$$

along with Exercise 43.2 and the known Fourier transform of Gaussians to show

$$e^{-|x|} = \int_0^\infty ds \frac{1}{\sqrt{\pi s}} e^{-s} e^{-\frac{x^2}{4s}} \text{ for all } x \in \mathbb{R}. \tag{43.17}$$

Thus we have written $e^{-|x|}$ as an average of Gaussians.

Exercise 43.4. Now let $x \in \mathbb{R}^n$ and $\|x\|^2 := \sum_{i=1}^n x_i^2$ be the standard Euclidean norm. Show for all $m > 0$ that

$$\mathcal{F}\left[e^{-m\|x\|}\right](\xi) = \frac{2^{n/2}}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right) \frac{m}{\left(m^2 + |\xi|^2\right)^{\frac{n+1}{2}}}, \tag{43.18}$$

where $\Gamma(x)$ in the gamma function defined as

$$\Gamma(x) := \int_0^\infty t^x e^{-t} \frac{dt}{t}.$$

Hint: By Exercise 43.3 with x replaced by $m\|x\|$ we know that

$$e^{-m\|x\|} = \int_0^\infty ds \frac{1}{\sqrt{\pi s}} e^{-s} e^{-\frac{m^2}{4s}\|x\|^2} \text{ for all } x \in \mathbb{R}^n.$$

Remark 43.12. This result can be used to show,

$$e^{-m\sqrt{-\Delta}} f(x) = \int_{\mathbb{R}^n} Q_m(x-y) f(y) dy$$

where

$$\begin{aligned} Q_m(x) &= 2^{n/2} \frac{\Gamma((n+1)/2)}{(2\pi)^{n/2}\sqrt{\pi}} \frac{m}{(m^2 + |x|^2)^{(n+1)/2}} = \frac{\Gamma((n+1)/2)}{\pi^{n/2}\sqrt{\pi}} \frac{m}{(m^2 + |x|^2)^{(n+1)/2}} \\ &= \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{m}{(m^2 + |x|^2)^{(n+1)/2}}. \end{aligned}$$

The extra factors of $\sqrt{2\pi}$ come from the normalized convolution.

Corollary 43.13. *By the B.L.T. Theorem 32.4, the maps $\mathcal{F}|_{\mathcal{S}}$ and $\mathcal{F}^{-1}|_{\mathcal{S}}$ extend to bounded linear maps $\bar{\mathcal{F}}$ and $\bar{\mathcal{F}}^{-1}$ from $L^2 \to L^2$. These maps satisfy the following properties:*

1. *$\bar{\mathcal{F}}$ and $\bar{\mathcal{F}}^{-1}$ are unitary and are inverses to one another as the notation suggests.*
2. *If $f \in L^2$, then $\bar{\mathcal{F}} f$ is uniquely characterized as the function, $G \in L^2$ such that*
$$\langle G, \psi \rangle = \langle f, \hat{\psi} \rangle \text{ for all } \psi \in C_c^\infty(\mathbb{R}^n).$$
3. *If $f \in L^1 \cap L^2$, then $\bar{\mathcal{F}} f = \hat{f}$ a.e.*
4. *For $f \in L^2$ we may compute $\bar{\mathcal{F}}$ and $\bar{\mathcal{F}}^{-1}$ by*
$$\bar{\mathcal{F}} f(\xi) = L^2\text{–}\lim_{R\to\infty} \int_{|x|\le R} f(x) e^{-ix\cdot\xi} \mathbf{d}x \text{ and} \tag{43.19}$$
$$\bar{\mathcal{F}}^{-1} f(\xi) = L^2\text{–}\lim_{R\to\infty} \int_{|x|\le R} f(x) e^{ix\cdot\xi} \mathbf{d}x. \tag{43.20}$$
5. *We may further extend $\bar{\mathcal{F}}$ to a map from $L^1 + L^2 \to C_0 + L^2$ (still denote by $\bar{\mathcal{F}}$) defined by $\bar{\mathcal{F}} f = \hat{h} + \bar{\mathcal{F}} g$ where $f = h + g \in L^1 + L^2$. For $f \in L^1 + L^2$, $\bar{\mathcal{F}} f$ may be characterized as the unique function $F \in L^1_{loc}(\mathbb{R}^n)$ such that*
$$\langle F, \varphi \rangle = \langle f, \hat{\varphi} \rangle \text{ for all } \varphi \in C_c^\infty(\mathbb{R}^n). \tag{43.21}$$
Moreover if Eq. (43.21) holds then $F \in C_0 + L^2 \subset L^1_{loc}(\mathbb{R}^n)$ and Eq.(43.21) is valid for all $\varphi \in \mathcal{S}$.

Proof. 1. and 2. If $f \in L^2$ and $\varphi_n \in \mathcal{S}$ such that $\varphi_n \to f$ in L^2, then $\bar{\mathcal{F}} f := \lim_{n\to\infty} \hat{\varphi}_n$. Since $\hat{\varphi}_n \in \mathcal{S} \subset L^1$, we may concluded that $\|\hat{\varphi}_n\|_2 = \|\varphi_n\|_2$ for all n. Thus
$$\left\|\bar{\mathcal{F}} f\right\|_2 = \lim_{n\to\infty} \|\hat{\varphi}_n\|_2 = \lim_{n\to\infty} \|\varphi_n\|_2 = \|f\|_2$$
which shows that $\bar{\mathcal{F}}$ is an isometry from L^2 to L^2 and similarly $\bar{\mathcal{F}}^{-1}$ is an isometry. Since $\bar{\mathcal{F}}^{-1}\bar{\mathcal{F}} = \mathcal{F}^{-1}\mathcal{F} = id$ on the dense set $\mathcal{S}$, it follows by continuity that $\bar{\mathcal{F}}^{-1}\bar{\mathcal{F}} = id$ on all of L^2. Hence $\bar{\mathcal{F}}\bar{\mathcal{F}}^{-1} = id$, and thus $\bar{\mathcal{F}}^{-1}$ is the inverse of $\bar{\mathcal{F}}$. This proves item 1. Moreover, if $\psi \in C_c^\infty(\mathbb{R}^n)$, then
$$\langle \bar{\mathcal{F}} f, \psi \rangle = \lim_{n\to\infty} \langle \hat{\varphi}_n, \psi \rangle = \lim_{n\to\infty} \langle \varphi_n, \hat{\psi} \rangle = \langle f, \psi \rangle \tag{43.22}$$
and this equation uniquely characterizes $\bar{\mathcal{F}} f$ by Corollary 31.41. Notice that Eq. (43.22) also holds for all $\psi \in \mathcal{S}$.

3. If $f \in L^1 \cap L^2$, we have already seen that $\hat{f} \in C_0(\mathbb{R}^n) \subset L^1_{loc}$ and that $\langle \hat{f}, \psi \rangle = \langle f, \hat{\psi} \rangle$ for all $\psi \in C_c^\infty(\mathbb{R}^n)$. Combining this with item 2. shows $\langle \hat{f} - \bar{\mathcal{F}} f, \psi \rangle = 0$ or all $\psi \in C_c^\infty(\mathbb{R}^n)$ and so again by Corollary 31.41 we conclude that $\hat{f} - \bar{\mathcal{F}} f = 0$ a.e.

4. Let $f \in L^2$ and $R < \infty$ and set $f_R(x) := f(x) 1_{|x|\le R}$. Then $f_R \in L^1 \cap L^2$ and therefore $\bar{\mathcal{F}} f_R = \hat{f}_R$. Since $\bar{\mathcal{F}}$ is an isometry and (by the dominated convergence theorem) $f_R \to f$ in L^2, it follows that
$$\bar{\mathcal{F}} f = L^2\text{–}\lim_{R\to\infty} \bar{\mathcal{F}} f_R = L^2\text{–}\lim_{R\to\infty} \hat{f}_R.$$

5. If $f = h + g \in L^1 + L^2$ and $\varphi \in \mathcal{S}$, then by Eq. (43.22) and item 4. of Theorem 43.3,
$$\langle \hat{h} + \bar{\mathcal{F}} g, \varphi \rangle = \langle h, \hat{\varphi} \rangle + \langle g, \hat{\varphi} \rangle = \langle h + g, \hat{\varphi} \rangle. \tag{43.23}$$
In particular if $h + g = 0$ a.e., then $\langle \hat{h} + \bar{\mathcal{F}} g, \varphi \rangle = 0$ for all $\varphi \in \mathcal{S}$ and since $\hat{h} + \bar{\mathcal{F}} g \in L^1_{loc}$ it follows from Corollary 31.41 that $\hat{h} + \bar{\mathcal{F}} g = 0$ a.e. This shows that $\bar{\mathcal{F}} f$ is well defined independent of how $f \in L^1 + L^2$ is decomposed into the sum of an L^1 and an L^2 function. Moreover Eq. (43.23) shows Eq. (43.21) holds with $F = \hat{h} + \bar{\mathcal{F}} g \in C_0 + L^2$ and $\varphi \in \mathcal{S}$. Now suppose $G \in L^1_{loc}$ and $\langle G, \varphi \rangle = \langle f, \hat{\varphi} \rangle$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$. Then by what we just proved, $\langle G, \varphi \rangle = \langle F, \varphi \rangle$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$ and so another application of Corollary 31.41 shows $G = F \in C_0 + L^2$. ■

Notation 43.14 *Given the results of Corollary 43.13, there is little danger in writing $\hat{f}$ or $\mathcal{F} f$ for $\bar{\mathcal{F}} f$ when $f \in L^1 + L^2$.*

Corollary 43.15. *If f and g are L^1 functions such that $\hat{f}, \hat{g} \in L^1$, then*
$$\mathcal{F}(fg) = \hat{f} \bigstar \hat{g} \text{ and } \mathcal{F}^{-1}(fg) = f^\vee \bigstar g^\vee.$$
Since $\mathcal{S}$ is closed under pointwise products and $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is an isomorphism it follows that $\mathcal{S}$ is closed under convolution as well.

Proof. By Theorem 43.11, $f, g, \hat{f}, \hat{g} \in L^1 \cap L^\infty$ and hence $f \cdot g \in L^1 \cap L^\infty$ and $\hat{f} \bigstar \hat{g} \in L^1 \cap L^\infty$. Since
$$\mathcal{F}^{-1}\left(\hat{f} \bigstar \hat{g}\right) = \mathcal{F}^{-1}\left(\hat{f}\right) \cdot \mathcal{F}^{-1}\left(\hat{g}\right) = f \cdot g \in L^1$$
we may conclude from Theorem 43.11 that
$$\hat{f} \bigstar \hat{g} = \mathcal{F}\mathcal{F}^{-1}\left(\hat{f} \bigstar \hat{g}\right) = \mathcal{F}(f \cdot g).$$
Similarly one shows $\mathcal{F}^{-1}(fg) = f^\vee \bigstar g^\vee$. ■

Corollary 43.16. *Let $p(x,\xi)$ and $p(x,D_x)$ be as in Notation 43.8 with each function $a_\alpha(x)$ being a smooth function of $x \in \mathbb{R}^n$. Then for $f \in \mathcal{S}$,*

$$p(x,D_x)f(x) = \int_{\mathbb{R}^n} p(x,\xi)\hat{f}(\xi)\, e^{ix\cdot\xi}\mathbf{d}\xi. \tag{43.24}$$

Proof. For $f \in \mathcal{S}$, we have

$$\begin{aligned} p(x,D_x)f(x) &= p(x,D_x)\left(\mathcal{F}^{-1}\hat{f}\right)(x) = p(x,D_x)\int_{\mathbb{R}^n}\hat{f}(\xi)\, e^{ix\cdot\xi}\mathbf{d}\xi \\ &= \int_{\mathbb{R}^n}\hat{f}(\xi)\, p(x,D_x)e^{ix\cdot\xi}\mathbf{d}\xi = \int_{\mathbb{R}^n}\hat{f}(\xi)\, p(x,\xi)e^{ix\cdot\xi}\mathbf{d}\xi. \end{aligned}$$

■

If $p(x,\xi)$ is a more general function of (x,ξ) then that given in Notation 43.8, the right member of Eq. (43.24) may still make sense, in which case we may use it as a definition of $p(x,D_x)$. A linear operator defined this way is called a **pseudo differential operator** and they turn out to be a useful class of operators to study when working with partial differential equations.

Corollary 43.17. *Suppose $p(\xi) = \sum_{|\alpha|\le N} a_\alpha \xi^\alpha$ is a polynomial in $\xi \in \mathbb{R}^n$ and $f \in L^2$. Then $p(\partial)f$ exists weakly in L^2 (see Definition 31.50 or 41.3) iff $\xi \to p(i\xi)\hat{f}(\xi) \in L^2$ in which case*

$$(p(\partial)f)\hat{}\,(\xi) = p(i\xi)\hat{f}(\xi) \text{ for a.e. } \xi.$$

In particular, if $g \in L^2$ then $f \in L^2$ solves the equation, $p(\partial)f = g$ iff $p(i\xi)\hat{f}(\xi) = \hat{g}(\xi)$ for a.e. ξ.

Proof. By Exercise 31.20 or Proposition 41.12, if $p(\partial)f = g$ in L^2, then there exists $f_n \in C_c^\infty(\mathbb{R}^n)$ such that $f_n \to f$ and $p(\partial)f_n \to g$ in $L^2(\mathbb{R}^n)$. Therefore if $\varphi \in \mathcal{S}$,[1] then by standard integration by parts,

$$\begin{aligned} \langle g,\varphi\rangle &= \lim_{n\to\infty}\langle p(\partial)f_n,\varphi\rangle = \lim_{n\to\infty}\langle p(\partial)f_n,\varphi\rangle \\ &= \lim_{n\to\infty}\langle f_n, p(-\partial)\varphi\rangle = \langle f, p(-\partial)\varphi\rangle. \end{aligned}$$

Using this fact, the fact that $\mathcal{S}$ is preserved by the Fourier transform, and

$$\begin{aligned} p(-\partial)\hat{\varphi}(\xi) &= p(-\partial_\xi)\int_{\mathbb{R}^n}\varphi(x)\, e^{-i\xi\cdot x}\mathbf{d}x \\ &= \int_{\mathbb{R}^n} p(ix)\varphi(x)\, e^{-i\xi\cdot x}\mathbf{d}x = (p(ix)\varphi(x))\hat{}\,, \end{aligned}$$

[1] This actually holds more generally for any $\varphi \in C^N(\mathbb{R}^n)$ such that φ and $p(-\partial)\varphi$ are in $L^2(\mathbb{R}^n)$.

we learn

$$\begin{aligned} \langle\hat{g},\varphi\rangle &= \langle g,\hat{\varphi}\rangle = \langle f, p(-\partial)\hat{\varphi}\rangle = \left\langle f, (p(ix)\varphi(x))\hat{}\right\rangle \\ &= \left\langle \hat{f}, p(i(\cdot))\varphi\right\rangle = \left\langle p(i(\cdot))\hat{f},\varphi\right\rangle \end{aligned}$$

from which it follows we find, $p(i(\cdot))\hat{f} = \hat{g}$ a.e.

Conversely if $p(i\xi)\hat{f}(\xi) = \hat{g}(\xi)$, then using $\langle\hat{g},\varphi^\vee\rangle = \left\langle g,(\varphi^\vee)\hat{}\right\rangle = \langle g,\varphi\rangle$ for all $g \in L^2$, we find

$$\begin{aligned} \langle f, p(-\partial)\varphi\rangle &= \langle\hat{f}, [p(-\partial)\varphi]^\vee\rangle = \langle\hat{f}(\xi), p(i\xi)\varphi^\vee(\xi)\rangle \\ &= \left\langle p(i\xi)\hat{f}(\xi),\varphi^\vee(\xi)\right\rangle = \langle\hat{g},\varphi^\vee(\xi)\rangle = \langle g,\varphi\rangle \end{aligned}$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^n)$. This shows that $p(\partial)f = g$ weakly in $L^2(\mathbb{R}^n)$. ■

Exercise 43.5 (Variation on the proof of Corollary 43.17). Suppose $p(\xi)$ is a polynomial in $\xi \in \mathbb{R}^d$ and $u \in L^2$ such that $p(\partial)u \in L^2$.[2] Show

$$\mathcal{F}(p(\partial)u)(\xi) = p(i\xi)\hat{u}(\xi) \in L^2.$$

Conversely if $u \in L^2$ such that $p(i\xi)\hat{u}(\xi) \in L^2$, show $p(\partial)u \in L^2$.

43.5 Summary of Basic Properties of $\mathcal{F}$ and $\mathcal{F}^{-1}$

The following table summarizes some of the basic properties of the Fourier transform and its inverse.

f	$\longleftrightarrow$	$\hat{f}$ or $f^\vee$
Smoothness	$\longleftrightarrow$	Decay at infinity
∂^α	$\longleftrightarrow$	Multiplication by $(\pm i\xi)^\alpha$
$\mathcal{S}$	$\longleftrightarrow$	$\mathcal{S}$
$L^2(\mathbb{R}^n)$	$\longleftrightarrow$	$L^2(\mathbb{R}^n)$
Convolution	$\longleftrightarrow$	Products.

43.6 Fourier Transforms of Measures and Bochner's Theorem

To motivate the next definition suppose that μ is a finite measure on $\mathbb{R}^n$ which is absolutely continuous relative to Lebesgue measure, $d\mu(x) = \rho(x)\mathbf{d}x$. Then it is reasonable to require

[2] Here we say that $p(\partial)u = g$ exists in $L^2(\mathbb{R}^n)$ iff $\langle g,\varphi\rangle = \langle f, p(-\partial)\varphi\rangle$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$.

$$\hat{\mu}(\xi) := \hat{\rho}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x}\rho(x)\mathbf{d}x = \int_{\mathbb{R}^n} e^{-i\xi\cdot x}d\mu(x)$$

and

$$(\mu\bigstar g)(x) := \rho\bigstar g(x) = \int_{\mathbb{R}^n} g(x-y)\rho(x)dx = \int_{\mathbb{R}^n} g(x-y)d\mu(y)$$

when $g:\mathbb{R}^n\to\mathbb{C}$ is a function such that the latter integral is defined, for example assume g is bounded. These considerations lead to the following definitions.

Definition 43.18. *The Fourier transform, $\hat{\mu}$, of a complex measure μ on $\mathcal{B}_{\mathbb{R}^n}$ is defined by*

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x}d\mu(x) \tag{43.25}$$

and the convolution with a function g is defined by

$$(\mu\bigstar g)(x) = \int_{\mathbb{R}^n} g(x-y)d\mu(y)$$

when the integral is defined.

It follows from the dominated convergence theorem that $\hat{\mu}$ is continuous. Also by a variant of Exercise 16.2, if μ and ν are two complex measure on $\mathcal{B}_{\mathbb{R}^n}$ such that $\hat{\mu}=\hat{\nu}$, then $\mu=\nu$. The reader is asked to give another proof of this fact in Exercise 43.6 below.

Example 43.19. Let σ_t be the surface measure on the sphere S_t of radius t centered at zero in $\mathbb{R}^3$. Then

$$\hat{\sigma}_t(\xi) = 4\pi t\frac{\sin t|\xi|}{|\xi|}.$$

Indeed,

$$\begin{aligned}\hat{\sigma}_t(\xi) &= \int_{tS^2} e^{-ix\cdot\xi}d\sigma(x) = t^2\int_{S^2} e^{-itx\cdot\xi}d\sigma(x)\\ &= t^2\int_{S^2} e^{-itx_3|\xi|}d\sigma(x) = t^2\int_0^{2\pi}d\theta\int_0^{\pi}d\varphi\sin\varphi e^{-it\cos\varphi|\xi|}\\ &= 2\pi t^2\int_{-1}^{1} e^{-itu|\xi|}du = 2\pi t^2\frac{1}{-it|\xi|}e^{-itu|\xi|}|_{u=-1}^{u=1} = 4\pi t^2\frac{\sin t|\xi|}{t|\xi|}.\end{aligned}$$

Definition 43.20. *A function $\chi:\mathbb{R}^n\to\mathbb{C}$ is said to be* ***positive (semi) definite*** *iff the matrices $A:=\{\chi(\xi_k-\xi_j)\}_{k,j=1}^m$ are positive definite for all $m\in\mathbb{N}$ and $\{\xi_j\}_{j=1}^m\subset\mathbb{R}^n$.*

Lemma 43.21. *If $\chi\in C(\mathbb{R}^n,\mathbb{C})$ is a positive definite function, then*

1. *$\chi(0)\ge 0$.*
2. *$\chi(-\xi)=\overline{\chi(\xi)}$ for all $\xi\in\mathbb{R}^n$.*
3. *$|\chi(\xi)|\le\chi(0)$ for all $\xi\in\mathbb{R}^n$.*
4. *For all $f\in\mathbb{S}(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}d\xi d\eta\ge 0. \tag{43.26}$$

Proof. Taking $m=1$ and $\xi_1=0$ we learn $\chi(0)|\lambda|^2\ge 0$ for all $\lambda\in\mathbb{C}$ which proves item 1. Taking $m=2$, $\xi_1=\xi$ and $\xi_2=\eta$, the matrix

$$A := \begin{bmatrix}\chi(0) & \chi(\xi-\eta)\\ \chi(\eta-\xi) & \chi(0)\end{bmatrix}$$

is positive definite from which we conclude $\chi(\xi-\eta)=\overline{\chi(\eta-\xi)}$ (since $A=A^*$ by definition) and

$$0\le\det\begin{bmatrix}\chi(0) & \chi(\xi-\eta)\\ \chi(\eta-\xi) & \chi(0)\end{bmatrix} = |\chi(0)|^2-|\chi(\xi-\eta)|^2.$$

and hence $|\chi(\xi)|\le\chi(0)$ for all ξ. This proves items 2. and 3. Item 4. follows by approximating the integral in Eq. (43.26) by Riemann sums,

$$\begin{aligned}&\int_{\mathbb{R}^n\times\mathbb{R}^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}d\xi d\eta\\ &\quad=\lim_{\varepsilon\downarrow 0}\varepsilon^{-2n}\sum_{\xi,\eta\in(\varepsilon\mathbb{Z}^n)\cap[-\varepsilon^{-1},\varepsilon^{-1}]^n}\chi(\xi-\eta)f(\xi)\overline{f(\eta)}\ge 0.\end{aligned}$$

The details are left to the reader. ■

Lemma 43.22. *If μ is a finite positive measure on $\mathcal{B}_{\mathbb{R}^n}$, then $\chi:=\hat{\mu}\in C(\mathbb{R}^n,\mathbb{C})$ is a positive definite function.*

Proof. As has already been observed after Definition 43.18, the dominated convergence theorem implies $\hat{\mu}\in C(\mathbb{R}^n,\mathbb{C})$. Since μ is a positive measure (and hence real),

$$\hat{\mu}(-\xi) = \int_{\mathbb{R}^n} e^{i\xi\cdot x}d\mu(x) = \overline{\int_{\mathbb{R}^n} e^{-i\xi\cdot x}d\mu(x)} = \overline{\hat{\mu}(-\xi)}.$$

From this it follows that for any $m\in\mathbb{N}$ and $\{\xi_j\}_{j=1}^m\subset\mathbb{R}^n$, the matrix $A:=\{\hat{\mu}(\xi_k-\xi_j)\}_{k,j=1}^m$ is self-adjoint. Moreover if $\lambda\in\mathbb{C}^m$,

$$\sum_{k,j=1}^m \hat{\mu}(\xi_k-\xi_j)\lambda_k\bar{\lambda}_j = \int_{\mathbb{R}^n}\sum_{k,j=1}^m e^{-i(\xi_k-\xi_j)\cdot x}\lambda_k\bar{\lambda}_j d\mu(x)$$
$$= \int_{\mathbb{R}^n}\sum_{k,j=1}^m e^{-i\xi_k\cdot x}\lambda_k \overline{e^{-i\xi_j\cdot x}\lambda_j} d\mu(x)$$
$$= \int_{\mathbb{R}^n}\left|\sum_{k=1}^m e^{-i\xi_k\cdot x}\lambda_k\right|^2 d\mu(x) \geq 0$$

showing A is positive definite. ■

Theorem 43.23 (Bochner's Theorem). *Suppose $\chi \in C(\mathbb{R}^n,\mathbb{C})$ is positive definite function, then there exists a unique positive measure μ on $\mathcal{B}_{\mathbb{R}^n}$ such that $\chi = \hat{\mu}$.*

Proof. If $\chi(\xi) = \hat{\mu}(\xi)$, then for $f \in \mathcal{S}$ we would have

$$\int_{\mathbb{R}^n} f d\mu = \int_{\mathbb{R}^n} \left(f^\vee\right)^{\hat{}} d\mu = \int_{\mathbb{R}^n} f^\vee(\xi)\hat{\mu}(\xi)\mathbf{d}\xi.$$

This suggests that we define

$$I(f) := \int_{\mathbb{R}^n} \chi(\xi) f^\vee(\xi)\mathbf{d}\xi \text{ for all } f \in \mathcal{S}.$$

We will now show I is positive in the sense if $f \in \mathcal{S}$ and $f \geq 0$ then $I(f) \geq 0$. For general $f \in \mathcal{S}$ we have

$$I(|f|^2) = \int_{\mathbb{R}^n}\chi(\xi)\left(|f|^2\right)^\vee(\xi)\mathbf{d}\xi = \int_{\mathbb{R}^n}\chi(\xi)\left(f^\vee \bigstar \bar{f}^\vee\right)(\xi)\mathbf{d}\xi$$
$$= \int_{\mathbb{R}^n}\chi(\xi) f^\vee(\xi-\eta)\bar{f}^\vee(\eta)\mathbf{d}\eta\mathbf{d}\xi = \int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi-\eta)\overline{f^\vee(-\eta)}\mathbf{d}\eta\mathbf{d}\xi$$
$$= \int_{\mathbb{R}^n}\chi(\xi-\eta)f^\vee(\xi)\overline{f^\vee(\eta)}\mathbf{d}\eta\mathbf{d}\xi \geq 0. \tag{43.27}$$

For $t > 0$ let $p_t(x) := t^{-n/2}e^{-|x|^2/2t} \in \mathcal{S}$ and define

$$I_t(x) := I\bigstar p_t(x) := I(p_t(x-\cdot)) = I(\left|\sqrt{p_t(x-\cdot)}\right|^2)$$

which is non-negative by Eq. (43.27) and the fact that $\sqrt{p_t(x-\cdot)} \in \mathcal{S}$. Using

$$[p_t(x-\cdot)]^\vee(\xi) = \int_{\mathbb{R}^n} p_t(x-y)e^{iy\cdot\xi}\mathbf{d}y = \int_{\mathbb{R}^n} p_t(y)e^{i(y+x)\cdot\xi}\mathbf{d}y$$
$$= e^{ix\cdot\xi}p_t^\vee(\xi) = e^{ix\cdot\xi}e^{-t|\xi|^2/2},$$

$$\langle I_t,\psi\rangle = \int_{\mathbb{R}^n} I(p_t(x-\cdot))\psi(x)\mathbf{d}x$$
$$= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\chi(\xi)\left[p_t(x-\cdot)\right]^\vee(\xi)\psi(x)\mathbf{d}\xi\right)\mathbf{d}x$$
$$= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\chi(\xi)e^{ix\cdot\xi}e^{-t|\xi|^2/2}\psi(x)\mathbf{d}\xi\right)\mathbf{d}x$$
$$= \int_{\mathbb{R}^n}\chi(\xi)\psi^\vee(\xi)e^{-t|\xi|^2/2}\mathbf{d}\xi$$

which coupled with the dominated convergence theorem shows

$$\langle I\bigstar p_t,\psi\rangle \to \int_{\mathbb{R}^n}\chi(\xi)\psi^\vee(\xi)\mathbf{d}\xi = I(\psi) \text{ as } t\downarrow 0.$$

Hence if $\psi \geq 0$, then $I(\psi) = \lim_{t\downarrow 0}\langle I_t,\psi\rangle \geq 0$.

Let $K \subset \mathbb{R}$ be a compact set and $\psi \in C_c(\mathbb{R},[0,\infty))$ be a function such that $\psi = 1$ on K. If $f \in C_c^\infty(\mathbb{R},\mathbb{R})$ is a smooth function with $\operatorname{supp}(f) \subset K$, then $0 \leq \|f\|_\infty \psi - f \in \mathcal{S}$ and hence

$$0 \leq \langle I, \|f\|_\infty\psi - f\rangle = \|f\|_\infty\langle I,\psi\rangle - \langle I,f\rangle$$

and therefore $\langle I,f\rangle \leq \|f\|_\infty\langle I,\psi\rangle$. Replacing f by $-f$ implies, $-\langle I,f\rangle \leq \|f\|_\infty\langle I,\psi\rangle$ and hence we have proved

$$|\langle I,f\rangle| \leq C(\operatorname{supp}(f))\|f\|_\infty \tag{43.28}$$

for all $f \in \mathcal{D}_{\mathbb{R}^n} := C_c^\infty(\mathbb{R}^n,\mathbb{R})$ where $C(K)$ is a finite constant for each compact subset of $\mathbb{R}^n$. Because of the estimate in Eq. (43.28), it follows that $I|_{\mathcal{D}_{\mathbb{R}^n}}$ has a unique extension I to $C_c(\mathbb{R}^n,\mathbb{R})$ still satisfying the estimates in Eq. (43.28) and moreover this extension is still positive. So by the Riesz – Markov Theorem **??**, there exists a unique Radon – measure μ on $\mathbb{R}^n$ such that such that $\langle I,f\rangle = \mu(f)$ for all $f \in C_c(\mathbb{R}^n,\mathbb{R})$.

To finish the proof we must show $\hat{\mu}(\eta) = \chi(\eta)$ for all $\eta \in \mathbb{R}^n$ given

$$\mu(f) = \int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi)\mathbf{d}\xi \text{ for all } f \in C_c^\infty(\mathbb{R}^n,\mathbb{R}). \tag{43.29}$$

Let $f \in C_c^\infty(\mathbb{R}^n,\mathbb{R}_+)$ be a radial function such $f(0) = 1$ and $f(x)$ is decreasing as $|x|$ increases. Let $f_\varepsilon(x) := f(\varepsilon x)$, then by Theorem 43.3,

$$\mathcal{F}^{-1}\left[e^{-i\eta x}f_\varepsilon(x)\right](\xi) = \varepsilon^{-n}f^\vee(\frac{\xi-\eta}{\varepsilon})$$

and therefore, from Eq. (43.29),

$$\int_{\mathbb{R}^n} e^{-i\eta x} f_\varepsilon(x) d\mu(x) = \int_{\mathbb{R}^n} \chi(\xi)\varepsilon^{-n} f^\vee(\frac{\xi-\eta}{\varepsilon})\mathbf{d}\xi. \tag{43.30}$$

Because $\int_{\mathbb{R}^n} f^\vee(\xi)\mathbf{d}\xi = \mathcal{F}f^\vee(0) = f(0) = 1$, we may apply the approximate δ – function Theorem 31.33 to Eq. (43.30) to find

$$\int_{\mathbb{R}^n} e^{-i\eta x} f_\varepsilon(x) d\mu(x) \to \chi(\eta) \text{ as } \varepsilon \downarrow 0. \tag{43.31}$$

On the the other hand, when $\eta = 0$, the monotone convergence theorem implies $\mu(f_\varepsilon) \uparrow \mu(1) = \mu(\mathbb{R}^n)$ and therefore $\mu(\mathbb{R}^n) = \mu(1) = \chi(0) < \infty$. Now knowing the μ is a finite measure we may use the dominated convergence theorem to concluded

$$\mu(e^{-i\eta x} f_\varepsilon(x)) \to \mu(e^{-i\eta x}) = \hat{\mu}(\eta) \text{ as } \varepsilon \downarrow 0$$

for all η. Combining this equation with Eq. (43.31) shows $\hat{\mu}(\eta) = \chi(\eta)$ for all $\eta \in \mathbb{R}^n$. ■

43.7 Supplement: Heisenberg Uncertainty Principle

Suppose that H is a Hilbert space and A, B are two densely defined symmetric operators on H. More explicitly, A is a densely defined symmetric linear operator on H means there is a dense subspace $\mathcal{D}_A \subset H$ and a linear map $A : \mathcal{D}_A \to H$ such that $\langle A\varphi|\psi\rangle = \langle\varphi|A\psi\rangle$ for all $\varphi, \psi \in \mathcal{D}_A$. Let

$$\mathcal{D}_{AB} := \{\varphi \in H : \varphi \in \mathcal{D}_B \text{ and } B\varphi \in \mathcal{D}_A\}$$

and for $\varphi \in \mathcal{D}_{AB}$ let $(AB)\varphi = A(B\varphi)$ with a similar definition of $\mathcal{D}_{BA}$ and BA. Moreover, let $\mathcal{D}_C := \mathcal{D}_{AB} \cap \mathcal{D}_{BA}$ and for $\varphi \in \mathcal{D}_C$, let

$$C\varphi = \frac{1}{i}[A, B]\varphi = \frac{1}{i}(AB - BA)\varphi.$$

Notice that for $\varphi, \psi \in \mathcal{D}_C$ we have

$$\begin{aligned}\langle C\varphi|\psi\rangle &= \frac{1}{i}\{\langle AB\varphi|\psi\rangle - \langle BA\varphi|\psi\rangle\} = \frac{1}{i}\{\langle B\varphi|A\psi\rangle - \langle A\varphi|B\psi\rangle\} \\ &= \frac{1}{i}\{\langle\varphi|BA\psi\rangle - \langle\varphi|AB\psi\rangle\} = \langle\varphi|C\psi\rangle,\end{aligned}$$

so that C is symmetric as well.

Theorem 43.24 (Heisenberg Uncertainty Principle). *Continue the above notation and assumptions,*

$$\frac{1}{2}|\langle\psi|C\psi\rangle| \le \sqrt{\|A\psi\|^2 - \langle\psi|A\psi\rangle} \cdot \sqrt{\|B\psi\|^2 - \langle\psi|B\psi\rangle} \tag{43.32}$$

for all $\psi \in \mathcal{D}_C$. Moreover if $\|\psi\| = 1$ and equality holds in Eq. (43.32), then

$$\begin{aligned}(A - \langle\psi|A\psi\rangle I)\psi &= i\lambda(B - \langle\psi|B\psi\rangle I)\psi \text{ or} \\ (B - \langle\psi|B\psi\rangle I)\psi &= i\lambda(A - \langle\psi|A\psi\rangle I)\psi\end{aligned} \tag{43.33}$$

for some $\lambda \in \mathbb{R}$.

Proof. By homogeneity (43.32) we may assume that $\|\psi\| = 1$. Let $a := \langle\psi|A\psi\rangle$, $b = \langle\psi|B\psi\rangle$, $\tilde{A} = A - aI$, and $\tilde{B} = B - bI$. Then we have still have

$$[\tilde{A}, \tilde{B}] = [A - aI, B - bI] = iC.$$

Now

$$\begin{aligned}i\langle\psi|C\psi\rangle &= -\langle\psi|iC\psi\rangle = -\langle\psi|[\tilde{A}, \tilde{B}]\psi\rangle = -\langle\psi|\tilde{A}\tilde{B}\psi\rangle + \langle\psi|\tilde{B}\tilde{A}\psi\rangle \\ &= -\left(\langle\tilde{A}\psi|\tilde{B}\psi\rangle - \langle\tilde{B}\psi|\tilde{A}\psi\rangle\right) = -2i\,\mathrm{Im}\langle\tilde{A}\psi|\tilde{B}\psi\rangle\end{aligned}$$

from which we learn

$$|\langle\psi|C\psi\rangle| = 2\left|\mathrm{Im}\langle\tilde{A}\psi|\tilde{B}\psi\rangle\right| \le 2\left|\langle\tilde{A}\psi|\tilde{B}\psi\rangle\right| \le 2\left\|\tilde{A}\psi\right\|\left\|\tilde{B}\psi\right\| \tag{43.34}$$

with equality iff $\mathrm{Re}\langle\tilde{A}\psi|\tilde{B}\psi\rangle = 0$ and $\tilde{A}\psi$ and $\tilde{B}\psi$ are linearly dependent, i.e. iff Eq. (43.33) holds. Equation (43.32) now follows from the inequality in Eq. (43.34) and the identities,

$$\begin{aligned}\left\|\tilde{A}\psi\right\|^2 &= \|A\psi - a\psi\|^2 = \|A\psi\|^2 + a^2\|\psi\|^2 - 2a\,\mathrm{Re}\langle A\psi|\psi\rangle \\ &= \|A\psi\|^2 + a^2 - 2a^2 = \|A\psi\|^2 - \langle A\psi|\psi\rangle\end{aligned}$$

and similarly

$$\left\|\tilde{B}\psi\right\| = \|B\psi\|^2 - \langle B\psi|\psi\rangle.$$

■

Example 43.25. As an example, take $H = L^2(\mathbb{R})$, $A = \frac{1}{i}\partial_x$ and $B = M_x$ with $\mathcal{D}_A := \{f \in H : f' \in H\}$ (f' is the weak derivative) and $\mathcal{D}_B := \left\{f \in H : \int_{\mathbb{R}} |xf(x)|^2 dx < \infty\right\}$. In this case,

$$\mathcal{D}_C = \{f \in H : f', xf \text{ and } xf' \text{ are in } H\}$$

and $C = -I$ on $\mathcal{D}_C$. Therefore for a **unit** vector $\psi \in \mathcal{D}_C$,

$$\frac{1}{2} \le \left\|\frac{1}{i}\psi' - a\psi\right\|_2 \cdot \|x\psi - b\psi\|_2$$

where $a = i\int_{\mathbb{R}} \psi\bar{\psi}' dm$ [3] and $b = \int_{\mathbb{R}} x\,|\psi(x)|^2\,dm(x)$. Thus we have

$$\frac{1}{4} = \frac{1}{4}\int_{\mathbb{R}} |\psi|^2\,dm \le \int_{\mathbb{R}} (k-a)^2 \left|\hat{\psi}(k)\right|^2 dk \cdot \int_{\mathbb{R}} (x-b)^2\,|\psi(x)|^2\,dx. \qquad (43.35)$$

Equality occurs if there exists $\lambda \in \mathbb{R}$ such that

$$i\lambda\,(x-b)\,\psi(x) = (\frac{1}{i}\partial_x - a)\psi(x) \text{ a.e.}$$

Working formally, this gives rise to the ordinary differential equation (in weak form),

$$\psi_x = [-\lambda(x-b) + ia]\,\psi \qquad (43.36)$$

which has solutions (see Exercise 43.7 below)

$$\psi = C\exp\left(\int_{\mathbb{R}} [-\lambda(x-b)+ia]\,dx\right) = C\exp\left(-\frac{\lambda}{2}(x-b)^2 + iax\right). \qquad (43.37)$$

Let $\lambda = \frac{1}{2t}$ and choose C so that $\|\psi\|_2 = 1$ to find

$$\psi_{t,a,b}(x) = \left(\frac{1}{2t}\right)^{1/4} \exp\left(-\frac{1}{4t}(x-b)^2 + iax\right)$$

are the functions (called **coherent states)** which saturate the Heisenberg uncertainty principle in Eq. (43.35).

43.7.1 Exercises

Exercise 43.6. Suppose μ is a complex measure on $\mathbb{R}^n$ and

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x} d\mu(x)$$

is its Fourier transform as defined in Definition 43.18. Show μ satisfies,

$$\langle\hat{\mu}, \varphi\rangle := \int_{\mathbb{R}^n} \hat{\mu}(\xi)\varphi(\xi)\mathbf{d}\xi = \mu(\hat{\varphi}) := \int_{\mathbb{R}^n} \hat{\varphi}\,d\mu \text{ for all } \varphi \in \mathcal{S}$$

and use this to show if μ is a complex measure such that $\hat{\mu} \equiv 0$, then $\mu \equiv 0$.

[3] The constant a may also be described as

$$a = i\int_{\mathbb{R}} \psi\bar{\psi}' dm = \sqrt{2\pi}i\int_{\mathbb{R}} \hat{\psi}(\xi)\overline{(\bar{\psi}')^{\hat{}}}(\xi)d\xi = \int_{\mathbb{R}} \xi \left|\hat{\psi}(\xi)\right|^2 dm(\xi).$$

Exercise 43.7. Show that ψ described in Eq. (43.37) is the general solution to Eq. (43.36). **Hint:** Suppose that φ is any solution to Eq. (43.36) and ψ is given as in Eq. (43.37) with $C = 1$. Consider the weak – differential equation solved by φ/ψ.

43.7.2 More Proofs of the Fourier Inversion Theorem

Exercise 43.8. Suppose that $f \in L^1(\mathbb{R})$ and assume that f continuously differentiable in a neighborhood of 0, show

$$\lim_{M\to\infty}\int_{-\infty}^{\infty} \frac{\sin Mx}{x} f(x)dx = \pi f(0) \qquad (43.38)$$

using the following steps.

1. Use Example 12.12 to deduce,
$$\lim_{M\to\infty}\int_{-1}^{1} \frac{\sin Mx}{x}dx = \lim_{M\to\infty}\int_{-M}^{M} \frac{\sin x}{x}dx = \pi.$$
2. Explain why
$$0 = \lim_{M\to\infty}\int_{|x|\ge 1} \sin Mx \cdot \frac{f(x)}{x}dx \text{ and}$$
$$0 = \lim_{M\to\infty}\int_{|x|\le 1} \sin Mx \cdot \frac{f(x)-f(0)}{x}dx.$$
3. Adding identities, making use part (1), proves Eq. (43.38).

Exercise 43.9 (Fourier Inversion Formula). Suppose that $f \in L^1(\mathbb{R})$ such that $\hat{f} \in L^1(\mathbb{R})$.

1. Further assume that f is continuously differentiable in a neighborhood of 0. Show that
$$\Lambda := \int_{\mathbb{R}} \hat{f}(\xi)\mathbf{d}\xi = f(0).$$
Hint: by the dominated convergence theorem, $\Lambda := \lim_{M\to\infty}\int_{|\xi|\le M} \hat{f}(\xi)\mathbf{d}\xi$. Now use the definition of $\hat{f}(\xi)$, Fubini's theorem and Exercise 43.8.
2. Apply part 1. of this exercise with f replace by $\tau_{-y}f := f(\cdot + y)$ for some $y \in \mathbb{R}$ to prove
$$f(y) = \int_{\mathbb{R}} \hat{f}(\xi)e^{iy\cdot\xi}\mathbf{d}\xi \qquad (43.39)$$
provided f is now continuously differentiable near y.

The goal of the next exercises is to give yet another proof of the Fourier inversion formula.

Notation 43.26 *For $L>0$, let $C_L^k(\mathbb{R})$ denote the space of C^k – $2\pi L$ periodic functions:*

$$C_L^k(\mathbb{R}):=\left\{f\in C^k(\mathbb{R}):f(x+2\pi L)=f(x)\text{ for all }x\in\mathbb{R}\right\}.$$

Also let $\langle\cdot,\cdot\rangle_L$ denote the inner product on the Hilbert space $H_L:=L^2([-\pi L,\pi L])$ given by

$$\langle f|g\rangle_L:=\frac{1}{2\pi L}\int_{[-\pi L,\pi L]}f(x)\bar{g}(x)dx.$$

Exercise 43.10. Recall that $\left\{\chi_k^L(x):=e^{ikx/L}:k\in\mathbb{Z}\right\}$ is an orthonormal basis for H_L and in particular for $f\in H_L$,

$$f=\sum_{k\in\mathbb{Z}}\langle f|\chi_k^L\rangle_L\chi_k^L \tag{43.40}$$

where the convergence takes place in $L^2([-\pi L,\pi L])$. Suppose now that $f\in C_L^2(\mathbb{R})$[4]. Show (by two integration by parts)

$$\left|\langle f|\chi_k^L\rangle_L\right|\le\frac{L^2}{k^2}\left\|f''\right\|_\infty$$

where $\|g\|_\infty$ denote the uniform norm of a function g. Use this to conclude that the sum in Eq. (43.40) is uniformly convergent and from this conclude that Eq. (43.40) holds pointwise.

Note: it is enough to assume $f\in C_L^1(\mathbb{R})$ by making use of the identity,

$$\left|\langle f|\chi_k^L\rangle_L\right|=\frac{L}{|k|}\left|\langle f'|\chi_k^L\rangle_L\right|$$

along with the Cauchy Schwarz inequality to see

$$\left(\sum_{k\neq0}\left|\langle f|\chi_k^L\rangle_L\right|\right)^2\le\sum_{k\neq0}\left|\langle f'|\chi_k^L\rangle_L\right|^2\cdot\sum_{k\neq0}\left(\frac{L}{|k|}\right)^2.$$

Exercise 43.11 (Fourier Inversion Formula on $\mathcal{S}$). Let $f\in\mathcal{S}(\mathbb{R})$, $L>0$ and

$$f_L(x):=\sum_{k\in\mathbb{Z}}f(x+2\pi kL). \tag{43.41}$$

Show:

1. The sum defining f_L is convergent and moreover that $f_L\in C_L^\infty(\mathbb{R})$.
2. Show $\langle f_L|\chi_k^L\rangle_L=\frac{1}{\sqrt{2\pi}L}\hat{f}(k/L)$.
3. Conclude from Exercise 43.10 that
$$f_L(x)=\frac{1}{\sqrt{2\pi}L}\sum_{k\in\mathbb{Z}}\hat{f}(k/L)e^{ikx/L}\text{ for all }x\in\mathbb{R}. \tag{43.42}$$
4. Show, by passing to the limit, $L\to\infty$, in Eq. (43.42) that Eq. (43.39) holds for all $x\in\mathbb{R}$. **Hint:** Recall that $\hat{f}\in\mathcal{S}$.

Exercise 43.12 (See Exercise 18.15). Folland 8.13 on p. 254.

Exercise 43.13 (Wirtinger's inequality). Given $a>0$ and $f\in C^1([0,a],\mathbb{C})$ such that $f(0)=f(a)=0$, show

$$\int_0^a|f(x)|^2dx\le\left(\frac{a}{\pi}\right)^2\int_0^a|f'(x)|^2dx.$$

Hint: to use the notation above, let $L=\pi/a$ and extend f to $[-a,0]$ by setting $f(-x)=-f(x)$ for $0\le x\le a$ and then extend f. Now compute $\int_0^a|f(x)|^2dx$ and $\int_0^a|f'(x)|^2dx$ in terms of their Fourier coefficients, $\left\langle f|\chi_k^L\right\rangle_L$ and $\left\langle f'|\chi_k^L\right\rangle_L$ respectively.

Exercise 43.14 (Sampling Theorem). Let

$$\operatorname{sinc}x=\begin{cases}\frac{\sin\pi x}{\pi x}&\text{if }x\neq0\\1&\text{if }x=0\end{cases}$$

and for any $a\in(0,\infty)$, let

$$\mathcal{H}_a=\{f\in L^2(m):\hat{f}(\xi)=0\text{ a.e. when }|\xi|>\pi a\}.$$

Show

1. Show that every $f\in\mathcal{H}_a$ has a version[5] $f_0\in C_0(\mathbb{R})$ and moreover,
$$\|f_0\|_u\le\sqrt{a}\,\|f\|_{L^2(m)}. \tag{43.43}$$
[We now identify f with this continuous version.] **Hint:** after identifying $L^2([-\pi a,\pi a],\lambda)$ as a subspace of $L^2(\mathbb{R},\lambda)$ one has
$$\mathcal{H}_a=\mathcal{F}^{-1}L^2([-\pi a,\pi a],\lambda).$$

[4] We view $C_L^2(\mathbb{R})$ as a subspace of H_L by identifying $f\in C_L^2(\mathbb{R})$ with $f|_{[-\pi L,\pi L]}\in H_L$.

[5] We say that f_0 is a version of f if $f(x)=f_0(x)$ for m – a.e. x.

2. Show by direct computation that

$$\mathcal{F}^{-1}\left[\frac{1}{\sqrt{2\pi a}}e^{-in\xi/a}1_{|\xi|\le\pi a}\right](x)=\operatorname{sinc}(ax-n).$$

3. If $f\in\mathcal{H}_a$ then (assuming f is the C_0 – version as in part a), show

$$f(x)=\sum_{k=-\infty}^{\infty}f(k/a)\operatorname{sinc}(ax-k),$$

where the series converges both uniformly and in L^2. [**Hint:** Start by writing $\hat{f}(\xi)$ for $|\xi|\le\pi a$ as a Fourier expansion in the orthonormal basis $\left\{e^{-in\xi/a}\right\}_{n=-\infty}^{\infty}$ for $L^2\left([-\pi a,\pi a],\frac{m}{2\pi a}\right)$.]

In the terminology of signal analysis, a signal of band width $2\pi a$ is completely determined by sampling its value at a sequence of points $\{k/2\pi a\}$ whose spacing is the reciprocal of the bandwidth.

Exercise 43.15. Folland 8.16 on p. 255.

Exercise 43.16. Folland 8.17 on p. 255.

Exercise 43.17. Let $\lambda:=(2\pi)^{-n/2}m$ where m is Lebesgue measure on $\mathbb{R}^n$. Suppose that $f\in L^2(\lambda)$ such that $f=f1_S$ a.e. for some $S\in\mathcal{B}_{\mathbb{R}^n}$ with $\lambda(S)<\infty$. Show for ever $E\in\mathcal{B}_{\mathbb{R}^n}$ that

$$\int_E\left|\hat{f}\right|^2d\lambda\le\|f\|_{L^2(\lambda)}^2\lambda(S)\cdot\lambda(E).$$

(The Fourier transform of a function whose support has finite measure.)

Exercise 43.18. Folland 8.22 on p. 256. (Bessel functions.)

Exercise 43.19. Folland 8.23 on p. 256. (Hermite Polynomial problems and Harmonic oscillators.)

Exercise 43.20. Folland 8.31 on p. 263. (Poisson Summation formula problem.)

43.7.3 Disregard this subsection on old variants of Corollary 43.17

Proof. (Old proof of Corollary 43.17._ By definition $p(\partial)f=g$ in L^2 iff

$$\langle g,\varphi\rangle=\langle f,p(-\partial)\varphi\rangle\text{ for all }\varphi\in C_c^\infty(\mathbb{R}^n).\tag{43.44}$$

If follows from repeated use of Lemma 41.23 that the previous equation is equivalent to

$$\langle g,\varphi\rangle=\langle f,p(-\partial)\varphi\rangle\text{ for all }\varphi\in\mathcal{S}(\mathbb{R}^n).\tag{43.45}$$

This may also be easily proved directly as well as follows. Choose $\psi\in C_c^\infty(\mathbb{R}^n)$ such that $\psi(x)=1$ for $x\in B_0(1)$ and for $\varphi\in\mathcal{S}(\mathbb{R}^n)$ let $\varphi_n(x):=\psi(x/n)\varphi(x)$. By the chain rule and the product rule (Eq. **??** of Appendix **??**),

$$\partial^\alpha\varphi_n(x)=\sum_{\beta\le\alpha}\binom{\alpha}{\beta}n^{-|\beta|}\left(\partial^\beta\psi\right)(x/n)\cdot\partial^{\alpha-\beta}\varphi(x)$$

along with the dominated convergence theorem shows $\varphi_n\to\varphi$ and $\partial^\alpha\varphi_n\to\partial^\alpha\varphi$ in L^2 as $n\to\infty$. Therefore if Eq. (43.44) holds, we find Eq. (43.45) holds because

$$\langle g,\varphi\rangle=\lim_{n\to\infty}\langle g,\varphi_n\rangle=\lim_{n\to\infty}\langle f,p(-\partial)\varphi_n\rangle=\langle f,p(-\partial)\varphi\rangle.$$

To complete the proof simply observe that $\langle g,\varphi\rangle=\langle\hat{g},\varphi^\vee\rangle$ and

$$\begin{aligned}\langle f,p(-\partial)\varphi\rangle&=\langle\hat{f},[p(-\partial)\varphi]^\vee\rangle=\langle\hat{f}(\xi),p(i\xi)\varphi^\vee(\xi)\rangle\\&=\langle p(i\xi)\hat{f}(\xi),\varphi^\vee(\xi)\rangle\end{aligned}$$

for all $\varphi\in\mathcal{S}(\mathbb{R}^n)$. From these two observations and the fact that $\mathcal{F}$ is bijective on $\mathcal{S}$, one sees that Eq. (43.45) holds iff $\xi\to p(i\xi)\hat{f}(\xi)\in L^2$ and $\hat{g}(\xi)=p(i\xi)\hat{f}(\xi)$ for a.e. ξ. ■

Exercise 43.21 (See Corollary 43.17). Let $f\in L^2(\mathbb{R}^n)$ and α be a multi-index. If $\partial^\alpha f$ exists in $L^2(\mathbb{R}^n)$[6] then $\mathcal{F}(\partial^\alpha f)=(i\xi)^\alpha\hat{f}(\xi)$ in $L^2(\mathbb{R}^n)$ and conversely if $\left(\xi\to\xi^\alpha\hat{f}(\xi)\right)\in L^2(\mathbb{R}^n)$ then $\partial^\alpha f$ exists.

[6] Here we say that $\partial^\alpha f=g$ exists in $L^2(\mathbb{R}^n)$ iff $\langle g,\varphi\rangle=(-1)^{|\alpha|}\langle f,\partial^\alpha\varphi\rangle$ for all $\varphi\in C_c^\infty(\mathbb{R}^n)$.

Constant Coefficient partial differential equations

Suppose that $p(\xi) = \sum_{|\alpha| \le k} a_\alpha \xi^\alpha$ with $a_\alpha \in \mathbb{C}$ and

$$L = p(D_x) := \Sigma_{|\alpha| \le N} a_\alpha D_x^\alpha = \Sigma_{|\alpha| \le N} a_\alpha \left(\frac{1}{i}\partial_x\right)^\alpha. \tag{44.1}$$

Then for $f \in \mathcal{S}$

$$\widehat{Lf}(\xi) = p(\xi)\hat{f}(\xi),$$

that is to say the Fourier transform takes a constant coefficient partial differential operator to multiplication by a polynomial. This fact can often be used to solve constant coefficient partial differential equation. For example suppose $g : \mathbb{R}^n \to \mathbb{C}$ is a given function and we want to find a solution to the equation $Lf = g$. Taking the Fourier transform of both sides of the equation $Lf = g$ would imply $p(\xi)\hat{f}(\xi) = \hat{g}(\xi)$ and therefore $\hat{f}(\xi) = \hat{g}(\xi)/p(\xi)$ provided $p(\xi)$ is never zero. (We will discuss what happens when $p(\xi)$ has zeros a bit more later on.) So we should expect

$$f(x) = \mathcal{F}^{-1}\left(\frac{1}{p(\xi)}\hat{g}(\xi)\right)(x) = \mathcal{F}^{-1}\left(\frac{1}{p(\xi)}\right) \bigstar g(x).$$

Definition 44.1. *Let $L = p(D_x)$ as in Eq. (44.1). Then we let $\sigma(L) :=$Ran$(p) \subset \mathbb{C}$ and call $\sigma(L)$ the* ***spectrum*** *of L. Given a measurable function $G : \sigma(L) \to \mathbb{C}$, we define (a possibly unbounded operator) $G(L) : L^2(\mathbb{R}^n, m) \to L^2(\mathbb{R}^n, m)$ by*

$$G(L)f := \mathcal{F}^{-1} M_{G\circ p} \mathcal{F}$$

where $M_{G\circ p}$ denotes the operation on $L^2(\mathbb{R}^n, m)$ of multiplication by $G \circ p$, i.e.

$$M_{G\circ p} f = (G \circ p) f$$

with domain given by those $f \in L^2$ such that $(G \circ p) f \in L^2$.

At a formal level we expect

$$G(L)f = \mathcal{F}^{-1}(G \circ p) \bigstar g.$$

44.1 Elliptic examples

As a specific example consider the equation

$$\left(-\Delta + m^2\right) f = g \tag{44.2}$$

where $f, g : \mathbb{R}^n \to \mathbb{C}$ and $\Delta = \sum_{i=1}^n \partial^2/\partial x_i^2$ is the usual Laplacian on $\mathbb{R}^n$. By Corollary 43.17 (i.e. taking the Fourier transform of this equation), solving Eq. (44.2) with $f, g \in L^2$ is equivalent to solving

$$\left(|\xi|^2 + m^2\right) \hat{f}(\xi) = \hat{g}(\xi). \tag{44.3}$$

The unique solution to this latter equation is

$$\hat{f}(\xi) = \left(|\xi|^2 + m^2\right)^{-1} \hat{g}(\xi)$$

and therefore,

$$f(x) = \mathcal{F}^{-1}\left(\left(|\xi|^2 + m^2\right)^{-1} \hat{g}(\xi)\right)(x) =: \left(-\Delta + m^2\right)^{-1} g(x).$$

We expect

$$\mathcal{F}^{-1}\left(\left(|\xi|^2 + m^2\right)^{-1} \hat{g}(\xi)\right)(x) = G_m \bigstar g(x) = \int_{\mathbb{R}^n} G_m(x-y) g(y) \mathbf{d}y,$$

where

$$G_m(x) := \mathcal{F}^{-1}\left(|\xi|^2 + m^2\right)^{-1}(x) = \int_{\mathbb{R}^n} \frac{1}{m^2 + |\xi|^2} e^{i\xi\cdot x} \mathbf{d}\xi.$$

At the moment $\mathcal{F}^{-1}\left(|\xi|^2 + m^2\right)^{-1}$ only makes sense when $n = 1, 2,$ or 3 because only then is $\left(|\xi|^2 + m^2\right)^{-1} \in L^2(\mathbb{R}^n)$.

For now we will restrict our attention to the one dimensional case, $n = 1$, in which case

$$G_m(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{(\xi + mi)(\xi - mi)} e^{i\xi x} d\xi. \tag{44.4}$$

The function G_m may be computed using standard complex variable contour integration methods to find, for $x \geq 0$,

$$G_m(x) = \frac{1}{\sqrt{2\pi}} 2\pi i \frac{e^{i^2 mx}}{2im} = \frac{1}{2m}\sqrt{2\pi} e^{-mx}$$

and since G_m is an even function,

$$G_m(x) = \mathcal{F}^{-1}\left(|\xi|^2 + m^2\right)^{-1}(x) = \frac{\sqrt{2\pi}}{2m} e^{-m|x|}. \tag{44.5}$$

This result is easily verified to be correct, since

$$\begin{aligned}\mathcal{F}\left[\frac{\sqrt{2\pi}}{2m} e^{-m|x|}\right](\xi) &= \frac{\sqrt{2\pi}}{2m}\int_{\mathbb{R}} e^{-m|x|} e^{-ix\cdot\xi} \mathbf{d}x \\ &= \frac{1}{2m}\left(\int_0^\infty e^{-mx} e^{-ix\cdot\xi} dx + \int_{-\infty}^0 e^{mx} e^{-ix\cdot\xi} dx\right) \\ &= \frac{1}{2m}\left(\frac{1}{m+i\xi} + \frac{1}{m-i\xi}\right) = \frac{1}{m^2+\xi^2}.\end{aligned}$$

Hence in conclusion we find that $\left(-\Delta + m^2\right) f = g$ has solution given by

$$f(x) = G_m \bigstar g(x) = \frac{\sqrt{2\pi}}{2m}\int_{\mathbb{R}} e^{-m|x-y|} g(y)\mathbf{d}y = \frac{1}{2m}\int_{\mathbb{R}} e^{-m|x-y|} g(y) dy.$$

Question. Why do we get a unique answer here given that $f(x) = A\sinh(x) + B\cosh(x)$ solves

$$\left(-\Delta + m^2\right) f = 0?$$

The answer is that such an f is not in L^2 unless $f = 0$! More generally it is worth noting that $A\sinh(x) + B\cosh(x)$ is not in $\mathcal{P}$ unless $A = B = 0$.

What about when $m = 0$ in which case $m^2 + \xi^2$ becomes ξ^2 which has a zero at 0. Noting that constants are solutions to $\Delta f = 0$, we might look at

$$\lim_{m\downarrow 0}(G_m(x) - 1) = \lim_{m\downarrow 0}\frac{\sqrt{2\pi}}{2m}(e^{-m|x|} - 1) = -\frac{\sqrt{2\pi}}{2}|x|.$$

as a solution, i.e. we might conjecture that

$$f(x) := -\frac{1}{2}\int_{\mathbb{R}} |x-y| \, g(y) dy$$

solves the equation $-f'' = g$. To verify this we have

$$f(x) := -\frac{1}{2}\int_{-\infty}^x (x-y)\, g(y) dy - \frac{1}{2}\int_x^\infty (y-x)\, g(y) dy$$

so that

$$\begin{aligned}f'(x) &= -\frac{1}{2}\int_{-\infty}^x g(y) dy + \frac{1}{2}\int_x^\infty g(y) dy \text{ and} \\ f''(x) &= -\frac{1}{2} g(x) - \frac{1}{2} g(x).\end{aligned}$$

44.2 Poisson Semi-Group

Let us now consider the problems of finding a function $(x_0, x) \in [0,\infty) \times \mathbb{R}^n \to u(x_0, x) \in \mathbb{C}$ such that

$$\left(\frac{\partial^2}{\partial x_0^2} + \Delta\right) u = 0 \text{ with } u(0,\cdot) = f \in L^2(\mathbb{R}^n). \tag{44.6}$$

Let $\hat{u}(x_0,\xi) := \int_{\mathbb{R}^n} u(x_0,x) e^{-ix\cdot\xi} \mathbf{d}x$ denote the Fourier transform of u in the $x \in \mathbb{R}^n$ variable. Then Eq. (44.6) becomes

$$\left(\frac{\partial^2}{\partial x_0^2} - |\xi|^2\right) \hat{u}(x_0,\xi) = 0 \text{ with } \hat{u}(0,\xi) = \hat{f}(\xi) \tag{44.7}$$

and the general solution to this differential equation ignoring the initial condition is of the form

$$\hat{u}(x_0,\xi) = A(\xi) e^{-x_0|\xi|} + B(\xi) e^{x_0|\xi|} \tag{44.8}$$

for some function $A(\xi)$ and $B(\xi)$. Let us now impose the extra condition that $u(x_0,\cdot) \in L^2(\mathbb{R}^n)$ or equivalently that $\hat{u}(x_0,\cdot) \in L^2(\mathbb{R}^n)$ for all $x_0 \geq 0$. The solution in Eq. (44.8) will not have this property unless $B(\xi)$ decays very rapidly at ∞. The simplest way to achieve this is to assume $B = 0$ in which case we now get a unique solution to Eq. (44.7), namely

$$\hat{u}(x_0,\xi) = \hat{f}(\xi) e^{-x_0|\xi|}.$$

Applying the inverse Fourier transform gives

$$u(x_0,x) = \mathcal{F}^{-1}\left[\hat{f}(\xi) e^{-x_0|\xi|}\right](x) =: \left(e^{-x_0\sqrt{-\Delta}} f\right)(x)$$

and moreover

$$\left(e^{-x_0\sqrt{-\Delta}} f\right)(x) = P_{x_0} * f(x)$$

where $P_{x_0}(x) = (2\pi)^{-n/2}\left(\mathcal{F}^{-1}e^{-x_0|\xi|}\right)(x)$. From Exercise 43.4,

$$P_{x_0}(x) = (2\pi)^{-n/2}\left(\mathcal{F}^{-1}e^{-x_0|\xi|}\right)(x) = c_n \frac{x_0}{(x_0^2 + |x|^2)^{(n+1)/2}}$$

where

$$c_n = (2\pi)^{-n/2}\frac{\Gamma((n+1)/2)}{\sqrt{\pi}2^{n/2}} = \frac{\Gamma((n+1)/2)}{2^n\pi^{(n+1)/2}}.$$

Hence we have proved the following proposition.

Proposition 44.2. *For $f \in L^2(\mathbb{R}^n)$,*

$$e^{-x_0\sqrt{-\Delta}}f = P_{x_0} * f \text{ for all } x_0 \geq 0$$

and the function $u(x_0,x) := e^{-x_0\sqrt{-\Delta}}f(x)$ is C^∞ for $(x_0,x) \in (0,\infty)\times\mathbb{R}^n$ and solves Eq. (44.6).

44.3 Heat Equation on $\mathbb{R}^n$

The heat equation for a function $u : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{C}$ is the partial differential equation

$$\left(\partial_t - \frac{1}{2}\Delta\right)u = 0 \text{ with } u(0,x) = f(x), \tag{44.9}$$

where f is a given function on $\mathbb{R}^n$. By Fourier transforming Eq. (44.9) in the x – variables only, one finds that (44.9) implies that

$$\left(\partial_t + \frac{1}{2}|\xi|^2\right)\hat{u}(t,\xi) = 0 \text{ with } \hat{u}(0,\xi) = \hat{f}(\xi). \tag{44.10}$$

and hence that $\hat{u}(t,\xi) = e^{-t|\xi|^2/2}\hat{f}(\xi)$. Inverting the Fourier transform then shows that

$$u(t,x) = \mathcal{F}^{-1}\left(e^{-t|\xi|^2/2}\hat{f}(\xi)\right)(x) = \left(\mathcal{F}^{-1}\left(e^{-t|\xi|^2/2}\right)\bigstar f\right)(x) =: e^{t\Delta/2}f(x).$$

From Example 43.4,

$$\mathcal{F}^{-1}\left(e^{-t|\xi|^2/2}\right)(x) = p_t(x) = t^{-n/2}e^{-\frac{1}{2t}|x|^2}$$

and therefore,

$$u(t,x) = \int_{\mathbb{R}^n} p_t(x-y)f(y)\mathbf{d}y.$$

This suggests the following theorem.

Theorem 44.3. *Let*

$$\rho(t,x,y) := (2\pi t)^{-n/2}\, e^{-|x-y|^2/2t} \tag{44.11}$$

*be the **heat kernel** on $\mathbb{R}^n$. Then*

$$\left(\partial_t - \frac{1}{2}\Delta_x\right)\rho(t,x,y) = 0 \text{ and } \lim_{t\downarrow 0}\rho(t,x,y) = \delta_x(y), \tag{44.12}$$

where δ_x is the δ – function at x in $\mathbb{R}^n$. More precisely, if f is a continuous bounded (can be relaxed considerably) function on $\mathbb{R}^n$, then

$$u(t,x) = \int_{\mathbb{R}^n}\rho(t,x,y)f(y)dy$$

is a solution to Eq. (44.9) where $u(0,x) := \lim_{t\downarrow 0}u(t,x)$.

Proof. Direct computations show that $\left(\partial_t - \frac{1}{2}\Delta_x\right)\rho(t,x,y) = 0$ and an application of Theorem 31.33 shows $\lim_{t\downarrow 0}\rho(t,x,y) = \delta_x(y)$ or equivalently that $\lim_{t\downarrow 0}\int_{\mathbb{R}^n}\rho(t,x,y)f(y)dy = f(x)$ uniformly on compact subsets of $\mathbb{R}^n$. This shows that $\lim_{t\downarrow 0}u(t,x) = f(x)$ uniformly on compact subsets of $\mathbb{R}^n$. ■

This notation suggests that we should be able to compute the solution to g to $(\Delta - m^2)g = f$ using

$$\begin{aligned} g(x) = \left(m^2 - \Delta\right)^{-1}f(x) &= \int_0^\infty \left(e^{-(m^2-\Delta)t}f\right)(x)dt \\ &= \int_0^\infty\left(e^{-m^2t}p_{2t}\bigstar f\right)(x)dt, \end{aligned}$$

a fact which is easily verified using the Fourier transform. This gives us a method to compute $G_m(x)$ from the previous section, namely

$$G_m(x) = \int_0^\infty e^{-m^2t}p_{2t}(x)dt = \int_0^\infty (2t)^{-n/2}e^{-m^2t-\frac{1}{4t}|x|^2}dt.$$

We make the change of variables, $\lambda = |x|^2/4t$ ($t = |x|^2/4\lambda$, $dt = -\frac{|x|^2}{4\lambda^2}d\lambda$) to find

$$\begin{aligned} G_m(x) &= \int_0^\infty (2t)^{-n/2}e^{-m^2t-\frac{1}{4t}|x|^2}dt = \int_0^\infty\left(\frac{|x|^2}{2\lambda}\right)^{-n/2}e^{-m^2|x|^2/4\lambda-\lambda}\frac{|x|^2}{(2\lambda)^2}d\lambda \\ &= \frac{2^{(n/2-2)}}{|x|^{n-2}}\int_0^\infty\lambda^{n/2-2}e^{-\lambda}e^{-m^2|x|^2/4\lambda}d\lambda. \end{aligned} \tag{44.13}$$

In case $n = 3$, Eq. (44.13) becomes

$$G_m(x) = \frac{\sqrt{\pi}}{\sqrt{2}\,|x|}\int_0^\infty \frac{1}{\sqrt{\pi\lambda}} e^{-\lambda} e^{-m^2|x|^2/4\lambda} d\lambda = \frac{\sqrt{\pi}}{\sqrt{2}\,|x|} e^{-m|x|}$$

where the last equality follows from Exercise 43.3. Hence when $n = 3$ we have found

$$\begin{aligned}\left(m^2 - \Delta\right)^{-1} f(x) = G_m \bigstar f(x) &= (2\pi)^{-3/2} \int_{\mathbb{R}^3} \frac{\sqrt{\pi}}{\sqrt{2}\,|x-y|} e^{-m|x-y|} f(y) dy \\ &= \int_{\mathbb{R}^3} \frac{1}{4\pi\,|x-y|} e^{-m|x-y|} f(y) dy. \qquad (44.14)\end{aligned}$$

The function $\frac{1}{4\pi|x|} e^{-m|x|}$ is called the Yukawa potential.

Let us work out $G_m(x)$ for n odd. By differentiating Eq. (43.17) of Exercise 43.3 we find

$$\begin{aligned}\int_0^\infty d\lambda \lambda^{k-1/2} e^{-\frac{1}{4\lambda}x^2} e^{-\lambda m^2} &= \int_0^\infty d\lambda \frac{1}{\sqrt{\lambda}} e^{-\frac{1}{4\lambda}x^2} \left(-\frac{d}{da}\right)^k e^{-\lambda a}|_{a=m^2} \\ &= \left(-\frac{d}{da}\right)^k \frac{\sqrt{\pi}}{\sqrt{a}} e^{-\sqrt{a}x} = p_{m,k}(x) e^{-mx}\end{aligned}$$

where $p_{m,k}(x)$ is a polynomial in x with $\deg p_m = k$ with

$$\begin{aligned}p_{m,k}(0) &= \sqrt{\pi}\left(-\frac{d}{da}\right)^k a^{-1/2}|_{a=m^2} = \sqrt{\pi}(\frac{1}{2}\frac{3}{2}\dots\frac{2k-1}{2})m^{2k+1} \\ &= m^{2k+1}\sqrt{\pi}2^{-k}(2k-1)!!.\end{aligned}$$

Letting $k - 1/2 = n/2 - 2$ and $m = 1$ we find $k = \frac{n-1}{2} - 2 \in \mathbb{N}$ for $n = 3, 5, \dots$. and we find

$$\int_0^\infty \lambda^{n/2-2} e^{-\frac{1}{4\lambda}x^2} e^{-\lambda} d\lambda = p_{1,k}(x) e^{-x} \text{ for all } x > 0.$$

Therefore,

$$G_m(x) = \frac{2^{(n/2-2)}}{|x|^{n-2}} \int_0^\infty \lambda^{n/2-2} e^{-\lambda} e^{-m^2|x|^2/4\lambda} d\lambda = \frac{2^{(n/2-2)}}{|x|^{n-2}} p_{1,n/2-2}(m\,|x|) e^{-m|x|}.$$

Now for even m, I think we get Bessel functions in the answer. (BRUCE: look this up.) Let us at least work out the asymptotics of $G_m(x)$ for $x \to \infty$. To this end let

$$\psi(y) := \int_0^\infty \lambda^{n/2-2} e^{-(\lambda+\lambda^{-1}y^2)} d\lambda = y^{n-2}\int_0^\infty \lambda^{n/2-2} e^{-(\lambda y^2 + \lambda^{-1})} d\lambda$$

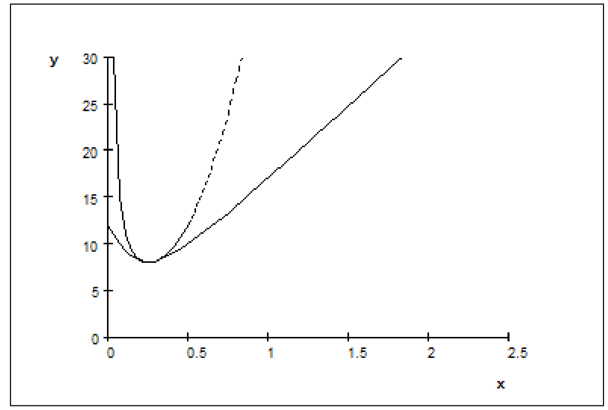
The function $f_y(\lambda) := (y^2\lambda + \lambda^{-1})$ satisfies,

$$f_y'(\lambda) = \left(y^2 - \lambda^{-2}\right) \text{ and } f_y''(\lambda) = 2\lambda^{-3} \text{ and } f_y'''(\lambda) = -6\lambda^{-4}$$

so by Taylor's theorem with remainder we learn

$$f_y(\lambda) \cong 2y + y^3(\lambda - y^{-1})^2 \text{ for all } \lambda > 0,$$

see Figure 44.3 below.



Plot of f_4 and its second order Taylor approximation.

So by the usual asymptotics arguments,

$$\begin{aligned}\psi(y) &\cong y^{n-2}\int_{(-\varepsilon+y^{-1},y^{-1}+\varepsilon)} \lambda^{n/2-2} e^{-(\lambda y^2+\lambda^{-1})} d\lambda \\ &\cong y^{n-2}\int_{(-\varepsilon+y^{-1},y^{-1}+\varepsilon)} \lambda^{n/2-2} \exp\left(-2y - y^3(\lambda - y^{-1})^2\right) d\lambda \\ &\cong y^{n-2}e^{-2y}\int_{\mathbb{R}} \lambda^{n/2-2} \exp\left(-y^3(\lambda - y^{-1})^2\right) d\lambda \text{ (let } \lambda \to \lambda y^{-1}) \\ &= e^{-2y}y^{n-2}y^{-n/2+1}\int_{\mathbb{R}} \lambda^{n/2-2} \exp\left(-y(\lambda-1)^2\right) d\lambda \\ &= e^{-2y}y^{n-2}y^{-n/2+1}\int_{\mathbb{R}} (\lambda+1)^{n/2-2} \exp\left(-y\lambda^2\right) d\lambda.\end{aligned}$$

The point is we are still going to get exponential decay at ∞.

When $m = 0$, Eq. (44.13) becomes

$$G_0(x) = \frac{2^{(n/2-2)}}{|x|^{n-2}}\int_0^\infty \lambda^{n/2-1} e^{-\lambda}\frac{d\lambda}{\lambda} = \frac{2^{(n/2-2)}}{|x|^{n-2}}\Gamma(n/2-1)$$

where $\Gamma(x)$ in the gamma function defined in Eq. (**??**). Hence for "reasonable" functions f (and $n \neq 2$) we expect that (see Proposition 44.4 below)

$$(-\Delta)^{-1}f(x) = G_0 \bigstar f(x) = 2^{(n/2-2)}\Gamma(n/2-1)(2\pi)^{-n/2}\int_{\mathbb{R}^n}\frac{1}{|x-y|^{n-2}}f(y)dy$$
$$= \frac{1}{4\pi^{n/2}}\Gamma(n/2-1)\int_{\mathbb{R}^n}\frac{1}{|x-y|^{n-2}}f(y)dy.$$

The function

$$G(x) := \frac{1}{4\pi^{n/2}}\Gamma(n/2-1)\frac{1}{|x|^{n-2}} \tag{44.15}$$

is a "Green's function" for $-\Delta$. Recall from Exercise 12.12 that, for $n = 2k$, $\Gamma(\frac{n}{2}-1) = \Gamma(k-1) = (k-2)!$, and for $n = 2k+1$,

$$\Gamma(\frac{n}{2}-1) = \Gamma(k-1/2) = \Gamma(k-1+1/2) = \sqrt{\pi}\frac{1\cdot 3\cdot 5\cdot\dots\cdot(2k-3)}{2^{k-1}}$$
$$= \sqrt{\pi}\frac{(2k-3)!!}{2^{k-1}} \text{ where } (-1)!! =: 1.$$

Hence

$$G(x) = \frac{1}{4}\frac{1}{|x|^{n-2}}\begin{cases}\frac{1}{\pi^k}(k-2)! \text{ if } & n = 2k\\ \frac{1}{\pi^k}\frac{(2k-3)!!}{2^{k-1}} \text{ if } n = 2k+1\end{cases}$$

and in particular when $n = 3$,

$$G(x) = \frac{1}{4\pi}\frac{1}{|x|}$$

which is consistent with Eq. (44.14) with $m = 0$.

Proposition 44.4. *Let $n \geq 3$ and for $x \in \mathbb{R}^n$, let $\rho_t(x) = \rho(t,x,0) := \left(\frac{1}{2\pi t}\right)^{n/2} e^{-\frac{1}{2t}|x|^2}$ (see Eq. (44.11))and $G(x)$ be as in Eq. (44.15) so that*

$$G(x) := \frac{C_n}{|x|^{n-2}} = \frac{1}{2}\int_0^\infty \rho_t(x)\,dt \text{ for } x \neq 0.$$

Then

$$-\Delta(G*u) = -G*\Delta u = u$$

for all $u \in C_c^2(\mathbb{R}^n)$.

Proof. For $f \in C_c(\mathbb{R}^n)$,

$$G*f(x) = C_n\int_{\mathbb{R}^n} f(x-y)\frac{1}{|y|^{n-2}}dy$$

is well defined, since

$$\int_{\mathbb{R}^n}|f(x-y)|\frac{1}{|y|^{n-2}}dy \leq M\int_{|y|\leq R+|x|}\frac{1}{|y|^{n-2}}dy < \infty$$

where M is a bound on f and $\text{supp}(f) \subset B(0,R)$. Similarly, $|x| \leq r$, we have

$$\sup_{|x|\leq r}|f(x-y)|\frac{1}{|y|^{n-2}} \leq M1_{\{|y|\leq R+r\}}\frac{1}{|y|^{n-2}} \in L^1(dy),$$

from which it follows that $G*f$ is a continuous function. Similar arguments show if $f \in C_c^2(\mathbb{R}^n)$, then $G*f \in C^2(\mathbb{R}^n)$ and $\Delta(G*f) = G*\Delta f$. So to finish the proof it suffices to show $G*\Delta u = u$.

For this we now write, making use of Fubini-Tonelli, integration by parts, the fact that $\partial_t\rho_t(y) = \frac{1}{2}\Delta\rho_t(y)$ and the dominated convergence theorem,

$$\begin{aligned}G*\Delta u(x) &= \frac{1}{2}\int_{\mathbb{R}^n}\Delta u(x-y)\left(\int_0^\infty\rho_t(y)\,dt\right)dy\\
&= \frac{1}{2}\int_0^\infty dt\int_{\mathbb{R}^n}\Delta u(x-y)\rho_t(y)\,dy\\
&= \frac{1}{2}\int_0^\infty dt\int_{\mathbb{R}^n}\Delta_y u(x-y)\rho_t(y)\,dy\\
&= \frac{1}{2}\int_0^\infty dt\int_{\mathbb{R}^n}u(x-y)\Delta_y\rho_t(y)\,dy\\
&= \int_0^\infty dt\int_{\mathbb{R}^n}u(x-y)\frac{d}{dt}\rho_t(y)\,dy\\
&= \lim_{\varepsilon\downarrow 0}\int_\varepsilon^\infty dt\int_{\mathbb{R}^n}u(x-y)\frac{d}{dt}\rho_t(y)\,dy\\
&= \lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^n}u(x-y)\left(\int_\varepsilon^\infty\frac{d}{dt}\rho_t(y)\,dt\right)dy\\
&= -\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^n}u(x-y)\rho_\varepsilon(y)\,dy = u(x),\end{aligned}$$

where in the last equality we have used the fact that ρ_t is an approximate δ – sequence. ∎

44.4 Wave Equation on $\mathbb{R}^n$

Let us now consider the wave equation on $\mathbb{R}^n$,

$$\begin{aligned}0 &= \left(\partial_t^2 - \Delta\right)u(t,x) \text{ with}\\
u(0,x) &= f(x) \text{ and } u_t(0,x) = g(x).\end{aligned} \tag{44.16}$$

Taking the Fourier transform in the x variables gives the following equation

$$\begin{aligned}0 &= \hat{u}_{tt}(t,\xi) + |\xi|^2\hat{u}(t,\xi) \text{ with}\\
\hat{u}(0,\xi) &= \hat{f}(\xi) \text{ and } \hat{u}_t(0,\xi) = \hat{g}(\xi).\end{aligned} \tag{44.17}$$

The solution to these equations is

$$\hat{u}(t,\xi) = \hat{f}(\xi)\cos(t|\xi|) + \hat{g}(\xi)\frac{\sin t|\xi|}{|\xi|}$$

and hence we should have

$$\begin{aligned} u(t,x) &= \mathcal{F}^{-1}\left(\hat{f}(\xi)\cos(t|\xi|) + \hat{g}(\xi)\frac{\sin t|\xi|}{|\xi|}\right)(x) \\ &= \mathcal{F}^{-1}\cos(t|\xi|) \bigstar f(x) + \mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} \bigstar g(x) \\ &= \frac{d}{dt}\mathcal{F}^{-1}\left[\frac{\sin t|\xi|}{|\xi|}\right] \bigstar f(x) + \mathcal{F}^{-1}\left[\frac{\sin t|\xi|}{|\xi|}\right] \bigstar g(x). \end{aligned} \tag{44.18}$$

The question now is how interpret this equation. In particular what are the inverse Fourier transforms of $\mathcal{F}^{-1}\cos(t|\xi|)$ and $\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|}$. Since $\frac{d}{dt}\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} \bigstar f(x) = \mathcal{F}^{-1}\cos(t|\xi|) \bigstar f(x)$, it really suffices to understand $\mathcal{F}^{-1}\left[\frac{\sin t|\xi|}{|\xi|}\right]$. The problem we immediately run into here is that $\frac{\sin t|\xi|}{|\xi|} \in L^2(\mathbb{R}^n)$ iff $n = 1$ so that is the case we should start with.

Again by complex contour integration methods one can show

$$\begin{aligned} \left(\mathcal{F}^{-1}\xi^{-1}\sin t\xi\right)(x) &= \frac{\pi}{\sqrt{2\pi}}\left(1_{x+t>0} - 1_{(x-t)>0}\right) \\ &= \frac{\pi}{\sqrt{2\pi}}\left(1_{x>-t} - 1_{x>t}\right) = \frac{\pi}{\sqrt{2\pi}} 1_{[-t,t]}(x) \end{aligned}$$

where in writing the last line we have assume that $t \geq 0$. Again this easily seen to be correct because

$$\begin{aligned} \mathcal{F}\left[\frac{\pi}{\sqrt{2\pi}} 1_{[-t,t]}(x)\right](\xi) &= \frac{1}{2}\int_{\mathbb{R}} 1_{[-t,t]}(x) e^{-i\xi\cdot x} dx = \frac{1}{-2i\xi} e^{-i\xi\cdot x}\big|_{-t}^{t} \\ &= \frac{1}{2i\xi}\left[e^{i\xi t} - e^{-i\xi t}\right] = \xi^{-1}\sin t\xi. \end{aligned}$$

Therefore,

$$\left(\mathcal{F}^{-1}\xi^{-1}\sin t\xi\right) \bigstar f(x) = \frac{1}{2}\int_{-t}^{t} f(x-y)dy$$

and the solution to the one dimensional wave equation is

$$\begin{aligned} u(t,x) &= \frac{d}{dt}\frac{1}{2}\int_{-t}^{t} f(x-y)dy + \frac{1}{2}\int_{-t}^{t} g(x-y)dy \\ &= \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \frac{1}{2}\int_{-t}^{t} g(x-y)dy \\ &= \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \frac{1}{2}\int_{x-t}^{x+t} g(y)dy. \end{aligned}$$

We can arrive at this same solution by more elementary means as follows. We first note in the one dimensional case that wave operator factors, namely

$$0 = \left(\partial_t^2 - \partial_x^2\right) u(t,x) = \left(\partial_t - \partial_x\right)\left(\partial_t + \partial_x\right) u(t,x).$$

Let $U(t,x) := \left(\partial_t + \partial_x\right) u(t,x)$, then the wave equation states $\left(\partial_t - \partial_x\right) U = 0$ and hence by the chain rule $\frac{d}{dt}U(t, x-t) = 0$. So

$$U(t, x-t) = U(0,x) = g(x) + f'(x)$$

and replacing x by $x+t$ in this equation shows

$$\left(\partial_t + \partial_x\right) u(t,x) = U(t,x) = g(x+t) + f'(x+t).$$

Working similarly, we learn that

$$\frac{d}{dt}u(t, x+t) = g(x+2t) + f'(x+2t)$$

which upon integration implies

$$\begin{aligned} u(t,x+t) &= u(0,x) + \int_0^t \left\{g(x+2\tau) + f'(x+2\tau)\right\} d\tau \\ &= f(x) + \int_0^t g(x+2\tau)d\tau + \frac{1}{2} f(x+2\tau)|_0^t \\ &= \frac{1}{2}\left(f(x) + f(x+2t)\right) + \int_0^t g(x+2\tau)d\tau. \end{aligned}$$

Replacing $x \to x-t$ in this equation gives

$$u(t,x) = \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \int_0^t g(x-t+2\tau)d\tau$$

and then letting $y = x - t + 2\tau$ in the last integral shows again that

$$u(t,x) = \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \frac{1}{2}\int_{x-t}^{x+t} g(y)dy.$$

When $n > 3$ it is necessary to treat $\mathcal{F}^{-1}\left[\frac{\sin t|\xi|}{|\xi|}\right]$ as a "distribution" or "generalized function," see Section 45 below. So for now let us take $n = 3$, in which case from Example 43.19 it follows that

$$\mathcal{F}^{-1}\left[\frac{\sin t|\xi|}{|\xi|}\right] = \frac{t}{4\pi t^2}\sigma_t = t\bar{\sigma}_t \tag{44.19}$$

where $\bar{\sigma}_t$ is $\frac{1}{4\pi t^2}\sigma_t$, the surface measure on S_t normalized to have total measure one. Hence from Eq. (44.18) the solution to the three dimensional wave equation should be given by

$$u(t,x) = \frac{d}{dt}\left(t\bar{\sigma}_t \bigstar f(x)\right) + t\bar{\sigma}_t \bigstar g\left(x\right). \tag{44.20}$$

Using this definition in Eq. (44.20) gives

$$\begin{aligned} u(t,x) &= \frac{d}{dt}\left\{t\int_{S_t} f(x-y)d\bar{\sigma}_t(y)\right\} + t\int_{S_t} g(x-y)d\bar{\sigma}_t(y) \\ &= \frac{d}{dt}\left\{t\int_{S_1} f(x-t\omega)d\bar{\sigma}_1(\omega)\right\} + t\int_{S_1} g(x-t\omega)d\bar{\sigma}_1(\omega) \\ &= \frac{d}{dt}\left\{t\int_{S_1} f(x+t\omega)d\bar{\sigma}_1(\omega)\right\} + t\int_{S_1} g(x+t\omega)d\bar{\sigma}_1(\omega). \end{aligned} \tag{44.21}$$

Proposition 44.5. *Suppose $f \in C^3(\mathbb{R}^3)$ and $g \in C^2(\mathbb{R}^3)$, then $u(t,x)$ defined by Eq. (44.21) is in $C^2\left(\mathbb{R}\times\mathbb{R}^3\right)$ and is a classical solution of the wave equation in Eq. (44.16).*

Proof. The fact that $u \in C^2\left(\mathbb{R}\times\mathbb{R}^3\right)$ follows by the usual differentiation under the integral arguments. Suppose we can prove the proposition in the special case that $f \equiv 0$. Then for $f \in C^3(\mathbb{R}^3)$, the function $v(t,x) = +t\int_{S_1} g(x+t\omega)d\bar{\sigma}_1(\omega)$ solves the wave equation $0 = \left(\partial_t^2 - \Delta\right)v(t,x)$ with $v(0,x) = 0$ and $v_t(0,x) = g(x)$. Differentiating the wave equation in t shows $u = v_t$ also solves the wave equation with $u(0,x) = g(x)$ and $u_t(0,x) = v_{tt}(0,x) = -\Delta_x v(0,x) = 0$. These remarks reduced the problems to showing u in Eq. (44.21) with $f \equiv 0$ solves the wave equation. So let

$$u(t,x) := t\int_{S_1} g(x+t\omega)d\bar{\sigma}_1(\omega). \tag{44.22}$$

We now give two proofs the u solves the wave equation.

Proof 1. Since solving the wave equation is a local statement and $u(t,x)$ only depends on the values of g in $B(x,t)$ we it suffices to consider the case where $g \in C_c^2\left(\mathbb{R}^3\right)$. Taking the Fourier transform of Eq. (44.22) in the x variable shows

$$\begin{aligned} \hat{u}(t,\xi) &= t\int_{S_1} d\bar{\sigma}_1(\omega)\int_{\mathbb{R}^3} g(x+t\omega)e^{-i\xi\cdot x}\mathbf{d}x \\ &= t\int_{S_1} d\bar{\sigma}_1(\omega)\int_{\mathbb{R}^3} g(x)e^{-i\xi\cdot x}e^{it\omega\cdot\xi}\mathbf{d}x = \hat{g}(\xi)t\int_{S_1} e^{it\omega\cdot\xi}d\bar{\sigma}_1(\omega) \\ &= \hat{g}(\xi)t\frac{\sin|tk|}{|tk|} = \hat{g}(\xi)\frac{\sin\left(t\left|\xi\right|\right)}{|\xi|} \end{aligned}$$

wherein we have made use of Example 43.19. This completes the proof since $\hat{u}(t,\xi)$ solves Eq. (44.17) as desired.

Proof 2. Differentiating

$$S(t,x) := \int_{S_1} g(x+t\omega)d\bar{\sigma}_1(\omega)$$

in t gives

$$\begin{aligned} S_t(t,x) &= \frac{1}{4\pi}\int_{S_1} \nabla g(x+t\omega)\cdot\omega d\sigma(\omega) \\ &= \frac{1}{4\pi}\int_{B(0,1)} \nabla_\omega\cdot\nabla g(x+t\omega)dm(\omega) \\ &= \frac{t}{4\pi}\int_{B(0,1)} \Delta g(x+t\omega)dm(\omega) \\ &= \frac{1}{4\pi t^2}\int_{B(0,t)} \Delta g(x+y)dm(y) \\ &= \frac{1}{4\pi t^2}\int_0^t dr\ r^2\int_{|y|=r} \Delta g(x+y)d\sigma(y) \end{aligned}$$

where we have used the divergence theorem, made the change of variables $y = t\omega$ and used the disintegration formula in Eq. (**??**),

$$\int\limits_{\mathbb{R}^d} f(x)dm(x) = \int\limits_{[0,\infty)\times S^{n-1}} f(r\ \omega)\ d\sigma(\omega)r^{n-1}dr = \int_0^\infty dr\int_{|y|=r} f(y)d\sigma(y).$$

Since $u(t,x) = tS(t,x)$ if follows that

$$\begin{aligned} u_{tt}(t,x) &= \frac{\partial}{\partial t}\left[S(t,x) + tS_t(t,x)\right] \\ &= S_t(t,x) + \frac{\partial}{\partial t}\left[\frac{1}{4\pi t}\int_0^t dr\ r^2\int_{|y|=r} \Delta g(x+y)d\sigma(y)\right] \\ &= S_t(t,x) - \frac{1}{4\pi t^2}\int_0^t dr\int_{|y|=r} \Delta g(x+y)d\sigma(y) \\ &\quad + \frac{1}{4\pi t}\int_{|y|=t} \Delta g(x+y)d\sigma(y) \\ &= S_t(t,x) - S_t(t,x) + \frac{t}{4\pi t^2}\int_{|y|=1} \Delta g(x+t\omega)d\sigma(\omega) \\ &= t\Delta u(t,x) \end{aligned}$$

as required. ∎

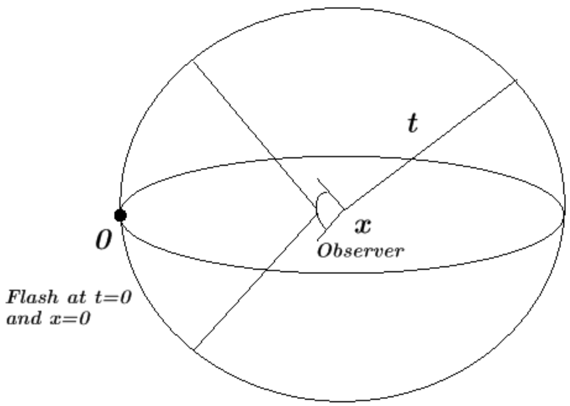


Fig. 44.1. The geometry of the solution to the wave equation in three dimensions. The observer sees a flash at $t = 0$ and $x = 0$ only at time $t = |x|$. The wave progates sharply with speed 1.

The solution in Eq. (44.21) exhibits a basic property of wave equations, namely finite propagation speed. To exhibit the finite propagation speed, suppose that $f = 0$ (for simplicity) and g has compact support near the origin, for example think of $g = \delta_0(x)$. Then $x + tw = 0$ for some w iff $|x| = t$. Hence the "wave front" propagates at unit speed and the wave front is sharp. See Figure 44.1 below.

The solution of the two dimensional wave equation may be found using "Hadamard's method of decent" which we now describe. Suppose now that f and g are functions on $\mathbb{R}^2$ which we may view as functions on $\mathbb{R}^3$ which happen not to depend on the third coordinate. We now go ahead and solve the three dimensional wave equation using Eq. (44.21) and f and g as initial conditions. It is easily seen that the solution $u(t, x, y, z)$ is again independent of z and hence is a solution to the two dimensional wave equation. See figure 44.2 below.

Notice that we still have finite speed of propagation but no longer sharp propagation. The explicit formula for u is given in the next proposition.

Proposition 44.6. *Suppose $f \in C^3(\mathbb{R}^2)$ and $g \in C^2(\mathbb{R}^2)$, then*

$$u(t,x) := \frac{\partial}{\partial t}\left[\frac{t}{2\pi}\iint_{D_1}\frac{f(x+tw)}{\sqrt{1-|w|^2}}dm(w)\right] + \frac{t}{2\pi}\iint_{D_1}\frac{g(x+tw)}{\sqrt{1-|w|^2}}dm(w)$$

is in $C^2\left(\mathbb{R}\times\mathbb{R}^2\right)$ and solves the wave equation in Eq. (44.16).

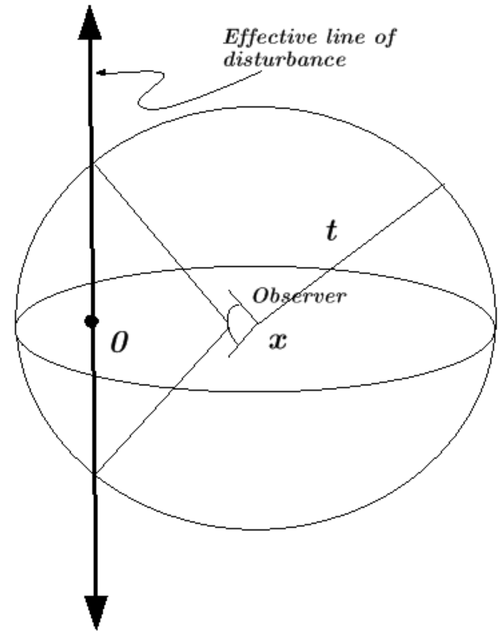


Fig. 44.2. The geometry of the solution to the wave equation in two dimensions. A flash at $0 \in \mathbb{R}^2$ looks like a line of flashes to the fictitious 3 – d observer and hence she sees the effect of the flash for $t \geq |x|$. The wave still propagates with speed 1. However there is no longer sharp propagation of the wave front, similar to water waves.

Proof. As usual it suffices to consider the case where $f \equiv 0$. By symmetry u may be written as

$$u(t,x) = 2t\int_{S_t^+} g(x-y)d\bar{\sigma}_t(y) = 2t\int_{S_t^+} g(x+y)d\bar{\sigma}_t(y)$$

where S_t^+ is the portion of S_t with $z \geq 0$. The surface S_t^+ may be parametrized by $R(u,v) = (u, v, \sqrt{t^2-u^2-v^2})$ with $(u,v) \in D_t := \left\{(u,v) : u^2+v^2 \leq t^2\right\}$. In these coordinates we have

$$\begin{aligned}4\pi t^2 d\bar{\sigma}_t &= \left|\left(-\partial_u\sqrt{t^2-u^2-v^2}, -\partial_v\sqrt{t^2-u^2-v^2}, 1\right)\right| dudv \\ &= \left|\left(\frac{u}{\sqrt{t^2-u^2-v^2}}, \frac{v}{\sqrt{t^2-u^2-v^2}}, 1\right)\right| dudv \\ &= \sqrt{\frac{u^2+v^2}{t^2-u^2-v^2}+1dudv} = \frac{|t|}{\sqrt{t^2-u^2-v^2}}dudv\end{aligned}$$

and therefore,

$$u(t,x) = \frac{2t}{4\pi t^2}\int_{D_t} g(x + (u, v, \sqrt{t^2 - u^2 - v^2}))\frac{|t|}{\sqrt{t^2 - u^2 - v^2}}dudv$$
$$= \frac{1}{2\pi}\operatorname{sgn}(t)\int_{D_t}\frac{g(x + (u, v))}{\sqrt{t^2 - u^2 - v^2}}dudv.$$

This may be written as

$$u(t,x) = \frac{1}{2\pi}\operatorname{sgn}(t)\iint_{D_t}\frac{g(x + w)}{\sqrt{t^2 - |w|^2}}dm(w)$$
$$= \frac{1}{2\pi}\operatorname{sgn}(t)\frac{t^2}{|t|}\iint_{D_1}\frac{g(x + tw)}{\sqrt{1 - |w|^2}}dm(w)$$
$$= \frac{1}{2\pi}t\iint_{D_1}\frac{g(x + tw)}{\sqrt{1 - |w|^2}}dm(w)$$

■

44.5 Elliptic Regularity

The following theorem is a special case of the main theorem (Theorem 44.11) of this section.

Theorem 44.7. *Suppose that $M \subset_o \mathbb{R}^n$, $v \in C^\infty(M)$ and $u \in L^1_{loc}(M)$ satisfies $\Delta u = v$ weakly, then u has a (necessarily unique) version $\tilde{u} \in C^\infty(M)$.*

Proof. We may always assume $n \geq 3$, by embedding the $n = 1$ and $n = 2$ cases in the $n = 3$ cases. For notational simplicity, assume $0 \in M$ and we will show u is smooth near 0. To this end let $\theta \in C_c^\infty(M)$ such that $\theta = 1$ in a neighborhood of 0 and $\alpha \in C_c^\infty(M)$ such that $\operatorname{supp}(\alpha) \subset \{\theta = 1\}$ and $\alpha = 1$ in a neighborhood of 0 as well, see Figure 44.3 Then formally, we have with $\beta := 1 - \alpha$,

$$G * (\theta v) = G * (\theta\Delta u) = G * (\theta\Delta(\alpha u + \beta u))$$
$$= G * (\Delta(\alpha u) + \theta\Delta(\beta u)) = \alpha u + G * (\theta\Delta(\beta u))$$

so that

$$u(x) = G * (\theta v)(x) - G * (\theta\Delta(\beta u))(x)$$

for $x \in \operatorname{supp}(\alpha)$. The last term is formally given by

$$G * (\theta\Delta(\beta u))(x) = \int_{\mathbb{R}^n} G(x - y)\theta(y)\Delta(\beta(y)u(y))dy$$
$$= \int_{\mathbb{R}^n}\beta(y)\Delta_y\left[G(x - y)\theta(y)\right]\cdot u(y)dy$$

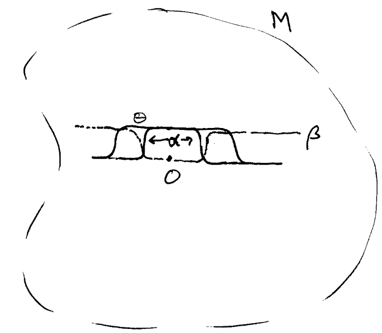


Fig. 44.3. The region M and the cutoff functions, θ and α.

which makes sense for x near 0. Therefore we find

$$u(x) = G * (\theta v)(x) - \int_{\mathbb{R}^n}\beta(y)\Delta_y\left[G(x - y)\theta(y)\right]\cdot u(y)dy.$$

Clearly all of the above manipulations were correct if we know u were C^2 to begin with. So for the general case, let $u_n = u * \delta_n$ with $\{\delta_n\}_{n=1}^\infty$ – the usual sort of δ – sequence approximation. Then $\Delta u_n = v * \delta_n =: v_n$ away from ∂M and

$$u_n(x) = G * (\theta v_n)(x) - \int_{\mathbb{R}^n}\beta(y)\Delta_y\left[G(x - y)\theta(y)\right]\cdot u_n(y)dy. \tag{44.23}$$

Since $u_n \to u$ in $L^1_{loc}(\mathcal{O})$ where $\mathcal{O}$ is a sufficiently small neighborhood of 0, we may pass to the limit in Eq. (44.23) to find $u(x) = \tilde{u}(x)$ for a.e. $x \in \mathcal{O}$ where

$$\tilde{u}(x) := G * (\theta v)(x) - \int_{\mathbb{R}^n}\beta(y)\Delta_y\left[G(x - y)\theta(y)\right]\cdot u(y)dy.$$

This concluded the proof since $\tilde{u}$ is smooth for x near 0. ■

Definition 44.8. *We say $L = p(D_x)$ as defined in Eq. (44.1) is **elliptic** if $p_k(\xi) := \sum_{|\alpha|=k} a_\alpha\xi^\alpha$ is zero iff $\xi = 0$. We will also say the polynomial $p(\xi) := \sum_{|\alpha|\leq k} a_\alpha\xi^\alpha$ is **elliptic** if this condition holds.*

Remark 44.9. If $p(\xi) := \sum_{|\alpha|\leq k} a_\alpha\xi^\alpha$ is an elliptic polynomial, then there exists $A < \infty$ such that $\inf_{|\xi|\geq A}|p(\xi)| > 0$. Since $p_k(\xi)$ is everywhere non-zero for $\xi \in S^{n-1}$ and $S^{n-1} \subset \mathbb{R}^n$ is compact, $\varepsilon := \inf_{|\xi|=1}|p_k(\xi)| > 0$. By homogeneity this implies

$$|p_k(\xi)| \geq \varepsilon|\xi|^k \text{ for all } \xi \in \mathbb{A}^n.$$

Since

$$|p(\xi)| = \left|p_k(\xi) + \sum_{|\alpha|<k} a_\alpha \xi^\alpha\right| \geq |p_k(\xi)| - \left|\sum_{|\alpha|<k} a_\alpha \xi^\alpha\right| \geq \varepsilon |\xi|^k - C\left(1 + |\xi|^{k-1}\right)$$

for some constant $C < \infty$ from which it is easily seen that for A sufficiently large,

$$|p(\xi)| \geq \frac{\varepsilon}{2} |\xi|^k \text{ for all } |\xi| \geq A.$$

For the rest of this section, let $L = p(D_x)$ be an elliptic operator and $M \subset_0 \mathbb{R}^n$. As mentioned at the beginning of this section, the formal solution to $Lu = v$ for $v \in L^2(\mathbb{R}^n)$ is given by

$$u = L^{-1}v = G * v$$

where

$$G(x) := \int_{\mathbb{R}^n} \frac{1}{p(\xi)} e^{ix\cdot\xi} \mathbf{d}\xi.$$

Of course this integral may not be convergent because of the possible zeros of p and the fact $\frac{1}{p(\xi)}$ may not decay fast enough at infinity. We we will introduce a smooth cut off function $\chi(\xi)$ which is 1 on $C_0(A) := \{x \in \mathbb{R}^n : |x| \leq A\}$ and $\operatorname{supp}(\chi) \subset C_0(2A)$ where A is as in Remark 44.9. Then for $M > 0$ let

$$G_M(x) = \int_{\mathbb{R}^n} \frac{(1-\chi(\xi))\,\chi(\xi/M)}{p(\xi)} e^{ix\cdot\xi} \mathbf{d}\xi, \tag{44.24}$$

$$\delta(x) := \chi^\vee(x) = \int_{\mathbb{R}^n} \chi(\xi) e^{ix\cdot\xi} \mathbf{d}\xi, \text{ and } \delta_M(x) = M^n \delta(Mx). \tag{44.25}$$

Notice $\int_{\mathbb{R}^n} \delta(x)dx = \mathcal{F}\delta(0) = \chi(0) = 1$, $\delta \in \mathcal{S}$ since $\chi \in \mathcal{S}$ and

$$LG_M(x) = \int_{\mathbb{R}^n} (1-\chi(\xi))\,\chi(\xi/M) e^{ix\cdot\xi} \mathbf{d}\xi = \int_{\mathbb{R}^n} [\chi(\xi/M) - \chi(\xi)]\, e^{ix\cdot\xi} \mathbf{d}\xi = \delta_M(x) - \delta(x)$$

provided $M > 2$.

Proposition 44.10. *Let p be an elliptic polynomial of degree m. The function G_M defined in Eq. (44.24) satisfies the following properties,*

1. *$G_M \in \mathcal{S}$ for all $M > 0$.*
2. *$LG_M(x) = M^n \delta(Mx) - \delta(x)$.*
3. *There exists $G \in C_c^\infty(\mathbb{R}^n \setminus \{0\})$ such that for all multi-indices α, $\lim_{M\to\infty} \partial^\alpha G_M(x) = \partial^\alpha G(x)$ uniformly on compact subsets in $\mathbb{R}^n \setminus \{0\}$.*

Proof. We have already proved the first two items. For item 3., we notice that

$$\begin{aligned}(-x)^\beta D^\alpha G_M(x) &= \int_{\mathbb{R}^n} \frac{(1-\chi(\xi))\,\chi(\xi/M)\xi^\alpha}{p(\xi)} (-D)_\xi^\beta e^{ix\cdot\xi} \mathbf{d}\xi \\ &= \int_{\mathbb{R}^n} D_\xi^\beta \left[\frac{(1-\chi(\xi))\,\xi^\alpha}{p(\xi)} \chi(\xi/M)\right] e^{ix\cdot\xi} \mathbf{d}\xi \\ &= \int_{\mathbb{R}^n} D_\xi^\beta \frac{(1-\chi(\xi))\,\xi^\alpha}{p(\xi)} \cdot \chi(\xi/M) e^{ix\cdot\xi} \mathbf{d}\xi + R_M(x)\end{aligned}$$

where

$$R_M(x) = \sum_{\gamma<\beta} \binom{\beta}{\gamma} M^{|\gamma|-|\beta|} \int_{\mathbb{R}^n} D_\xi^\gamma \frac{(1-\chi(\xi))\,\xi^\alpha}{p(\xi)} \cdot \left(D^{\beta-\gamma}\chi\right)(\xi/M) e^{ix\cdot\xi} \mathbf{d}\xi.$$

Using

$$\left|D_\xi^\gamma \left[\frac{\xi^\alpha}{p(\xi)} (1-\chi(\xi))\right]\right| \leq C |\xi|^{|\alpha|-m-|\gamma|}$$

and the fact that

$$\begin{aligned}\operatorname{supp}(\left(D^{\beta-\gamma}\chi\right)(\xi/M)) &\subset \{\xi \in \mathbb{R}^n : A \leq |\xi|/M \leq 2A\} \\ &= \{\xi \in \mathbb{R}^n : AM \leq |\xi| \leq 2AM\}\end{aligned}$$

we easily estimate

$$\begin{aligned}|R_M(x)| &\leq C \sum_{\gamma<\beta} \binom{\beta}{\gamma} M^{|\gamma|-|\beta|} \int_{\{\xi\in\mathbb{R}^n : AM\leq|\xi|\leq 2AM\}} |\xi|^{|\alpha|-m-|\gamma|} \mathbf{d}\xi \\ &\leq C \sum_{\gamma<\beta} \binom{\beta}{\gamma} M^{|\gamma|-|\beta|} M^{|\alpha|-m-|\gamma|+n} = CM^{|\alpha|-|\beta|-m+n}.\end{aligned}$$

Therefore, $R_M \to 0$ uniformly in x as $M \to \infty$ provided $|\beta| > |\alpha| - m + n$. It follows easily now that $G_M \to G$ in $C_c^\infty(\mathbb{R}^n \setminus \{0\})$ and furthermore that

$$(-x)^\beta D^\alpha G(x) = \int_{\mathbb{R}^n} D_\xi^\beta \frac{(1-\chi(\xi))\,\xi^\alpha}{p(\xi)} \cdot e^{ix\cdot\xi} \mathbf{d}\xi$$

provided β is sufficiently large. In particular we have shown,

$$D^\alpha G(x) = \frac{1}{|x|^{2k}} \int_{\mathbb{R}^n} (-\Delta_\xi)^k \frac{(1-\chi(\xi))\,\xi^\alpha}{p(\xi)} \cdot e^{ix\cdot\xi} \mathbf{d}\xi$$

provided $m - |\alpha| + 2k > n$, i.e. $k > (n - m + |\alpha|)/2$. We are now ready to use this result to prove elliptic regularity for the constant coefficient case. ■

Theorem 44.11. *Suppose $L = p(D_\xi)$ is an elliptic differential operator on $\mathbb{R}^n$, $M \subset_o \mathbb{R}^n$, $v \in C^\infty(M)$ and $u \in L^1_{loc}(M)$ satisfies $Lu = v$ weakly, then u has a (necessarily unique) version $\tilde{u} \in C^\infty(M)$.*

Proof. For notational simplicity, assume $0 \in M$ and we will show u is smooth near 0. To this end let $\theta \in C_c^\infty(M)$ such that $\theta = 1$ in a neighborhood of 0 and $\alpha \in C_c^\infty(M)$ such that $\operatorname{supp}(\alpha) \subset \{\theta = 1\}$, and $\alpha = 1$ in a neighborhood of 0 as well. Then formally, we have with $\beta := 1 - \alpha$,

$$\begin{aligned} G_M * (\theta v) &= G_M * (\theta L u) = G_M * (\theta L(\alpha u + \beta u)) \\ &= G_M * (L(\alpha u) + \theta L(\beta u)) \\ &= \delta_M * (\alpha u) - \delta * (\alpha u) + G_M * (\theta L(\beta u)) \end{aligned}$$

so that

$$\delta_M * (\alpha u)(x) = G_M * (\theta v)(x) - G_M * (\theta L(\beta u))(x) + \delta * (\alpha u). \tag{44.26}$$

Since

$$\begin{aligned} \mathcal{F}\left[G_M * (\theta v)\right](\xi) &= \hat{G}_M(\xi)\,(\theta v)\hat{}\,(\xi) = \frac{(1-\chi(\xi))\,\chi(\xi/M)}{p(\xi)}\,(\theta v)\hat{}\,(\xi) \\ &\to \frac{(1-\chi(\xi))}{p(\xi)}\,(\theta v)\hat{}\,(\xi) \text{ as } M \to \infty \end{aligned}$$

with the convergence taking place in L^2 (actually in $\mathcal{S}$), it follows that

$$\begin{aligned} G_M * (\theta v) \to \text{“}G * (\theta v)\text{”}(x) &:= \int_{\mathbb{R}^n} \frac{(1-\chi(\xi))}{p(\xi)}\,(\theta v)\hat{}\,(\xi) e^{ix\cdot\xi} \mathbf{d}\xi \\ &= \mathcal{F}^{-1}\left[\frac{(1-\chi(\xi))}{p(\xi)}\,(\theta v)\hat{}\,(\xi)\right](x) \in \mathcal{S}. \end{aligned}$$

So passing the the limit, $M \to \infty$, in Eq. (44.26) we learn for almost every $x \in \mathbb{R}^n$,

$$u(x) = G * (\theta v)(x) - \lim_{M\to\infty} G_M * (\theta L(\beta u))(x) + \delta * (\alpha u)(x)$$

for a.e. $x \in \operatorname{supp}(\alpha)$. Using the support properties of θ and β we see for x near 0 that $(\theta L(\beta u))(y) = 0$ unless $y \in \operatorname{supp}(\theta)$ and $y \notin \{\alpha = 1\}$, i.e. unless y is in an annulus centered at 0. So taking x sufficiently close to 0, we find $x - y$ stays away from 0 as y varies through the above mentioned annulus, and therefore

$$\begin{aligned} G_M * (\theta L(\beta u))(x) &= \int_{\mathbb{R}^n} G_M(x-y)(\theta L(\beta u))(y)\mathbf{d}y \\ &= \int_{\mathbb{R}^n} L_y^* \{\theta(y) G_M(x-y)\} \cdot (\beta u)(y)\mathbf{d}y \\ &\to \int_{\mathbb{R}^n} L_y^* \{\theta(y) G(x-y)\} \cdot (\beta u)(y)\mathbf{d}y \text{ as } M \to \infty. \end{aligned}$$

Therefore we have shown,

$$u(x) = G * (\theta v)(x) - \int_{\mathbb{R}^n} L_y^* \{\theta(y) G(x-y)\} \cdot (\beta u)(y)\mathbf{d}y + \delta * (\alpha u)(x)$$

for almost every x in a neighborhood of 0. (Again it suffices to prove this equation and in particular Eq. (44.26) assuming $u \in C^2(M)$ because of the same convolution argument we have use above.) Since the right side of this equation is the linear combination of smooth functions we have shown u has a smooth version in a neighborhood of 0. ∎

Remarks 44.12 *We could avoid introducing $G_M(x)$ if $\deg(p) > n$, in which case $\frac{(1-\chi(\xi))}{p(\xi)} \in L^1$ and so*

$$G(x) := \int_{\mathbb{R}^n} \frac{(1-\chi(\xi))}{p(\xi)} e^{ix\cdot\xi} \mathbf{d}\xi$$

is already well defined function with $G \in C^\infty(\mathbb{R}^n \setminus \{0\}) \cap BC(\mathbb{R}^n)$. If $\deg(p) < n$, we may consider the operator $L^k = [p(D_x)]^k = p^k(D_x)$ where k is chosen so that $k \cdot \deg(p) > n$. Since $Lu = v$ implies $L^k u = L^{k-1} v$ weakly, we see to prove the hypoellipticity of L it suffices to prove the hypoellipticity of L^k.

Elementary Generalized Functions / Distribution Theory

This chapter has been highly influenced by Friedlander's book [6].

45.1 Distributions on $U \subset_o \mathbb{R}^n$

Let U be an open subset of $\mathbb{R}^n$ and

$$C_c^\infty(U) = \cup_{K \sqsubset\sqsubset U} C^\infty(K) \tag{45.1}$$

denote the set of smooth functions on U with compact support in U.

Definition 45.1. *A sequence $\{\varphi_k\}_{k=1}^\infty \subset \mathcal{D}(U)$ converges to $\varphi \in \mathcal{D}(U)$, iff there is a compact set $K \sqsubset\sqsubset U$ such that $\operatorname{supp}(\varphi_k) \subset K$ for all k and $\varphi_k \to \varphi$ in $C^\infty(K)$.*

Definition 45.2 (Distributions on $U \subset_o \mathbb{R}^n$). *A generalized function T on $U \subset_o \mathbb{R}^n$ is a continuous linear functional on $\mathcal{D}(U)$, i.e. $T : \mathcal{D}(U) \to \mathbb{C}$ is linear and $\lim_{n\to\infty}\langle T, \varphi_k\rangle = 0$ for all $\{\varphi_k\} \subset \mathcal{D}(U)$ such that $\varphi_k \to 0$ in $\mathcal{D}(U)$. We denote the space of generalized functions by $\mathcal{D}'(U)$.*

Proposition 45.3. *Let $T : \mathcal{D}(U) \to \mathbb{C}$ be a linear functional. Then $T \in \mathcal{D}'(U)$ iff for all $K \sqsubset\sqsubset U$, there exist $n \in \mathbb{N}$ and $C < \infty$ such that*

$$|T(\varphi)| \le C p_n(\varphi) \text{ for all } \varphi \in C^\infty(K). \tag{45.2}$$

Proof. Suppose that $\{\varphi_k\} \subset \mathcal{D}(U)$ such that $\varphi_k \to 0$ in $\mathcal{D}(U)$. Let K be a compact set such that $\operatorname{supp}(\varphi_k) \subset K$ for all k. Since $\lim_{k\to\infty} p_n(\varphi_k) = 0$, it follows that if Eq. (45.2) holds that $\lim_{n\to\infty}\langle T, \varphi_k\rangle = 0$. Conversely, suppose that there is a compact set $K \sqsubset\sqsubset U$ such that for no choice of $n \in \mathbb{N}$ and $C < \infty$, Eq. (45.2) holds. Then we may choose non-zero $\varphi_n \in C^\infty(K)$ such that

$$|T(\varphi_n)| \ge n p_n(\varphi_n) \text{ for all } n.$$

Let $\psi_n = \frac{1}{n p_n(\varphi_n)}\varphi_n \in C^\infty(K)$, then $p_n(\psi_n) = 1/n \to 0$ as $n \to \infty$ which shows that $\psi_n \to 0$ in $\mathcal{D}(U)$. On the other hence $|T(\psi_n)| \ge 1$ so that $\lim_{n\to\infty}\langle T, \psi_n\rangle \ne 0$. **Alternate Proof:**The definition of T being continuous is equivalent to $T|_{C^\infty(K)}$ being sequentially continuous for all $K \sqsubset\sqsubset U$. Since $C^\infty(K)$ is a metric space, sequential continuity and continuity are the same thing. Hence T is continuous iff $T|_{C^\infty(K)}$ is continuous for all $K \sqsubset\sqsubset U$. Now $T|_{C^\infty(K)}$ is continuous iff a bound like Eq. (45.2) holds. ∎

Definition 45.4. *Let Y be a topological space and $T_y \in \mathcal{D}'(U)$ for all $y \in Y$. We say that $T_y \to T \in \mathcal{D}'(U)$ as $y \to y_0$ iff*

$$\lim_{y\to y_0}\langle T_y, \varphi\rangle = \langle T, \varphi\rangle \text{ for all } \varphi \in \mathcal{D}(U).$$

45.2 Examples of distributions and related computations

Example 45.5. Let μ be a positive Radon measure on U and $f \in L^1_{loc}(U)$. Define $T \in \mathcal{D}'(U)$ by $\langle T_f, \varphi\rangle = \int_U \varphi f d\mu$ for all $\varphi \in \mathcal{D}(U)$. Notice that if $\varphi \in C^\infty(K)$ then

$$|\langle T_f, \varphi\rangle| \le \int_U |\varphi f|\, d\mu = \int_K |\varphi f|\, d\mu \le C_K \|\varphi\|_\infty$$

where $C_K := \int_K |f|\, d\mu < \infty$. Hence $T_f \in \mathcal{D}'(U)$. Furthermore, the map

$$f \in L^1_{loc}(U) \to T_f \in \mathcal{D}'(U)$$

is injective. Indeed, $T_f = 0$ is equivalent to

$$\int_U \varphi f d\mu = 0 \text{ for all } \varphi \in \mathcal{D}(U). \tag{45.3}$$

for all $\varphi \in C^\infty(K)$. By the dominated convergence theorem and the usual convolution argument, this is equivalent to

$$\int_U \varphi f d\mu = 0 \text{ for all } \varphi \in C_c(U). \tag{45.4}$$

Now fix a compact set $K \sqsubset\sqsubset U$ and $\varphi_n \in C_c(U)$ such that $\varphi_n \to \overline{\operatorname{sgn}(f)}1_K$ in $L^1(\mu)$. By replacing φ_n by $\chi(\varphi_n)$ if necessary, where

$$\chi(z) = \begin{cases} z & \text{if } |z| \le 1 \\ \frac{z}{|z|} & \text{if } |z| \ge 1, \end{cases}$$

we may assume that $|\varphi_n| \le 1$. By passing to a further subsequence, we may assume that $\varphi_n \to \overline{\operatorname{sgn}(f)}1_K$ a.e.. Thus we have

$$0 = \lim_{n\to\infty} \int_U \varphi_n f d\mu = \int_U \overline{\operatorname{sgn}(f)} 1_K f d\mu = \int_K |f| \, d\mu.$$

This shows that $|f(x)| = 0$ for μ -a.e. $x \in K$. Since K is arbitrary and U is the countable union of such compact sets K, it follows that $f(x) = 0$ for μ -a.e. $x \in U$.

The injectivity may also be proved slightly more directly as follows. As before, it suffices to prove Eq. (45.4) implies that $f(x) = 0$ for μ − a.e. x. We may further assume that f is real by considering real and imaginary parts separately. Let $K \sqsubset\sqsubset U$ and $\varepsilon > 0$ be given. Set $A = \{f > 0\} \cap K$, then $\mu(A) < \infty$ and hence since all σ finite measure on U are Radon, there exists $F \subset A \subset V$ with F compact and $V \subset_o U$ such that $\mu(V \setminus F) < \delta$. By Uryshon's lemma, there exists $\varphi \in C_c(V)$ such that $0 \le \varphi \le 1$ and $\varphi = 1$ on F. Then by Eq. (45.4)

$$0 = \int_U \varphi f d\mu = \int_F \varphi f d\mu + \int_{V\setminus F} \varphi f d\mu = \int_F \varphi f d\mu + \int_{V\setminus F} \varphi f d\mu$$

so that

$$\int_F f d\mu = \left| \int_{V\setminus F} \varphi f d\mu \right| \le \int_{V\setminus F} |f| \, d\mu < \varepsilon$$

provided that δ is chosen sufficiently small by the $\varepsilon - \delta$ definition of absolute continuity. Similarly, it follows that

$$0 \le \int_A f d\mu \le \int_F f d\mu + \varepsilon \le 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\int_A f d\mu = 0$. Since K was arbitrary, we learn that

$$\int_{\{f>0\}} f d\mu = 0$$

which shows that $f \le 0$ μ − a.e. Similarly, one shows that $f \ge 0$ μ − a.e. and hence $f = 0$ μ − a.e.

Example 45.6. Let us now assume that $\mu = m$ and write $\langle T_f, \varphi\rangle = \int_U \varphi f dm$. For the moment let us also assume that $U = \mathbb{R}$. Then we have

1. $\lim_{M\to\infty} T_{\sin Mx} = 0$
2. $\lim_{M\to\infty} T_{M^{-1}\sin Mx} = \pi\delta_0$ where δ_0 is the point measure at 0.
3. If $f \in L^1(\mathbb{R}^n, dm)$ with $\int_{\mathbb{R}^n} f dm = 1$ and $f_\varepsilon(x) = \varepsilon^{-n} f(x/\varepsilon)$, then $\lim_{\varepsilon\downarrow 0} T_{f_\varepsilon} = \delta_0$. As a special case, consider $\lim_{\varepsilon\downarrow 0} \frac{\varepsilon}{\pi(x^2+\varepsilon^2)} = \delta_0$.

Definition 45.7 (Multiplication by smooth functions). *Suppose that $g \in C^\infty(U)$ and $T \in \mathcal{D}'(U)$ then we define $gT \in \mathcal{D}'(U)$ by*

$$\langle gT, \varphi\rangle = \langle T, g\varphi\rangle \text{ for all } \varphi \in \mathcal{D}(U).$$

It is easily checked that gT is continuous.

Definition 45.8 (Differentiation). *For $T \in \mathcal{D}'(U)$ and $i \in \{1, 2, \dots, n\}$ let $\partial_i T \in \mathcal{D}'(U)$ be the distribution defined by*

$$\langle \partial_i T, \varphi\rangle = -\langle T, \partial_i \varphi\rangle \text{ for all } \varphi \in \mathcal{D}(U).$$

Again it is easy to check that $\partial_i T$ is a distribution.

More generally if $L = \sum_{|\alpha|\le m} a_\alpha \partial^\alpha$ with $a_\alpha \in C^\infty(U)$ for all α, then LT is the distribution defined by

$$\langle LT, \varphi\rangle = \langle T, \sum_{|\alpha|\le m} (-1)^{|\alpha|} \partial^\alpha (a_\alpha \varphi)\rangle \text{ for all } \varphi \in \mathcal{D}(U).$$

Hence we can talk about distributional solutions to differential equations of the form $LT = S$.

Example 45.9. Suppose that $f \in L^1_{loc}$ and $g \in C^\infty(U)$, then $gT_f = T_{gf}$. If further $f \in C^1(U)$, then $\partial_i T_f = T_{\partial_i f}$. If $f \in C^m(U)$, then $LT_f = T_{Lf}$.

Example 45.10. Suppose that $a \in U$, then

$$\langle \partial_i \delta_a, \varphi\rangle = -\partial_i \varphi(a)$$

and more generally we have

$$\langle L\delta_a, \varphi\rangle = \sum_{|\alpha|\le m} (-1)^{|\alpha|} \partial^\alpha (a_\alpha \varphi)(a).$$

Example 45.11. Consider the distribution $T := T_{|x|}$ for $x \in \mathbb{R}$, i.e. take $U = \mathbb{R}$. Then

$$\frac{d}{dx} T = T_{\operatorname{sgn}(x)} \text{ and } \frac{d^2}{d^2x} T = 2\delta_0.$$

More generally, suppose that f is piecewise C^1, the

$$\frac{d}{dx} T_f = T_{f'} + \sum (f(x+) - f(x-)) \, \delta_x.$$

Example 45.12. Consider $T = T_{\ln|x|}$ on $\mathcal{D}(\mathbb{R})$. Then

$$\begin{aligned}\langle T', \varphi\rangle &= -\int_{\mathbb{R}} \ln|x|\,\varphi'(x)dx = -\lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \ln|x|\,\varphi'(x)dx \\ &= -\lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \ln|x|\,\varphi'(x)dx \\ &= \lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \frac{1}{x}\varphi(x)dx - \lim_{\varepsilon\downarrow 0}\left[\ln\varepsilon(\varphi(\varepsilon)-\varphi(-\varepsilon))\right] \\ &= \lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \frac{1}{x}\varphi(x)dx.\end{aligned}$$

We will write $T' = PV\frac{1}{x}$ in the future. Here is another formula for T',

$$\begin{aligned}\langle T', \varphi\rangle &= \lim_{\varepsilon\downarrow 0}\int_{1\geq|x|>\varepsilon} \frac{1}{x}\varphi(x)dx + \int_{|x|>1}\frac{1}{x}\varphi(x)dx \\ &= \lim_{\varepsilon\downarrow 0}\int_{1\geq|x|>\varepsilon} \frac{1}{x}[\varphi(x)-\varphi(0)]dx + \int_{|x|>1}\frac{1}{x}\varphi(x)dx \\ &= \int_{1\geq|x|} \frac{1}{x}[\varphi(x)-\varphi(0)]dx + \int_{|x|>1}\frac{1}{x}\varphi(x)dx.\end{aligned}$$

Please notice in the last example that $\frac{1}{x}\notin L^1_{loc}(\mathbb{R})$ so that $T_{1/x}$ is not well defined. This is an example of the so called division problem of distributions. Here is another possible interpretation of $\frac{1}{x}$ as a distribution.

Example 45.13. Here we try to define $1/x$ as $\lim_{y\downarrow 0}\frac{1}{x\pm iy}$, that is we want to define a distribution $T_\pm$ by

$$\langle T_\pm, \varphi\rangle := \lim_{y\downarrow 0}\int \frac{1}{x\pm iy}\varphi(x)dx.$$

Let us compute T_+ explicitly,

$$\begin{aligned}&\lim_{y\downarrow 0}\int_{\mathbb{R}} \frac{1}{x+iy}\varphi(x)dx \\ &\quad= \lim_{y\downarrow 0}\int_{|x|\leq 1}\frac{1}{x+iy}\varphi(x)dx + \lim_{y\downarrow 0}\int_{|x|>1}\frac{1}{x+iy}\varphi(x)dx \\ &\quad= \lim_{y\downarrow 0}\int_{|x|\leq 1}\frac{1}{x+iy}\left[\varphi(x)-\varphi(0)\right]dx + \varphi(0)\lim_{y\downarrow 0}\int_{|x|\leq 1}\frac{1}{x+iy}dx \\ &\quad+ \int_{|x|>1}\frac{1}{x}\varphi(x)dx \\ &\quad= PV\int_{\mathbb{R}}\frac{1}{x}\varphi(x)dx + \varphi(0)\lim_{y\downarrow 0}\int_{|x|\leq 1}\frac{1}{x+iy}dx.\end{aligned}$$

Now by deforming the contour we have

$$\int_{|x|\leq 1}\frac{1}{x+iy}dx = \int_{\varepsilon<|x|\leq 1}\frac{1}{x+iy}dx + \int_{C_\varepsilon}\frac{1}{z+iy}dz$$

where $C_\varepsilon : z = \varepsilon e^{i\theta}$ with $\theta : \pi \to 0$. Therefore,

$$\begin{aligned}\lim_{y\downarrow 0}\int_{|x|\leq 1}\frac{1}{x+iy}dx &= \lim_{y\downarrow 0}\int_{\varepsilon<|x|\leq 1}\frac{1}{x+iy}dx + \lim_{y\downarrow 0}\int_{C_\varepsilon}\frac{1}{z+iy}dz \\ &= \int_{\varepsilon<|x|\leq 1}\frac{1}{x}dx + \int_{C_\varepsilon}\frac{1}{z}dz = 0 - \pi.\end{aligned}$$

Hence we have shown that $T_+ = PV\frac{1}{x} - i\pi\delta_0$. Similarly, one shows that $T_- = PV\frac{1}{x} + i\pi\delta_0$. Notice that it follows from these computations that $T_- - T_+ = i2\pi\delta_0$. Notice that

$$\frac{1}{x-iy} - \frac{1}{x+iy} = \frac{2iy}{x^2+y^2}$$

and hence we conclude that $\lim_{y\downarrow 0}\frac{y}{x^2+y^2} = \pi\delta_0$ – a result that we saw in Example 45.6, item 3.

Example 45.14. Suppose that μ is a complex measure on $\mathbb{R}$ and $F(x) = \mu((-\infty,x])$, then $T'_F = \mu$. Moreover, if $f\in L^1_{loc}(\mathbb{R})$ and $T'_f = \mu$, then $f = F + C$ a.e. for some constant C.

Proof. Let $\varphi\in\mathcal{D} := \mathcal{D}(\mathbb{R})$, then

$$\begin{aligned}\langle T'_F, \varphi\rangle &= -\langle T_F, \varphi'\rangle = -\int_{\mathbb{R}} F(x)\varphi'(x)dx = -\int_{\mathbb{R}}dx\int_{\mathbb{R}}d\mu(y)\varphi'(x)1_{y\leq x} \\ &= -\int_{\mathbb{R}}d\mu(y)\int_{\mathbb{R}}dx\varphi'(x)1_{y\leq x} = \int_{\mathbb{R}}d\mu(y)\varphi(y) = \langle\mu,\varphi\rangle\end{aligned}$$

by Fubini's theorem and the fundamental theorem of calculus. If $T'_f = \mu$, then $T'_{f-F} = 0$ and the result follows from Corollary 45.16 below. ■

Lemma 45.15. *Suppose that* $T\in\mathcal{D}'(\mathbb{R}^n)$ *is a distribution such that* $\partial_i T = 0$ *for some* i*, then there exists a distribution* $S\in\mathcal{D}'(\mathbb{R}^{n-1})$ *such that* $\langle T,\varphi\rangle = \langle S,\bar{\varphi}_i\rangle$ *for all* $\varphi\in\mathcal{D}(\mathbb{R}^n)$ *where*

$$\bar{\varphi}_i = \int_{\mathbb{R}}\tau_{te_i}\varphi dt \in \mathcal{D}(\mathbb{R}^{n-1}).$$

Proof. To simplify notation, assume that $i = n$ and write $x\in\mathbb{R}^n$ as $x = (y,z)$ with $y\in\mathbb{R}^{n-1}$ and $z\in\mathbb{R}$. Let $\theta\in C_c^\infty(\mathbb{R})$ such that $\int_{\mathbb{R}}\theta(z)dz = 1$ and for $\psi\in\mathcal{D}(\mathbb{R}^{n-1})$, let $\psi\otimes\theta(x) = \psi(y)\theta(z)$. The mapping

$$\psi \in \mathcal{D}(\mathbb{R}^{n-1}) \in \psi \otimes \theta \in \mathcal{D}(\mathbb{R}^n)$$

is easily seen to be sequentially continuous and therefore $\langle S, \psi\rangle := \langle T, \psi \otimes \theta\rangle$ defined a distribution in $\mathcal{D}'(\mathbb{R}^n)$. Now suppose that $\varphi \in \mathcal{D}(\mathbb{R}^n)$. If $\varphi = \partial_n f$ for some $f \in \mathcal{D}(\mathbb{R}^n)$ we would have to have $\int \varphi(y,z)dz = 0$. This is not generally true, however the function $\varphi - \bar{\varphi} \otimes \theta$ does have this property. Define

$$f(y,z) := \int_{-\infty}^{z} \left[\varphi(y,z') - \bar{\varphi}(y)\theta(z')\right] dz',$$

then $f \in \mathcal{D}(\mathbb{R}^n)$ and $\partial_n f = \varphi - \bar{\varphi} \otimes \theta$. Therefore,

$$0 = -\langle \partial_n T, f\rangle = \langle T, \partial_n f\rangle = \langle T, \varphi\rangle - \langle T, \bar{\varphi} \otimes \theta\rangle = \langle T, \varphi\rangle - \langle S, \bar{\varphi}\rangle.$$

■

Corollary 45.16. *Suppose that $T \in \mathcal{D}'(\mathbb{R}^n)$ is a distribution such that there exists $m \geq 0$ such that*

$$\partial^\alpha T = 0 \text{ for all } |\alpha| = m,$$

then $T = T_p$ where $p(x)$ is a polynomial on $\mathbb{R}^n$ of degree less than or equal to $m-1$, where by convention if $\deg(p) = -1$ then $p := 0$.

Proof. The proof will be by induction on n and m. The corollary is trivially true when $m = 0$ and n is arbitrary. Let $n = 1$ and assume the corollary holds for $m = k-1$ with $k \geq 1$. Let $T \in \mathcal{D}'(\mathbb{R})$ such that $0 = \partial^k T = \partial^{k-1}\partial T$. By the induction hypothesis, there exists a polynomial, q, of degree $k-2$ such that $T' = T_q$. Let $p(x) = \int_0^x q(z)dz$, then p is a polynomial of degree at most $k-1$ such that $p' = q$ and hence $T_p' = T_q = T'$. So $(T - T_p)' = 0$ and hence by Lemma 45.15, $T - T_p = T_C$ where $C = \langle T - T_p, \theta\rangle$ and θ is as in the proof of Lemma 45.15. This proves the he result for $n = 1$. For the general induction, suppose there exists $(m,n) \in \mathbb{N}^2$ with $m \geq 0$ and $n \geq 1$ such that assertion in the corollary holds for pairs (m', n') such that either $n' < n$ of $n' = n$ and $m' \leq m$. Suppose that $T \in \mathcal{D}'(\mathbb{R}^n)$ is a distribution such that

$$\partial^\alpha T = 0 \text{ for all } |\alpha| = m+1.$$

In particular this implies that $\partial^\alpha \partial_n T = 0$ for all $|\alpha| = m-1$ and hence by induction $\partial_n T = T_{q_n}$ where q_n is a polynomial of degree at most $m-1$ on $\mathbb{R}^n$. Let $p_n(x) = \int_0^z q_n(y,z')dz'$ a polynomial of degree at most m on $\mathbb{R}^n$. The polynomial p_n satisfies, 1) $\partial^\alpha p_n = 0$ if $|\alpha| = m$ and $\alpha_n = 0$ and 2) $\partial_n p_n = q_n$. Hence $\partial_n(T - T_{p_n}) = 0$ and so by Lemma 45.15,

$$\langle T - T_{p_n}, \varphi\rangle = \langle S, \bar{\varphi}_n\rangle$$

for some distribution $S \in \mathcal{D}'(\mathbb{R}^{n-1})$. If α is a multi-index such that $\alpha_n = 0$ and $|\alpha| = m$, then

$$\begin{aligned} 0 &= \langle \partial^\alpha T - \partial^\alpha T_{p_n}, \varphi\rangle = \langle T - T_{p_n}, \partial^\alpha \varphi\rangle = \langle S, \overline{(\partial^\alpha\varphi)}_n\rangle \\ &= \langle S, \partial^\alpha \bar{\varphi}_n\rangle = (-1)^{|\alpha|}\langle \partial^\alpha S, \bar{\varphi}_n\rangle. \end{aligned}$$

and in particular by taking $\varphi = \psi \otimes \theta$, we learn that $\langle \partial^\alpha S, \psi\rangle = 0$ for all $\psi \in \mathcal{D}(\mathbb{R}^{n-1})$. Thus by the induction hypothesis, $S = T_r$ for some polynomial (r) of degree at most m on $\mathbb{R}^{n-1}$. Letting $p(y,z) = p_n(y,z) + r(y)$ – a polynomial of degree at most m on $\mathbb{R}^n$, it is easily checked that $T = T_p$. ■

Example 45.17. Consider the wave equation

$$(\partial_t - \partial_x)(\partial_t + \partial_x)u(t,x) = \left(\partial_t^2 - \partial_x^2\right)u(t,x) = 0.$$

From this equation one learns that $u(t,x) = f(x+t) + g(x-t)$ solves the wave equation for $f, g \in C^2$. Suppose that f is a bounded Borel measurable function on $\mathbb{R}$ and consider the function $f(x+t)$ as a distribution on $\mathbb{R}$. We compute

$$\begin{aligned} \langle (\partial_t - \partial_x) f(x+t), \varphi(x,t)\rangle &= \int_{\mathbb{R}^2} f(x+t)\left(\partial_x - \partial_t\right)\varphi(x,t)dxdt \\ &= \int_{\mathbb{R}^2} f(x)\left[\left(\partial_x - \partial_t\right)\varphi\right](x-t,t)dxdt \\ &= -\int_{\mathbb{R}^2} f(x)\frac{d}{dt}\left[\varphi(x-t,t)\right]dxdt \\ &= -\int_{\mathbb{R}} f(x)\left[\varphi(x-t,t)\right]|_{t=-\infty}^{t=\infty}dx = 0. \end{aligned}$$

This shows that $(\partial_t - \partial_x) f(x+t) = 0$ in the distributional sense. Similarly, $(\partial_t + \partial_x) g(x-t) = 0$ in the distributional sense. Hence $u(t,x) = f(x+t) + g(x-t)$ solves the wave equation in the distributional sense whenever f and g are bounded Borel measurable functions on $\mathbb{R}$.

Example 45.18. Consider $f(x) = \ln|x|$ for $x \in \mathbb{R}^2$ and let $T = T_f$. Then, pointwise we have

$$\nabla \ln|x| = \frac{x}{|x|^2} \text{ and } \Delta \ln|x| = \frac{2}{|x|^2} - 2x \cdot \frac{x}{|x|^4} = 0.$$

Hence $\Delta f(x) = 0$ for all $x \in \mathbb{R}^2$ except at $x = 0$ where it is not defined. Does this imply that $\Delta T = 0$? **No**, in fact $\Delta T = 2\pi\delta$ as we shall now prove. By definition of ΔT and the dominated convergence theorem,

$$\langle \Delta T, \varphi\rangle = \langle T, \Delta\varphi\rangle = \int_{\mathbb{R}^2} \ln|x|\,\Delta\varphi(x)dx = \lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \ln|x|\,\Delta\varphi(x)dx.$$

Using the divergence theorem,

$$\begin{aligned}
\int_{|x|>\varepsilon} \ln|x|\,\Delta\varphi(x)dx
&= -\int_{|x|>\varepsilon} \nabla \ln|x| \cdot \nabla\varphi(x)dx + \int_{\partial\{|x|>\varepsilon\}} \ln|x|\,\nabla\varphi(x)\cdot n(x)dS(x) \\
&= \int_{|x|>\varepsilon} \Delta \ln|x|\,\varphi(x)dx - \int_{\partial\{|x|>\varepsilon\}} \nabla \ln|x| \cdot n(x)\varphi(x)dS(x) \\
&\quad + \int_{\partial\{|x|>\varepsilon\}} \ln|x| \left(\nabla\varphi(x)\cdot n(x)\right) dS(x) \\
&= \int_{\partial\{|x|>\varepsilon\}} \ln|x| \left(\nabla\varphi(x)\cdot n(x)\right) dS(x) \\
&\qquad - \int_{\partial\{|x|>\varepsilon\}} \nabla \ln|x| \cdot n(x)\varphi(x)dS(x),
\end{aligned}$$

where $n(x)$ is the outward pointing normal, i.e. $n(x) = -\hat{x} := x/|x|$. Now

$$\left| \int_{\partial\{|x|>\varepsilon\}} \ln|x| \left(\nabla\varphi(x)\cdot n(x)\right) dS(x) \right| \le C \left(\ln \varepsilon^{-1}\right) 2\pi\varepsilon \to 0 \text{ as } \varepsilon \downarrow 0$$

where C is a bound on $(\nabla\varphi(x)\cdot n(x))$. While

$$\begin{aligned}
\int_{\partial\{|x|>\varepsilon\}} \nabla \ln|x| \cdot n(x)\varphi(x)dS(x) &= \int_{\partial\{|x|>\varepsilon\}} \frac{\hat{x}}{|x|} \cdot (-\hat{x})\varphi(x)dS(x) \\
&= -\frac{1}{\varepsilon} \int_{\partial\{|x|>\varepsilon\}} \varphi(x)dS(x) \\
&\to -2\pi\varphi(0) \text{ as } \varepsilon \downarrow 0.
\end{aligned}$$

Combining these results shows

$$\langle \Delta T, \varphi \rangle = 2\pi\varphi(0).$$

Exercise 45.1. Carry out a similar computation to that in Example 45.18 to show

$$\Delta T_{1/|x|} = -4\pi\delta$$

where now $x \in \mathbb{R}^3$.

Example 45.19. Let $z = x + iy$, and $\bar{\partial} = \frac{1}{2}(\partial_x + i\partial_y)$. Let $T = T_{1/z}$, then

$$\bar{\partial} T = \pi\delta_0 \text{ or imprecisely } \bar{\partial}\frac{1}{z} = \pi\delta(z).$$

Proof. Pointwise we have $\bar{\partial}\frac{1}{z} = 0$ so we shall work as above. We then have

$$\begin{aligned}
\langle \bar{\partial} T, \varphi \rangle &= -\langle T, \bar{\partial}\varphi \rangle = -\int_{\mathbb{R}^2} \frac{1}{z} \bar{\partial}\varphi(z)dm(z) \\
&= -\lim_{\varepsilon\downarrow 0} \int_{|z|>\varepsilon} \frac{1}{z} \bar{\partial}\varphi(z)dm(z) \\
&= \lim_{\varepsilon\downarrow 0} \int_{|z|>\varepsilon} \bar{\partial}\frac{1}{z} \varphi(z)dm(z) \\
&\qquad - \lim_{\varepsilon\downarrow 0} \int_{\partial\{|z|>\varepsilon\}} \frac{1}{z}\varphi(z)\frac{1}{2}\left(n_1(z) + i n_2(z)\right) d\sigma(z) \\
&= 0 - \lim_{\varepsilon\downarrow 0} \int_{\partial\{|z|>\varepsilon\}} \frac{1}{z}\varphi(z)\frac{1}{2}\left(\frac{-z}{|z|}\right) d\sigma(z) \\
&= \frac{1}{2}\lim_{\varepsilon\downarrow 0} \int_{\partial\{|z|>\varepsilon\}} \frac{1}{|z|}\varphi(z)d\sigma(z) \\
&= \pi \lim_{\varepsilon\downarrow 0} \frac{1}{2\pi\varepsilon} \int_{\partial\{|z|>\varepsilon\}} \varphi(z)d\sigma(z) = \pi\varphi(0).
\end{aligned}$$

■

45.3 Other classes of test functions

(For what follows, see Exercises 35.34 and 35.35 of Chapter **??**.

Notation 45.20 *Suppose that X is a vector space and $\{p_n\}_{n=0}^\infty$ is a family of semi-norms on X such that $p_n \le p_{n+1}$ for all n and with the property that $p_n(x) = 0$ for all n implies that $x = 0$. (We allow for $p_n = p_0$ for all n in which case X is a normed vector space.) Let τ be the smallest topology on X such that $p_n(x - \cdot) : X \to [0, \infty)$ is continuous for all $n \in \mathbb{N}$ and $x \in X$. For $n \in \mathbb{N}$, $x \in X$ and $\varepsilon > 0$ let $B_n(x, \varepsilon) := \{y \in X : p_n(x - y) < \varepsilon\}$.*

Proposition 45.21. *The balls $\mathcal{B} := \{B_n(x, \varepsilon) : n \in \mathbb{N},\ x \in X \text{ and } \varepsilon > 0\}$ for a basis for the topology τ. This topology is the same as the topology induced by the metric d on X defined by*

$$d(x, y) = \sum_{n=0}^\infty 2^{-n} \frac{p_n(x - y)}{1 + p_n(x - y)}.$$

Moreover, a sequence $\{x_k\} \subset X$ is convergent to $x \in X$ iff $\lim_{k\to\infty} d(x, x_k) = 0$ iff $\lim_{n\to\infty} p_n(x, x_k) = 0$ for all $n \in \mathbb{N}$ and $\{x_k\} \subset X$ is Cauchy in X iff $\lim_{k,l\to\infty} d(x_l, x_k) = 0$ iff $\lim_{k,l\to\infty} p_n(x_l, x_k) = 0$ for all $n \in \mathbb{N}$.

Proof. Suppose that $z \in B_n(x,\varepsilon) \cap B_m(y,\delta)$ and assume with out loss of generality that $m \geq n$. Then if $p_m(w-z) < \alpha$, we have

$$p_m(w-y) \leq p_m(w-z) + p_m(z-y) < \alpha + p_m(z-y) < \delta$$

provided that $\alpha \in (0, \delta - p_m(z-y))$ and similarly

$$p_n(w-x) \leq p_m(w-x) \leq p_m(w-z) + p_m(z-x) < \alpha + p_m(z-x) < \varepsilon$$

provided that $\alpha \in (0, \varepsilon - p_m(z-x))$. So choosing

$$\delta = \frac{1}{2} \min\left(\delta - p_m(z-y), \varepsilon - p_m(z-x)\right),$$

we have shown that $B_m(z,\alpha) \subset B_n(x,\varepsilon) \cap B_m(y,\delta)$. This shows that $\mathcal{B}$ forms a basis for a topology. In detail, $V \subset_o X$ iff for all $x \in V$ there exists $n \in \mathbb{N}$ and $\varepsilon > 0$ such that $B_n(x,\varepsilon) := \{y \in X : p_n(x-y) < \varepsilon\} \subset V$. Let $\tau(\mathcal{B})$ be the topology generated by $\mathcal{B}$. Since$|p_n(x-y) - p_n(x-z)| \leq p_n(y-z)$, we see that $p_n(x-\cdot)$ is continuous on relative to $\tau(\mathcal{B})$ for each $x \in X$ and $n \in \mathbb{N}$. This shows that $\tau \subset \tau(\mathcal{B})$. On the other hand, since $p_n(x-\cdot)$ is $\tau-$ continuous, it follows that $B_n(x,\varepsilon) = \{y \in X : p_n(x-y) < \varepsilon\} \in \tau$ for all $x \in X$, $\varepsilon > 0$ and $n \in \mathbb{N}$. This shows that $\mathcal{B} \subset \tau$ and therefore that $\tau(\mathcal{B}) \subset \tau$. Thus $\tau = \tau(\mathcal{B})$. Given $x \in X$ and $\varepsilon > 0$, let $B_d(x,\varepsilon) = \{y \in X : d(x,y) < \varepsilon\}$ be a $d-$ ball. Choose N large so that $\sum_{n=N+1}^{\infty} 2^{-n} < \varepsilon/2$. Then $y \in B_N(x,\varepsilon/4)$ we have

$$d(x,y) = p_N(x-y) \sum_{n=0}^{N} 2^{-n} + \varepsilon/2 < 2\frac{\varepsilon}{4} + \varepsilon/2 < \varepsilon$$

which shows that $B_N(x,\varepsilon/4) \subset B_d(x,\varepsilon)$. Conversely, if $d(x,y) < \varepsilon$, then

$$2^{-n} \frac{p_n(x-y)}{1 + p_n(x-y)} < \varepsilon$$

which implies that

$$p_n(x-y) < \frac{2^{-n}\varepsilon}{1 - 2^{-n}\varepsilon} =: \delta$$

when $2^{-n}\varepsilon < 1$ which shows that $B_n(x,\delta)$ contains $B_d(x,\varepsilon)$ with ε and δ as above. This shows that τ and the topology generated by d are the same. The moreover statements are now easily proved and are left to the reader. ■

Exercise 45.2. Keeping the same notation as Proposition 45.21 and further assume that $\{p'_n\}_{n\in\mathbb{N}}$ is another family of semi-norms as in Notation 45.20. Then the topology τ' determined by $\{p'_n\}_{n\in\mathbb{N}}$ is weaker then the topology τ determined by $\{p_n\}_{n\in\mathbb{N}}$ (i.e. $\tau' \subset \tau$) iff for every $n \in \mathbb{N}$ there is an $m \in \mathbb{N}$ and $C < \infty$ such that $p'_n \leq C p_m$.

Lemma 45.22. *Suppose that X and Y are vector spaces equipped with sequences of norms $\{p_n\}$ and $\{q_n\}$ as in Notation 45.20. Then a linear map $T : X \to Y$ is continuous if for all $n \in \mathbb{N}$ there exists $C_n < \infty$ and $m_n \in \mathbb{N}$ such that $q_n(Tx) \leq C_n p_{m_n}(x)$ for all $x \in X$. In particular, $f \in X^*$ iff $|f(x)| \leq C p_m(x)$ for some $C < \infty$ and $m \in \mathbb{N}$. (We may also characterize continuity by sequential convergence since both X and Y are metric spaces.)*

Proof. Suppose that T is continuous, then $\{x : q_n(Tx) < 1\}$ is an open neighborhood of 0 in X. Therefore, there exists $m \in \mathbb{N}$ and $\varepsilon > 0$ such that $B_m(0,\varepsilon) \subset \{x : q_n(Tx) < 1\}$. So for $x \in X$ and $\alpha < 1$, $\alpha\varepsilon x / p_m(x) \in B_m(0,\varepsilon)$ and thus

$$q_n\left(\frac{\alpha\varepsilon}{p_m(x)} Tx\right) < 1 \Longrightarrow q_n(Tx) < \frac{1}{\alpha\varepsilon} p_m(x)$$

for all x. Letting $\alpha \uparrow 1$ shows that $q_n(Tx) \leq \frac{1}{\varepsilon} p_m(x)$ for all $x \in X$. Conversely, if T satisfies

$$q_n(Tx) \leq C_n p_{m_n}(x) \text{ for all } x \in X,$$

then

$$q_n(Tx - Tx') = q_n(T(x-x')) \leq C_n p_{m_n}(x-x') \text{ for all } x, y \in X.$$

This shows $Tx' \to Tx$ as $x' \to x$, i.e. that T is continuous. ■

Definition 45.23. *A Frechét space is a vector space X equipped with a family $\{p_n\}$ of semi-norms such that X is complete in the associated metric d.*

Example 45.24. Let $K \sqsubset\sqsubset \mathbb{R}^n$ and $C^\infty(K) := \{f \in C_c^\infty(\mathbb{R}^n) : \operatorname{supp}(f) \subset K\}$. For $m \in \mathbb{N}$, let

$$p_m(f) := \sum_{|\alpha| \leq m} \|\partial^\alpha f\|_\infty.$$

Then $(C^\infty(K), \{p_m\}_{m=1}^\infty)$ is a Frechét space. Moreover the derivative operators $\{\partial_k\}$ and multiplication by smooth functions are continuous linear maps from $C^\infty(K)$ to $C^\infty(K)$. If μ is a finite measure on K, then $T(f) := \int_K \partial^\alpha f d\mu$ is an element of $C^\infty(K)^*$ for any multi index α.

Example 45.25. Let $U \subset_o \mathbb{R}^n$ and for $m \in \mathbb{N}$, and a compact set $K \sqsubset\sqsubset U$ let

$$p_m^K(f) := \sum_{|\alpha| \leq m} \|\partial^\alpha f\|_{\infty,K} := \sum_{|\alpha| \leq m} \max_{x \in K} |\partial^\alpha f(x)|.$$

Choose a sequence $K_m \sqsubset\sqsubset U$ such that $K_m \subset K_{m+1}^o \subset K_{m+1} \sqsubset\sqsubset U$ for all m and set $q_m(f) = p_m^{K_m}(f)$. Then $(C^\infty(K), \{p_m\}_{m=1}^\infty)$ is a Frechét space and the topology in independent of the choice of sequence of compact sets K exhausting U. Moreover the derivative operators $\{\partial_k\}$ and multiplication by smooth functions are continuous linear maps from $C^\infty(U)$ to $C^\infty(U)$. If μ is a finite measure with compact support in U, then $T(f) := \int_K \partial^\alpha f d\mu$ is an element of $C^\infty(U)^*$ for any multi index α.

Proposition 45.26. *A linear functional T on $C^\infty(U)$ is continuous, i.e. $T \in C^\infty(U)^*$ iff there exists a compact set $K \sqsubset\sqsubset U$, $m \in \mathbb{N}$ and $C < \infty$ such that*

$$|\langle T, \varphi\rangle| \le C p_m^K(\varphi) \text{ for all } \varphi \in C^\infty(U).$$

Notation 45.27 *Let $\nu_s(x) := (1+|x|)^s$ (or change to $\nu_s(x) = (1+|x|^2)^{s/2} = \langle x\rangle^s$?) for $x \in \mathbb{R}^n$ and $s \in \mathbb{R}$.*

Example 45.28. Let $\mathcal{S}$ denote the space of functions $f \in C^\infty(\mathbb{R}^n)$ such that f and all of its partial derivatives decay faster that $(1+|x|)^{-m}$ for all $m > 0$ as in Definition 43.6. Define

$$p_m(f) = \sum_{|\alpha|\le m} \|(1+|\cdot|)^m \partial^\alpha f(\cdot)\|_\infty = \sum_{|\alpha|\le m} \|(\mu_m \partial^\alpha f(\cdot)\|_\infty,$$

then $(\mathcal{S}, \{p_m\})$ is a Frechét space. Again the derivative operators $\{\partial_k\}$ and multiplication by function $f \in \mathcal{P}$ are examples of continuous linear operators on $\mathcal{S}$. For an example of an element $T \in \mathcal{S}^*$, let μ be a measure on $\mathbb{R}^n$ such that

$$\int (1+|x|)^{-N} d|\mu|(x) < \infty$$

for some $N \in \mathbb{N}$. Then $T(f) := \int_K \partial^\alpha f d\mu$ defines and element of $\mathcal{S}^*$.

Proposition 45.29. *The Fourier transform $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is a continuous linear transformation.*

Proof. For the purposes of this proof, it will be convenient to use the semi-norms

$$p_m'(f) = \sum_{|\alpha|\le m} \left\|(1+|\cdot|^2)^m \partial^\alpha f(\cdot)\right\|_\infty.$$

This is permissible, since by Exercise 45.2 they give rise to the same topology on $\mathcal{S}$. Let $f \in \mathcal{S}$ and $m \in \mathbb{N}$, then

$$\begin{aligned}(1+|\xi|^2)^m \partial^\alpha \hat{f}(\xi) &= (1+|\xi|^2)^m \mathcal{F}\left((-ix)^\alpha f\right)(\xi) \\ &= \mathcal{F}\left[(1-\Delta)^m \left((-ix)^\alpha f\right)\right](\xi)\end{aligned}$$

and therefore if we let $g = (1-\Delta)^m \left((-ix)^\alpha f\right) \in \mathcal{S}$,

$$\begin{aligned}\left|(1+|\xi|^2)^m \partial^\alpha \hat{f}(\xi)\right| &\le \|g\|_1 = \int_{\mathbb{R}^n} |g(x)|\, dx \\ &= \int_{\mathbb{R}^n} |g(x)| (1+|x|^2)^n \frac{1}{(1+|x|^2)^n} d\xi \\ &\le C \left\||g(\cdot)| (1+|\cdot|^2)^n\right\|_\infty\end{aligned}$$

where $C = \int_{\mathbb{R}^n} \frac{1}{(1+|x|^2)^n} d\xi < \infty$. Using the product rule repeatedly, it is not hard to show

$$\begin{aligned}\left\||g(\cdot)| (1+|\cdot|^2)^n\right\|_\infty &= \left\|(1+|\cdot|^2)^n (1-\Delta)^m \left((-ix)^\alpha f\right)\right\|_\infty \\ &\le k \sum_{|\beta|\le 2m} \left\|(1+|\cdot|^2)^{n+|\alpha|/2} \partial^\beta f\right\|_\infty \\ &\le k p_{2m+n}'(f)\end{aligned}$$

for some constant $k < \infty$. Combining the last two displayed equations implies that $p_m'(\hat{f}) \le C k p_{2m+n}'(f)$ for all $f \in \mathcal{S}$, and thus $\mathcal{F}$ is continuous. ■

Proposition 45.30. *The subspace $C_c^\infty(\mathbb{R}^n)$ is dense in $\mathcal{S}(\mathbb{R}^n)$.*

Proof. Let $\theta \in C_c^\infty(\mathbb{R}^n)$ such that $\theta = 1$ in a neighborhood of 0 and set $\theta_m(x) = \theta(x/m)$ for all $m \in \mathbb{N}$. We will now show for all $f \in \mathcal{S}$ that $\theta_m f$ converges to f in $\mathcal{S}$. The main point is by the product rule,

$$\begin{aligned}\partial^\alpha \left(\theta_m f - f\right)(x) &= \sum_{\beta\le\alpha} \binom{\alpha}{\beta} \partial^{\alpha-\beta}\theta_m(x) \partial^\beta f(x) - f \\ &= \sum_{\beta\le\alpha:\beta\ne\alpha} \binom{\alpha}{\beta} \frac{1}{m^{|\alpha-\beta|}} \partial^{\alpha-\beta}\theta(x/m) \partial^\beta f(x).\end{aligned}$$

Since $\max\left\{\left\|\partial^\beta\theta\right\|_\infty : \beta \le \alpha\right\}$ is bounded it then follows from the last equation that $\|\mu_t \partial^\alpha (\theta_m f - f)\|_\infty = O(1/m)$ for all $t > 0$ and α. That is to say $\theta_m f \to f$ in $\mathcal{S}$. ■

Lemma 45.31 (Peetre's Inequality). *For all $x, y \in \mathbb{R}^n$ and $s \in \mathbb{R}$,*

$$(1+|x+y|)^s \le \min\left\{(1+|y|)^{|s|}(1+|x|)^s, (1+|y|)^s(1+|x|)^{|s|}\right\} \tag{45.5}$$

that is to say $\nu_s(x+y) \le \nu_{|s|}(x)\nu_s(y)$ and $\nu_s(x+y) \le \nu_s(x)\nu_{|s|}(y)$ for all $s \in \mathbb{R}$, where $\nu_s(x) = (1+|x|)^s$ as in Notation 45.27. We also have the same results for $\langle x\rangle$, namely

$$\langle x+y\rangle^s \le 2^{|s|/2} \min\left\{\langle x\rangle^{|s|}\langle y\rangle^s, \langle x\rangle^s\langle y\rangle^{|s|}\right\}. \tag{45.6}$$

Proof. By elementary estimates,

$$(1+|x+y|) \le 1+|x|+|y| \le (1+|x|)(1+|y|)$$

and so for Eq. (45.5) holds if $s \ge 0$. Now suppose that $s < 0$, then

$$(1+|x+y|)^s \ge (1+|x|)^s(1+|y|)^s$$

and letting $x \to x - y$ and $y \to -y$ in this inequality implies

$$(1+|x|)^s \geq (1+|x+y|)^s (1+|y|)^s.$$

This inequality is equivalent to

$$(1+|x+y|)^s \leq (1+|x|)^s (1+|y|)^{-s} = (1+|x|)^s (1+|y|)^{|s|}.$$

By symmetry we also have

$$(1+|x+y|)^s \leq (1+|x|)^{|s|} (1+|y|)^s.$$

For the proof of Eq. (45.6

$$\begin{aligned}\langle x+y\rangle^2 &= 1+|x+y|^2 \leq 1+(|x|+|y|)^2 = 1+|x|^2+|y|^2+2|x||y| \\ &\leq 1+2|x|^2+2|y|^2 \leq 2(1+|x|^2)(1+|y|^2) = 2\langle x\rangle^2\langle y\rangle^2.\end{aligned}$$

From this it follows that $\langle x\rangle^{-2} \leq 2\langle x+y\rangle^{-2}\langle y\rangle^2$ and hence

$$\langle x+y\rangle^{-2} \leq 2\langle x\rangle^{-2}\langle y\rangle^2.$$

So if $s \geq 0$, then

$$\langle x+y\rangle^s \leq 2^{s/2}\langle x\rangle^s\langle y\rangle^s$$

and

$$\langle x+y\rangle^{-s} \leq 2^{s/2}\langle x\rangle^{-s}\langle y\rangle^s.$$

■

Proposition 45.32. *Suppose that $f, g \in \mathcal{S}$ then $f * g \in \mathcal{S}$.*

Proof. First proof. Since $\mathcal{F}(f*g) = \hat{f}\hat{g} \in \mathcal{S}$ it follows that $f*g = \mathcal{F}^{-1}(\hat{f}\hat{g}) \in \mathcal{S}$ as well. For the second proof we will make use of Peetre's inequality. We have for any $k, l \in \mathbb{N}$ that

$$\begin{aligned}\nu_t(x)\,|\partial^\alpha(f*g)(x)| &= \nu_t(x)\,|\partial^\alpha f * g(x)| \leq \nu_t(x)\int |\partial^\alpha f(x-y)|\,|g(y)|\,dy \\ &\leq C\nu_t(x)\int \nu_{-k}(x-y)\nu_{-l}(y)dy \leq C\nu_t(x)\int \nu_{-k}(x)\nu_k(y)\nu_{-l}(y)dy \\ &= C\nu_{t-k}(x)\int \nu_{k-l}(y)dy.\end{aligned}$$

Choosing $k = t$ and $l > t + n$ we learn that

$$\nu_t(x)\,|\partial^\alpha(f*g)(x)| \leq C\int \nu_{k-l}(y)dy < \infty$$

showing $\|\nu_t\partial^\alpha(f*g)\|_\infty < \infty$ for all $t \geq 0$ and $\alpha \in \mathbb{N}^n$. ■

45.4 Compactly supported distributions

Definition 45.33. *For a distribution $T \in \mathcal{D}'(U)$ and $V \subset_o U$, we say $T|_V = 0$ if $\langle T, \varphi\rangle = 0$ for all $\varphi \in \mathcal{D}(V)$.*

Proposition 45.34. *Suppose that $\mathcal{V} := \{V_\alpha\}_{\alpha\in A}$ is a collection of open subset of U such that $T|_{V_\alpha} = 0$ for all α, then $T|_W = 0$ where $W = \cup_{\alpha\in A}V_\alpha$.*

Proof. Let $\{\psi_\alpha\}_{\alpha\in A}$ be a smooth partition of unity subordinate to $\mathcal{V}$, i.e. $\operatorname{supp}(\psi_\alpha) \subset V_\alpha$ for all $\alpha \in A$, for each point $x \in W$ there exists a neighborhood $N_x \subset_o W$ such that $\#\{\alpha \in A : \operatorname{supp}(\psi_\alpha) \cap N_x \neq \emptyset\} < \infty$ and $1_W = \sum_{\alpha\in A}\psi_\alpha$. Then for $\varphi \in \mathcal{D}(W)$, we have $\varphi = \sum_{\alpha\in A}\varphi\psi_\alpha$ and there are only a finite number of nonzero terms in the sum since $\operatorname{supp}(\varphi)$ is compact. Since $\varphi\psi_\alpha \in \mathcal{D}(V_\alpha)$ for all α,

$$\langle T, \varphi\rangle = \langle T, \sum_{\alpha\in A}\varphi\psi_\alpha\rangle = \sum_{\alpha\in A}\langle T, \varphi\psi_\alpha\rangle = 0.$$

■

Definition 45.35. *The support,* $\operatorname{supp}(T)$, *of a distribution $T \in \mathcal{D}'(U)$ is the relatively closed subset of U determined by*

$$U \setminus \operatorname{supp}(T) = \cup\{V \subset_o U : T|_V = 0\}.$$

By Proposition 45.26, $\operatorname{supp}(T)$ *may described as the smallest (relatively) closed set F such that $T|_{U\setminus F} = 0$.*

Proposition 45.36. *If $f \in L^1_{loc}(U)$, then* $\operatorname{supp}(T_f) = \operatorname{ess\,sup}(f)$, *where*

$$\operatorname{ess\,sup}(f) := \{x \in U : m(\{y \in V : f(y) \neq 0\}) > 0 \text{ for all neighborhoods } V \text{ of } x\}$$

as in Definition 31.26.

Proof. The key point is that $T_f|_V = 0$ iff $f = 0$ a.e. on V and therefore

$$U \setminus \operatorname{supp}(T_f) = \cup\{V \subset_o U : f1_V = 0 \text{ a.e.}\}.$$

On the other hand,

$$\begin{aligned}U \setminus \operatorname{ess\,sup}(f) &= \{x \in U : m(\{y \in V : f(y) \neq 0\}) = 0 \text{ for some neighborhood } V \text{ of } x\} \\ &= \cup\{x \in U : f1_V = 0 \text{ a.e. for some neighborhood } V \text{ of } x\} \\ &= \cup\{V \subset_o U : f1_V = 0 \text{ a.e.}\}\end{aligned}$$

■

Definition 45.37. *Let $\mathcal{E}'(U) := \{T \in \mathcal{D}'(U) : \operatorname{supp}(T) \subset U$ is compact$\}$ – the compactly supported distributions in $\mathcal{D}'(U)$.*

Lemma 45.38. *Suppose that $T \in \mathcal{D}'(U)$ and $f \in C^\infty(U)$ is a function such that $K := \operatorname{supp}(T) \cap \operatorname{supp}(f)$ is a compact subset of U. Then we may define $\langle T, f\rangle := \langle T, \theta f\rangle$, where $\theta \in \mathcal{D}(U)$ is any function such that $\theta = 1$ on a neighborhood of K. Moreover, if $K \sqsubset\sqsubset U$ is a given compact set and $F \sqsubset\sqsubset U$ is a compact set such that $K \subset F^o$, then there exists $m \in \mathbb{N}$ and $C < \infty$ such that*

$$|\langle T, f\rangle| \le C \sum_{|\beta| \le m} \left\|\partial^\beta f\right\|_{\infty, F} \tag{45.7}$$

for all $f \in C^\infty(U)$ such that $\operatorname{supp}(T) \cap \operatorname{supp}(f) \subset K$. In particular if $T \in \mathcal{E}'(U)$ then T extends uniquely to a linear functional on $C^\infty(U)$ and there is a compact subset $F \sqsubset\sqsubset U$ such that the estimate in Eq. (45.7) holds for all $f \in C^\infty(U)$.

Proof. Suppose that $\tilde\theta$ is another such cutoff function and let V be an open neighborhood of K such that $\theta = \tilde\theta = 1$ on V. Setting $g := \left(\theta - \tilde\theta\right) f \in \mathcal{D}(U)$ we see that

$$\operatorname{supp}(g) \subset \operatorname{supp}(f) \setminus V \subset \operatorname{supp}(f) \setminus K = \operatorname{supp}(f) \setminus \operatorname{supp}(T) \subset U \setminus \operatorname{supp}(T),$$

see Figure 45.1 below. Therefore,

$$0 = \langle T, g\rangle = \langle T, \left(\theta - \tilde\theta\right) f\rangle = \langle T, \theta f\rangle - \langle T, \tilde\theta f\rangle$$

which shows that $\langle T, f\rangle$ is well defined. Moreover, if $F \sqsubset\sqsubset U$ is a compact set

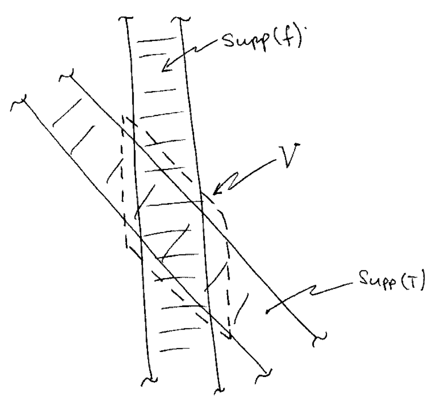


Fig. 45.1. Intersecting the supports.

such that $K \subset F^o$ and $\theta \in C_c^\infty(F^0)$ is a function which is 1 on a neighborhood of K, we have

$$|\langle T, f\rangle| = |\langle T, \theta f\rangle| = C \sum_{|\alpha| \le m} \|\partial^\alpha(\theta f)\|_\infty \le C \sum_{|\beta| \le m} \left\|\partial^\beta f\right\|_{\infty, F}$$

and this estimate holds for all $f \in C^\infty(U)$ such that $\operatorname{supp}(T) \cap \operatorname{supp}(f) \subset K$. ■

Theorem 45.39. *The restriction of $T \in C^\infty(U)^*$ to $C_c^\infty(U)$ defines an element in $\mathcal{E}'(U)$. Moreover the map*

$$T \in C^\infty(U)^* \xrightarrow{i} T|_{\mathcal{D}(U)} \in \mathcal{E}'(U)$$

is a linear isomorphism of vector spaces. The inverse map is defined as follows. Given $S \in \mathcal{E}'(U)$ and $\theta \in C_c^\infty(U)$ such that $\theta = 1$ on $K = \operatorname{supp}(S)$ then $i^{-1}(S) = \theta S$, where $\theta S \in C^\infty(U)^*$ defined by

$$\langle \theta S, \varphi\rangle = \langle S, \theta\varphi\rangle \text{ for all } \varphi \in C^\infty(U).$$

Proof. Suppose that $T \in C^\infty(U)^*$ then there exists a compact set $K \sqsubset\sqsubset U$, $m \in \mathbb{N}$ and $C < \infty$ such that

$$|\langle T, \varphi\rangle| \le C p_m^K(\varphi) \text{ for all } \varphi \in C^\infty(U)$$

where p_m^K is defined in Example 45.25. It is clear using the sequential notion of continuity that $T|_{\mathcal{D}(U)}$ is continuous on $\mathcal{D}(U)$, i.e. $T|_{\mathcal{D}(U)} \in \mathcal{D}'(U)$. Moreover, if $\theta \in C_c^\infty(U)$ such that $\theta = 1$ on a neighborhood of K then

$$|\langle T, \theta\varphi\rangle - \langle T, \varphi\rangle| = |\langle T, (\theta - 1)\varphi\rangle| \le C p_m^K((\theta - 1)\varphi) = 0,$$

which shows $\theta T = T$. Hence $\operatorname{supp}(T) = \operatorname{supp}(\theta T) \subset \operatorname{supp}(\theta) \sqsubset\sqsubset U$ showing that $T|_{\mathcal{D}(U)} \in \mathcal{E}'(U)$. Therefore the map i is well defined and is clearly linear. I also claim that i is injective because if $T \in C^\infty(U)^*$ and $i(T) = T|_{\mathcal{D}(U)} \equiv 0$, then $\langle T, \varphi\rangle = \langle \theta T, \varphi\rangle = \langle T|_{\mathcal{D}(U)}, \theta\varphi\rangle = 0$ for all $\varphi \in C^\infty(U)$. To show i is surjective suppose that $S \in \mathcal{E}'(U)$. By Lemma 45.38 we know that S extends uniquely to an element $\tilde S$ of $C^\infty(U)^*$ such that $\tilde S|_{\mathcal{D}(U)} = S$, i.e. $i(\tilde S) = S$. and $K = \operatorname{supp}(S)$. ■

Lemma 45.40. *The space $\mathcal{E}'(U)$ is a sequentially dense subset of $\mathcal{D}'(U)$.*

Proof. Choose $K_n \sqsubset\sqsubset U$ such that $K_n \subset K_{n+1}^o \subset K_{n+1} \uparrow U$ as $n \to \infty$. Let $\theta_n \in C_c^\infty(K_{n+1}^0)$ such that $\theta_n = 1$ on K. Then for $T \in \mathcal{D}'(U)$, $\theta_n T \in \mathcal{E}'(U)$ and $\theta_n T \to T$ as $n \to \infty$. ■

45.5 Tempered Distributions and the Fourier Transform

The space of tempered distributions $\mathcal{S}'(\mathbb{R}^n)$ is the continuous dual to $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$. A linear functional T on $\mathcal{S}$ is continuous iff there exists $k \in \mathbb{N}$ and $C < \infty$ such that

$$|\langle T,\varphi\rangle| \le Cp_k(\varphi) := C\sum_{|\alpha|\le k}\|\nu_k\partial^\alpha\varphi\|_\infty \tag{45.8}$$

for all $\varphi\in\mathcal{S}$. Since $\mathcal{D}=\mathcal{D}(\mathbb{R}^n)$ is a dense subspace of $\mathcal{S}$ any element $T\in\mathcal{S}'$ is determined by its restriction to $\mathcal{D}$. Moreover, if $T\in\mathcal{S}'$ it is easy to see that $T|_{\mathcal{D}}\in\mathcal{D}'$. Conversely and element $T\in\mathcal{D}'$ satisfying an estimate of the form in Eq. (45.8) for all $\varphi\in\mathcal{D}$ extend uniquely to an element of $\mathcal{S}'$. In this way we may view $\mathcal{S}'$ as a subspace of $\mathcal{D}'$.

Example 45.41. Any compactly supported distribution is tempered, i.e. $\mathcal{E}'(U)\subset\mathcal{S}'(\mathbb{R}^n)$ for any $U\subset_o\mathbb{R}^n$.

One of the virtues of $\mathcal{S}'$ is that we may extend the Fourier transform to $\mathcal{S}'$. Recall that for L^1 functions f and g we have the identity,

$$\langle\hat f,g\rangle=\langle f,\hat g\rangle.$$

This suggests the following definition.

Definition 45.42. *The Fourier and inverse Fourier transform of a tempered distribution $T\in\mathcal{S}'$ are the distributions $\hat T=\mathcal{F}T\in\mathcal{S}'$ and $T^\vee=\mathcal{F}^{-1}T\in\mathcal{S}'$ defined by*

$$\langle\hat T,\varphi\rangle=\langle T,\hat\varphi\rangle \text{ and } \langle T^\vee,\varphi\rangle=\langle T,\varphi^\vee\rangle \text{ for all } \varphi\in\mathcal{S}.$$

Since $\mathcal{F}:\mathcal{S}\to\mathcal{S}$ is a continuous isomorphism with inverse $\mathcal{F}^{-1}$, one easily checks that $\hat T$ and $T^\vee$ are well defined elements of $\mathcal{S}$ and that $\mathcal{F}^{-1}$ is the inverse of $\mathcal{F}$ on $\mathcal{S}'$.

Example 45.43. Suppose that μ is a complex measure on $\mathbb{R}^n$. Then we may view μ as an element of $\mathcal{S}'$ via $\langle\mu,\varphi\rangle=\int\varphi d\mu$ for all $\varphi\in\mathcal{S}'$. Then by Fubini-Tonelli,

$$\begin{aligned}\langle\hat\mu,\varphi\rangle=\langle\mu,\hat\varphi\rangle&=\int\hat\varphi(x)d\mu(x)=\int\left[\int\varphi(\xi)e^{-ix\cdot\xi}d\xi\right]d\mu(x)\\&=\int\left[\int\varphi(\xi)e^{-ix\cdot\xi}d\mu(x)\right]d\xi\end{aligned}$$

which shows that $\hat\mu$ is the distribution associated to the continuous function $\xi\to\int e^{-ix\cdot\xi}d\mu(x)$. $\int e^{-ix\cdot\xi}d\mu(x)$We will somewhat abuse notation and identify the distribution $\hat\mu$ with the function $\xi\to\int e^{-ix\cdot\xi}d\mu(x)$. When $d\mu(x)=f(x)dx$ with $f\in L^1$, we have $\hat\mu=\hat f$, so the definitions are all consistent.

Corollary 45.44. *Suppose that μ is a complex measure such that $\hat\mu=0$, then $\mu=0$. So complex measures on $\mathbb{R}^n$ are uniquely determined by their Fourier transform.*

Proof. If $\hat\mu=0$, then $\mu=0$ as a distribution, i.e. $\int\varphi d\mu=0$ for all $\varphi\in\mathcal{S}$ and in particular for all $\varphi\in\mathcal{D}$. By Example 45.5 this implies that μ is the zero measure. ■

More generally we have the following analogous theorem for compactly supported distributions.

Theorem 45.45. *Let $S\in\mathcal{E}'(\mathbb{R}^n)$, then $\hat S$ is an analytic function and $\hat S(z)=\langle S(x),e^{-ix\cdot z}\rangle$. Also if $\operatorname{supp}(S)\sqsubset\sqsubset B(0,M)$, then $\hat S(z)$ satisfies a bound of the form*

$$\left|\hat S(z)\right|\le C(1+|z|)^m e^{M|\operatorname{Im} z|}$$

for some $m\in\mathbb{N}$ and $C<\infty$. If $S\in\mathcal{D}(\mathbb{R}^n)$, i.e. if S is assumed to be smooth, then for all $m\in\mathbb{N}$ there exists $C_m<\infty$ such that

$$\left|\hat S(z)\right|\le C_m(1+|z|)^{-m}e^{M|\operatorname{Im} z|}.$$

Proof. The function $h(z)=\langle S(\xi),e^{-iz\cdot\xi}\rangle$ for $z\in\mathbb{C}^n$ is analytic since the map $z\in\mathbb{C}^n\to e^{-iz\cdot\xi}\in C^\infty(\xi\in\mathbb{R}^n)$ is analytic and S is complex linear. Moreover, we have the bound

$$\begin{aligned}|h(z)|=\left|\langle S(\xi),e^{-iz\cdot\xi}\rangle\right|&\le C\sum_{|\alpha|\le m}\left\|\partial_\xi^\alpha e^{-iz\cdot\xi}\right\|_{\infty,B(0,M)}\\&=C\sum_{|\alpha|\le m}\left\|z^\alpha e^{-iz\cdot\xi}\right\|_{\infty,B(0,M)}\\&\le C\sum_{|\alpha|\le m}|z|^{|\alpha|}\left\|e^{-iz\cdot\xi}\right\|_{\infty,B(0,M)}\le C(1+|z|)^m e^{M|\operatorname{Im} z|}.\end{aligned}$$

If we now assume that $S\in\mathcal{D}(\mathbb{R}^n)$, then

$$\begin{aligned}\left|z^\alpha\hat S(z)\right|&=\left|\int_{\mathbb{R}^n}S(\xi)z^\alpha e^{-iz\cdot\xi}d\xi\right|=\left|\int_{\mathbb{R}^n}S(\xi)(i\partial_\xi)^\alpha e^{-iz\cdot\xi}d\xi\right|\\&=\left|\int_{\mathbb{R}^n}(-i\partial_\xi)^\alpha S(\xi)e^{-iz\cdot\xi}d\xi\right|\le e^{M|\operatorname{Im} z|}\int_{\mathbb{R}^n}\left|\partial_\xi{}^\alpha S(\xi)\right|d\xi\end{aligned}$$

showing

$$|z^\alpha|\left|\hat S(z)\right|\le e^{M|\operatorname{Im} z|}\|\partial^\alpha S\|_1$$

and therefore

$$(1+|z|)^m\left|\hat S(z)\right|\le Ce^{M|\operatorname{Im} z|}\sum_{|\alpha|\le m}\|\partial^\alpha S\|_1\le Ce^{M|\operatorname{Im} z|}.$$

So to finish the proof it suffices to show $h = \hat{S}$ in the sense of distributions[1]. For this let $\varphi \in \mathcal{D}$, $K \sqsubset\sqsubset \mathbb{R}^n$ be a compact set for $\varepsilon > 0$ let

$$\hat{\varphi}_\varepsilon(\xi) = (2\pi)^{-n/2}\varepsilon^n \sum_{x\in\varepsilon\mathbb{Z}^n} \varphi(x)e^{-ix\cdot\xi}.$$

This is a finite sum and

$$\begin{aligned}
&\sup_{\xi\in K} |\partial^\alpha (\hat{\varphi}_\varepsilon(\xi) - \hat{\varphi}(\xi))| \\
&\quad = \sup_{\xi\in K} \left| \sum_{y\in\varepsilon\mathbb{Z}^n} \int_{y+\varepsilon(0,1]^n} \left((-iy)^\alpha \varphi(y)e^{-iy\cdot\xi} - (-ix)^\alpha \varphi(x)e^{-ix\cdot\xi}\right) dx \right| \\
&\quad \le \sum_{y\in\varepsilon\mathbb{Z}^n} \int_{y+\varepsilon(0,1]^n} \sup_{\xi\in K} \left|y^\alpha\varphi(y)e^{-iy\cdot\xi} - x^\alpha\varphi(x)e^{-ix\cdot\xi}\right| dx
\end{aligned}$$

By uniform continuity of $x^\alpha\varphi(x)e^{-ix\cdot\xi}$ for $(\xi, x) \in K \times \mathbb{R}^n$ (φ has compact support),

$$\delta(\varepsilon) = \sup_{\xi\in K} \sup_{y\in\varepsilon\mathbb{Z}^n} \sup_{x\in y+\varepsilon(0,1]^n} \left|y^\alpha\varphi(y)e^{-iy\cdot\xi} - x^\alpha\varphi(x)e^{-ix\cdot\xi}\right| \to 0 \text{ as } \varepsilon \downarrow 0$$

which shows

$$\sup_{\xi\in K} |\partial^\alpha (\hat{\varphi}_\varepsilon(\xi) - \hat{\varphi}(\xi))| \le C\delta(\varepsilon)$$

where C is the volume of a cube in $\mathbb{R}^n$ which contains the support of φ. This shows that $\hat{\varphi}_\varepsilon \to \hat{\varphi}$ in $C^\infty(\mathbb{R}^n)$. Therefore,

$$\begin{aligned}
\langle \hat{S}, \varphi\rangle &= \langle S, \hat{\varphi}\rangle = \lim_{\varepsilon\downarrow 0}\langle S, \hat{\varphi}_\varepsilon\rangle = \lim_{\varepsilon\downarrow 0}(2\pi)^{-n/2}\varepsilon^n \sum_{x\in\varepsilon\mathbb{Z}^n} \varphi(x)\langle S(\xi), e^{-ix\cdot\xi}\rangle \\
&= \lim_{\varepsilon\downarrow 0}(2\pi)^{-n/2}\varepsilon^n \sum_{x\in\varepsilon\mathbb{Z}^n} \varphi(x)h(x) = \int_{\mathbb{R}^n} \varphi(x)h(x)dx = \langle h, \varphi\rangle.
\end{aligned}$$

■

[1] This is most easily done using Fubini's Theorem 46.2 for distributions proved below. This proof goes as follows. Let $\theta, \eta \in \mathcal{D}(\mathbb{R}^n)$ such that $\theta = 1$ on a neighborhood of $\operatorname{supp}(S)$ and $\eta = 1$ on a neighborhood of $\operatorname{supp}(\phi)$ then

$$\begin{aligned}
\langle h, \phi\rangle &= \langle \phi(x), \langle S(\xi), e^{-ix\cdot\xi}\rangle\rangle = \langle \eta(x)\phi(x), \langle S(\xi), \theta(\xi)e^{-ix\cdot\xi}\rangle\rangle \\
&= \langle \phi(x), \langle S(\xi), \eta(x)\theta(\xi)e^{-ix\cdot\xi}\rangle\rangle.
\end{aligned}$$

We may now apply Theorem 46.2 to conclude,

$$\begin{aligned}
\langle h, \phi\rangle &= \langle S(\xi), \langle \phi(x), \eta(x)\theta(\xi)e^{-ix\cdot\xi}\rangle\rangle = \langle S(\xi), \theta(\xi)\langle \phi(x), e^{-ix\cdot\xi}\rangle\rangle = \langle S(\xi), \langle \phi(x), e^{-ix\cdot\xi}\rangle\rangle \\
&= \langle S(\xi), \hat{\phi}(\xi)\rangle.
\end{aligned}$$

Remark 45.46. Notice that

$$\partial^\alpha \hat{S}(z) = \langle S(x), \partial_z^\alpha e^{-ix\cdot z}\rangle = \langle S(x), (-ix)^\alpha e^{-ix\cdot z}\rangle = \langle (-ix)^\alpha S(x), e^{-ix\cdot z}\rangle$$

and $(-ix)^\alpha S(x) \in \mathcal{E}'(\mathbb{R}^n)$. Therefore, we find a bound of the form

$$\left|\partial^\alpha \hat{S}(z)\right| \le C(1 + |z|)^{m'} e^{M|\operatorname{Im} z|}$$

where C and m' depend on α. In particular, this shows that $\hat{S} \in \mathcal{P}$, i.e. $\mathcal{S}'$ is preserved under multiplication by $\hat{S}$.

The converse of this theorem holds as well. For the moment we only have the tools to prove the smooth converse. The general case will follow by using the notion of convolution to regularize a distribution to reduce the question to the smooth case.

Theorem 45.47. *Let $S \in \mathcal{S}(\mathbb{R}^n)$ and assume that $\hat{S}$ is an analytic function and there exists an $M < \infty$ such that for all $m \in \mathbb{N}$ there exists $C_m < \infty$ such that*

$$\left|\hat{S}(z)\right| \le C_m(1 + |z|)^{-m} e^{M|\operatorname{Im} z|}.$$

Then $\operatorname{supp}(S) \subset \overline{B(0, M)}$.

Proof. By the Fourier inversion formula,

$$S(x) = \int_{\mathbb{R}^n} \hat{S}(\xi)e^{i\xi\cdot x}d\xi$$

and by deforming the contour, we may express this integral as

$$S(x) = \int_{\mathbb{R}^n + i\eta} \hat{S}(\xi)e^{i\xi\cdot x}d\xi = \int_{\mathbb{R}^n} \hat{S}(\xi + i\eta)e^{i(\xi+i\eta)\cdot x}d\xi$$

for any $\eta \in \mathbb{R}^n$. From this last equation it follows that

$$\begin{aligned}
|S(x)| &\le e^{-\eta\cdot x} \int_{\mathbb{R}^n} \left|\hat{S}(\xi + i\eta)\right| d\xi \le C_m e^{-\eta\cdot x} e^{M|\eta|} \int_{\mathbb{R}^n} (1 + |\xi + i\eta|)^{-m} d\xi \\
&\le C_m e^{-\eta\cdot x} e^{M|\eta|} \int_{\mathbb{R}^n} (1 + |\xi|)^{-m} d\xi \le \tilde{C}_m e^{-\eta\cdot x} e^{M|\eta|}
\end{aligned}$$

where $\tilde{C}_m < \infty$ if $m > n$. Letting $\eta = \lambda x$ with $\lambda > 0$ we learn

$$|S(x)| \le \tilde{C}_m \exp\left(-\lambda |x|^2 + M|x|\right) = \tilde{C}_m e^{\lambda|x|(M-|x|)}. \tag{45.9}$$

Hence if $|x| > M$, we may let $\lambda \to \infty$ in Eq. (45.9) to show $S(x) = 0$. That is to say $\operatorname{supp}(S) \subset \overline{B(0, M)}$. ■

Let us now pause to work out some specific examples of Fourier transform of measures.

Example 45.48 (Delta Functions). Let $a \in \mathbb{R}^n$ and δ_a be the point mass measure at a, then

$$\hat{\delta}_a(\xi) = e^{-ia\cdot\xi}.$$

In particular it follows that

$$\mathcal{F}^{-1}e^{-ia\cdot\xi} = \delta_a.$$

To see the content of this formula, let $\varphi \in \mathcal{S}$. Then

$$\int e^{-ia\cdot\xi}\varphi^\vee(\xi)d\xi = \langle e^{-ia\cdot\xi}, \mathcal{F}^{-1}\varphi\rangle = \langle \mathcal{F}^{-1}e^{-ia\cdot\xi}, \varphi\rangle = \langle \delta_a, \varphi\rangle = \varphi(a)$$

which is precisely the Fourier inversion formula.

Example 45.49. Suppose that $p(x)$ is a polynomial. Then

$$\langle \hat{p}, \varphi\rangle = \langle p, \hat{\varphi}\rangle = \int p(\xi)\hat{\varphi}(\xi)d\xi.$$

Now

$$\begin{aligned} p(\xi)\hat{\varphi}(\xi) &= \int \varphi(x)p(\xi)e^{-i\xi\cdot x}dx = \int \varphi(x)p(i\partial_x)e^{-i\xi\cdot x}dx \\ &= \int p(-i\partial_x)\varphi(x)e^{-i\xi\cdot x}dx = \mathcal{F}\left(p(-i\partial)\varphi\right)(\xi) \end{aligned}$$

which combined with the previous equation implies

$$\begin{aligned} \langle \hat{p}, \varphi\rangle &= \int \mathcal{F}\left(p(-i\partial)\varphi\right)(\xi)d\xi = \left(\mathcal{F}^{-1}\mathcal{F}\left(p(-i\partial)\varphi\right)\right)(0) = p(-i\partial)\varphi(0) \\ &= \langle \delta_0, p(-i\partial)\varphi\rangle = \langle p(i\partial)\delta_0, \varphi\rangle. \end{aligned}$$

Thus we have shown that $\hat{p} = p(i\partial)\delta_0$.

Lemma 45.50. *Let $p(\xi)$ be a polynomial in $\xi \in \mathbb{R}^n$, $L = p(-i\partial)$ (a constant coefficient partial differential operator) and $T \in \mathcal{S}'$, then*

$$\mathcal{F}p(-i\partial)T = p\hat{T}.$$

In particular if $T = \delta_0$, we have

$$\mathcal{F}p(-i\partial)\delta_0 = p\cdot\hat{\delta}_0 = p.$$

Proof. By definition,

$$\langle \mathcal{F}LT, \varphi\rangle = \langle LT, \hat{\varphi}\rangle = \langle p(-i\partial)T, \hat{\varphi}\rangle = \langle T, p(i\partial)\hat{\varphi}\rangle$$

and

$$p(i\partial_\xi)\hat{\varphi}(\xi) = p(i\partial_\xi)\int \varphi(x)e^{-ix\cdot\xi}dx = \int p(x)\varphi(x)e^{-ix\cdot\xi}dx = (p\varphi)\,\hat{}\,.$$

Thus

$$\langle \mathcal{F}LT, \varphi\rangle = \langle T, p(i\partial)\hat{\varphi}\rangle = \langle T, (p\varphi)\,\hat{}\,\rangle = \langle \hat{T}, p\varphi\rangle = \langle p\hat{T}, \varphi\rangle$$

which proves the lemma. ■

Example 45.51. Let $n = 1$, $-\infty < a < b < \infty$, and $d\mu(x) = 1_{[a,b]}(x)dx$. Then

$$\begin{aligned} \hat{\mu}(\xi) &= \int_a^b e^{-ix\cdot\xi}dx = \frac{1}{\sqrt{2\pi}}\frac{e^{-ix\cdot\xi}}{-i\xi}|_a^b = \frac{1}{\sqrt{2\pi}}\frac{e^{-ib\cdot\xi} - e^{-ia\cdot\xi}}{-i\xi} \\ &= \frac{1}{\sqrt{2\pi}}\frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi}. \end{aligned}$$

So by the inversion formula we may conclude that

$$\mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}}\frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi}\right)(x) = 1_{[a,b]}(x) \tag{45.10}$$

in the sense of distributions. This also true at the Level of L^2 – functions. When $a = -b$ and $b > 0$ these formula reduce to

$$\mathcal{F}1_{[-b,b]} = \frac{1}{\sqrt{2\pi}}\frac{e^{ib\cdot\xi} - e^{-ib\cdot\xi}}{i\xi} = \frac{2}{\sqrt{2\pi}}\frac{\sin b\xi}{\xi}$$

and

$$\mathcal{F}^{-1}\frac{2}{\sqrt{2\pi}}\frac{\sin b\xi}{\xi} = 1_{[-b,b]}.$$

Let us pause to work out Eq. (45.10) by first principles. For $M \in (0,\infty)$ let ν_N be the complex measure on $\mathbb{R}^n$ defined by

$$d\nu_M(\xi) = \frac{1}{\sqrt{2\pi}}1_{|\xi|\le M}\frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi}d\xi,$$

then

$$\frac{1}{\sqrt{2\pi}}\frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi} = \lim_{M\to\infty}\nu_M \text{ in the } \mathcal{S}' \text{ topology.}$$

Hence

$$\mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}}\frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi}\right)(x) = \lim_{M\to\infty}\mathcal{F}^{-1}\nu_M$$

and

$$\mathcal{F}^{-1}\nu_M(\xi) = \int_{-M}^{M} \frac{1}{\sqrt{2\pi}} \frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi} e^{i\xi x} d\xi.$$

Since is $\xi \to \frac{1}{\sqrt{2\pi}} \frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi} e^{i\xi x}$ is a holomorphic function on $\mathbb{C}$ we may deform the contour to any contour in $\mathbb{C}$ starting at $-M$ and ending at M. Let Γ_M denote the straight line path from $-M$ to -1 along the real axis followed by the contour $e^{i\theta}$ for θ going from π to 2π and then followed by the straight line path from 1 to M. Then

$$\begin{aligned}\int_{|\xi|\le M} \frac{1}{\sqrt{2\pi}} \frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi} e^{i\xi x} d\xi &= \int_{\Gamma_M} \frac{1}{\sqrt{2\pi}} \frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi} e^{i\xi x} d\xi \\ &= \int_{\Gamma_M} \frac{1}{\sqrt{2\pi}} \frac{e^{i(x-a)\cdot\xi} - e^{i(x-b)\cdot\xi}}{i\xi} d\xi \\ &= \frac{1}{2\pi i} \int_{\Gamma_M} \frac{e^{i(x-a)\cdot\xi} - e^{i(x-b)\cdot\xi}}{i\xi} dm(\xi).\end{aligned}$$

By the usual contour methods we find

$$\lim_{M\to\infty} \frac{1}{2\pi i} \int_{\Gamma_M} \frac{e^{iy\xi}}{\xi} dm(\xi) = \begin{cases} 1 \text{ if } & y > 0 \\ 0 \text{ if } & y < 0 \end{cases}$$

and therefore we have

$$\mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}} \frac{e^{-ia\cdot\xi} - e^{-ib\cdot\xi}}{i\xi}\right)(x) = \lim_{M\to\infty} \mathcal{F}^{-1}\nu_M(x) = 1_{x>a} - 1_{x>b} = 1_{[a,b]}(x).$$

Example 45.52. Let σ_t be the surface measure on the sphere S_t of radius t centered at zero in $\mathbb{R}^3$. Then

$$\hat{\sigma}_t(\xi) = 4\pi t \frac{\sin t|\xi|}{|\xi|}.$$

Indeed,

$$\begin{aligned}\hat{\sigma}_t(\xi) &= \int_{tS^2} e^{-ix\cdot\xi} d\sigma(x) = t^2 \int_{S^2} e^{-itx\cdot\xi} d\sigma(x) \\ &= t^2 \int_{S^2} e^{-itx_3|\xi|} d\sigma(x) = t^2 \int_0^{2\pi} d\theta \int_0^{\pi} d\varphi \sin\varphi e^{-it\cos\varphi|\xi|} \\ &= 2\pi t^2 \int_{-1}^{1} e^{-itu|\xi|} du = 2\pi t^2 \frac{1}{-it|\xi|} e^{-itu|\xi|}|_{u=-1}^{u=1} = 4\pi t^2 \frac{\sin t|\xi|}{t|\xi|}.\end{aligned}$$

By the inversion formula, it follows that

$$\mathcal{F}^{-1} \frac{\sin t|\xi|}{|\xi|} = \frac{t}{4\pi t^2} \sigma_t = t\bar{\sigma}_t$$

where $\bar{\sigma}_t$ is $\frac{1}{4\pi t^2}\sigma_t$, the surface measure on S_t normalized to have total measure one.

Let us again pause to try to compute this inverse Fourier transform directly. To this end, let $f_M(\xi) := \frac{\sin t|\xi|}{t|\xi|} 1_{|\xi|\le M}$. By the dominated convergence theorem, it follows that $f_M \to \frac{\sin t|\xi|}{t|\xi|}$ in $\mathcal{S}'$, i.e. pointwise on $\mathcal{S}$. Therefore,

$$\langle \mathcal{F}^{-1} \frac{\sin t|\xi|}{t|\xi|}, \varphi\rangle = \langle \frac{\sin t|\xi|}{t|\xi|}, \mathcal{F}^{-1}\varphi\rangle = \lim_{M\to\infty} \langle f_M, \mathcal{F}^{-1}\varphi\rangle = \lim_{M\to\infty} \langle \mathcal{F}^{-1} f_M, \varphi\rangle$$

and

$$\begin{aligned}(2\pi)^{3/2}\mathcal{F}^{-1} f_M(x) &= (2\pi)^{3/2} \int_{\mathbb{R}^3} \frac{\sin t|\xi|}{t|\xi|} 1_{|\xi|\le M} e^{i\xi\cdot x} d\xi \\ &= \int_{r=0}^{M} \int_{\theta=0}^{2\pi} \int_{\varphi=0}^{\pi} \frac{\sin tr}{tr} e^{ir|x|\cos\varphi} r^2 \sin\varphi dr d\varphi d\theta \\ &= \int_{r=0}^{M} \int_{\theta=0}^{2\pi} \int_{u=-1}^{1} \frac{\sin tr}{tr} e^{ir|x|u} r^2 dr du d\theta \\ &= 2\pi \int_{r=0}^{M} \frac{\sin tr}{t} \frac{e^{ir|x|} - e^{-ir|x|}}{ir|x|} r dr \\ &= \frac{4\pi}{t|x|} \int_{r=0}^{M} \sin tr \sin r|x| dr \\ &= \frac{4\pi}{t|x|} \int_{r=0}^{M} \frac{1}{2} (\cos(r(t+|x|) - \cos(r(t-|x|)) dr \\ &= \frac{4\pi}{t|x|} \frac{1}{2(t+|x|)} (\sin(r(t+|x|) - \sin(r(t-|x|)) |_{r=0}^{M} \\ &= \frac{4\pi}{t|x|} \frac{1}{2} \left(\frac{\sin(M(t+|x|)}{t+|x|} - \frac{\sin(M(t-|x|)}{t-|x|} \right)\end{aligned}$$

Now make use of the fact that $\frac{\sin Mx}{x} \to \pi\delta(x)$ in one dimension to finish the proof.

45.6 Wave Equation

Given a distribution T and a test function φ, we wish to define $T * \varphi \in C^\infty$ by the formula

$$T * \varphi(x) = \text{“}\int T(y)\varphi(x-y)dy\text{”} = \langle T, \varphi(x - \cdot)\rangle.$$

As motivation for wanting to understand convolutions of distributions let us reconsider the wave equation in $\mathbb{R}^n$,

$$0 = \left(\partial_t^2 - \Delta\right) u(t,x) \text{ with}$$
$$u(0,x) = f(x) \text{ and } u_t(0,x) = g(x).$$

Taking the Fourier transform in the x variables gives the following equation

$$0 = \hat{u}_{t\,t}(t,\xi) + |\xi|^2 \hat{u}(t,\xi)\text{with}$$
$$\hat{u}(0,\xi) = \hat{f}(\xi) \text{ and } \hat{u}_t(0,\xi) = \hat{g}(\xi).$$

The solution to these equations is

$$\hat{u}(t,\xi) = \hat{f}(\xi)\cos(t|\xi|) + \hat{g}(\xi)\frac{\sin t|\xi|}{|\xi|}$$

and hence we should have

$$\begin{aligned} u(t,x) &= \mathcal{F}^{-1}\left(\hat{f}(\xi)\cos(t|\xi|) + \hat{g}(\xi)\frac{\sin t|\xi|}{|\xi|}\right)(x) \\ &= \mathcal{F}^{-1}\cos(t|\xi|) * f(x) + \mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} * g(x) \\ &= \frac{d}{dt}\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} * f(x) + \mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} * g(x). \end{aligned}$$

The question now is how interpret this equation. In particular what are the inverse Fourier transforms of $\mathcal{F}^{-1}\cos(t|\xi|)$ and $\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|}$. Since $\frac{d}{dt}\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} * f(x) = \mathcal{F}^{-1}\cos(t|\xi|) * f(x)$, it really suffices to understand $\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|}$. This was worked out in Example 45.51 when $n = 1$ where we found

$$\begin{aligned} \left(\mathcal{F}^{-1}\xi^{-1}\sin t\xi\right)(x) &= \frac{\pi}{\sqrt{2\pi}}\left(1_{x+t>0} - 1_{(x-t)>0}\right) \\ &= \frac{\pi}{\sqrt{2\pi}}\left(1_{x>-t} - 1_{x>t}\right) = \frac{\pi}{\sqrt{2\pi}}1_{[-t,t]}(x) \end{aligned}$$

where in writing the last line we have assume that $t \geq 0$. Therefore,

$$\left(\mathcal{F}^{-1}\xi^{-1}\sin t\xi\right) * f(x) = \frac{1}{2}\int_{-t}^{t} f(x-y)dy$$

Therefore the solution to the one dimensional wave equation is

$$\begin{aligned} u(t,x) &= \frac{d}{dt}\frac{1}{2}\int_{-t}^{t} f(x-y)dy + \frac{1}{2}\int_{-t}^{t} g(x-y)dy \\ &= \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \frac{1}{2}\int_{-t}^{t} g(x-y)dy \\ &= \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \frac{1}{2}\int_{x-t}^{x+t} g(y)dy. \end{aligned}$$

We can arrive at this same solution by more elementary means as follows. We first note in the one dimensional case that wave operator factors, namely

$$0 = \left(\partial_t^2 - \partial_x^2\right) u(t,x) = \left(\partial_t - \partial_x\right)\left(\partial_t + \partial_x\right) u(t,x).$$

Let $U(t,x) := \left(\partial_t + \partial_x\right) u(t,x)$, then the wave equation states $\left(\partial_t - \partial_x\right) U = 0$ and hence by the chain rule $\frac{d}{dt}U(t,x-t) = 0$. So

$$U(t,x-t) = U(0,x) = g(x) + f'(x)$$

and replacing x by $x+t$ in this equation shows

$$\left(\partial_t + \partial_x\right) u(t,x) = U(t,x) = g(x+t) + f'(x+t).$$

Working similarly, we learn that

$$\frac{d}{dt}u(t,x+t) = g(x+2t) + f'(x+2t)$$

which upon integration implies

$$\begin{aligned} u(t,x+t) &= u(0,x) + \int_0^t \left\{g(x+2\tau) + f'(x+2\tau)\right\} d\tau. \\ &= f(x) + \int_0^t g(x+2\tau)d\tau + \frac{1}{2}f(x+2\tau)|_0^t \\ &= \frac{1}{2}\left(f(x) + f(x+2t)\right) + \int_0^t g(x+2\tau)d\tau. \end{aligned}$$

Replacing $x \to x - t$ in this equation then implies

$$u(t,x) = \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \int_0^t g(x-t+2\tau)d\tau.$$

Finally, letting $y = x - t + 2\tau$ in the last integral gives

$$u(t,x) = \frac{1}{2}\left(f(x-t) + f(x+t)\right) + \frac{1}{2}\int_{x-t}^{x+t} g(y)dy$$

as derived using the Fourier transform.

For the three dimensional case we have

$$\begin{aligned} u(t,x) &= \frac{d}{dt}\mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} * f(x) + \mathcal{F}^{-1}\frac{\sin t|\xi|}{|\xi|} * g(x) \\ &= \frac{d}{dt}\left(t\bar{\sigma}_t * f(x)\right) + t\bar{\sigma}_t * g(x). \end{aligned}$$

The question is what is $\mu * g(x)$ where μ is a measure. To understand the definition, suppose first that $d\mu(x) = \rho(x)dx$, then we should have

$$\mu * g(x) = \rho * g(x) = \int_{\mathbb{R}^n} g(x-y)\rho(x)dx = \int_{\mathbb{R}^n} g(x-y)d\mu(y).$$

Thus we expect our solution to the wave equation should be given by

$$\begin{aligned} u(t,x) &= \frac{d}{dt}\left\{t\int_{S_t} f(x-y)d\bar{\sigma}_t(y)\right\} + t\int_{S_t} g(x-y)d\bar{\sigma}_t(y) \\ &= \frac{d}{dt}\left\{t\int_{S_1} f(x-t\omega)d\omega\right\} + t\int_{S_1} g(x-t\omega)d\omega \\ &= \frac{d}{dt}\left\{t\int_{S_1} f(x+t\omega)d\omega\right\} + t\int_{S_1} g(x+t\omega)d\omega \qquad (45.11) \end{aligned}$$

where $d\omega := d\bar{\sigma}_1(\omega)$. Notice the sharp propagation of speed. To understand this suppose that $f = 0$ for simplicity and g has compact support near the origin, for example think of $g = \delta_0(x)$, the $x + tw = 0$ for some w iff $|x| = t$. Hence the wave front propagates at unit speed in a sharp way. See figure below.

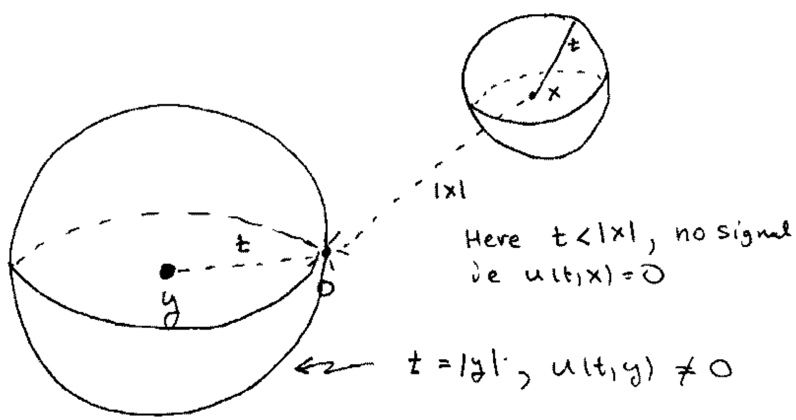


Fig. 45.2. The geometry of the solution to the wave equation in three dimensions.

We may also use this solution to solve the two dimensional wave equation using Hadamard's method of decent. Indeed, suppose now that f and g are function on $\mathbb{R}^2$ which we may view as functions on $\mathbb{R}^3$ which do not depend on the third coordinate say. We now go ahead and solve the three dimensional wave equation using Eq. (45.11) and f and g as initial conditions. It is easily seen

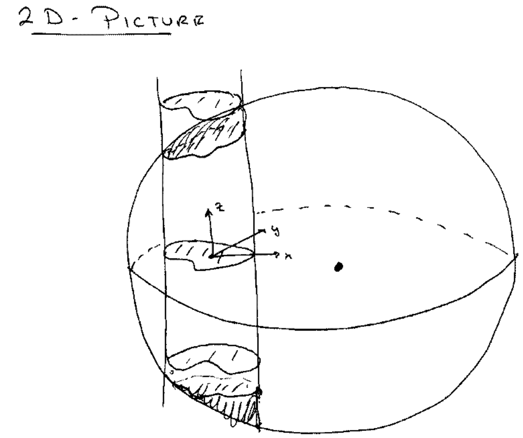


Fig. 45.3. The geometry of the solution to the wave equation in two dimensions.

that the solution $u(t,x,y,z)$ is again independent of z and hence is a solution to the two dimensional wave equation. See figure below.

Notice that we still have finite speed of propagation but no longer sharp propagation. In fact we can work out the solution analytically as follows. Again for simplicity assume that $f \equiv 0$. Then

$$\begin{aligned} u(t,x,y) &= \frac{t}{4\pi}\int_0^{2\pi} d\theta \int_0^{\pi} d\varphi \sin\varphi g((x,y) + t(\sin\varphi\cos\theta, \sin\varphi\sin\theta)) \\ &= \frac{t}{2\pi}\int_0^{2\pi} d\theta \int_0^{\pi/2} d\varphi \sin\varphi g((x,y) + t(\sin\varphi\cos\theta, \sin\varphi\sin\theta)) \end{aligned}$$

and letting $u = \sin\varphi$, so that $du = \cos\varphi d\varphi = \sqrt{1-u^2}d\varphi$ we find

$$u(t,x,y) = \frac{t}{2\pi}\int_0^{2\pi} d\theta \int_0^1 \frac{du}{\sqrt{1-u^2}} u g((x,y) + ut(\cos\theta, \sin\theta))$$

and then letting $r = ut$ we learn,

$$\begin{aligned} u(t,x,y) &= \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^t \frac{dr}{\sqrt{1-r^2/t^2}} \frac{r}{t} g((x,y) + r(\cos\theta, \sin\theta)) \\ &= \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^t \frac{dr}{\sqrt{t^2-r^2}} r g((x,y) + r(\cos\theta, \sin\theta)) \\ &= \frac{1}{2\pi}\iint_{D_t} \frac{g((x,y)+w))}{\sqrt{t^2-|w|^2}} dm(w). \end{aligned}$$

Here is a better alternative derivation of this result. We begin by using symmetry to find

$$u(t,x) = 2t\int_{S_t^+} g(x-y)d\bar{\sigma}_t(y) = 2t\int_{S_t^+} g(x+y)d\bar{\sigma}_t(y)$$

where S_t^+ is the portion of S_t with $z \geq 0$. This sphere is parametrized by $R(u,v) = (u, v, \sqrt{t^2-u^2-v^2})$ with $(u,v) \in D_t := \left\{(u,v) : u^2+v^2 \leq t^2\right\}$. In these coordinates we have

$$\begin{aligned} 4\pi t^2 d\bar{\sigma}_t &= \left|\left(-\partial_u\sqrt{t^2-u^2-v^2}, -\partial_v\sqrt{t^2-u^2-v^2}, 1\right)\right| dudv \\ &= \left|\left(\frac{u}{\sqrt{t^2-u^2-v^2}}, \frac{v}{\sqrt{t^2-u^2-v^2}}, 1\right)\right| dudv \\ &= \sqrt{\frac{u^2+v^2}{t^2-u^2-v^2}+1}dudv = \frac{|t|}{\sqrt{t^2-u^2-v^2}}dudv \end{aligned}$$

and therefore,

$$\begin{aligned} u(t,x) &= \frac{2t}{4\pi t^2}\int_{S_t^+} g(x+(u,v,\sqrt{t^2-u^2-v^2}))\frac{|t|}{\sqrt{t^2-u^2-v^2}}dudv \\ &= \frac{1}{2\pi}\operatorname{sgn}(t)\int_{S_t^+}\frac{g(x+(u,v))}{\sqrt{t^2-u^2-v^2}}dudv. \end{aligned}$$

This may be written as

$$u(t,x) = \frac{1}{2\pi}\operatorname{sgn}(t)\iint_{D_t}\frac{g((x,y)+w))}{\sqrt{t^2-|w|^2}}dm(w)$$

as before. (I should check on the $\operatorname{sgn}(t)$ term.)

45.7 Appendix: Topology on $C_c^\infty(U)$

Let U be an open subset of $\mathbb{R}^n$ and

$$C_c^\infty(U) = \cup_{K\sqsubset\sqsubset U} C^\infty(K) \tag{45.12}$$

denote the set of smooth functions on U with compact support in U. Our goal is to topologize $C_c^\infty(U)$ in a way which is compatible with he topologies defined in Example 45.24 above. This leads us to the inductive limit topology which we now pause to introduce.

Definition 45.53 (Indcutive Limit Topology). *Let X be a set, $X_\alpha \subset X$ for $\alpha \in A$ (A is an index set) and assume that $\tau_\alpha \subset 2^{X_\alpha}$ is a topology on X_α for each α. Let $i_\alpha : X_\alpha \to X$ denote the inclusion maps. The inductive limit topology on X is the largest topology τ on X such that i_α is continuous for all $\alpha \in A$. That is to say, $\tau = \cap_{\alpha\in A} i_{\alpha *}(\tau_\alpha)$, i.e. a set $U \subset X$ is open ($U \in \tau$) iff $i_\alpha^{-1}(A) = A \cap X_\alpha \in \tau_\alpha$ for all $\alpha \in A$.*

Notice that $C \subset X$ is closed iff $C \cap X_\alpha$ is closed in X_α for all α. Indeed, $C \subset X$ is closed iff $C^c = X \setminus C \subset X$ is open, iff $C^c \cap X_\alpha = X_\alpha \setminus C$ is open in X_α iff $X_\alpha \cap C = X_\alpha \setminus (X_\alpha \setminus C)$ is closed in X_α for all $\alpha \in A$.

Definition 45.54. *Let $\mathcal{D}(U)$ denote $C_c^\infty(U)$ equipped with the inductive limit topology arising from writing $C_c^\infty(U)$ as in Eq. (45.12) and using the Frechét topologies on $C^\infty(K)$ as defined in Example 45.24.*

For each $K \sqsubset\sqsubset U$, $C^\infty(K)$ is a closed subset of $\mathcal{D}(U)$. Indeed if F is another compact subset of U, then $C^\infty(K) \cap C^\infty(F) = C^\infty(K \cap F)$, which is a closed subset of $C^\infty(F)$. The set $\mathcal{U} \subset \mathcal{D}(U)$ defined by

$$\mathcal{U} = \left\{\psi \in \mathcal{D}(U) : \sum_{|\alpha|\leq m}\|\partial^\alpha(\psi-\varphi)\|_\infty < \varepsilon\right\} \tag{45.13}$$

for some $\varphi \in \mathcal{D}(U)$ and $\varepsilon > 0$ is an open subset of $\mathcal{D}(U)$. Indeed, if $K \sqsubset\sqsubset U$, then

$$\mathcal{U} \cap C^\infty(K) = \left\{\psi \in C^\infty(K) : \sum_{|\alpha|\leq m}\|\partial^\alpha(\psi-\varphi)\|_\infty < \varepsilon\right\}$$

is easily seen to be open in $C^\infty(K)$.

Proposition 45.55. *Let (X,τ) be as described in Definition 45.53 and $f : X \to Y$ be a function where Y is another topological space. Then f is continuous iff $f \circ i_\alpha : X_\alpha \to Y$ is continuous for all $\alpha \in A$.*

Proof. Since the composition of continuous maps is continuous, it follows that $f \circ i_\alpha : X_\alpha \to Y$ is continuous for all $\alpha \in A$ if $f : X \to Y$ is continuous. Conversely, if $f \circ i_\alpha$ is continuous for all $\alpha \in A$, then for all $V \subset_o Y$ we have

$$\tau_\alpha \ni (f\circ i_\alpha)^{-1}(V) = i_\alpha^{-1}(f^{-1}(V)) = f^{-1}(V) \cap X_\alpha \text{ for all } \alpha \in A$$

showing that $f^{-1}(V) \in \tau$. ■

Lemma 45.56. *Let us continue the notation introduced in Definition 45.53. Suppose further that there exists $\alpha_k \in A$ such that $X_k' := X_{\alpha_k} \uparrow X$ as $k \to \infty$ and for each $\alpha \in A$ there exists an $k \in \mathbb{N}$ such that $X_\alpha \subset X_k'$ and the inclusion* *map is continuous. Then* $\tau = \{A \subset X : A \cap X'_k \subset_o X'_k \text{ for all } k\}$ *and a function* $f : X \to Y$ *is continuous iff* $f|_{X'_k} : X'_k \to Y$ *is continuous for all* k. *In short the inductive limit topology on* X *arising from the two collections of subsets* $\{X_\alpha\}_{\alpha\in A}$ *and* $\{X'_k\}_{k\in\mathbb{N}}$ *are the same.*

Proof. Suppose that $A \subset X$, if $A \in \tau$ then $A \cap X'_k = A \cap X_{\alpha_k} \subset_o X'_k$ by definition. Now suppose that $A \cap X'_k \subset_o X'_k$ for all k. For $\alpha \in A$ choose k such that $X_\alpha \subset X'_k$, then $A \cap X_\alpha = (A \cap X'_k) \cap X_\alpha \subset_o X_\alpha$ since $A \cap X'_k$ is open in X'_k and by assumption that X_α is continuously embedded in X'_k, $V \cap X_\alpha \subset_o X_\alpha$ for all $V \subset_o X'_k$. The characterization of continuous functions is prove similarly. ■

Let $K_k \sqsubset\sqsubset U$ for $k \in \mathbb{N}$ such that $K_k^o \subset K_k \subset K_{k+1}^o \subset K_{k+1}$ for all k and $K_k \uparrow U$ as $k \to \infty$. Then it follows for any $K \sqsubset\sqsubset U$, there exists an k such that $K \subset K_k^o \subset K_k$. One now checks that the map $C^\infty(K)$ embeds continuously into $C^\infty(K_k)$ and moreover, $C^\infty(K)$ is a closed subset of $C^\infty(K_{k+1})$. Therefore we may describe $\mathcal{D}(U)$ as $C_c^\infty(U)$ with the inductively limit topology coming from $\cup_{k\in\mathbb{N}} C^\infty(K_k)$.

Lemma 45.57. *Suppose that* $\{\varphi_k\}_{k=1}^\infty \subset \mathcal{D}(U)$, *then* $\varphi_k \to \varphi \in \mathcal{D}(U)$ *iff* $\varphi_k - \varphi \to 0 \in \mathcal{D}(U)$.

Proof. Let $\varphi \in \mathcal{D}(U)$ and $\mathcal{U} \subset \mathcal{D}(U)$ be a set. We will begin by showing that $\mathcal{U}$ is open in $\mathcal{D}(U)$ iff $\mathcal{U} - \varphi$ is open in $\mathcal{D}(U)$. To this end let K_k be the compact sets described above and choose k_0 sufficiently large so that $\varphi \in C^\infty(K_k)$ for all $k \ge k_0$. Now $\mathcal{U} - \varphi \subset \mathcal{D}(U)$ is open iff $(\mathcal{U} - \varphi) \cap C^\infty(K_k)$ is open in $C^\infty(K_k)$ for all $k \ge k_0$. Because $\varphi \in C^\infty(K_k)$, we have $(\mathcal{U} - \varphi) \cap C^\infty(K_k) = \mathcal{U} \cap C^\infty(K_k) - \varphi$ which is open in $C^\infty(K_k)$ iff $\mathcal{U} \cap C^\infty(K_k)$ is open $C^\infty(K_k)$. Since this is true for all $k \ge k_0$ we conclude that $\mathcal{U} - \varphi$ is an open subset of $\mathcal{D}(U)$ iff $\mathcal{U}$ is open in $\mathcal{D}(U)$. Now $\varphi_k \to \varphi$ in $\mathcal{D}(U)$ iff for all $\varphi \in \mathcal{U} \subset_o \mathcal{D}(U)$, $\varphi_k \in \mathcal{U}$ for almost all k which happens iff $\varphi_k - \varphi \in \mathcal{U} - \varphi$ for almost all k. Since $\mathcal{U} - \varphi$ ranges over all open neighborhoods of 0 when $\mathcal{U}$ ranges over the open neighborhoods of φ, the result follows. ■

Lemma 45.58. *A sequence* $\{\varphi_k\}_{k=1}^\infty \subset \mathcal{D}(U)$ *converges to* $\varphi \in \mathcal{D}(U)$, *iff there is a compact set* $K \sqsubset\sqsubset U$ *such that* $\operatorname{supp}(\varphi_k) \subset K$ *for all* k *and* $\varphi_k \to \varphi$ *in* $C^\infty(K)$.

Proof. If $\varphi_k \to \varphi$ in $C^\infty(K)$, then for any open set $\mathcal{V} \subset \mathcal{D}(U)$ with $\varphi \in \mathcal{V}$ we have $\mathcal{V} \cap C^\infty(K)$ is open in $C^\infty(K)$ and hence $\varphi_k \in \mathcal{V} \cap C^\infty(K) \subset \mathcal{V}$ for almost all k. This shows that $\varphi_k \to \varphi \in \mathcal{D}(U)$. For the converse, suppose that there exists $\{\varphi_k\}_{k=1}^\infty \subset \mathcal{D}(U)$ which converges to $\varphi \in \mathcal{D}(U)$ yet there is no compact set K such that $\operatorname{supp}(\varphi_k) \subset K$ for all k. Using Lemma45.57, we may replace φ_k by $\varphi_k - \varphi$ if necessary so that we may assume $\varphi_k \to 0$ in $\mathcal{D}(U)$. By passing to a subsequences of $\{\varphi_k\}$ and $\{K_k\}$ if necessary, we may also assume there $x_k \in K_{k+1} \setminus K_k$ such that $\varphi_k(x_k) \neq 0$ for all k. Let p denote the semi-norm on $C_c^\infty(U)$ defined by

$$p(\varphi) = \sum_{k=0}^\infty \sup\left\{\frac{|\varphi(x)|}{|\varphi_k(x_k)|} : x \in K_{k+1} \setminus K_k^o\right\}.$$

One then checks that

$$p(\varphi) \le \left(\sum_{k=0}^N \frac{1}{|\varphi_k(x_k)|}\right)\|\varphi\|_\infty$$

for $\varphi \in C^\infty(K_{N+1})$. This shows that $p|_{C^\infty(K_{N+1})}$ is continuous for all N and hence p is continuous on $\mathcal{D}(U)$. Since p is continuous on $\mathcal{D}(U)$ and $\varphi_k \to 0$ in $\mathcal{D}(U)$, it follows that $\lim_{k\to\infty} p(\varphi_k) = p(\lim_{k\to\infty}\varphi_k) = p(0) = 0$. While on the other hand, $p(\varphi_k) \ge 1$ by construction and hence we have arrived at a contradiction. Thus for any convergent sequence $\{\varphi_k\}_{k=1}^\infty \subset \mathcal{D}(U)$ there is a compact set $K \sqsubset\sqsubset U$ such that $\operatorname{supp}(\varphi_k) \subset K$ for all k. We will now show that $\{\varphi_k\}_{k=1}^\infty$ is convergent to φ in $C^\infty(K)$. To this end let $\mathcal{U} \subset \mathcal{D}(U)$ be the open set described in Eq. (45.13), then $\varphi_k \in \mathcal{U}$ for almost all k and in particular, $\varphi_k \in \mathcal{U} \cap C^\infty(K)$ for almost all k. (Letting $\varepsilon > 0$ tend to zero shows that $\operatorname{supp}(\varphi) \subset K$, i.e. $\varphi \in C^\infty(K)$.) Since sets of the form $\mathcal{U} \cap C^\infty(K)$ with $\mathcal{U}$ as in Eq. (45.13) form a neighborhood base for the $C^\infty(K)$ at φ, we concluded that $\varphi_k \to \varphi$ in $C^\infty(K)$. ■

Definition 45.59 (Distributions on $U \subset_o \mathbb{R}^n$). *A generalized function on* $U \subset_o \mathbb{R}^n$ *is a continuous linear functional on* $\mathcal{D}(U)$. *We denote the space of generalized functions by* $\mathcal{D}'(U)$.

Proposition 45.60. *Let* $f : \mathcal{D}(U) \to \mathbb{C}$ *be a linear functional. Then the following are equivalent.*

1. *f is continuous, i.e.* $f \in \mathcal{D}'(U)$.
2. *For all* $K \sqsubset\sqsubset U$, *there exist* $n \in \mathbb{N}$ *and* $C < \infty$ *such that*
$$|f(\varphi)| \le C p_n(\varphi) \text{ for all } \varphi \in C^\infty(K). \tag{45.14}$$
3. *For all sequences* $\{\varphi_k\} \subset \mathcal{D}(U)$ *such that* $\varphi_k \to 0$ *in* $\mathcal{D}(U)$, $\lim_{k\to\infty} f(\varphi_k) = 0$.

Proof. 1) $\iff$ 2). If f is continuous, then by definition of the inductive limit topology $f|_{C^\infty(K)}$ is continuous. Hence an estimate of the type in Eq. (45.14) must hold. Conversely if estimates of the type in Eq. (45.14) hold for all compact sets K, then $f|_{C^\infty(K)}$ is continuous for all $K \sqsubset\sqsubset U$ and again by the definition of the inductive limit topologies, f is continuous on $\mathcal{D}'(U)$. 1) $\iff$ 3) By Lemma 45.58, the assertion in item 3. is equivalent to saying that $f|_{C^\infty(K)}$ is sequentially continuous for all $K \sqsubset\sqsubset U$. Since the topology on $C^\infty(K)$ is first countable (being a metric topology), sequential continuity and continuity are the same think. Hence item 3. is equivalent to the assertion that $f|_{C^\infty(K)}$ is continuous for all $K \sqsubset\sqsubset U$ which is equivalent to the assertion that f is continuous on $\mathcal{D}'(U)$. ■

Proposition 45.61. *The maps* $(\lambda, \varphi) \in \mathbb{C} \times \mathcal{D}(U) \to \lambda\varphi \in \mathcal{D}(U)$ *and* $(\varphi, \psi) \in \mathcal{D}(U) \times \mathcal{D}(U) \to \varphi + \psi \in \mathcal{D}(U)$ *are continuous. (Actually, I will have to look up how to decide to this.) What is obvious is that all of these operations are sequentially continuous, which is enough for our purposes.*

Convolutions involving distributions

46.1 Tensor Product of Distributions

Let $X\subset_o\mathbb{R}^n$ and $Y\subset_o\mathbb{R}^m$ and $S\in\mathcal{D}'(X)$ and $T\in\mathcal{D}'(Y)$. We wish to define $S\otimes T\in\mathcal{D}'(X\times Y)$. Informally, we should have

$$\begin{aligned}\langle S\otimes T,\varphi\rangle &= \int_{X\times Y} S(x)T(y)\varphi(x,y)dxdy\\ &=\int_X dxS(x)\int_Y dyT(y)\varphi(x,y)=\int_Y dyT(y)\int_X dxS(x)\varphi(x,y).\end{aligned}$$

Of course we should interpret this last equation as follows,

$$\langle S\otimes T,\varphi\rangle=\langle S(x),\langle T(y),\varphi(x,y)\rangle\rangle=\langle T(y),\langle S(x),\varphi(x,y)\rangle\rangle. \tag{46.1}$$

This formula takes on particularly simple form when $\varphi=u\otimes v$ with $u\in\mathcal{D}(X)$ and $v\in\mathcal{D}(Y)$ in which case

$$\langle S\otimes T,u\otimes v\rangle=\langle S,u\rangle\langle T,v\rangle. \tag{46.2}$$

We begin with the following smooth version of the Weierstrass approximation theorem which will be used to show Eq. (46.2) uniquely determines $S\otimes T$.

Theorem 46.1 (Density Theorem). *Suppose that $X\subset_o\mathbb{R}^n$ and $Y\subset_o\mathbb{R}^m$, then $\mathcal{D}(X)\otimes\mathcal{D}(Y)$ is dense in $\mathcal{D}(X\times Y)$.*

Proof. First let us consider the special case where $X=(0,1)^n$ and $Y=(0,1)^m$ so that $X\times Y=(0,1)^{m+n}$. To simplify notation, let $m+n=k$ and $\Omega=(0,1)^k$ and $\pi_i:\Omega\to(0,1)$ be projection onto the ith factor of Ω. Suppose that $\varphi\in C_c^\infty(\Omega)$ and $K=\operatorname{supp}(\varphi)$. We will view $\varphi\in C_c^\infty(\mathbb{R}^k)$ by setting $\varphi=0$ outside of Ω. Since K is compact $\pi_i(K)\subset[a_i,b_i]$ for some $0<a_i<b_i<1$. Let $a=\min\{a_i:i=1,\dots,k\}$ and $b=\max\{b_i:i=1,\dots,k\}$. Then $\operatorname{supp}(\varphi)=K\subset[a,b]^k\subset\Omega$. As in the proof of the Weierstrass approximation theorem, let $q_n(t)=c_n(1-t^2)^n1_{|t|\le1}$ where c_n is chosen so that $\int_{\mathbb{R}}q_n(t)dt=1$. Also set $Q_n=q_n\otimes\cdots\otimes q_n$, i.e. $Q_n(x)=\prod_{i=1}^k q_n(x_i)$ for $x\in\mathbb{R}^k$. Let

$$f_n(x):=Q_n*\varphi(x)=c_n^k\int_{\mathbb{R}^k}\varphi(y)\prod_{i=1}^k(1-(x_i-y_i)^2)^n1_{|x_i-y_i|\le1}dy_i. \tag{46.3}$$

By standard arguments, we know that $\partial^\alpha f_n\to\partial^\alpha\varphi$ uniformly on $\mathbb{R}^k$ as $n\to\infty$. Moreover for $x\in\Omega$, it follows from Eq. (46.3) that

$$f_n(x):=c_n^k\int_\Omega\varphi(y)\prod_{i=1}^k(1-(x_i-y_i)^2)^n dy_i=p_n(x)$$

where $p_n(x)$ is a polynomial in x. Notice that $p_n\in C^\infty((0,1))\otimes\cdots\otimes C^\infty((0,1))$ so that we are almost there.[1] We need only cutoff these functions so that they have compact support. To this end, let $\theta\in C_c^\infty((0,1))$ be a function such that $\theta=1$ on a neighborhood of $[a,b]$ and define

$$\begin{aligned}\varphi_n&=(\theta\otimes\cdots\otimes\theta)\,f_n\\ &=(\theta\otimes\cdots\otimes\theta)\,p_n\in C_c^\infty((0,1))\otimes\cdots\otimes C_c^\infty((0,1)).\end{aligned}$$

I claim now that $\varphi_n\to\varphi$ in $\mathcal{D}(\Omega)$. Certainly by construction $\operatorname{supp}(\varphi_n)\subset[a,b]^k\sqsubset\sqsubset\Omega$ for all n. Also

$$\begin{aligned}\partial^\alpha(\varphi-\varphi_n)&=\partial^\alpha(\varphi-(\theta\otimes\cdots\otimes\theta)\,f_n)\\ &=(\theta\otimes\cdots\otimes\theta)\,(\partial^\alpha\varphi-\partial^\alpha f_n)+R_n\end{aligned} \tag{46.4}$$

where R_n is a sum of terms of the form $\partial^\beta(\theta\otimes\cdots\otimes\theta)\cdot\partial^\gamma f_n$ with $\beta\ne0$. Since $\partial^\beta(\theta\otimes\cdots\otimes\theta)=0$ on $[a,b]^k$ and $\partial^\gamma f_n$ converges uniformly to zero on $\mathbb{R}^k\setminus[a,b]^k$, it follows that $R_n\to0$ uniformly as $n\to\infty$. Combining this with Eq. (46.4) and the fact that $\partial^\alpha f_n\to\partial^\alpha\varphi$ uniformly on $\mathbb{R}^k$ as $n\to\infty$, we see that $\varphi_n\to\varphi$ in $\mathcal{D}(\Omega)$. This finishes the proof in the case $X=(0,1)^n$ and $Y=(0,1)^m$. For the general case, let $K=\operatorname{supp}(\varphi)\sqsubset\sqsubset X\times Y$ and

[1] One could also construct $f_n\in C^\infty(\mathbb{R})^{\otimes k}$ such that $\partial^\alpha f_n\to\partial^\alpha f$ uniformlly as $n\to\infty$ using Fourier series. To this end, let $\tilde\phi$ be the 1 – periodic extension of ϕ to $\mathbb{R}^k$. Then $\tilde\phi\in C^\infty_{\text{periodic}}(\mathbb{R}^k)$ and hence it may be written as

$$\tilde\phi(x)=\sum_{m\in\mathbb{Z}^k}c_me^{i2\pi m\cdot x}$$

where the $\{c_m:m\in\mathbb{Z}^k\}$ are the Fourier coefficients of $\tilde\phi$ which decay faster that $(1+|m|)^{-l}$ for any $l>0$. Thus $f_n(x):=\sum_{m\in\mathbb{Z}^k:|m|\le n}c_me^{i2\pi m\cdot x}\in C^\infty(\mathbb{R})^{\otimes k}$ and $\partial^\alpha f_n\to\partial^\alpha\phi$ unifromly on Ω as $n\to\infty$.

$K_1 = \pi_1(K) \sqsubset\sqsubset X$ and $K_2 = \pi_2(K) \sqsubset\sqsubset Y$ where π_1 and π_2 are projections from $X \times Y$ to X and Y respectively. Then $K \sqsubset K_1 \times K_2 \sqsubset\sqsubset X \times Y$. Let $\{V_i\}_{i=1}^a$ and $\{U_j\}_{j=1}^b$ be finite covers of K_1 and K_2 respectively by open sets $V_i = (a_i, b_i)$ and $U_j = (c_j, d_j)$ with $a_i, b_i \in X$ and $c_j, d_j \in Y$. Also let $\alpha_i \in C_c^\infty(V_i)$ for $i = 1, \ldots, a$ and $\beta_j \in C_c^\infty(U_j)$ for $j = 1, \ldots, b$ be functions such that $\sum_{i=1}^a \alpha_i = 1$ on a neighborhood of K_1 and $\sum_{j=1}^b \beta_j = 1$ on a neighborhood of K_2. Then $\varphi = \sum_{i=1}^a \sum_{j=1}^b (\alpha_i \otimes \beta_j)\varphi$ and by what we have just proved (after scaling and translating) each term in this sum, $(\alpha_i \otimes \beta_j)\varphi$, may be written as a limit of elements in $\mathcal{D}(X) \otimes \mathcal{D}(Y)$ in the $\mathcal{D}(X \times Y)$ topology. ∎

Theorem 46.2 (Distribution-Fubini-Theorem). *Let $S \in \mathcal{D}'(X)$, $T \in \mathcal{D}'(Y)$, $h(x) := \langle T(y), \varphi(x,y)\rangle$ and $g(y) := \langle S(x), \varphi(x,y)\rangle$. Then $h = h_\varphi \in \mathcal{D}(X)$, $g = g_\varphi \in \mathcal{D}(Y)$, $\partial^\alpha h(x) = \langle T(y), \partial_x^\alpha \varphi(x,y)\rangle$ and $\partial^\beta g(y) = \langle S(x), \partial_y^\beta \varphi(x,y)\rangle$ for all multi-indices α and β. Moreover*

$$\langle S(x), \langle T(y), \varphi(x,y)\rangle\rangle = \langle S, h\rangle = \langle T, g\rangle = \langle T(y), \langle S(x), \varphi(x,y)\rangle\rangle. \tag{46.5}$$

We denote this common value by $\langle S \otimes T, \varphi\rangle$ and call $S \otimes T$ the tensor product of S and T. This distribution is uniquely determined by its values on $\mathcal{D}(X) \otimes \mathcal{D}(Y)$ and for $u \in \mathcal{D}(X)$ and $v \in \mathcal{D}(Y)$ we have

$$\langle S \otimes T, u \otimes v\rangle = \langle S, u\rangle\langle T, v\rangle.$$

Proof. Let $K = \operatorname{supp}(\varphi) \sqsubset\sqsubset X \times Y$ and $K_1 = \pi_1(K)$ and $K_2 = \pi_2(K)$. Then $K_1 \sqsubset\sqsubset X$ and $K_2 \sqsubset\sqsubset Y$ and $K \subset K_1 \times K_2 \subset X \times Y$. If $x \in X$ and $y \notin K_2$, then $\varphi(x,y) = 0$ and more generally $\partial_x^\alpha \varphi(x,y) = 0$ so that $\{y : \partial_x^\alpha \varphi(x,y) \neq 0\} \subset K_2$. Thus for all $x \in X$, $\operatorname{supp}(\partial^\alpha \varphi(x,\cdot)) \subset K_2 \subset Y$. By the fundamental theorem of calculus,

$$\partial_y^\beta \varphi(x+v, y) - \partial_y^\beta \varphi(x,y) = \int_0^1 \partial_v^x \partial_y^\beta \varphi(x + \tau v, y) d\tau \tag{46.6}$$

and therefore

$$\begin{aligned}\left\|\partial_y^\beta \varphi(x+v,\cdot) - \partial_y^\beta \varphi(x,\cdot)\right\|_\infty &\le |v| \int_0^1 \left\|\nabla_x \partial_y^\beta \varphi(x+\tau v,\cdot)\right\|_\infty d\tau \\ &\le |v| \left\|\nabla_x \partial_y^\beta \varphi\right\|_\infty \to 0 \text{ as } \nu \to 0.\end{aligned}$$

This shows that $x \in X \to \varphi(x,\cdot) \in \mathcal{D}(Y)$ is continuous. Thus h is continuous being the composition of continuous functions. Letting $v = te_i$ in Eq. (46.6) we find

$$\begin{aligned}&\frac{\partial_y^\beta \varphi(x+te_i, y) - \partial_y^\beta \varphi(x,y)}{t} - \frac{\partial}{\partial x_i}\partial_y^\beta \varphi(x,y) \\ &\quad = \int_0^1 \left[\frac{\partial}{\partial x_i}\partial_y^\beta \varphi(x+\tau te_i, y) - \frac{\partial}{\partial x_i}\partial_y^\beta \varphi(x,y)\right] d\tau\end{aligned}$$

and hence

$$\begin{aligned}&\left\|\frac{\partial_y^\beta \varphi(x+te_i,\cdot) - \partial_y^\beta \varphi(x,\cdot)}{t} - \frac{\partial}{\partial x_i}\partial_y^\beta \varphi(x,\cdot)\right\|_\infty \\ &\quad \le \int_0^1 \left\|\frac{\partial}{\partial x_i}\partial_y^\beta \varphi(x+\tau te_i,\cdot) - \frac{\partial}{\partial x_i}\partial_y^\beta \varphi(x,\cdot)\right\|_\infty d\tau\end{aligned}$$

which tends to zero as $t \to 0$. Thus we have checked that

$$\frac{\partial}{\partial x_i}\varphi(x,\cdot) = \mathcal{D}'(Y)\text{–}\lim_{t\to 0}\frac{\varphi(x+te_i,\cdot) - \varphi(x,\cdot)}{t}$$

and therefore,

$$\frac{h(x+te_i) - h(x)}{t} = \langle T, \frac{\varphi(x+te_i,\cdot) - \varphi(x,\cdot)}{t}\rangle \to \langle T, \frac{\partial}{\partial x_i}\varphi(x,\cdot)\rangle$$

as $t \to 0$ showing $\partial_i h(x)$ exists and is given by $\langle T, \frac{\partial}{\partial x_i}\varphi(x,\cdot)\rangle$. By what we have proved above, it follows that $\partial_i h(x) = \langle T, \frac{\partial}{\partial x_i}\varphi(x,\cdot)\rangle$ is continuous in x. By induction on $|\alpha|$, it follows that $\partial^\alpha h(x)$ exists and is continuous and $\partial^\alpha h(x) = \langle T(y), \partial_x^\alpha \varphi(x,y)\rangle$ for all α. Now if $x \notin K_1$, then $\varphi(x,\cdot) \equiv 0$ showing that $\{x \in X : h(x) \neq 0\} \subset K_1$ and hence $\operatorname{supp}(h) \subset K_1 \sqsubset\sqsubset X$. Thus h has compact support. This proves all of the assertions made about h. The assertions pertaining to the function g are prove analogously. Let $\langle \Gamma, \varphi\rangle = \langle S(x), \langle T(y), \varphi(x,y)\rangle\rangle = \langle S, h_\varphi\rangle$ for $\varphi \in \mathcal{D}(X \times Y)$. Then Γ is clearly linear and we have

$$\begin{aligned}|\langle \Gamma, \varphi\rangle| &= |\langle S, h_\varphi\rangle| \\ &\le C \sum_{|\alpha|\le m} \|\partial_x^\alpha h_\varphi\|_{\infty,K_1} = C \sum_{|\alpha|\le m} \|\langle T(y), \partial_x^\alpha \varphi(\cdot,y)\rangle\|_{\infty,K_1}\end{aligned}$$

which combined with the estimate

$$|\langle T(y), \partial_x^\alpha \varphi(x,y)\rangle| \le C \sum_{|\beta|\le p} \left\|\partial_y^\beta \partial_x^\alpha \varphi(x,y)\rangle\right\|_{\infty,K_2}$$

shows

$$|\langle \Gamma, \varphi\rangle| \le C \sum_{|\alpha|\le m}\sum_{|\beta|\le p} \left\|\partial_y^\beta \partial_x^\alpha \varphi(x,y)\rangle\right\|_{\infty,K_1\times K_2}.$$

So Γ is continuous, i.e. $\Gamma \in \mathcal{D}'(X \times Y)$, i.e.

$$\varphi \in \mathcal{D}(X \times Y) \to \langle S(x), \langle T(y), \varphi(x,y)\rangle\rangle$$

defines a distribution. Similarly,

$$\varphi \in \mathcal{D}(X \times Y) \to \langle T(y), \langle S(x), \varphi(x,y)\rangle\rangle$$

also defines a distribution and since both of these distributions agree on the dense subspace $\mathcal{D}(X) \otimes \mathcal{D}(Y)$, it follows they are equal. ■

Theorem 46.3. *If (T, φ) is a distribution test function pair satisfying one of the following three conditions*

1. $T \in \mathcal{E}'(\mathbb{R}^n)$ *and* $\varphi \in C^\infty(\mathbb{R}^n)$
2. $T \in \mathcal{D}'(\mathbb{R}^n)$ *and* $\varphi \in \mathcal{D}(\mathbb{R}^n)$ *or*
3. $T \in \mathcal{S}'(\mathbb{R}^n)$ *and* $\varphi \in \mathcal{S}(\mathbb{R}^n)$,

let

$$T * \varphi(x) = \text{“}\int T(y)\varphi(x-y)dy\text{”} = \langle T, \varphi(x-\cdot)\rangle. \tag{46.7}$$

Then $T * \varphi \in C^\infty(\mathbb{R}^n)$, $\partial^\alpha(T * \varphi) = (\partial^\alpha T * \varphi) = (T * \partial^\alpha \varphi)$ for all α and $\operatorname{supp}(T * \varphi) \subset \overline{\operatorname{supp}(T) + \operatorname{supp}(\varphi)}$. *Moreover if (3) holds then $T * \varphi \in \mathcal{P}$ – the space of smooth functions with slow decrease.*

Proof. I will supply the proof for case (3) since the other cases are similar and easier. Let $h(x) := T * \varphi(x)$. Since $T \in \mathcal{S}'(\mathbb{R}^n)$, there exists $m \in \mathbb{N}$ and $C < \infty$ such that $|\langle T, \varphi\rangle| \le C p_m(\varphi)$ for all $\varphi \in \mathcal{S}$, where p_m is defined in Example 45.28. Therefore,

$$\begin{aligned}|h(x) - h(y)| &= |\langle T, \varphi(x-\cdot) - \varphi(y-\cdot)\rangle| \le C p_m(\varphi(x-\cdot) - \varphi(y-\cdot)) \\ &= C \sum_{|\alpha| \le m} \|\mu_m(\partial^\alpha \varphi(x-\cdot) - \partial^\alpha \varphi(y-\cdot))\|_\infty .\end{aligned}$$

Let $\psi := \partial^\alpha \varphi$, then

$$\psi(x-z) - \psi(y-z) = \int_0^1 \nabla\psi(y + \tau(x-y) - z) \cdot (x-y) d\tau \tag{46.8}$$

and hence

$$\begin{aligned}|\psi(x-z) - \psi(y-z)| &\le |x-y| \cdot \int_0^1 |\nabla\psi(y + \tau(x-y) - z)| \, d\tau \\ &\le C\,|x-y| \int_0^1 \mu_{-M}(y + \tau(x-y) - z) d\tau\end{aligned}$$

for any $M < \infty$. By Peetre's inequality,

$$\mu_{-M}(y + \tau(x-y) - z) \le \mu_{-M}(z)\mu_M(y + \tau(x-y))$$

so that

$$\begin{aligned}|\partial^\alpha\varphi(x-z) - \partial^\alpha\varphi(y-z)| &\le C\,|x-y|\,\mu_{-M}(z) \int_0^1 \mu_M(y + \tau(x-y)) d\tau \\ &\le C(x,y)\,|x-y|\,\mu_{-M}(z)\end{aligned} \tag{46.9}$$

where $C(x,y)$ is a continuous function of (x,y). Putting all of this together we see that

$$|h(x) - h(y)| \le \tilde{C}(x,y)\,|x-y| \to 0 \text{ as } x \to y,$$

showing h is continuous. Let us now compute a partial derivative of h. Suppose that $v \in \mathbb{R}^n$ is a fixed vector, then by Eq. (46.8),

$$\begin{aligned}&\frac{\varphi(x+tv-z) - \varphi(x-z)}{t} - \partial_v\varphi(x-z) \\ &\quad = \int_0^1 \nabla\varphi(x + \tau t v - z) \cdot v d\tau - \partial_v\varphi(x-z) \\ &\quad = \int_0^1 \left[\partial_v\varphi(x + \tau t v - z) - \partial_v\varphi(x-z)\right] d\tau.\end{aligned}$$

This then implies

$$\begin{aligned}&\left|\partial_z^\alpha \left\{\frac{\varphi(x+tv-z) - \varphi(x-z)}{t} - \partial_v\varphi(x-z)\right\}\right| \\ &\quad = \left|\int_0^1 \partial_z^\alpha \left[\partial_v\varphi(x + \tau t v - z) - \partial_v\varphi(x-z)\right] d\tau\right| \\ &\quad \le \int_0^1 \left|\partial_z^\alpha \left[\partial_v\varphi(x + \tau t v - z) - \partial_v\varphi(x-z)\right]\right| d\tau.\end{aligned}$$

But by the same argument as above, it follows that

$$\left|\partial_z^\alpha \left[\partial_v\varphi(x + \tau t v - z) - \partial_v\varphi(x-z)\right]\right| \le C(x + \tau t v, x)\,|\tau t v|\,\mu_{-M}(z)$$

and thus

$$\begin{aligned}&\left|\partial_z^\alpha \left\{\frac{\varphi(x+tv-z) - \varphi(x-z)}{t} - \partial_v\varphi(x-z)\right\}\right| \\ &\quad \le t\mu_{-M}(z) \int_0^1 C(x + \tau t v, x)\tau d\tau\,|v|\,\mu_{-M}(z).\end{aligned}$$

Putting this all together shows

$$\begin{aligned}\left\|\mu_M \partial_z^\alpha \left\{\frac{\varphi(x+tv-z) - \varphi(x-z)}{t} - \partial_v\varphi(x-z)\right\}\right\|_\infty &= O(t) \\ &\to 0 \text{ as } t \to 0.\end{aligned}$$

That is to say $\frac{\varphi(x+tv-\cdot)-\varphi(x-\cdot)}{t} \to \partial_v\varphi(x-\cdot)$ in $\mathcal{S}$ as $t\to 0$. Hence since T is continuous on $\mathcal{S}$, we learn

$$\begin{aligned}\partial_v\left(T*\varphi\right)(x) &= \partial_v\langle T,\varphi(x-\cdot)\rangle = \lim_{t\to 0}\langle T,\frac{\varphi(x+tv-\cdot)-\varphi(x-\cdot)}{t}\rangle \\ &= \langle T,\partial_v\varphi(x-\cdot)\rangle = T*\partial_v\varphi(x).\end{aligned}$$

By the first part of the proof, we know that $\partial_v(T*\varphi)$ is continuous and hence by induction it now follows that $T*\varphi$ is C^∞ and $\partial^\alpha T*\varphi = T*\partial^\alpha\varphi$. Since

$$\begin{aligned}T*\partial^\alpha\varphi(x) &= \langle T(z),\left(\partial^\alpha\varphi\right)(x-z)\rangle = (-1)^\alpha\langle T(z),\partial_z^\alpha\varphi(x-z)\rangle \\ &= \langle\partial_z^\alpha T(z),\varphi(x-z)\rangle = \partial^\alpha T*\varphi(x)\end{aligned}$$

the proof is complete except for showing $T*\varphi\in\mathcal{P}$. For the last statement, it suffices to prove $|T*\varphi(x)|\le C\mu_M(x)$ for some $C<\infty$ and $M<\infty$. This goes as follows

$$|h(x)| = |\langle T,\varphi(x-\cdot)\rangle| \le Cp_m(\varphi(x-\cdot)) = C\sum_{|\alpha|\le m}\|\mu_m(\partial^\alpha\varphi(x-\cdot)\|_\infty$$

and using Peetre's inequality, $|\partial^\alpha\varphi(x-z)|\le C\mu_{-m}(x-z)\le C\mu_{-m}(z)\mu_m(x)$ so that

$$\|\mu_m(\partial^\alpha\varphi(x-\cdot)\|_\infty \le C\mu_m(x).$$

Thus it follows that $|T*\varphi(x)|\le C\mu_m(x)$ for some $C<\infty$. If $x\in\mathbb{R}^n\setminus(\mathrm{supp}(T)+\mathrm{supp}(\varphi))$ and $y\in\mathrm{supp}(\varphi)$ then $x-y\notin\mathrm{supp}(T)$ for otherwise $x=x-y+y\in\mathrm{supp}(T)+\mathrm{supp}(\varphi)$. Thus

$$\mathrm{supp}(\varphi(x-\cdot)) = x-\mathrm{supp}(\varphi)\subset\mathbb{R}^n\setminus\mathrm{supp}(T)$$

and hence $h(x)=\langle T,\varphi(x-\cdot)\rangle = 0$ for all $x\in\mathbb{R}^n\setminus(\mathrm{supp}(T)+\mathrm{supp}(\varphi))$. This implies that $\{h\neq 0\}\subset\mathrm{supp}(T)+\mathrm{supp}(\varphi)$ and hence

$$\mathrm{supp}(h) = \overline{\{h\neq 0\}}\subset\overline{\mathrm{supp}(T)+\mathrm{supp}(\varphi)}.$$

■

As we have seen in the previous theorem, $T*\varphi$ is a smooth function and hence may be used to define a distribution in $\mathcal{D}'(\mathbb{R}^n)$ by

$$\langle T*\varphi,\psi\rangle = \int T*\varphi(x)\psi(x)dx = \int\langle T,\varphi(x-\cdot)\rangle\psi(x)dx.$$

Using the linearity of T we might expect that

$$\int\langle T,\varphi(x-\cdot)\rangle\psi(x)dx = \langle T,\int\varphi(x-\cdot)\psi(x)dx\rangle$$

or equivalently that

$$\langle T*\varphi,\psi\rangle = \langle T,\varphi*\psi\rangle \tag{46.10}$$

where $\varphi(x) := \varphi(-x)$.

Theorem 46.4. *Suppose that if (T,φ) is a distribution test function pair satisfying one the three condition in Theorem 46.3, then $T*\varphi$ as a distribution may be characterized by*

$$\langle T*\varphi,\psi\rangle = \langle T,\varphi*\psi\rangle \tag{46.11}$$

for all $\psi\in\mathcal{D}(\mathbb{R}^n)$. Moreover, if $T\in\mathcal{S}'$ and $\varphi\in\mathcal{S}$ then Eq. (46.11) holds for all $\psi\in\mathcal{S}$.

Proof. Let us first assume that $T\in\mathcal{D}'$ and $\varphi,\psi\in\mathcal{D}$ and $\theta\in\mathcal{D}$ be a function such that $\theta=1$ on a neighborhood of the support of ψ. Then

$$\begin{aligned}\langle T*\varphi,\psi\rangle &= \int_{\mathbb{R}^n}\langle T,\varphi(x-\cdot)\rangle\psi(x)dx = \langle\psi(x),\langle T(y),\varphi(x-y)\rangle\rangle \\ &= \langle\theta(x)\psi(x),\langle T(y),\varphi(x-y)\rangle\rangle \\ &= \langle\psi(x),\theta(x)\langle T(y),\varphi(x-y)\rangle\rangle \\ &= \langle\psi(x),\langle T(y),\theta(x)\varphi(x-y)\rangle\rangle.\end{aligned}$$

Now the function, $\theta(x)\varphi(x-y)\in\mathcal{D}(\mathbb{R}^n\times\mathbb{R}^n)$, so we may apply Fubini's theorem for distributions to conclude that

$$\begin{aligned}\langle T*\varphi,\psi\rangle &= \langle\psi(x),\langle T(y),\theta(x)\varphi(x-y)\rangle\rangle \\ &= \langle T(y),\langle\psi(x),\theta(x)\varphi(x-y)\rangle\rangle \\ &= \langle T(y),\langle\theta(x)\psi(x),\varphi(x-y)\rangle\rangle \\ &= \langle T(y),\langle\psi(x),\varphi(x-y)\rangle\rangle \\ &= \langle T(y),\psi*\varphi(y)\rangle = \langle T,\psi*\varphi\rangle\end{aligned}$$

as claimed. If $T\in\mathcal{E}'$, let $\alpha\in\mathcal{D}(\mathbb{R}^n)$ be a function such that $\alpha=1$ on a neighborhood of $\mathrm{supp}(T)$, then working as above,

$$\begin{aligned}\langle T*\varphi,\psi\rangle &= \langle\psi(x),\langle T(y),\theta(x)\varphi(x-y)\rangle\rangle \\ &= \langle\psi(x),\langle T(y),\alpha(y)\theta(x)\varphi(x-y)\rangle\rangle\end{aligned}$$

and since $\alpha(y)\theta(x)\varphi(x-y)\in\mathcal{D}(\mathbb{R}^n\times\mathbb{R}^n)$ we may apply Fubini's theorem for distributions to conclude again that

$$\begin{aligned}\langle T*\varphi,\psi\rangle &= \langle T(y),\langle\psi(x),\alpha(y)\theta(x)\varphi(x-y)\rangle\rangle \\ &= \langle\alpha(y)T(y),\langle\theta(x)\psi(x),\varphi(x-y)\rangle\rangle \\ &= \langle T(y),\langle\psi(x),\varphi(x-y)\rangle\rangle = \langle T,\psi*\varphi\rangle.\end{aligned}$$

Now suppose that $T\in\mathcal{S}'$ and $\varphi,\psi\in\mathcal{S}$. Let $\varphi_n,\psi_n\in\mathcal{D}$ be a sequences such that $\varphi_n\to\varphi$ and $\psi_n\to\psi$ in $\mathcal{S}$, then using arguments similar to those in the proof of Theorem 46.3, one shows

$$\langle T*\varphi,\psi\rangle = \lim_{n\to\infty}\langle T*\varphi_n,\psi_n\rangle = \lim_{n\to\infty}\langle T,\psi_n*\varphi_n\rangle = \langle T,\psi*\varphi\rangle.$$

■

Theorem 46.5. *Let $U \subset_o \mathbb{R}^n$, then $\mathcal{D}(U)$ is sequentially dense in $\mathcal{E}'(U)$. When $U = \mathbb{R}^n$ we have $\mathcal{E}'(\mathbb{R}^n)$ is a dense subspace of $\mathcal{S}'(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n)$. Hence we have the following inclusions,*

$$\begin{aligned}&\mathcal{D}(U) \subset \mathcal{E}'(U) \subset \mathcal{D}'(U),\\ &\mathcal{D}(\mathbb{R}^n) \subset \mathcal{E}'(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n) \text{ and}\\ &\mathcal{D}(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n)\end{aligned}$$

with all inclusions being dense in the next space up.

Proof. The key point is to show $\mathcal{D}(U)$ is dense in $\mathcal{E}'(U)$. Choose $\theta \in C_c^\infty(\mathbb{R}^n)$ such that $\operatorname{supp}(\theta) \subset B(0,1)$, $\theta = \theta$ and $\int \theta(x)dx = 1$. Let $\theta_m(x) = m^{-n}\theta(mx)$ so that $\operatorname{supp}(\theta_m) \subset B(0,1/m)$. An element in $T \in \mathcal{E}'(U)$ may be viewed as an element in $\mathcal{E}'(\mathbb{R}^n)$ in a natural way. Namely if $\chi \in C_c^\infty(U)$ such that $\chi = 1$ on a neighborhood of $\operatorname{supp}(T)$, and $\varphi \in C^\infty(\mathbb{R}^n)$, let $\langle T, \varphi\rangle = \langle T, \chi\varphi\rangle$. Define $T_m = T * \theta_m$. It is easily seen that $\operatorname{supp}(T_n) \subset \operatorname{supp}(T) + B(0,1/m) \subset U$ for all m sufficiently large. Hence $T_m \in \mathcal{D}(U)$ for large enough m. Moreover, if $\psi \in \mathcal{D}(U)$, then

$$\langle T_m, \psi\rangle = \langle T * \theta_m, \psi\rangle = \langle T, \theta_m * \psi\rangle = \langle T, \theta_m * \psi\rangle \to \langle T, \psi\rangle$$

since $\theta_m * \psi \to \psi$ in $\mathcal{D}(U)$ by standard arguments. If $U = \mathbb{R}^n$, $T \in \mathcal{E}'(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n)$ and $\psi \in \mathcal{S}$, the same argument goes through to show $\langle T_m, \psi\rangle \to \langle T, \psi\rangle$ provided we show $\theta_m * \psi \to \psi$ in $\mathcal{S}(\mathbb{R}^n)$ as $m \to \infty$. This latter is proved by showing for all α and $t > 0$, I

$$\|\mu_t (\partial^\alpha \theta_m * \psi - \partial^\alpha \psi)\|_\infty \to 0 \text{ as } m \to \infty,$$

which is a consequence of the estimates:

$$\begin{aligned}|\partial^\alpha \theta_m * \psi(x) - \partial^\alpha \psi(x)| &= |\theta_m * \partial^\alpha \psi(x) - \partial^\alpha \psi(x)|\\ &= \left|\int \theta_m(y)\left[\partial^\alpha \psi(x-y) - \partial^\alpha \psi(x)\right]dy\right|\\ &\le \sup_{|y|\le 1/m} |\partial^\alpha \psi(x-y) - \partial^\alpha \psi(x)|\\ &\le \frac{1}{m}\sup_{|y|\le 1/m} |\nabla \partial^\alpha \psi(x-y)|\\ &\le \frac{1}{m} C \sup_{|y|\le 1/m} \mu_{-t}(x-y)\\ &\le \frac{1}{m} C \mu_{-t}(x-y) \sup_{|y|\le 1/m} \mu_t(y)\\ &\le \frac{1}{m} C \left(1 + m^{-1}\right)^t \mu_{-t}(x).\end{aligned}$$

■

Definition 46.6 (Convolution of Distributions). *Suppose that $T \in \mathcal{D}'$ and $S \in \mathcal{E}'$, then define $T * S \in \mathcal{D}'$ by*

$$\langle T * S, \varphi\rangle = \langle T \otimes S, \varphi_+\rangle$$

where $\varphi_+(x,y) = \varphi(x+y)$ for all $x, y \in \mathbb{R}^n$. More generally we may define $T * S$ for any two distributions having the property that $\operatorname{supp}(T \otimes S) \cap \operatorname{supp}(\varphi_+) = [\operatorname{supp}(T) \times \operatorname{supp}(S)] \cap \operatorname{supp}(\varphi_+)$ is compact for all $\varphi \in \mathcal{D}$.

Proposition 46.7. *Suppose that $T \in \mathcal{D}'$ and $S \in \mathcal{E}'$ then $T * S$ is well defined and*

$$\langle T * S, \varphi\rangle = \langle T(x), \langle S(y), \varphi(x+y)\rangle\rangle = \langle S(y), \langle T(x), \varphi(x+y)\rangle\rangle. \tag{46.12}$$

Moreover, if $T \in \mathcal{S}'$ then $T * S \in \mathcal{S}'$ and $\mathcal{F}(T * S) = \hat{S}\hat{T}$. Recall from Remark 45.46 that $\hat{S} \in \mathcal{P}$ so that $\hat{S}\hat{T} \in \mathcal{S}'$.

Proof. Let $\theta \in \mathcal{D}$ be a function such that $\theta = 1$ on a neighborhood of $\operatorname{supp}(S)$, then by Fubini's theorem for distributions,

$$\begin{aligned}\langle T \otimes S, \varphi_+\rangle &= \langle T \otimes S(x,y), \theta(y)\varphi(x+y)\rangle = \langle T(x)S(y), \theta(y)\varphi(x+y)\rangle\\ &= \langle T(x), \langle S(y), \theta(y)\varphi(x+y)\rangle\rangle = \langle T(x), \langle S(y), \varphi(x+y)\rangle\rangle\end{aligned}$$

and

$$\begin{aligned}\langle T \otimes S, \varphi_+\rangle &= \langle T(x)S(y), \theta(y)\varphi(x+y)\rangle = \langle S(y), \langle T(x), \theta(y)\varphi(x+y)\rangle\rangle\\ &= \langle S(y), \theta(y)\langle T(x), \varphi(x+y)\rangle\rangle = \langle S(y), \langle T(x), \varphi(x+y)\rangle\rangle\end{aligned}$$

proving Eq. (46.12). Suppose that $T \in \mathcal{S}'$, then

$$\begin{aligned}|\langle T * S, \varphi\rangle| = |\langle T(x), \langle S(y), \varphi(x+y)\rangle\rangle| &\le C \sum_{|\alpha|\le m} \|\mu_m \partial_x^\alpha \langle S(y), \varphi(\cdot + y)\rangle\|_\infty\\ &= C \sum_{|\alpha|\le m} \|\mu_m \langle S(y), \partial^\alpha \varphi(\cdot + y)\rangle\|_\infty\end{aligned}$$

and

$$\begin{aligned}|\langle S(y), \partial^\alpha \varphi(x+y)\rangle| &\le C \sum_{|\beta|\le p} \sup_{y\in K} \left|\partial^\beta \partial^\alpha \varphi(x+y)\right|\\ &\le C p_{m+p}(\varphi) \sup_{y\in K} \mu_{-m-p}(x+y)\\ &\le C p_{m+p}(\varphi) \mu_{-m-p}(x) \sup_{y\in K} \mu_{m+p}(y)\\ &= \tilde{C} \mu_{-m-p}(x) p_{m+p}(\varphi).\end{aligned}$$

Combining the last two displayed equations shows

$$|\langle T * S, \varphi\rangle| \le C p_{m+p}(\varphi)$$

which shows that $T * S \in \mathcal{S}'$. We still should check that

$$\langle T * S, \varphi\rangle = \langle T(x), \langle S(y), \varphi(x+y)\rangle\rangle = \langle S(y), \langle T(x), \varphi(x+y)\rangle\rangle$$

still holds for all $\varphi \in \mathcal{S}$. This is a matter of showing that all of the expressions are continuous in $\mathcal{S}$ when restricted to $\mathcal{D}$. Explicitly, let $\varphi_m \in \mathcal{D}$ be a sequence of functions such that $\varphi_m \to \varphi$ in $\mathcal{S}$, then

$$\langle T * S, \varphi\rangle = \lim_{n\to\infty} \langle T * S, \varphi_n\rangle = \lim_{n\to\infty} \langle T(x), \langle S(y), \varphi_n(x+y)\rangle\rangle \tag{46.13}$$

and

$$\langle T * S, \varphi\rangle = \lim_{n\to\infty} \langle T * S, \varphi_n\rangle = \lim_{n\to\infty} \langle S(y), \langle T(x), \varphi_n(x+y)\rangle\rangle. \tag{46.14}$$

So it suffices to show the map $\varphi \in \mathcal{S} \to \langle S(y), \varphi(\cdot + y)\rangle \in \mathcal{S}$ is continuous and $\varphi \in \mathcal{S} \to \langle T(x), \varphi(x + \cdot)\rangle \in C^\infty(\mathbb{R}^n)$ are continuous maps. These may verified by methods similar to what we have been doing, so I will leave the details to the reader. Given these continuity assertions, we may pass to the limits in Eq. (46.13d (46.14) to learn

$$\langle T * S, \varphi\rangle = \langle T(x), \langle S(y), \varphi(x+y)\rangle\rangle = \langle S(y), \langle T(x), \varphi(x+y)\rangle\rangle$$

still holds for all $\varphi \in \mathcal{S}$. The last and most important point is to show $\mathcal{F}(T*S) = \hat{S}\hat{T}$. Using

$$\begin{aligned}\hat{\varphi}(x+y) &= \int_{\mathbb{R}^n} \varphi(\xi) e^{-i\xi\cdot(x+y)} d\xi = \int_{\mathbb{R}^n} \varphi(\xi) e^{-i\xi\cdot y} e^{-i\xi\cdot x} d\xi \\ &= \mathcal{F}\left(\varphi(\xi) e^{-i\xi\cdot y}\right)(x)\end{aligned}$$

and the definition of $\mathcal{F}$ on $\mathcal{S}'$ we learn

$$\begin{aligned}\langle \mathcal{F}(T * S), \varphi\rangle &= \langle T * S, \hat{\varphi}\rangle = \langle S(y), \langle T(x), \hat{\varphi}(x+y)\rangle\rangle \\ &= \langle S(y), \langle T(x), \mathcal{F}\left(\varphi(\xi)e^{-i\xi\cdot y}\right)(x)\rangle\rangle \\ &= \langle S(y), \langle \hat{T}(\xi), \varphi(\xi)e^{-i\xi\cdot y}\rangle\rangle.\end{aligned} \tag{46.15}$$

Let $\theta \in \mathcal{D}$ be a function such that $\theta = 1$ on a neighborhood of $\operatorname{supp}(S)$ and assume $\varphi \in \mathcal{D}$ for the moment. Then from Eq. (46.15) and Fubini's theorem for distributions we find

$$\begin{aligned}\langle \mathcal{F}(T * S), \varphi\rangle &= \langle S(y), \theta(y)\langle \hat{T}(\xi), \varphi(\xi)e^{-i\xi\cdot y}\rangle\rangle \\ &= \langle S(y), \langle \hat{T}(\xi), \varphi(\xi)\theta(y)e^{-i\xi\cdot y}\rangle\rangle \\ &= \langle \hat{T}(\xi), \langle S(y), \varphi(\xi)\theta(y)e^{-i\xi\cdot y}\rangle\rangle \\ &= \langle \hat{T}(\xi), \varphi(\xi)\langle S(y), e^{-i\xi\cdot y}\rangle\rangle \\ &= \langle \hat{T}(\xi), \varphi(\xi)\hat{S}(\xi)\rangle = \langle \hat{S}(\xi)\hat{T}(\xi), \varphi(\xi)\rangle.\end{aligned} \tag{46.16}$$

Since $\mathcal{F}(T * S) \in \mathcal{S}'$ and $\hat{S}\hat{T} \in \mathcal{S}'$, we conclude that *Eq.* (46.16) holds for all $\varphi \in \mathcal{S}$ and hence $\mathcal{F}(T * S) = \hat{S}\hat{T}$ as was to be proved. ■

46.2 Elliptic Regularity

Theorem 46.8 (Hypoellipticity). *Suppose that $p(x) = \sum_{|\alpha|\le m} a_\alpha \xi^\alpha$ is a polynomial on $\mathbb{R}^n$ and L is the constant coefficient differential operator*

$$L = p(\frac{1}{i}\partial) = \sum_{|\alpha|\le m} a_\alpha (\frac{1}{i}\partial)^\alpha = \sum_{|\alpha|\le m} a_\alpha (-i\partial)^\alpha.$$

Also assume there exists a distribution $T \in \mathcal{D}'(\mathbb{R}^n)$ such that $R := \delta - LT \in C^\infty(\mathbb{R}^n)$ and $T|_{\mathbb{R}^n\setminus\{0\}} \in C^\infty(\mathbb{R}^n \setminus \{0\})$. Then if $v \in C^\infty(U)$ and $u \in \mathcal{D}'(U)$ solves $Lu = v$ then $u \in C^\infty(U)$. In particular, all solutions u to the equation $Lu = 0$ are smooth.

Proof. We must show for each $x_0 \in U$ that u is smooth on a neighborhood of x_0. So let $x_0 \in U$ and $\theta \in \mathcal{D}(U)$ such that $0 \le \theta \le 1$ and $\theta = 1$ on neighborhood V of x_0. Also pick $\alpha \in \mathcal{D}(V)$ such that $0 \le \alpha \le 1$ and $\alpha = 1$ on a neighborhood of x_0. Then

$$\begin{aligned}\theta u &= \delta * (\theta u) = (LT + R) * (\theta u) = (LT) * (\theta u) + R * (\theta u) \\ &= T * L(\theta u) + R * (\theta u) \\ &= T * \{\alpha L(\theta u) + (1-\alpha)L(\theta u)\} + R * (\theta u) \\ &= T * \{\alpha L u + (1-\alpha)L(\theta u)\} + R * (\theta u) \\ &= T * (\alpha v) + R * (\theta u) + T * [(1-\alpha)L(\theta u)].\end{aligned}$$

Since $\alpha v \in \mathcal{D}(U)$ and $T \in \mathcal{D}'(\mathbb{R}^n)$ it follows that $R * (\theta u) \in C^\infty(\mathbb{R}^n)$. Also since $R \in C^\infty(\mathbb{R}^n)$ and $\theta u \in \mathcal{E}'(U)$, $R * (\theta u) \in C^\infty(\mathbb{R}^n)$. So to show θu, and hence u, is smooth near x_0 it suffices to show $T * g$ is smooth near x_0 where $g := (1-\alpha)L(\theta u)$. Working formally for the moment,

$$T * g(x) = \int_{\mathbb{R}^n} T(x-y)g(y)dy = \int_{\mathbb{R}^n\setminus\{\alpha=1\}} T(x-y)g(y)dy$$

which should be smooth for x near x_0 since in this case $x-y\neq 0$ when $g(y)\neq 0$. To make this precise, let $\delta>0$ be chosen so that $\alpha=1$ on a neighborhood of $\overline{B(x_0,\delta)}$ so that $\operatorname{supp}(g)\subset\overline{B(x_0,\delta)}^c$. For $\varphi\in\mathcal{D}(B(x_0,\delta/2)$,

$$\langle T*g,\varphi\rangle=\langle T(x),\langle g(y),\varphi(x+y)\rangle\rangle=\langle T,h\rangle$$

where $h(x):=\langle g(y),\varphi(x+y)\rangle$. If $|x|\leq\delta/2$

$$\operatorname{supp}(\varphi(x+\cdot))=\operatorname{supp}(\varphi)-x\subset B(x_0,\delta/2)-x\subset B(x_0,\delta)$$

so that $h(x)=0$ and hence $\operatorname{supp}(h)\subset\overline{B(x_0,\delta/2)}^c$. Hence if we let $\gamma\in\mathcal{D}(B(0,\delta/2))$ be a function such that $\gamma=1$ near 0, we have $\gamma h\equiv 0$, and thus

$$\langle T*g,\varphi\rangle=\langle T,h\rangle=\langle T,h-\gamma h\rangle=\langle(1-\gamma)T,h\rangle=\langle[(1-\gamma)T]*g,\varphi\rangle.$$

Since this last equation is true for all $\varphi\in\mathcal{D}(B(x_0,\delta/2))$, $T*g=[(1-\gamma)T]*g$ on $B(x_0,\delta/2)$ and this finishes the proof since $[(1-\gamma)T]*g\in C^\infty(\mathbb{R}^n)$ because $(1-\gamma)T\in C^\infty(\mathbb{R}^n)$. ■

Definition 46.9. *Suppose that* $p(x)=\sum_{|\alpha|\leq m}a_\alpha\xi^\alpha$ *is a polynomial on* $\mathbb{R}^n$ *and* L *is the constant coefficient differential operator*

$$L=p(\frac{1}{i}\partial)=\sum_{|\alpha|\leq m}a_\alpha(\frac{1}{i}\partial)^\alpha=\sum_{|\alpha|\leq m}a_\alpha(-i\partial)^\alpha.$$

Let $\sigma_p(L)(\xi):=\sum_{|\alpha|=m}a_\alpha\xi^\alpha$ *and call* $\sigma_p(L)$ *the principle symbol of* L. *The operator* L *is said to be elliptic provided that* $\sigma_p(L)(\xi)\neq 0$ *if* $\xi\neq 0$.

Theorem 46.10 (Existence of Parametrix). *Suppose that* $L=p(\frac{1}{i}\partial)$ *is an elliptic constant coefficient differential operator, then there exists a distribution* $T\in\mathcal{D}'(\mathbb{R}^n)$ *such that* $R:=\delta-LT\in C^\infty(\mathbb{R}^n)$ *and* $T|_{\mathbb{R}^n\setminus\{0\}}\in C^\infty(\mathbb{R}^n\setminus\{0\})$.

Proof. The idea is to try to find T such that $LT=\delta$. Taking the Fourier transform of this equation implies that $p(\xi)\hat{T}(\xi)=1$ and hence we should try to define $\hat{T}(\xi)=1/p(\xi)$. The main problem with this definition is that $p(\xi)$ may have zeros. However, these zeros can not occur for large ξ by the ellipticity assumption. Indeed, let $q(\xi):=\sigma_p(L)(\xi)=\sum_{|\alpha|=m}a_\alpha\xi^\alpha$, $r(\xi)=p(\xi)-q(\xi)=\sum_{|\alpha|<m}a_\alpha\xi^\alpha$ and let $c=\min\{|q(\xi)|:|\xi|=1\}\leq\max\{|q(\xi)|:|\xi|=1\}=:C$. Then because $|q(\cdot)|$ is a nowhere vanishing continuous function on the compact set $S:=\{\xi\in\mathbb{R}^n:|\xi|=1|\}$, $0<c\leq C<\infty$. For $\xi\in\mathbb{R}^n$, let $\hat{\xi}=\xi/|\xi|$ and notice

$$|p(\xi)|=|q(\xi)|-|r(\xi)|\geq c\,|\xi|^m-|r(\xi)|=|\xi|^m\,(c-\frac{|r(\xi)|}{|\xi|^m})>0$$

for all $|\xi|\geq M$ with M sufficiently large since $\lim_{\xi\to\infty}\frac{|r(\xi)|}{|\xi|^m}=0$. Choose $\theta\in\mathcal{D}(\mathbb{R}^n)$ such that $\theta=1$ on a neighborhood of $\overline{B(0,M)}$ and let

$$h(\xi)=\frac{1-\theta(\xi)}{p(\xi)}=\frac{\beta(\xi)}{p(\xi)}\in C^\infty(\mathbb{R}^n)$$

where $\beta=1-\theta$. Since $h(\xi)$ is bounded (in fact $\lim_{\xi\to\infty}h(\xi)=0$), $h\in\mathcal{S}'(\mathbb{R}^n)$ so there exists $T:=\mathcal{F}^{-1}h\in\mathcal{S}'(\mathbb{R}^n)$ is well defined. Moreover,

$$\mathcal{F}\left(\delta-LT\right)=1-p(\xi)h(\xi)=1-\beta(\xi)=\theta(\xi)\in\mathcal{D}(\mathbb{R}^n)$$

which shows that

$$R:=\delta-LT\in\mathcal{S}(\mathbb{R}^n)\subset C^\infty(\mathbb{R}^n).$$

So to finish the proof it suffices to show

$$T|_{\mathbb{R}^n\setminus\{0\}}\in C^\infty(\mathbb{R}^n\setminus\{0\}).$$

To prove this recall that

$$\mathcal{F}\left(x^\alpha T\right)=(i\partial)^\alpha\hat{T}=(i\partial)^\alpha h.$$

By the chain rule and the fact that any derivative of β is has compact support in $\overline{B(0,M)}^c$ and any derivative of $\frac{1}{p}$ is non-zero on this set,

$$\partial^\alpha h=\beta\partial^\alpha\frac{1}{p}+r_\alpha$$

where $r_\alpha\in\mathcal{D}(\mathbb{R}^n)$. Moreover,

$$\partial_i\frac{1}{p}=-\frac{\partial_i p}{p^2}\text{ and }\partial_j\partial_i\frac{1}{p}=-\partial_j\frac{\partial_i p}{p^2}=-\frac{\partial_j\partial_i p}{p^2}+2\frac{\partial_i p}{p^3}$$

from which it follows that

$$\left|\beta(\xi)\partial_i\frac{1}{p}(\xi)\right|\leq C\,|\xi|^{-(m+1)}\text{ and }\left|\beta(\xi)\partial_j\partial_i\frac{1}{p}\right|\leq C\,|\xi|^{-(m+2)}.$$

More generally, one shows by inductively that

$$\left|\beta(\xi)\partial^\alpha\frac{1}{p}\right|\leq C\,|\xi|^{-(m+|\alpha|)}. \tag{46.17}$$

In particular, if $k\in\mathbb{N}$ is given and α is chosen so that $|\alpha|+m>n+k$, then $|\xi|^k\,\partial^\alpha h(\xi)\in L^1(\xi)$ and therefore

$$x^\alpha T=\mathcal{F}^{-1}\left[(i\partial)^\alpha h\right]\in C^k(\mathbb{R}^n).$$

Hence we learn for any $k\in\mathbb{N}$, we may choose p sufficiently large so that

$$|x|^{2p}T\in C^k(\mathbb{R}^n).$$

This shows that $T|_{\mathbb{R}^n\setminus\{0\}}\in C^\infty(\mathbb{R}^n\setminus\{0\})$. ∎

Here is the induction argument that proves Eq. (46.17). Let $q_\alpha := p^{|\alpha|+1}\partial^\alpha p^{-1}$ with $q_0=1$, then

$$\partial_i\partial^\alpha p^{-1}=\partial_i\left(p^{-|\alpha|-1}q_\alpha\right)=(-|\alpha|-1)\,p^{-|\alpha|-2}q_\alpha\partial_i p+p^{-|\alpha|-1}\partial_i q_\alpha$$

so that

$$q_{\alpha+e_i}=p^{|\alpha|+2}\partial_i\partial^\alpha p^{-1}=(-|\alpha|-1)\,q_\alpha\partial_i p+p\partial_i q_\alpha.$$

It follows by induction that q_α is a polynomial in ξ and letting $d_\alpha:=\deg(q_\alpha)$, we have $d_{\alpha+e_i}\le d_\alpha+m-1$ with $d_0=1$. Again by induction this implies $d_\alpha\le|\alpha|\,(m-1)$. Therefore

$$\partial^\alpha p^{-1}=\frac{q_\alpha}{p^{|\alpha|+1}}\sim|\xi|^{d_\alpha-m(|\alpha|+1)}=|\xi|^{|\alpha|(m-1)-m(|\alpha|+1)}=|\xi|^{-(m+|\alpha|)}$$

as claimed in Eq. (46.17).

*** Beginning of WORK material. ***

46.3 Appendix: Old Proof of Theorem 46.4

This indeed turns out to be the case but is a bit painful to prove. The next theorem is the key ingredient to proving Eq. (46.10).

Theorem 46.11. *Let $\psi\in\mathcal{D}$ ($\psi\in\mathcal{S}$) $d\lambda(y)=\psi(y)dy$, and $\varphi\in C^\infty(\mathbb{R}^n)$ ($\varphi\in\mathcal{S}$). For $\varepsilon>0$ we may write $\mathbb{R}^n=\coprod_{m\in\mathbb{Z}^n}(m\varepsilon+\varepsilon Q)$ where $Q=(0,1]^n$. For $y\in(m\varepsilon+\varepsilon Q)$, let $y_\varepsilon\in m\varepsilon+\varepsilon\bar{Q}$ be the point closest to the origin in $m\varepsilon+\varepsilon\bar{Q}$. (This will be one of the corners of the translated cube.) In this way we define a function $y\in\mathbb{R}^n\to y_\varepsilon\in\varepsilon\mathbb{Z}^n$ which is constant on each cube $\varepsilon(m+Q)$. Let*

$$F_\varepsilon(x):=\int\varphi(x-y_\varepsilon)d\lambda(y)=\sum_{m\in\mathbb{Z}^n}\varphi(x-(m\varepsilon)_\varepsilon)\lambda(\varepsilon(m+Q)),\qquad(46.18)$$

then the above sum converges in $C^\infty(\mathbb{R}^n)$ ($\mathcal{S}$) and $F_\varepsilon\to\varphi*\psi$ in $C^\infty(\mathbb{R}^n)$ ($\mathcal{S}$) as $\varepsilon\downarrow0$. (In particular if $\varphi,\psi\in\mathcal{S}$ then $\varphi*\psi\in\mathcal{S}$.)

Proof. First suppose that $\psi\in\mathcal{D}$ the measure λ has compact support and hence the sum in Eq. (46.18) is finite and so is certainly convergent in $C^\infty(\mathbb{R}^n)$. To shows $F_\varepsilon\to\varphi*\psi$ in $C^\infty(\mathbb{R}^n)$, let K be a compact set and $m\in\mathbb{N}$. Then for $|\alpha|\le m$,

$$\begin{aligned}|\partial^\alpha F_\varepsilon(x)-\partial^\alpha\varphi*\psi(x)|&=\left|\int[\partial^\alpha\varphi(x-y_\varepsilon)-\partial^\alpha\varphi(x-y)]\,d\lambda(y)\right|\\&\le\int|\partial^\alpha\varphi(x-y_\varepsilon)-\partial^\alpha\varphi(x-y)|\,|\psi(y)|\,dy\qquad(46.19)\end{aligned}$$

and therefore,

$$\begin{aligned}\|\partial^\alpha F_\varepsilon-\partial^\alpha\varphi*\psi\|_{\infty,K}&\le\int\|\partial^\alpha\varphi(\cdot-y_\varepsilon)-\partial^\alpha\varphi(\cdot-y)\|_{\infty,K}\,|\psi(y)|\,dy\\&\le\sup_{y\in\mathrm{supp}(\psi)}\|\partial^\alpha\varphi(\cdot-y_\varepsilon)-\partial^\alpha\varphi(\cdot-y)\|_{\infty,K}\int|\psi(y)|\,dy.\end{aligned}$$

Since $\psi(y)$ has compact support, we may us the uniform continuity of $\partial^\alpha\varphi$ on compact sets to conclude

$$\sup_{y\in\mathrm{supp}(\psi)}\|\partial^\alpha\varphi(\cdot-y_\varepsilon)-\partial^\alpha\varphi(\cdot-y)\|_{\infty,K}\to0\text{ as }\varepsilon\downarrow0.$$

This finishes the proof for $\psi\in\mathcal{D}$ and $\varphi\in C^\infty(\mathbb{R}^n)$. Now suppose that both ψ and φ are in $\mathcal{S}$ in which case the sum in Eq. (46.18) is now an infinite sum in general so we need to check that it converges to an element in $\mathcal{S}$. For this we estimate each term in the sum. Given $s,t>0$ and a multi-index α, using Peetre's inequality and simple estimates,

$$\begin{aligned}|\partial^\alpha\varphi(x-(m\varepsilon)_\varepsilon)\lambda(\varepsilon(m+Q))|&\le C\nu_{-t}(x-(m\varepsilon)_\varepsilon)\int_{\varepsilon(m+Q)}|\psi(y)|\,dy\\&\le C\nu_{-t}(x)\nu_t((m\varepsilon)_\varepsilon)K\int_{\varepsilon(m+Q)}\nu_{-s}(y)dy\end{aligned}$$

for some finite constants K and C. Making the change of variables $y=m\varepsilon+\varepsilon z$, we find

$$\begin{aligned}\int_{\varepsilon(m+Q)}\nu_{-s}(y)dy&=\varepsilon^n\int_Q\nu_{-s}(m\varepsilon+\varepsilon z)dz\\&\le\varepsilon^n\nu_{-s}(m\varepsilon)\int_Q\nu_s(\varepsilon z)dy\\&=\varepsilon^n\nu_{-s}(m\varepsilon)\int_Q\frac{1}{(1+\varepsilon|z|)^s}dy\\&\le\varepsilon^n\nu_{-s}(m\varepsilon).\end{aligned}$$

Combining these two estimates shows

$$\begin{aligned}\|\nu_t\partial^\alpha\varphi(\cdot-(m\varepsilon)_\varepsilon)\lambda(\varepsilon(m+Q))\|_\infty&\le C\nu_t((m\varepsilon)_\varepsilon)\varepsilon^n\nu_{-s}(m\varepsilon)\\&\le C\nu_t(m\varepsilon)\nu_{-s}(m\varepsilon)\varepsilon^n\\&=C\nu_{t-s}((m\varepsilon)\varepsilon^n\end{aligned}$$

and therefore for some (different constant C)

$$\sum_{m\in\mathbb{Z}^n} p_k\left(\varphi(\cdot-(m\varepsilon)_\varepsilon)\lambda(\varepsilon(m+Q))\right) \le \sum_{m\in\mathbb{Z}^n} C\nu_{k-s}(m\varepsilon)\varepsilon^n = \sum_{m\in\mathbb{Z}^n} C\frac{1}{(1+\varepsilon|m|)^{k-s}}\varepsilon^n$$

which can be made finite by taking $s>k+n$ as can be seen by an comparison with the integral $\int\frac{1}{(1+\varepsilon|x|)^{k-s}}dx$. Therefore the sum is convergent in $\mathcal{S}$ as claimed. To finish the proof, we must show that $F_\varepsilon \to \varphi*\psi$ in $\mathcal{S}$. From Eq. (46.19) we still have

$$|\partial^\alpha F_\varepsilon(x)-\partial^\alpha\varphi*\psi(x)| \le \int|\partial^\alpha\varphi(x-y_\varepsilon)-\partial^\alpha\varphi(x-y)|\,|\psi(y)|\,dy.$$

The estimate in Eq. (46.9) gives

$$\begin{aligned}|\partial^\alpha\varphi(x-y_\varepsilon)-\partial^\alpha\varphi(x-y)| &\le C\int_0^1\nu_M(y_\varepsilon+\tau(y-y_\varepsilon))d\tau\,|y-y_\varepsilon|\,\nu_{-M}(x)\\ &\le C\varepsilon\nu_{-M}(x)\int_0^1\nu_M(y_\varepsilon+\tau(y-y_\varepsilon))d\tau\\ &\le C\varepsilon\nu_{-M}(x)\int_0^1\nu_M(y)d\tau = C\varepsilon\nu_{-M}(x)\nu_M(y)\end{aligned}$$

where in the last inequality we have used the fact that $|y_\varepsilon+\tau(y-y_\varepsilon)|\le|y|$. Therefore,

$$\|\nu_M\left(\partial^\alpha F_\varepsilon(x)-\partial^\alpha\varphi*\psi\right)\|_\infty \le C\varepsilon\int_{\mathbb{R}^n}\nu_M(y)\,|\psi(y)|\,dy \to 0 \text{ as } \varepsilon\to\infty$$

because $\int_{\mathbb{R}^n}\nu_M(y)\,|\psi(y)|\,dy<\infty$ for all $M<\infty$ since $\psi\in\mathcal{S}$. ■

We are now in a position to prove Eq. (46.10). Let us state this in the form of a theorem.

Theorem 46.12. *Suppose that if (T,φ) is a distribution test function pair satisfying one the three condition in Theorem 46.3, then $T*\varphi$ as a distribution may be characterized by*

$$\langle T*\varphi,\psi\rangle = \langle T,\varphi*\psi\rangle \tag{46.20}$$

for all $\psi\in\mathcal{D}(\mathbb{R}^n)$ and all $\psi\in\mathcal{S}$ when $T\in\mathcal{S}'$ and $\varphi\in\mathcal{S}$.

Proof. Let

$$\tilde{F}_\varepsilon = \int\varphi(x-y_\varepsilon)d\lambda(y) = \sum_{m\in\mathbb{Z}^n}\varphi(x-(m\varepsilon)_\varepsilon)\lambda(\varepsilon(m+Q))$$

then making use of Theorem 46.12 in all cases we find

$$\begin{aligned}\langle T,\varphi*\psi\rangle &= \lim_{\varepsilon\downarrow0}\langle T,\tilde{F}_\varepsilon\rangle\\ &= \lim_{\varepsilon\downarrow0}\langle T(x),\sum_{m\in\mathbb{Z}^n}\varphi(x-(m\varepsilon)_\varepsilon)\lambda(\varepsilon(m+Q))\rangle\\ &= \lim_{\varepsilon\downarrow0}\sum_{m\in\mathbb{Z}^n}\langle T(x),\varphi((m\varepsilon)_\varepsilon-x)\lambda(\varepsilon(m+Q))\rangle\\ &= \lim_{\varepsilon\downarrow0}\sum_{m\in\mathbb{Z}^n}\langle T*\varphi((m\varepsilon)_\varepsilon\rangle\lambda(\varepsilon(m+Q)).\end{aligned} \tag{46.21}$$

To compute this last limit, let $h(x)=T*\varphi(x)$ and let us do the hard case where $T\in\mathcal{S}'$. In this case we know that $h\in\mathcal{P}$, and in particular there exists $k<\infty$ and $C<\infty$ such that $\|\nu_k h\|_\infty<\infty$. So we have

$$\begin{aligned}&\left|\int_{\mathbb{R}^n}h(x)d\lambda(x)-\sum_{m\in\mathbb{Z}^n}\langle T*\varphi((m\varepsilon)_\varepsilon\rangle\lambda(\varepsilon(m+Q))\right|\\ &\quad=\left|\int_{\mathbb{R}^n}[h(x)-h(x_\varepsilon)]\,d\lambda(x)\right|\\ &\quad\le\int_{\mathbb{R}^n}|h(x)-h(x_\varepsilon)|\,|\psi(x)|\,dx.\end{aligned}$$

Now

$$|h(x)-h(x_\varepsilon)|\le C\left(\nu_k(x)+\nu_k(x_\varepsilon)\right)\le 2C\nu_k(x)$$

and since $\nu_k|\psi|\in L^1$ we may use the dominated convergence theorem to conclude

$$\lim_{\varepsilon\downarrow0}\left|\int_{\mathbb{R}^n}h(x)d\lambda(x)-\sum_{m\in\mathbb{Z}^n}\langle T*\varphi((m\varepsilon)_\varepsilon\rangle\lambda(\varepsilon(m+Q))\right|=0$$

which combined with Eq. (46.21) proves the theorem. ■

References

1. Henry B. Cohen, *Sequential denseness and the Eberlein-Šmulian theorem*, Math. Ann. **172** (1967), 209–210. MR MR0212549 (35 #3420)
2. Claude Dellacherie, *Capacités et processus stochastiques*, Springer-Verlag, Berlin, 1972, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67. MR MR0448504 (56 #6810)
3. L. E. Dubins and E. H. Spanier, *How to cut a cake fairly*, Amer. Math. Monthly **68** (1961), 1–17. MR MR0129031 (23 #B2068)
4. Nelson Dunford and Jacob T. Schwartz, *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988, General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. MR MR1009162 (90g:47001a)
5. John Elton, Theodore P. Hill, and Robert P. Kertz, *Optimal-partitioning inequalities for nonatomic probability measures*, Trans. Amer. Math. Soc. **296** (1986), no. 2, 703–725. MR MR846603 (87m:60012)
6. F. G. Friedlander, *Introduction to the theory of distributions*, second ed., Cambridge University Press, Cambridge, 1998, With additional material by M. Joshi. MR 2000g:46002: Call No. QA324 .F74 1998
7. Robert D. Gordon, *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument*, Ann. Math. Statistics **12** (1941), 364–366. MR MR0005558 (3,171e)
8. Theodore P. Hill, *Partitioning general probability measures*, Ann. Probab. **15** (1987), no. 2, 804–813. MR MR885145 (88d:60014)
9. Joe Howard, *Mackey compactness in Banach spaces*, Proc. Amer. Math. Soc. **37** (1973), 108–110. MR MR0312212 (47 #774)
10. Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. MR MR1474726 (99f:60082)
11. Martin L. Jones, *A note on a cake cutting algorithm of Banach and Knaster*, Amer. Math. Monthly **104** (1997), no. 4, 353–355. MR MR1450669 (99e:60012)
12. K. Kuratowski, *Topology. Vol. I*, New edition, revised and augmented. Translated from the French by J. Jaworowski, Academic Press, New York, 1966. MR MR0217751 (36 #840)
13. ———, *Topology. Vol. II*, New edition, revised and augmented. Translated from the French by A. Kirkor, Academic Press, New York, 1968. MR MR0259835 (41 #4467)
14. Lynn H. Loomis, *An introduction to abstract harmonic analysis*, D. Van Nostrand Company, Inc., Toronto-New York-London, 1953. MR 14,883c
15. J. G. Mauldon, *Continuous functions with zero derivative almost everywhere*, Quart. J. Math. Oxford Ser. (2) **17** (1966), 257–262. MR MR0197639 (33 #5804)
16. Vladimir G. Maz'ja, *Sobolev spaces*, Springer Series in Soviet Mathematics, Springer-Verlag, Berlin, 1985, Translated from the Russian by T. O. Shaposhnikova. MR 87g:46056
17. Hisashi Okamoto and Marcus Wunsch, *A geometric construction of continuous, strictly increasing singular functions*, Proc. Japan Acad. Ser. A Math. Sci. **83** (2007), no. 7, 114–118. MR MR2361422 (2008k:26008)
18. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
19. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR 2001g:60188

20. Michael Sharpe, *General theory of Markov processes*, Pure and Applied Mathematics, vol. 133, Academic Press Inc., Boston, MA, 1988. MR MR958914 (89m:60169)
21. Daniel W. Stroock, *Probability theory, an analytic view*, Cambridge University Press, Cambridge, 1993. MR MR1267569 (95f:60003)
22. Lajos Takács, *An increasing continuous singular function*, Amer. Math. Monthly **85** (1978), no. 1, 35–37. MR MR481327 (80f:26005)
23. Robert Whitley, *An elementary proof of the Eberlein-Šmulian theorem*, Math. Ann. **172** (1967), 116–118. MR MR0212548 (35 #3419)
24. Tudor Zamfirescu, *Typical monotone continuous functions*, Arch. Math. (Basel) **42** (1984), no. 2, 151–156. MR MR751487 (85m:26007)

Absolute continuity, 316, 331

Banach space, **163**
  reflexive, 257
  sums in, 171
Bounded variation, 316

Cauchy, 9
Cauchy sequence
  in a metric space, 157
  in a normed space, 163
Closed, *see* Sets
Coherent states, 566
Complete
  Metric space, 157
Continuous function, 156, 453, 456
Contraction Mapping Principle, 448

Function
  continuous, 155, 156, 453
  continuous at a point, 155
Fundamental theorem of calculus, 321, 333

Hahn-Banach Theorem, 256
Helly's selection principle, 521
Homeomorphism, 156, 454

Integration by parts, 321, 333
Isomorphic measure spaces, 521

Lusin Space, 521

Minikowski functional, 255

Neighborhood, **453**
  base, 453
  open, 453

Open, *see* Sets
Open cover, 223, 456

Polish Spaces, 493
Product topology, 457

Radon Measure
  Complex, 515
  Signed, 515
Reflexive, *see* Banach space

Sets
  closed, 453

Index

open, 453
Sub-base, *see* Topology
Summable, **19**

Topological Space, **453**
Topology, **453**
base, 454
discrete, 453
generated by functions, 457
induced / relative, 455
relative / induced, 455
sub-base, 454
trivial, 453
Total variation, 316

version, 251