22. Banach Spaces III: Calculus

In this section, X and Y will be Banach space and U will be an open subset of X.

Notation 22.1 (ϵ, O, and o notation)**.** Let $0 \in U \subset_o X$, and $f : U \longrightarrow Y$ be a function. We will write:

(1) $f(x) = \epsilon(x)$ 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 $\limsup_{x\to 0} \frac{\|f(x)\|}{\|x\|} < \infty$, where

$$\limsup_{x\to 0} \frac{\|f(x)\|}{\|x\|} \equiv \lim_{r\downarrow 0} \sup\{\|f(x)\| : 0 < \|x\| \le r\}.$$

(3) $f(x) = o(x)$ if $f(x) = \epsilon(x)O(x)$, i.e. $\lim_{x\to 0} \|f(x)\|/\|x\| = 0$.

Example 22.2. 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 $f(x_0 + h) = f(x_0) + \epsilon(h)$.
(2) If $f(x) = \epsilon(x)$ and $g(x) = \epsilon(x)$ then $f(x) + g(x) = \epsilon(x)$.
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 $f(x) = \epsilon(x)$ and $g(y) = \epsilon(y)$ then $g \circ f(x) = \epsilon(x)$.

22.1. **The Differential.**

Definition 22.3. A function $f : U \subset_o X \to Y$ is **differentiable** at $x_0 + h_0 \in U$ if there exists a linear transformation $\Lambda \in L(X,Y)$ such that

$$f(x_0 + h) - f(x_0 + h_0) - \Lambda h = o(h). \tag{22.1}$$

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 22.4. The linear transformation Λ in Definition 22.3 is necessarily unique. Indeed if Λ_1 is another linear transformation such that Eq. (22.1) 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 22.1. Show that a function $f : (a,b) \to X$ is a differentiable at $t \in (a,b)$ in the sense of Definition 4.6 iff it is differentiable in the sense of Definition 22.3. Also show $Df(t)v = v\dot{f}(t)$ for all $v \in \mathbb{R}$.

Example 22.5. Assume that $GL(X,Y)$ is non-empty. Then $f : GL(X,Y) \to GL(Y,X)$ defined by $f(A) \equiv A^{-1}$ is differentiable and

$$f'(A)B = -A^{-1}BA^{-1} \text{ for all } B \in L(X,Y).$$

Indeed (by Eq. (3.13)),

$$\begin{aligned} f(A+H) - f(A) &= (A+H)^{-1} - A^{-1} = (A\left(I + A^{-1}H\right))^{-1} - A^{-1} \\ &= \left(I + A^{-1}H\right))^{-1}A^{-1} - A^{-1} = \sum_{n=0}^{\infty}(-A^{-1}H)^n \cdot A^{-1} - A^{-1} \\ &= -A^{-1}HA^{-1} + \sum_{n=2}^{\infty}(-A^{-1}H)^n. \end{aligned}$$

Since

$$\|\sum_{n=2}^{\infty}(-A^{-1}H)^n\| \le \sum_{n=2}^{\infty}\|A^{-1}H\|^n \le \frac{\|A^{-1}\|^2\|H\|^2}{1-\|A^{-1}H\|},$$

we find that

$$f(A+H) - f(A) = -A^{-1}HA^{-1} + o(H).$$

22.2. **Product and Chain Rules.** The following theorem summarizes some basic properties of the differential.

Theorem 22.6. *The differential D has the following properties:*

Linearity: *D is linear, i.e. $D(f+\lambda g) = Df + \lambda Dg$.*

Product Rule: *If $f : U \subset_o X \to Y$ and $A : U \subset_o X \to L(X,Z)$ are differentiable at x_0 then so is $x \to (Af)(x) \equiv A(x)f(x)$ and*

$$D(Af)(x_0)h = (DA(x_0)h)f(x_0) + A(x_0)Df(x_0)h.$$

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 $y_0 \equiv f(h_o)$, then $g \circ f$ is differentiable at x_0 and $(g \circ f)'(x_0) = g'(y_0)f'(x_0)$.*

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 $y_0 \equiv f(h_o)$, $g'(y_0)$ is invertible, and $g \circ f$ is differentiable at x_0, then f is differentiable at x_0 and*

$$f'(x_0) \equiv [g'(x_0)]^{-1}(g \circ f)'(x_0). \tag{22.2}$$

Proof. For the proof of linearity, let $f, g : U \subset_o X \to Y$ be two functions which are differentiable at $x_0 \in U$ and $c \in \mathbb{R}$, then

$$\begin{aligned} (f+cg)(x_0+h) &= f(x_0) + Df(x_0)h + o(h) + c(g(x_0) + Dg(x_0)h + o(h) \\ &= (f+cg)(x_0) + (Df(x_0) + cDg(x_0))h + o(h), \end{aligned}$$

which implies that $(f+cg)$ is differentiable at x_0 and that

$$D(f+cg)(x_0) = Df(x_0) + cDg(x_0).$$

For item 2, we have

$$\begin{aligned} A(x_0+h)f(x_0+h) &= (A(x_0) + DA(x_0)h + o(h))(f(x_0) + f'(x_0)h + o(h)) \\ &= A(x_0)f(x_0) + A(x_0)f'(x_0)h + [DA(x_0)h]f(x_0) + o(h), \end{aligned}$$

which proves item 2.

Similarly for item 3,

$$\begin{aligned}(g\circ f)(x_0+h) &= g(f(x_0))+g'(f(x_0))(f(x_0+h)-f(x_0))+o(f(x_0+h)-f(x_0))\\ &= g(f(x_0))+g'(f(x_0))(Df(x_0)x_0+o(h))+o(f(x_0+h)-f(x_0)\\ &= g(f(x_0))+g'(f(x_0))Df(x_0)h+o(h),\end{aligned}$$

where in the last line we have used the fact that $f(x_0+h)-f(x_0)=O(h)$ (see Eq. (22.1)) and $o(O(h))=o(h)$.

Item 4. Since g is differentiable at $y_0=f(x_0)$,

$$g(f(x_0+h))-g(f(x_0))=g'(f(x_0))(f(x_0+h)-f(x_0))+o(f(x_0+h)-f(x_0)).$$

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

$$\begin{aligned}f(x_0+h)-f(x_0) &= g'(f(x_0))^{-1}\{(g\circ f)'(x_0)h+o(h)-o(f(x_0+h)-f(x_0))\}\\ &= g'(f(x_0))^{-1}(g\circ f)'(x_0)h+o(h)\\ &\quad -g'(f(x_0))^{-1}o(f(x_0+h)-f(x_0)).\end{aligned}\tag{22.3}$$

Using the continuity of f, $f(x_0+h)-f(x_0)$ is close to 0 if h is close to zero, and hence $\|o(f(x_0+h)-f(x_0))\|\le\frac{1}{2}\|f(x_0+h)-f(x_0)\|$ for all h sufficiently close to 0. (We may replace $\frac{1}{2}$ by any number $\alpha>0$ above.) Using this remark, we may take the norm of both sides of equation (22.3) to find

$$\|f(x_0+h)-f(x_0)\|\le\|g'(f(x_0))^{-1}(g\circ f)'(x_0)\|\|h\|+o(h)+\frac{1}{2}\|f(x_0+h)-f(x_0)\|$$

for h close to 0. Solving for $\|f(x_0+h)-f(x_0)\|$ in this last equation shows that

$$f(x_0+h)-f(x_0)=O(h).\tag{22.4}$$

(This is an improvement, since the continuity of f only guaranteed that $f(x_0+h)-f(x_0)=\epsilon(h)$.) Because of Eq. (22.4), we now know that $o(f(x_0+h)-f(x_0))=o(h)$, which combined with Eq. (22.3) 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 22.7. *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*

$$d(f\circ\sigma)(t)/dt=f'(\sigma(t))\dot\sigma(t).$$

Example 22.8. Let us continue on with Example 22.5 but now let $X=Y$ to simplify the notation. So $f:GL(X)\to GL(X)$ is the map $f(A)=A^{-1}$ and

$$f'(A)=-L_{A^{-1}}R_{A^{-1}},\text{ i.e. }f'=-L_fR_f.$$

where $L_AB=AB$ and $R_AB=AB$ 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

$$f''(A)B=-L_{f'(A)B}R_f-L_fR_{f'(A)B}.$$

More explicitly

$$[f''(A)B]\,C = A^{-1}BA^{-1}CA^{-1} + A^{-1}CA^{-1}BA^{-1}. \tag{22.5}$$

Working inductively one shows $f : GL(X) \to GL(X)$ defined by $f(A) \equiv A^{-1}$ is C^{∞}.

22.3. **Partial Derivatives.**

Definition 22.9 (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 22.7.

Proposition 22.10. *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 $x_0 \in U$, $\epsilon > 0$ such that $B(x_0, 2\epsilon) \subset U$ and $M \equiv \sup\{\|A(x)\| : x \in B(x_0, 2\epsilon)\} < \infty$[43]. For $x \in B(x_0, \epsilon) \cap D$ and $v \in D \cap B(0, \epsilon)$, by the fundamental theorem of calculus,

$$f(x + v) - f(x) = \int_0^1 \frac{df(x + tv)}{dt}\, dt = \int_0^1 (\partial_v f)(x + tv)\, dt = \int_0^1 A(x + tv)\, v\, dt. \tag{22.6}$$

For general $x \in B(x_0, \epsilon)$ and $v \in B(0, \epsilon)$, choose $x_n \in B(x_0, \epsilon) \cap D$ and $v_n \in D \cap B(0, \epsilon)$ such that $x_n \to x$ and $v_n \to v$. Then

$$f(x_n + v_n) - f(x_n) = \int_0^1 A(x_n + tv_n)\, v_n\, dt \tag{22.7}$$

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. (22.7) we have

$$\|\int_0^1 A(x + tv)\, v\, dt - \int_0^1 A(x_n + tv_n)\, v_n\, dt\| \le \int_0^1 \|A(x + tv) - A(x_n + tv_n)\,\|\|v\|\, dt + M\|v - v_n\|.$$

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. (22.7) converges to $\int_0^1 A(x + tv)\, v\, dt$. Hence Eq. (22.6) is valid for all $x \in B(x_0, \epsilon)$ and $v \in B(0, \epsilon)$. We also see that

$$f(x + v) - f(x) - A(x)v = \epsilon(v)v, \tag{22.8}$$

---

[43]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 counter example. 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 $n \to \infty$.

where $\epsilon(v) \equiv \int_0^1 [A(x+tv) - A(x)]\,dt$. Now

$$\|\epsilon(v)\| \le \int_0^1 \|A(x+tv) - A(x)\|\,dt \le \max_{t\in[0,1]} \|A(x+tv) - A(x)\| \to 0 \text{ as } v \to 0,$$

by the continuity of A. Thus, we have shown that f is differentiable and that $Df(x) = A(x)$. ■

22.4. **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 22.11. *If $Z \in C(J\times U, X)$ such that $D_xZ(t,x)$ exists for all $(t,x) \in J\times U$ and $D_xZ(t,x) \in C(J\times U, X)$ then Z is locally Lipschitz in x, see Definition 5.12.*

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 $\epsilon_t > 0$ such that $B(x,\epsilon_t) \subset U$ and

$$\sup\{\|D_xZ(t',x')\| : (t',x') \in N_t \times B(x,\epsilon_t)\} < \infty.$$

By the compactness of I, there exists a finite subset $\Lambda \subset I$ such that $I \subset \cup_{t\in I} N_t$. Let $\epsilon(x,I) := \min\{\epsilon_t : t \in \Lambda\}$ and

$$K(x,I) \equiv \sup\{\|DZ(t,x')\|(t,x') \in I \times B(x,\epsilon(x,I))\} < \infty.$$

Then by the fundamental theorem of calculus and the triangle inequality,

$$\|Z(t,x_1) - Z(t,x_0)\| \le \left(\int_0^1 \|D_xZ(t,x_0+s(x_1-x_0)\|\,ds\right)\|x_1-x_0\| \le K(x,I)\|x_1-x_0\|$$

for all $x_0, x_1 \in B(x,\epsilon(x,I))$ and $t \in I$. ■

Theorem 22.12 (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 $D_xZ \in C(\mathbb{R}\times U, X)$ and $\phi : \mathcal{D}(Z) \subset \mathbb{R}\times X \to X$ denote the maximal solution operator to the ordinary differential equation*

$$\dot{y}(t) = Z(t,y(t)) \text{ with } y(0) = x \in U, \tag{22.9}$$

see Notation 5.15 and Theorem 5.21. Then $\phi \in C^1(\mathcal{D}(Z),U)$, $\partial_t D_x\phi(t,x)$ exists and is continuous for $(t,x) \in \mathcal{D}(Z)$ and $D_x\phi(t,x)$ satisfies the linear differential equation,

$$\frac{d}{dt}D_x\phi(t,x) = [(D_xZ)(t,\phi(t,x))]D_x\phi(t,x) \text{ with } D_x\phi(0,x) = I_X \tag{22.10}$$

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

$$\mathcal{O}_\epsilon := \{y \in BC(J,U) : \|y - y_0\|_\infty < \epsilon\} \subset_o BC(J,X).$$

By Lemma 22.11, Z is locally Lipschitz and therefore Theorem 5.21 is applicable. By Eq. (5.30) of Theorem 5.21, there exists $\epsilon > 0$ and $\delta > 0$ such that $G : B(x_0,\delta) \to \mathcal{O}_\epsilon$ defined by $G(x) \equiv \phi(\cdot,x)|_J$ is continuous. By Lemma 22.13 below, for $\epsilon > 0$ sufficiently small the function $F : \mathcal{O}_\epsilon \to BC(J,X)$ defined by

$$F(y) \equiv y - \int_0^{\cdot} Z(t,y(t))dt. \tag{22.11}$$

is C^1 and

$$DF(y)v = v - \int_0^{\cdot} D_y Z(t, y(t))v(t)dt. \tag{22.12}$$

By the existence and uniqueness Theorem 5.5 for linear ordinary differential equations, $DF(y)$ 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 22.6 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

$$D_x\phi(t,x) - \int_0^t [DZ(\phi(\tau,x)]D_x\phi(\tau,x)\,d\tau = I_X.$$

As usual this equation implies $D_x\phi(t,x)$ is differentiable in t, $D_x\phi(t,x)$ is continuous in (t,x) and $D_x\phi(t,x)$ satisfies Eq. (22.10). ■

Lemma 22.13. *Continuing the notation used in the proof of Theorem 22.12 and further let*

$$f(y) \equiv \int_0^{\cdot} Z(\tau, y(\tau))\,d\tau \text{ for } y \in \mathcal{O}_\epsilon.$$

Then $f \in C^1(\mathcal{O}_\epsilon, Y)$ and for all $y \in \mathcal{O}_\epsilon$,

$$f'(y)h = \int_0^{\cdot} D_xZ(\tau, y(\tau))h(\tau)\,d\tau =: \Lambda_y h.$$

Proof. Let $h \in Y$ be sufficiently small and $\tau \in J$, then by fundamental theorem of calculus,

$$Z(\tau, y(\tau)+h(\tau)) - Z(\tau,y(\tau)) = \int_0^1 [D_xZ(\tau, y(\tau)+rh(\tau)) - D_xZ(\tau,y(\tau))]dr$$

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,

$$\|(f(y+h) - f(y) - \Lambda_y h)\|_\infty \le \|h\|_\infty \delta(h) \tag{22.13}$$

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 aide of Lemmas 22.11 and Lemma 5.13,

$$(r,\tau,h) \in [0,1]\times J \times Y \to \|D_xZ(\tau,y(\tau)+rh(\tau))\|$$

is bounded for small h provided $\epsilon > 0$ is sufficiently small. Thus it follows from the dominated convergence theorem that $\delta(h) \to 0$ as $h \to 0$ and hence Eq. (22.13) implies $f'(y)$ exists and is given by Λ_y. Similarly,

$$\|f'(y+h) - f'(y)\|_{op} \le \int_J \|D_xZ(\tau,y(\tau)+h(\tau)) - D_xZ(\tau,y(\tau))\|\,d\tau \to 0 \text{ as } h \to 0$$

showing f' is continuous. ■

Remark 22.14. If $Z \in C^k(U,X)$, then an inductive argument shows that $\phi \in C^k(\mathcal{D}(Z),X)$. For example if $Z \in C^2(U,X)$ then $(y(t),u(t)) := (\phi(t,x), D_x\phi(t,x))$ solves the ODE,

$$\frac{d}{dt}(y(t),u(t)) = \tilde{Z}\left((y(t),u(t))\right) \text{ with } (y(0),u(0)) = (x, Id_X)$$

where $\tilde{Z}$ is the C^1 – vector field defined by

$$\tilde{Z}\left(x,u\right) = \left(Z(x), D_x Z(x) u\right).$$

Therefore Theorem 22.12 may be applied to this equation to deduce: $D_x^2\phi(t,x)$ and $D_x^2\dot{\phi}(t,x)$ exist and are continuous. We may now differentiate Eq. (22.10) to find $D_x^2\phi(t,x)$ satisfies the ODE,

$$\frac{d}{dt}D_x^2\phi(t,x) = [\left(\partial_{D_x\phi(t,x)} D_x Z\right)(t,\phi(t,x))]D_x\phi(t,x) + [\left(D_x Z\right)(t,\phi(t,x))]D_x^2\phi(t,x)$$

with $D_x^2\phi(0,x) = 0$.

22.5. **Higher Order Derivatives.** As above, let $f : U \subset_o X \longrightarrow Y$ be a function. If f is differentiable on U, then the differential Df of f is a function from U to the Banach space $L(X,Y)$. If the function $Df : U \longrightarrow L(X,Y)$ is also differentiable on U, then its differential $D^2 f = D(Df) : U \longrightarrow L(X, L(X,Y))$. Similarly, $D^3 f = D(D(Df)) : U \longrightarrow L(X, L(X, L(X,Y)))$ if the differential of $D(Df)$ exists. In general, let $\mathcal{L}^1(X,Y) \equiv L(X,Y)$ and $\mathcal{L}^k(X,Y)$ be defined inductively by $\mathcal{L}^{k+1}(X,Y) = L(X, \mathcal{L}^k(X,Y))$. Then $(D^k f)(x) \in \mathcal{L}^k(X,Y)$ if it exists. It will be convenient to identify the space $\mathcal{L}^k(X,Y)$ with the Banach space defined in the next definition.

Definition 22.15. For $k \in \{1,2,3,\ldots\}$, let $M_k(X,Y)$ denote the set of functions $f : X^k \longrightarrow Y$ such that

(1) For $i \in \{1,2,\ldots,k\}$, $v \in X \longrightarrow f\langle v_1, v_2, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_k\rangle \in Y$ is linear [44] for all $\{v_i\}_{i=1}^n \subset X$.
(2) The norm $\|f\|_{M_k(X,Y)}$ should be finite, where

$$\|f\|_{M_k(X,Y)} \equiv \sup\{\frac{\|f\langle v_1, v_2, \ldots, v_k\rangle\|_Y}{\|v_1\|\|v_2\|\cdots\|v_k\|} : \{v_i\}_{i=1}^k \subset X \setminus \{0\}\}.$$

Lemma 22.16. *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 $(j_2 A)\langle v_1, v_2\rangle = (Av_1)v_2$ and if $A \in \mathcal{L}^3(X,Y) = L(X, L(X, L(X,Y)))$, $(j_3 A)\langle v_1, v_2, v_3\rangle = ((Av_1)v_2)v_3$ 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 $A \in \mathcal{L}^{k+1}(X,Y)$,

[44] I will routinely write $f\langle v_1, v_2, \ldots, v_k\rangle$ rather than $f(v_1, v_2, \ldots, v_k)$ when the function f depends on each of variables linearly, i.e. f is a multi-linear function.

$$\begin{aligned}\|j_{k+1}A\|_{M_{k+1}(X,Y)} &\equiv \sup\{\frac{\|(j_k(Av_0))\langle v_1,v_2,\dots,v_k\rangle\|_Y}{\|v_0\|\|v_1\|\|v_2\|\cdots\|v_k\|} : \{v_i\}_{i=0}^k \subset X\setminus\{0\}\} \\ &\equiv \sup\{\frac{\|(j_k(Av_0))\|_{M_k(X,Y)}}{\|v_0\|} : v_0\in X\setminus\{0\}\} \\ &= \sup\{\frac{\|Av_0\|_{\mathcal{L}^k(X,Y)}}{\|v_0\|} : v_0\in X\setminus\{0\}\} \\ &= \|A\|_{L(X,\mathcal{L}^k(X,Y))} \equiv \|A\|_{\mathcal{L}^{k+1}(X,Y)},\end{aligned}$$

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 shows 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^kf as function on U with values in $M_k(X,Y)$.

Theorem 22.17 (Differentiability). *Suppose $k\in\{1,2,\dots\}$ and D is a dense subspace of X, $f:U\subset_o X\longrightarrow Y$ 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 $l=1,2,\dots k$. Further assume there exists continuous functions $A_l:U\subset_o X\longrightarrow M_l(X,Y)$ such that such that $(\partial_{v_1}\partial_{v_2}\cdots\partial_{v_l}f)(x)=A_l(x)\langle v_1,v_2,\dots,v_l\rangle$ for all $x\in D\cap U$, $\{v_i\}_{i=1}^l\subset D$, and $l=1,2,\dots k$. Then $D^lf(x)$ exists and is equal to $A_l(x)$ for all $x\in U$ and $l=1,2,\dots,k$.*

Proof. We will prove the theorem by induction on k. We have already proved the theorem when $k=1$, see Proposition 22.10. Now suppose that $k>1$ and that the statement of the theorem holds when k is replaced by $k-1$. Hence we know that $D^lf(x)=A_l(x)$ for all $x\in U$ and $l=1,2,\dots,k-1$. We are also given that

$$(\partial_{v_1}\partial_{v_2}\cdots\partial_{v_k}f)(x)=A_k(x)\langle v_1,v_2,\dots,v_k\rangle\quad\forall x\in U\cap D,\{v_i\}\subset D. \tag{22.14}$$

Now we may write $(\partial_{v_2}\cdots\partial_{v_k}f)(x)$ as $(D^{k-1}f)(x)\langle v_2,v_3,\dots,v_k\rangle$ so that Eq. (22.14) may be written as

$$\partial_{v_1}(D^{k-1}f)(x)\langle v_2,v_3,\dots,v_k\rangle)=A_k(x)\langle v_1,v_2,\dots,v_k\rangle\quad\forall x\in U\cap D,\{v_i\}\subset D. \tag{22.15}$$

So by the fundamental theorem of calculus, we have that

$$((D^{k-1}f)(x+v_1)-(D^{k-1}f)(x))\langle v_2,v_3,\dots,v_k\rangle=\int_0^1 A_k(x+tv_1)\langle v_1,v_2,\dots,v_k\rangle\,dt \tag{22.16}$$

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 22.10, Eq. (22.16) 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,

$$(D^{k-1}f)(x+v_1)-(D^{k-1}f)(x)=\int_0^1 A_k(x+tv_1)\langle v_1,\cdots\rangle\,dt. \tag{22.17}$$

Hence

$$(D^{k-1}f)(x+v_1)-(D^{k-1}f)(x)-A_k(x)\langle v_1,\cdots\rangle = \int_0^1 [A_k(x+tv_1)-A_k(x)]\langle v_1,\cdots\rangle\, dt$$

from which we get the estimate,

$$\|(D^{k-1}f)(x+v_1)-(D^{k-1}f)(x)-A_k(x)\langle v_1,\cdots\rangle\| \le \epsilon(v_1)\|v_1\| \tag{22.18}$$

where $\epsilon(v_1) \equiv \int_0^1 \|A_k(x+tv_1)-A_k(x)\|\, dt$. Notice by the continuity of A_k that $\epsilon(v_1) \longrightarrow 0$ as $v_1 \longrightarrow 0$. Thus it follow from Eq. (22.18) that $D^{k-1}f$ is differentiable and that $(D^k f)(x) = A_k(x)$. ■

Example 22.18. Let $f : L^*(X,Y) \longrightarrow L^*(Y,X)$ be defined by $f(A) \equiv A^{-1}$. We assume that $L^*(X,Y)$ is not empty. Then f is infinitely differentiable and

$$(D^k f)(A)\langle V_1, V_2, \ldots, V_k\rangle = (-1)^k \sum_\sigma \{B^{-1}V_{\sigma(1)}B^{-1}V_{\sigma(2)}B^{-1}\cdots B^{-1}V_{\sigma(k)}B^{-1}\}, \tag{22.19}$$

where sum is over all permutations of σ of $\{1,2,\ldots,k\}$.

Let me check Eq. (22.19) 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 $A_2 : L^*(X,Y) \longrightarrow M_2(L(X,Y), L(Y,X))$. Also

$$\begin{aligned}
\|(A_2(B)-A_2(C))\langle V_1,V_2\rangle\| &\le 2\|B^{-1}V_2B^{-1}V_1B^{-1} - C^{-1}V_2C^{-1}V_1C^{-1}\| \\
&\le 2\|B^{-1}V_2B^{-1}V_1B^{-1} - B^{-1}V_2B^{-1}V_1C^{-1}\| \\
&\quad + 2\|B^{-1}V_2B^{-1}V_1C^{-1} - B^{-1}V_2C^{-1}V_1C^{-1}\| \\
&\quad\quad + 2\|B^{-1}V_2C^{-1}V_1C^{-1} - C^{-1}V_2C^{-1}V_1C^{-1}\| \\
&\le 2\|B^{-1}\|^2\|V_2\|\|V_1\|\|B^{-1}-C^{-1}\| \\
&\quad + 2\|B^{-1}\|\|C^{-1}\|\|V_2\|\|V_1\|\|B^{-1}-C^{-1}\| \\
&\quad\quad + 2\|C^{-1}\|^2\|V_2\|\|V_1\|\|B^{-1}-C^{-1}\|.
\end{aligned}$$

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 $B \longrightarrow B^{-1}$ is differentiable and hence continuous, it follows that $A_2(B)$ is also continuous in B. Hence by Theorem 22.17 $D^2f(A)$ exists and is given as in Eq. (22.19)

Example 22.19. Suppose that $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a $C^\infty-$ function and $F(x) \equiv \int_0^1 f(x(t))\, dt$ for $x \in X \equiv C([0,1],\mathbb{R})$ equipped with the norm $\|x\| \equiv \max_{t\in[0,1]}|x(t)|$. Then $F : X \longrightarrow \mathbb{R}$ is also infinitely differentiable and

$$(D^kF)(x)\langle v_1, v_2, \ldots, v_k\rangle = \int_0^1 f^{(k)}(x(t))v_1(t)\cdots v_k(t)\, dt, \tag{22.20}$$

for all $x \in X$ and $\{v_i\} \subset X$.

To verify this example, notice that

$$(\partial_v F)(x) \equiv \frac{d}{ds}|_0 F(x+sv) = \frac{d}{ds}|_0 \int_0^1 f(x(t)+sv(t))\,dt$$
$$= \int_0^1 \frac{d}{ds}|_0 f(x(t)+sv(t))\,dt = \int_0^1 f'(x(t))v(t)\,dt.$$

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 $x, y \in X$,

$$|A_k(x)\langle v_1, v_2, \dots, v_k\rangle - A_k(y)\langle v_1, v_2, \dots, v_k\rangle| \le \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))| \cdot |v_1(t)\cdots v_k(t)\,|dt$$
$$\le \prod_{i=1}^k \|v_i\| \int_0^1 |f^{(k)}(x(t)) - f^{(k)}(y(t))|dt,$$

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 22.17 to conclude that Eq. (22.20) is valid.

22.6. **Contraction Mapping Principle.**

Theorem 22.20. *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 $x \equiv \lim_{n\to\infty} x_n$ 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 $\alpha \in (0,1)$),

$$\rho(x_m, x_n) \leq \sum_{k=n}^{m-1} \alpha^k \rho(x_1, x_0) \leq \rho(x_1, x_0)\alpha^n \sum_{l=0}^{\infty} \alpha^l = \rho(x_1, x_0)\frac{\alpha^n}{1-\alpha}.$$

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 22.21 (Contraction Mapping Principle II)**.** *Suppose that (X, ρ) is a complete metric space and $S : X \longrightarrow X$ is a continuous map such that $S^{(n)}$ is a contraction for some $n \in \mathbb{N}$. Here*

$$S^{(n)} \equiv \overbrace{S \circ S \circ \ldots \circ S}^{n \text{ times}}$$

and we are assuming there exists $\alpha \in (0,1)$ such that $\rho(S^{(n)}(x), S^{(n)}(y)) \leq \alpha\rho(x, y)$ for all $x, y \in X$. Then S has a unique fixed point in X.

Proof. Let $T \equiv S^{(n)}$, then $T : X \longrightarrow X$ 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 22.22. *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: $\rho(S_z^{(n)}(x), S_z^{(n)}(y)) \leq \alpha\rho(x, y)$ *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 22.21 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 $T_z \equiv S_z^{(n)}$. If $z, w \in Z$, then

$$\begin{aligned}\rho(G(z), G(w)) &= \rho(T_z(G(z)), T_w(G(w))) \\ &\leq \rho(T_z(G(z)), T_w(G(z))) + \rho(T_w(G(z)), T_w(G(w))) \\ &\leq \rho(T_z(G(z)), T_w(G(z))) + \alpha\rho(G(z), G(w)).\end{aligned}$$

Solving this inequality for $\rho(G(z), G(w))$ gives

$$\rho(G(z), G(w)) \leq \frac{1}{1-\alpha}\rho(T_z(G(z)), T_w(G(z))).$$

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. ■

22.7. **Inverse and Implicit Function Theorems.** In this section, let X be a Banach space, $U \subset X$ be an open set, and $F : U \to X$ and $\epsilon : U \to X$ be continuous functions. **Question**: under what conditions on ϵ is $F(x) := x + \epsilon(x)$ a homeomorphism from $B_0(\delta)$ to $F(B_0(\delta))$ for some small $\delta > 0$? Let's start by looking at the one dimensional case first. So for the moment assume that $X = \mathbb{R}$, $U = (-1, 1)$, and $\epsilon : U \to \mathbb{R}$ is C^1. Then F will be one to one iff F is monotonic. This will be the case, for example, if $F' = 1 + \epsilon' > 0$. This in turn is guaranteed by assuming that $|\epsilon'| \le \alpha < 1$. (This last condition makes sense on a Banach space whereas assuming $1 + \epsilon' > 0$ is not as easily interpreted.)

Lemma 22.23. *Suppose that $U = B = B(0, r)$ $(r > 0)$ is a ball in X and $\epsilon : B \longrightarrow X$ is a C^1 function such that $\|D\epsilon\| \le \alpha < \infty$ on U. Then for all $x, y \in U$ we have:*

$$\|\epsilon(x) - \epsilon(y)\| \le \alpha\|x - y\|. \tag{22.21}$$

Proof. By the fundamental theorem of calculus and the chain rule:

$$\begin{aligned}\epsilon(y) - \epsilon(x) &= \int_0^1 \frac{d}{dt}\epsilon(x + t(y - x))dt \\ &= \int_0^1 [D\epsilon(x + t(y - x))](y - x)dt.\end{aligned}$$

Therefore, by the triangle inequality and the assumption that $\|D\epsilon(x)\| \le \alpha$ on B,

$$\|\epsilon(y) - \epsilon(x)\| \le \int_0^1 \|D\epsilon(x + t(y - x))\|dt \cdot \|(y - x)\| \le \alpha\|(y - x)\|.$$

■

Remark 22.24. It is easily checked that if $\epsilon : B = B(0, r) \to X$ is C^1 and satisfies (22.21) then $\|D\epsilon\| \le \alpha$ on B.

Using the above remark and the analogy to the one dimensional example, one is lead to the following proposition.

Proposition 22.25. *Suppose that $U = B = B(0, r)$ $(r > 0)$ is a ball in X, $\alpha \in (0, 1)$, $\epsilon : U \to X$ is continuous, $F(x) \equiv x + \epsilon(x)$ for $x \in U$, and ϵ satisfies:*

$$\|\epsilon(x) - \epsilon(y)\| \le \alpha\|x - y\| \quad \forall x, y \in B. \tag{22.22}$$

Then $F(B)$ is open in X and $F : B \to V := F(B)$ is a homeomorphism.

Proof. First notice from (22.22) that

$$\begin{aligned}\|x - y\| &= \|(F(x) - F(y)) - (\epsilon(x) - \epsilon(y))\| \\ &\le \|F(x) - F(y)\| + \|\epsilon(x) - \epsilon(y)\| \\ &\le \|F(x) - F(y)\| + \alpha\|(x - y)\|\end{aligned}$$

from which it follows that $\|x - y\| \le (1 - \alpha)^{-1}\|F(x) - F(y)\|$. Thus F is injective on B. Let $V \doteq F(B)$ and $G = F^{-1} : V \longrightarrow B$ denote the inverse function which exists since F is injective.

We will now show that V is open. For this let $x_0 \in B$ and $z_0 = F(x_0) = x_0 + \epsilon(x_0) \in V$. We wish to show for z close to z_0 that there is an $x \in B$ such that $F(x) = x + \epsilon(x) = z$ or equivalently $x = z - \epsilon(x)$. Set $S_z(x) \doteq z - \epsilon(x)$, then we are looking for $x \in B$ such that $x = S_z(x)$, i.e. we want to find a fixed point of S_z. We will show that such a fixed point exists by using the contraction mapping theorem.

Step 1. S_z is contractive for all $z \in X$. In fact for $x, y \in B$,

$$\|S_z(x) - S_z(y)\| = \|\epsilon(x) - \epsilon(y))\| \le \alpha\|x - y\|. \tag{22.23}$$

Step 2. For any $\delta > 0$ such the $C \doteq \overline{B(x_0, \delta)} \subset B$ and $z \in X$ such that $\|z - z_0\| < (1-\alpha)\delta$, we have $S_z(C) \subset C$. Indeed, let $x \in C$ and compute:

$$\begin{aligned}\|S_z(x) - x_0\| &= \|S_z(x) - S_{z_0}(x_0)\| \\ &= \|z - \epsilon(x) - (z_0 - \epsilon(x_0))\| \\ &= \|z - z_0 - (\epsilon(x) - \epsilon(x_0))\| \\ &\le \|z - z_0\| + \alpha\|x - x_0\| \\ &< (1-\alpha)\delta + \alpha\delta = \delta.\end{aligned}$$

wherein we have used $z_0 = F(x_0)$ and (22.22).

Since C is a closed subset of a Banach space X, we may apply the contraction mapping principle, Theorem 22.20 and Lemma 22.22, to S_z to show there is a continuous function $G : B(z_0, (1-\alpha)\delta) \to C$ such that

$$G(z) = S_z(G(z)) = z - \epsilon(G(z)) = z - F(G(z)) + G(z),$$

i.e. $F(G(z)) = z$. This shows that $B(z_0, (1-\alpha)\delta) \subset F(C) \subset F(B) = V$. That is z_0 is in the interior of V. Since $F^{-1}|_{B(z_0,(1-\alpha)\delta)}$ is necessarily equal to G which is continuous, we have also shown that F^{-1} is continuous in a neighborhood of z_0. Since $z_0 \in V$ was arbitrary, we have shown that V is open and that $F^{-1} : V \to U$ is continuous. ■

Theorem 22.26 (Inverse Function Theorem)**.** *Suppose X and Y are Banach spaces, $U \subset_o X$, $f \in C^k(U \to X)$ with $k \ge 1$, $x_0 \in U$ and $Df(x_0)$ is invertible. Then there is a ball $B = B(x_0, r)$ in U centered at x_0 such that*

1. *$V = f(B)$ is open,*
2. *$f|_B : B \to V$ is a homeomorphism,*
3. *$g \doteq (f|_B)^{-1} \in C^k(V, B)$ and*

$$g'(y) = [f'(g(y))]^{-1} \text{ for all } y \in V. \tag{22.24}$$

Proof. Define $F(x) \equiv [Df(x_0)]^{-1} f(x + x_0)$ and $\epsilon(x) \equiv x - F(x) \in X$ for $x \in (U - x_0)$. Notice that $0 \in U - x_0$, $DF(0) = I$, and that $D\epsilon(0) = I - I = 0$. Choose $r > 0$ such that $\tilde{B} \equiv B(0, r) \subset U - x_0$ and $\|D\epsilon(x)\| \le \frac{1}{2}$ for $x \in \tilde{B}$. By Lemma 22.23, ϵ satisfies (22.23) with $\alpha = 1/2$. By Proposition 22.25, $F(\tilde{B})$ is open and $F|_{\tilde{B}} : \tilde{B} \to F(\tilde{B})$ is a homeomorphism. Let $G \equiv F|_{\tilde{B}}^{-1}$ which we know to be a continuous map from $F(\tilde{B}) \to \tilde{B}$.

Since $\|D\epsilon(x)\| \le 1/2$ for $x \in \tilde{B}$, $DF(x) = I + D\epsilon(x)$ is invertible, see Corollary 3.70. Since $H(z) \doteq z$ is C^1 and $H = F \circ G$ on $F(\tilde{B})$, it follows from the converse to the chain rule, Theorem 22.6, that G is differentiable and

$$DG(z) = [DF(G(z))]^{-1} DH(z) = [DF(G(z))]^{-1}.$$

Since G, DF, and the map $A \in GL(X) \to A^{-1} \in GL(X)$ are all continuous maps, (see Example 22.5) the map $z \in F(\tilde{B}) \to DG(z) \in L(X)$ is also continuous, i.e. G is C^1.

Let $B = \tilde{B} + x_0 = B(x_0, r) \subset U$. Since $f(x) = [Df(x_0)]F(x - x_0)$ and $Df(x_0)$ is invertible (hence an open mapping), $V := f(B) = [Df(x_0)]F(\tilde{B})$ is open in X. It is also easily checked that $f|_B^{-1}$ exists and is given by

$$f|_B^{-1}(y) = x_0 + G([Df(x_0)]^{-1}y) \tag{22.25}$$

for $y \in V = f(B)$. This shows that $f|_B : B \to V$ is a homeomorphism and it follows from (22.25) that $g \doteq (f|_B)^{-1} \in C^1(V, B)$. Eq. (22.24) now follows from the chain rule and the fact that

$$f \circ g(y) = y \text{ for all } y \in B.$$

Since $f' \in C^{k-1}(B, L(X))$ and $i(A) := A^{-1}$ is a smooth map by Example 22.18, $g' = i \circ f' \circ g$ is C^1 if $k \geq 2$, i.e. g is C^2 if $k \geq 2$. Again using $g' = i \circ f' \circ g$, we may conclude g' is C^2 if $k \geq 3$, i.e. g is C^3 if $k \geq 3$. Continuing bootstrapping our way up we eventually learn $g \doteq (f|_B)^{-1} \in C^k(V, B)$ if f is C^k. ■

Theorem 22.27 (Implicit Function Theorem)**.** *Now 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 $D_2 f(x_0, y_0) \equiv D(f(x_0, \cdot))(y_0) : Y \to W$ 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*

$$Du(x) = -D_2 f(x, u(x))^{-1} D_1 f(x, u(x)) \text{ for all } x \in U. \tag{22.26}$$

Proof. Proof of 22.27. 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 $F(x, y) \equiv (x, f(x, y))$ 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 implicit 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 $\pi_2(x, y) \equiv y$ for all $(x, y) \in X \times Y$ and define C^k – function u_0 on U_0 by $u_0(x) \equiv \pi_2 \circ F^{-1}(x, 0)$. 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 $u \equiv u_0|_U$ 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

$$O \equiv \{x \in U | u(x) = u_1(x)\} = \{x \in U | u_0(x) = u_1(x)\}.$$

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 $x \in U$,

$$0 = 0 - 0 = f(\tilde{x}, u_0(\tilde{x})) - f(\tilde{x}, u_1(\tilde{x})) = R(\tilde{x})(u_1(\tilde{x}) - u_0(\tilde{x})) \tag{22.27}$$

where

$$R(\tilde{x}) \equiv \int_0^1 D_2 f((\tilde{x}, u_0(\tilde{x}) + t(u_1(\tilde{x}) - u_0(\tilde{x})))dt. \tag{22.28}$$

From Eq. (22.28) 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[45]. Thus $R(\tilde{x})$ is invertible for all $\tilde{x}$ sufficiently close to x. Using Eq. (22.27), this last remark 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. ■

22.8. **More on the Inverse Function Theorem.** In this section X and Y will denote two Banach spaces, $U \subset_o X$, $k \geq 1$, and $f \in C^k(U, Y)$. Suppose $x_0 \in U$, $h \in X$, and $f'(x_0)$ is invertible, then

$$f(x_0 + h) - f(x_0) = f'(x_0)h + o(h) = f'(x_0)\left[h + \epsilon(h)\right]$$

where

$$\epsilon(h) = f'(x_0)^{-1}\left[f(x_0 + h) - f(x_0)\right] - h = o(h).$$

In fact by the fundamental theorem of calculus,

$$\epsilon(h) = \int_0^1 \left(f'(x_0)^{-1} f'(x_0 + th) - I\right) h dt$$

but we will not use this here.

Let $h, h' \in B^X(0, R)$ and apply the fundamental theorem of calculus to $t \to f(x_0 + t(h' - h))$ to conclude

$$\begin{aligned}\epsilon(h') - \epsilon(h) &= f'(x_0)^{-1}\left[f(x_0 + h') - f(x_0 + h)\right] - (h' - h) \\ &= \left[\int_0^1 \left(f'(x_0)^{-1} f'(x_0 + t(h' - h)) - I\right) dt\right](h' - h).\end{aligned}$$

Taking norms of this equation gives

$$\|\epsilon(h') - \epsilon(h)\| \leq \left[\int_0^1 \left\|f'(x_0)^{-1} f'(x_0 + t(h' - h)) - I\right\| dt\right] \|h' - h\| \leq \alpha \|h' - h\|$$

where

$$\alpha := \sup_{x \in B^X(x_0, R)} \left\|f'(x_0)^{-1} f'(x) - I\right\|_{L(X)}. \tag{22.29}$$

We summarize these comments in the following lemma.

Lemma 22.28. *Suppose $x_0 \in U$, $R > 0$, $f : B^X(x_0, R) \to Y$ be a C^1 – function such that $f'(x_0)$ is invertible, α is as in Eq. (22.29) and $\epsilon \in C^1\left(B^X(0, R), X\right)$ is defined by*

$$f(x_0 + h) = f(x_0) + f'(x_0)\left(h + \epsilon(h)\right). \tag{22.30}$$

Then

$$\|\epsilon(h') - \epsilon(h)\| \leq \alpha \|h' - h\| \text{ for all } h, h' \in B^X(0, R). \tag{22.31}$$

[45] 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 $x \in U_0$.

Furthermore if $\alpha < 1$ (which may be achieved by shrinking R if necessary) then $f'(x)$ is invertible for all $x \in B^X(x_0, R)$ and

$$\sup_{x\in B^X(x_0,R)} \left\|f'(x)^{-1}\right\|_{L(Y,X)} \le \frac{1}{1-\alpha}\left\|f'(x_0)^{-1}\right\|_{L(Y,X)}. \tag{22.32}$$

Proof. It only remains to prove Eq. (22.32), so suppose now that $\alpha < 1$. Then by Proposition 3.69 $f'(x_0)^{-1}f'(x)$ is invertible and

$$\left\|\left[f'(x_0)^{-1}f'(x)\right]^{-1}\right\| \le \frac{1}{1-\alpha} \text{ for all } x \in B^X(x_0, R).$$

Since $f'(x) = f'(x_0)\left[f'(x_0)^{-1}f'(x)\right]$ this implies $f'(x)$ is invertible and

$$\left\|f'(x)^{-1}\right\| = \left\|\left[f'(x_0)^{-1}f'(x)\right]^{-1} f'(x_0)^{-1}\right\| \le \frac{1}{1-\alpha}\left\|f'(x_0)^{-1}\right\| \text{ for all } x \in B^X(x_0, R).$$

■

Theorem 22.29 (Inverse Function Theorem)**.** *Suppose $U \subset_o X$, $k \ge 1$ and $f \in C^k(U, Y)$ such that $f'(x)$ is invertible for all $x \in U$. Then:*

1. *$f : U \to Y$ is an open mapping, in particular $V := f(U) \subset_o Y$.*
2. *If f is injective, then $f^{-1} : V \to U$ is also a C^k – map and*
$$\left(f^{-1}\right)'(y) = \left[f'(f^{-1}(y))\right]^{-1} \text{ for all } y \in V.$$
3. *If $x_0 \in U$ and $R > 0$ such that $\overline{B^X(x_0, R)} \subset U$ and*
$$\sup_{x\in B^X(x_0,R)} \left\|f'(x_0)^{-1}f'(x) - I\right\| = \alpha < 1$$
(which may always be achieved by taking R sufficiently small by continuity of $f'(x)$) then $f|_{B^X(x_0,R)} : B^X(x_0, R) \to f(B^X(x_0, R))$ is invertible and $f|^{-1}_{B^X(x_0,R)} : f\left(B^X(x_0, R)\right) \to B^X(x_0, R)$ is C^k.
4. *Keeping the same hypothesis as in item 3. and letting $y_0 = f(x_0) \in Y$,*
$$f(B^X(x_0, r)) \subset B^Y(y_0, \|f'(x_0)\|(1 + \alpha)r) \text{ for all } r \le R$$
and
$$B^Y(y_0, \delta) \subset f(B^X(x_0, (1 - \alpha)^{-1}\left\|f'(x_0)^{-1}\right\|\delta))$$
for all $\delta < \delta(x_0) := (1 - \alpha) R/\left\|f'(x_0)^{-1}\right\|$.

Proof. Let x_0 and $R > 0$ be as in item 3. above and ϵ be as defined in Eq. (22.30) above, so that for $x, x' \in B^X(x_0, R)$,

$$f(x) = f(x_0) + f'(x_0)\left[(x - x_0) + \epsilon(x - x_0)\right] \text{ and}$$
$$f(x') = f(x_0) + f'(x_0)\left[(x' - x_0) + \epsilon(x' - x_0)\right].$$

Subtracting these two equations implies

$$f(x') - f(x) = f'(x_0)\left[x' - x + \epsilon(x' - x_0) - \epsilon(x - x_0)\right]$$

or equivalently

$$x' - x = f'(x_0)^{-1}\left[f(x') - f(x)\right] + \epsilon(x - x_0) - \epsilon(x' - x_0).$$

Taking norms of this equation and making use of Lemma 22.28 implies

$$\|x' - x\| \le \left\|f'(x_0)^{-1}\right\|\|f(x') - f(x)\| + \alpha\|x' - x\|$$

which implies

$$\|x'-x\| \le \frac{\left\|f'(x_0)^{-1}\right\|}{1-\alpha}\,\|f(x')-f(x)\| \text{ for all } x,x' \in B^X(x_0,R). \tag{22.33}$$

This shows that $f|_{B^X(x_0,R)}$ is injective and that $f|_{B^X(x_0,R)}^{-1} : f\left(B^X(x_0,R)\right) \to B^X(x_0,R)$ is Lipschitz continuous because

$$\left\|f|_{B^X(x_0,R)}^{-1}(y') - f|_{B^X(x_0,R)}^{-1}(y)\right\| \le \frac{\left\|f'(x_0)^{-1}\right\|}{1-\alpha}\,\|y'-y\| \text{ for all } y,y' \in f\left(B^X(x_0,R)\right).$$

Since $x_0 \in X$ was chosen arbitrarily, if we know $f : U \to Y$ is injective, we then know that $f^{-1} : V = f(U) \to U$ is necessarily continuous. The remaining assertions of the theorem now follow from the converse to the chain rule in Theorem 22.6 and the fact that f is an open mapping (as we shall now show) so that in particular $f\left(B^X(x_0,R)\right)$ is open.

Let $y \in B^Y(0,\delta)$, with δ to be determined later, we wish to solve the equation, for $x \in B^X(0,R)$,

$$f(x_0)+y = f(x_0+x) = f(x_0)+f'(x_0)\left(x+\epsilon(x)\right).$$

Equivalently we are trying to find $x \in B^X(0,R)$ such that

$$x = f'(x_0)^{-1}y - \epsilon(x) =: S_y(x).$$

Now using Lemma 22.28 and the fact that $\epsilon(0)=0$,

$$\begin{aligned}\|S_y(x)\| &\le \left\|f'(x_0)^{-1}y\right\| + \|\epsilon(x)\| \le \left\|f'(x_0)^{-1}\right\|\|y\| + \alpha\|x\| \\ &\le \left\|f'(x_0)^{-1}\right\|\delta + \alpha R.\end{aligned}$$

Therefore if we assume δ is chosen so that

$$\left\|f'(x_0)^{-1}\right\|\delta + \alpha R < R, \text{ i.e. } \delta < (1-\alpha)R/\left\|f'(x_0)^{-1}\right\| := \delta(x_0),$$

then $S_y : \overline{B^X(0,R)} \to B^X(0,R) \subset \overline{B^X(0,R)}$.

Similarly by Lemma 22.28, for all $x,z \in \overline{B^X(0,R)}$,

$$\|S_y(x)-S_y(z)\| = \|\epsilon(z)-\epsilon(x)\| \le \alpha\|x-z\|$$

which shows S_y is a contraction on $\overline{B^X(0,R)}$. Hence by the contraction mapping principle in Theorem 22.20, for every $y \in B^Y(0,\delta)$ there exits a unique solution $x \in B^X(0,R)$ such that $x = S_y(x)$ or equivalently

$$f(x_0+x) = f(x_0)+y.$$

Letting $y_0 = f(x_0)$, this last statement implies there exists a unique function $g : B^Y(y_0,\delta(x_0)) \to B^X(x_0,R)$ such that $f(g(y)) = y \in B^Y(y_0,\delta(x_0))$. From Eq. (22.33) it follows that

$$\begin{aligned}\|g(y)-x_0\| &= \|g(y)-g(y_0)\| \\ &\le \frac{\left\|f'(x_0)^{-1}\right\|}{1-\alpha}\,\|f(g(y))-f(g(y_0))\| = \frac{\left\|f'(x_0)^{-1}\right\|}{1-\alpha}\,\|y-y_0\|.\end{aligned}$$

This shows

$$g(B^Y(y_0,\delta)) \subset B^X(x_0,(1-\alpha)^{-1}\left\|f'(x_0)^{-1}\right\|\delta)$$

and therefore

$$B^Y(y_0,\delta) = f\left(g(B^Y(y_0,\delta))\right) \subset f\left(B^X(x_0,(1-\alpha)^{-1}\left\|f'(x_0)^{-1}\right\|\delta)\right)$$

for all $\delta < \delta(x_0)$.

This last assertion implies $f(x_0) \in f(W)^o$ for any $W \subset_o U$ with $x_0 \in W$. Since $x_0 \in U$ was arbitrary, this shows f is an open mapping. ■

22.8.1. *Alternate construction of* g. 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

$$\dot{x}(t) = [f'(x(t))]^{-1} h = Z(h, x(t)) \text{ with } x(0) = x_0 \tag{22.34}$$

where $Z(h,x) = [f'(x(t))]^{-1} h$. Conversely if x solves Eq. (22.34) 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.

22.9. **Applications.** A detailed discussion of the inverse function theorem on Banach and Fréchet 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.

Theorem 22.30 (Hamilton's Theorem on p. 110.)**.** *Let* $p : U := (a,b) \to V := (c,d)$ *be a smooth function with* $p' > 0$ *on* (a,b)*. For every* $g \in C^\infty_{2\pi}(\mathbb{R}, (c,d))$ *there exists a unique function* $y \in C^\infty_{2\pi}(\mathbb{R}, (a,b))$ *such that*

$$\dot{y}(t) + p(y(t)) = g(t).$$

Proof. Let $\tilde{V} := C^0_{2\pi}(\mathbb{R}, (c,d)) \subset_o C^0_{2\pi}(\mathbb{R}, \mathbb{R})$ and

$$\tilde{U} := \left\{ y \in C^1_{2\pi}(\mathbb{R}, \mathbb{R}) : a < y(t) < b \text{ and } c < \dot{y}(t) + p(y(t)) < d \text{ for all } t \right\} \subset_o C^1_{2\pi}(\mathbb{R}, (a,b)).$$

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.

Step 1. The differential of P is given by $P'(y)h = \dot{h} + p'(y)h$, see Exercise 22.7. 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 $h \in C^1_{2\pi}(\mathbb{R}, \mathbb{R})$ 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 implicit function theorem, $P : \tilde{U} \to \tilde{V}$ 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

$$\dot{y}_j(t) = g_j(t) - p(y_j(t)) \subset [\gamma, \delta] - [\gamma, \delta] \tag{22.35}$$

which is a compact subset of $\mathbb{R}$. The Ascoli-Arzela Theorem 3.59 now allows us to assume, by passing to a subsequence if necessary, that y_j is converging uniformly to $y \in C^0_{2\pi}(\mathbb{R}, [\alpha, \beta])$. It now follows that

$$\dot{y}_j(t) = g_j(t) - p(y_j(t)) \to g - p(y)$$

uniformly in t. Hence we concluded 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 $\tilde{V}$. ■

22.10. Exercises.

Exercise 22.2. 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

$$\dot{V}(t) = A(t)V(t) \text{ with } V(0) = I. \tag{22.36}$$

Assuming that $V(t)$ is invertible for all $t \in \mathbb{R}$, show that $V^{-1}(t) \equiv [V(t)]^{-1}$ must solve the differential equation

$$\frac{d}{dt}V^{-1}(t) = -V^{-1}(t)A(t) \text{ with } V^{-1}(0) = I. \tag{22.37}$$

See Exercise 5.14 as well.

Exercise 22.3 (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 $t \in J_{x,w} \to \phi(t, x, w)$ denote the maximal solution to the ODE

$$\dot{y}(t) = Z(y(t), w) \text{ with } y(0) = x \tag{22.38}$$

and

$$\mathcal{D} := \{(t, x, w) \in \mathbb{R} \times U \times V : t \in J_{x,w}\}$$

as in Exercise 5.18.

(1) Prove that ϕ is C^1 and that $D_w\phi(t, x, w)$ solves the differential equation:

$$\frac{d}{dt}D_w\phi(t, x, w) = (D_xZ)(\phi(t, x, w), w)D_w\phi(t, x, w) + (D_wZ)(\phi(t, x, w), w)$$

with $D_w\phi(0, x, w) = 0 \in L(W, X)$. **Hint:** See the hint for Exercise 5.18 with the reference to Theorem 5.21 being replace by Theorem 22.12.

(2) Also show with the aid of Duhamel's principle (Exercise 5.16) and Theorem 22.12 that

$$D_w\phi(t, x, w) = D_x\phi(t, x, w) \int_0^t D_x\phi(\tau, x, w)^{-1}(D_wZ)(\phi(\tau, x, w), w)d\tau$$

Exercise 22.4. (Differential of e^A) Let $f : L(X) \to L^*(X)$ be the exponential function $f(A) = e^A$. Prove that f is differentiable and that

$$Df(A)B = \int_0^1 e^{(1-t)A}Be^{tA}\, dt. \tag{22.39}$$

Hint: Let $B \in L(X)$ and define $w(t, s) = e^{t(A+sB)}$ for all $t, s \in \mathbb{R}$. Notice that

$$dw(t, s)/dt = (A + sB)w(t, s) \text{ with } w(0, s) = I \in L(X). \tag{22.40}$$

Use Exercise 22.3 to conclude that w is C^1 and that $w'(t, 0) \equiv dw(t, s)/ds|_{s=0}$ satisfies the differential equation,

$$\frac{d}{dt}w'(t, 0) = Aw'(t, 0) + Be^{tA} \text{ with } w(0, 0) = 0 \in L(X). \tag{22.41}$$

Solve this equation by Duhamel's principle (Exercise 5.16) and then apply Proposition 22.10 to conclude that f is differentiable with differential given by Eq. (22.39).

Exercise 22.5 (Local ODE Existence)**.** Let S_x be defined as in Eq. (5.22) from the proof of Theorem 5.10. Verify that S_x satisfies the hypothesis of Corollary 22.21. In particular we could have used Corollary 22.21 to prove Theorem 5.10.

Exercise 22.6 (Local ODE Existence Again)**.** Let $J = [-1, 1]$, $Z \in C^1(X, X)$, $Y := C(J, X)$ 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

$$\dot{y}(t) = Z(y(t)) \text{ with } y(0) = x \tag{22.42}$$

has a unique solution for small t and this solution is C^1 in x.

(1) If y solves Eq. (22.42) then y_s solves

$$\dot{y}_s(t) = sZ(y_s(t)) \text{ with } y_s(0) = x$$

or equivalently

$$y_s(t) = x + s\int_0^t Z(y_s(\tau))d\tau. \tag{22.43}$$

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 22.26 to conclude there exists $\epsilon > 0$ and a C^1 map $\phi : (-\epsilon, \epsilon) \times B(x_0, \epsilon) \to Y$ such that

$$F(s, \phi(s, x)) = (s, C_x) \text{ for all } (s, x) \in (-\epsilon, \epsilon) \times B(x_0, \epsilon).$$

(5) Show, for $s \leq \epsilon$ that $y_s(t) := \phi(s, x)(t)$ satisfies Eq. (22.43). Now define $y(t, x) = \phi(\epsilon/2, x)(2t/\epsilon)$ and show $y(t, x)$ solve Eq. (22.42) for $|t| < \epsilon/2$ and $x \in B(x_0, \epsilon)$.

Exercise 22.7. Show P defined in Theorem 22.30 is continuously differentiable and $P'(y)h = \dot{h} + p'(y)h$.