Alireza Salehi Golsefidy

Super-approximation

Contents

Chapter 1

What is super-approximation? 1st try.

1.1 Rough description of strong approximation

To understand the origin of the phrase *super-approximation*, we start with briefly formulating *strong approximation*. Let's start with the case of SL_2, the set of two-by-two matrices with entries in a given unital commutative ring and determinant 1. Strong approximation addresses questions of the following form.

Question 1. *Does the residue module n map $\pi_n : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ induces a surjective map from $\mathrm{SL}_2(\mathbb{Z})$ to $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$?*

In Exercise 3, you can find one approach for giving an affirmative answer to this question. Notice that for every unital commutative ring R, $\mathrm{SL}_2(R)$ can be identified with $V(R) := \{(a,b,c,d) \in R^4 |\ ad - bc = 1\}$. Question 1 is equivalent to asking if every solution of $ad - bc = 1$ in $\mathbb{Z}/n\mathbb{Z}$ has a *lift* to a solution of this equation in $\mathbb{Z}$.

One can think about Question 1 in terms of transitivity of certain subgroups of the group of automorphism of V as well. As you can see in Exercise 2, $\mathrm{SL}_2(\mathbb{Z})$ is generated by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This means

$$(a,b,c,d) \mapsto (a \pm b, b, c \pm d, d) \quad \text{and} \quad (a,b,c,d) \mapsto \pm(b, -a, d, -c)$$

induce a transitive action on $V(\mathbb{Z})$. The strong approximation is equivalent to saying that these maps induce a transitive action on $V(\mathbb{Z}/n\mathbb{Z})$ for every positive integer n.

Using the reduced row echelon process, one can show a similar result for $\mathrm{SL}_m(\mathbb{Z})$ for $m \geq 3$. This method is essentially based on using *unipotent elements* (u is called unipotent if all of its eigenvalues are 1). Following the same ideas, one can prove a similar result for symplectic groups. Let's recall that for every unital commutative ring R,

$$\mathrm{Sp}_{2n}(R) = \left\{ \gamma \in \mathrm{SL}_{2n}(R) \middle|\ \gamma \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \gamma^t = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \right\}.$$

This means $\pi_m : \mathrm{Sp}_{2n}(\mathbb{Z}) \to \mathrm{Sp}_{2n}(\mathbb{Z}/m\mathbb{Z})$ is surjective.

Next, we discuss that a similar statement does not hold for PGL_2. Let's define $\mathrm{PGL}_2(R) := \mathrm{GL}_2(R)/R^\times I$ where $R^\times$ is the group of units of R. Then

$$\overline{\det} : \mathrm{PGL}_2(R) \to R^\times/(R^\times)^2, \qquad \overline{\det}(\gamma R^\times I) := \det(\gamma)(R^\times)^2$$

is a well-defined surjective group homomorphism. Hence for every two distinct primes p and q, $\pi_{pq} : \mathrm{PGL}_2(\mathbb{Z}) \to \mathrm{PGL}_2(\mathbb{Z}/pq\mathbb{Z})$ is not surjective as

$$|\mathbb{Z}^\times| = 2 \quad \text{and} \quad |(\mathbb{Z}/pq\mathbb{Z})^\times/((\mathbb{Z}/pq\mathbb{Z})^\times)^2| = 4.$$

In technical terms, the big difference between SL_2 and PGL_2 is that SL_2 is *simply-connected* and PGL_2 is not. Notice that for every algebraically closed field F,

$$1 \to \mu_2(F) \to \mathrm{SL}_2(F) \to \mathrm{PGL}_2(F) \to 1$$

is a short exact sequence where $\mu_2(R) := \{x \in R |\ x^2 = 1\}$. Moreover $\mu_2(F)$ is a finite central subgroup. A group homomorphism with these properties is called *a central isogeny* (at least in characteristic zero). A(n) (algebraic) group which does not have a non-trivial (algebraic) isogeny is called a *simply-connected* (algebraic) group. For instance, SL_n and Sp_{2n} are simply connected algebraic groups, but PGL_n is not a simply connected algebraic group. Here we take a rudimentary approach and say that an *algebraic group* is a group which consists of solutions of certain polynomial equations and the group operations can be given by polynomial maps. Now we can formulate a version of strong approximation (due to Eichler, Kneser, and Platonov):

Theorem 2 (Strong approximation: the S-arithmetic case)**.** *Suppose $\mathbb{G}$ is a simply-connected algebraic group defined by polynomials with coefficients in $\mathbb{Z}$. Suppose $\mathbb{G}(\mathbb{C})$ is a product of almost simple groups (we say $\mathbb{G}$ is semisimple). Assume that $\mathbb{G}(\mathbb{Z}[1/q_0])$ is an infinite group. Then for every integer n with large enough prime factors the residue modulo n congruence map*

$$\pi_n : \mathbb{G}(\mathbb{Z}[1/q_0]) \to \mathbb{G}(\mathbb{Z}/n\mathbb{Z})$$

is surjective.

Next we want to see what happens if we restrict π_n to a subgroup Γ of $\mathbb{G}(\mathbb{Z}[1/q_0])$. Can we still get the entire $\mathbb{G}(\mathbb{Z}/n\mathbb{Z})$ (at least for integers n with large prime factors)?

We make one important observation: if there is an integer polynomial map $f : \mathbb{Q}(\mathbb{C}) \to \mathbb{C}$ such that $f(\mathbb{G}(\mathbb{Z}[1/q_0])) \neq 0$ but $f(\Gamma) = 0$, then it is not possible for $\pi_p(\Gamma) = \mathbb{G}(\mathbb{Z}/p\mathbb{Z})$ to hold for an arbitrarily large prime p. This is the case, because $f(\mathbb{G}(\mathbb{Z}[1/q_0])) \neq 0$ implies that for a large enough prime p, there is $\lambda \in \mathbb{G}(\mathbb{Z}[1/q_0])$ such that $\pi_p(f(\lambda)) \neq 0$, and so $f(\pi_p(\lambda)) \neq 0$. On the other hand, $f(\pi_p(\Gamma)) = 0$. Therefore $\pi_p(\lambda) \not\in \pi_p(\Gamma)$. We refer to this type of limitations as an *algebraic obstruction*.

If we refer to common solutions of a family of polynomials as *closed sets*, then to avoid the above algebraic obstruction we have to assume that *the smallest closed subset of $\mathbb{G}$ which contains Γ is $\mathbb{G}$*. We refer to the topology given by these closed sets as the *Zariski topology* of $\mathbb{G}$. In this language, the latest condition can be phrased as Γ *is Zariski-dense in* $\mathbb{G}$. Now we can formulate a stronger version of the strong approximation (due to Weisfeiler).

Theorem 3 (Strong approximation: the Zariski-dense case)**.** *Suppose $\mathbb{G}$ is a Zariski-connected simply-connected semisimple group given by integer polynomials. Suppose $\Gamma \subseteq \mathbb{G}(\mathbb{Z}[1/q_0])$ is a Zariski-dense subgroup. Then for every integer n with large enough prime factors the residue modulo n congruence map*

$$\pi_n : \Gamma \to \mathbb{G}(\mathbb{Z}/n\mathbb{Z})$$

is surjective.

1.2 Expanders and super-approximation

Suppose G is a group and Ω is a subset of G. We say Ω is a *symmetric subset* if the inverse of every element of Ω is in Ω. The Cayley graph of G with respect to Ω is an undirected graph whose set of vertices is G and $g_1, g_2 \in G$ are connected exactly when $g_1^{-1}g_2 \in \Omega$. The Cayley graph of G with respect to Ω is denoted by $\mathrm{Cay}(G;\Omega)$. Notice that $\mathrm{Cay}(G;\Omega)$ is a $|\Omega|$-regular graph; this means that the degree of every vertex is $|\Omega|$. The set of neighbors of g is $g\Omega = \{gw|\ w \in \Omega\}$. Continuing, we obtain that the connected component of g is the set

$$\{gw_1 \cdots w_n|\ n \in \mathbb{Z}^+, w_1, \ldots, w_n \in \Omega\}.$$

Since Ω is symmetric, $\{w_1 \cdots w_n|\ n \in \mathbb{Z}^+, w_1, \ldots, w_n \in \Omega\}$ is the subgroup generated by Ω. Hence $\mathrm{Cay}(G,\Omega)$ is connected if and only if Ω is a generating set of G. Therefore the strong approximation is equivalent to saying that if Ω is a symmetric generating set of a Zariski-dense subgroup of $\mathbb{G}(\mathbb{Z}[1/q_0])$, then under the right conditions on $\mathbb{G}$ and n, $\mathrm{Cay}(\mathbb{G}(\mathbb{Z}/n\mathbb{Z}), \pi_n(\Omega))$ is a connected graph. *Super-approximation* is about whether these graphs are *highly connected*.

Next we formulate what it means for a family of graphs to be highly connected. One way of thinking about the well-connectivity is in terms of people who live in a society. A society is well-connected if it is not consist of two or more communities that are not well-integrated. This means what links these communities together is much less than their sizes.

We can quantify this using the *Cheeger constant* of a graph. The Cheeger constant of a finite graph $\mathscr{G}$ is

$$h(\mathscr{G}) := \min\left\{ \frac{|E(A,A^c)|}{\min\{|A|,|A^c|\}} \,\middle|\, A \subseteq V_{\mathscr{G}} \right\},$$

where $E(A,A^c)$ is the set of all the edges that connect a vertex in A to a vertex in A^c. Notice that in a k-regular graph starting with a vertex v_0, the number of vertices that are of distance at most n from v_0 is at least

$$\min\{|V_{\mathscr{G}}|/2, (1+h(\mathscr{G})/k)^n\}.$$

This means these balls are *expanding* exponentially fast. motivated by this, we say a family $\{\mathscr{G}_i\}_i$ of k-regular graphs is a family of *expanders* if and only if $\inf_i h(\mathscr{G}_i) > 0$; this means there is a uniform positive constant for the Cheeger constants of all of these graphs. This implies that the number of vertices in balls of these graphs grow uniformly exponentially fast (till they contain at least half of the vertices).

The first explicit construction of a family of expanders is due to Margulis; this result was based on finitely generated groups with Kazhdan's property (T). A result of Selberg implies that $\{\mathrm{Cay}(\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z}), \Omega_i)\}_n$ is a family of expanders if $i = 1, 2$ and $\gcd(n, i) = 1$, where

$$\Omega_i = \left\{ \begin{pmatrix} 1 & \pm i \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm i & 1 \end{pmatrix} \right\}.$$

Selberg's proof was based on the Kloosterman sum, and his result can be applied to every finitely generated *congruence subgroup* of $\mathrm{SL}_2(\mathbb{Z})$. A subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is called a congruence subgroup if it contains $\ker \pi_n$ for some n, where

$$\pi_n : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$$

is the residue modulo n congruence map. The group generated by Ω_1 is $\mathrm{SL}_2(\mathbb{Z})$, and as you can see in Exercise 4, the subgroup generated by Ω_2 contains the kernel of π_4.

The group generated by Ω_3, however, is of infinite index in $\mathrm{SL}_2(\mathbb{Z})$ (see Exercise 8), and so it cannot be a congruence subgroup. Notice that the Zariski-closure of the group generated by Ω_3 contains the group generated by Ω_1, and so it is Zariski-dense in $\mathrm{SL}_2(\mathbb{Z})$. Peter Sarnak refer to this type of groups as *thin groups*; that means a *thin group* is a Zariski-dense subgroup of infinite index in $\mathbb{G}(\mathbb{Z})$ (or more generally $\mathbb{G}(\mathbb{Z}[1/q_0])$) for some algebraic group $\mathbb{G}$. Since the group generated by Ω_3 is Zariski-dense in $\mathrm{SL}_2(\mathbb{Z})$, by the strong approximation, for every large enough prime p (in fact it is enough to assume that $p \geq 5$), $\mathrm{Cay}(\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}), \Omega_3)$ is connected. Lubotzky asked whether these graphs form a family of expanders. This is referred to *Lubotzky's 1-2-3 problem*, and Bourgain and Gamburd in their seminal work gave an affirmative answer to this question.

Theorem 4 (Bourgain–Gamburd)**.** *Suppose Ω is a finite symmetric subset of $\mathrm{SL}_2(\mathbb{Q})$. Let Γ be the group generated by Ω. Suppose Γ is Zariski dense in $\mathrm{SL}_2(\mathbb{Q})$. Then there is p_0 such that the family of graphs $\{\mathrm{Cay}(\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}), \pi_p(\Omega)) |\ p \geq p_0, p$ prime$\}$ is a family of expanders.*

We refer to results of this type as *super-approximation*. The main goals of these notes are to cover the relevant general strategies, go over the type of tools involved, and survey the best known super-approximation results. This comes with the cost of not going into the details of most of the proofs.

1.3 Exercises

1. (Continued fraction) For a sequence of numbers $\{b_i\}_{i=0}^{\infty}$, we use $[b_0; b_1, \ldots, b_m]$ to denote

$$b_0 + \cfrac{1}{b_1 + \cfrac{1}{\cdots + \cfrac{1}{b_{m-1} + \cfrac{1}{b_m}}}}$$

and $[b_0; b_1, \ldots]$ to denote $\lim_{m\to\infty}[b_0; b_1, \ldots, b_m]$ (if this limit exists).

a) For a sequence of non-zero real numbers $\{b_i\}_{i=0}^{\infty}$, suppose

$$\begin{pmatrix} b_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} r_n(x) \\ s_n(x) \end{pmatrix}.$$

Prove that $\frac{r_n(x)}{s_n(x)} = [b_0; b_1, \ldots, b_n, x]$. (Hint: use induction on n.)

b) For a sequence of non-zero integers $\{b_i\}_{i=0}^{\infty}$, let $p_{-1} = 1$, $q_{-1} = 0$,

$$p_{n+1} := p_n b_{n+1} + p_{n-1} \qquad q_{n+1} := q_n b_{n+1} + q_{n-1}$$

for every non-negative integer n. Prove that

$$\begin{pmatrix} b_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix},$$

for every non-negative integer n. Deduce that $\frac{p_n}{q_n} = [b_0; b_1, \ldots, b_n]$, $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n+1}$, and $\gcd(p_n, q_n) = 1$.

c) For a sequence of positive integers $\{b_i\}_{i=0}^{\infty}$, suppose $\frac{p_n}{q_n}$ is the simple form of the rational number $[b_0; b_1, \ldots, b_n]$. Use the previous part to show that

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n+1}}{q_{n-1} q_n},$$

and deduce that $\lim_{n\to\infty} \frac{p_n}{q_n}$ exists, and so $[b_0; b_1, \ldots]$ is well-defined.

d) For a non-zero real number x, we let $x_0 := x$ and define the sequences $\{a_i\}_{i=0}^{\infty}$ and $\{x_i\}_{i=0}^{\infty}$ inductively as follows. We set $a_i := \lfloor x_i \rfloor$ for every integer i and $x_{i+1} := \frac{1}{\{x_i\}}$ where $\{y\} := y - \lfloor y \rfloor$ is the fractional part of y. We stop if x_i is an integer. suppose $\frac{p_n}{q_n}$ is the simple form of $[a_0; a_1, \ldots, a_n]$. Show that $x = [a_0; a_1, \ldots, a_n, x_{n+1}]$ for every non-negative integer n.

e) In the setting of the previous item, prove that

$$x = \frac{p_n x_{n+1} + p_{n-1}}{q_n x_{n+1} + q_{n-1}},$$

and deduce that

$$x - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(q_n x_{n+1} + q_{n-1})}.$$

f) In the above setting, prove that $x = [a_0; a_1, \ldots]$, and

$$\frac{1}{q_n(q_n + q_{n+1})} \leq \left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n q_{n+1}}.$$

g) For an irrational number $\alpha = [a_0; a_1, \ldots]$, let

$$M(\alpha) := \limsup_{n\to\infty} [a_n; a_{n+1}, \ldots] + [0; a_{n-1}, \ldots, a_1].$$

Prove that there are infinitely many rational numbers of simple form $\frac{p}{q}$ such that $|\alpha - \frac{p}{q}| \le \frac{1}{M(\alpha)q^2}$.

h) (Hurwitz's theorem) Prove that $M(\alpha) \ge \sqrt{5}$ for every rational number α, and equality holds for the Golden ratio $[1; 1, 1, \ldots]$.

2. (Generating $\mathrm{SL}_2(\mathbb{Z})$) Suppose $\gamma = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})$.

a) Suppose $\frac{a}{b} = [c_0; c_1, \ldots, c_n]$. Then

$$\gamma = \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & e \\ 0 & \pm 1 \end{pmatrix}$$

for some integer e.

b) Prove that $\mathrm{SL}_2(\mathbb{Z}) = \left\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\rangle$.

3. (Strong approximation: the $\mathrm{SL}_2(\mathbb{Z})$ case) Suppose n is a positive integer and $\bar{\gamma} = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$.

a) Prove that there are integers a and b such that $\pi_n(a) = \bar{a}$, $\pi_n(b) = \bar{b}$, and $\gcd(a, b) = 1$, where π_n is the residue modulo n congruence map.

b) Prove that there are $\lambda \in \mathrm{SL}_2(\mathbb{Z})$ and $\bar{e} \in \mathbb{Z}/n\mathbb{Z}$ such that

$$\pi_n(\lambda)^{-1}\bar{\gamma} = \begin{pmatrix} 1 & \bar{e} \\ 0 & 1 \end{pmatrix}.$$

c) Prove that $\pi_n : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ is surjective.

4. Let $\alpha := \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\beta := \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. Suppose a is odd and b is a non-zero even number. Let $v := \begin{pmatrix} a \\ b \end{pmatrix}$.

a) (The reduction process) Prove that there is $l \in \mathbb{Z}$ such that

$$\min\{\|\alpha^l v\|_\infty, \|\beta^l v\|_\infty\} < \|v\|_\infty.$$

b) Prove that there is $\gamma \in \langle \alpha, \beta \rangle$ such that $\gamma v = \pm \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

c) Prove that $\langle \alpha, \beta, -I \rangle = \ker \pi_2$ where $\pi_2 : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ is the residue modulo 2 congruence map.

d) Prove that $\langle \alpha, \beta \rangle$ contains the kernel of π_4.

5. (Ping-pong lemma) Suppose G is a group and it acts on a set X. Suppose G_1 and G_2 are two subsets of G, $|G_1| \geq 2$, and $|G_2| \geq 3$. Suppose X_1 and X_2 are two subsets of X such that $X_1 \not\subseteq X_2$ and $X_2 \not\subseteq X_1$. Suppose
$$(G_1 \setminus \{1\}) \cdot X_2 \subseteq X_1 \quad \text{and} \quad (G_2 \setminus \{1\}) \cdot X_1 \subseteq X_2.$$
Prove that $\langle G_1 \cup G_2 \rangle \simeq G_1 * G_2$.

6. Suppose $a \geq 2$. Let $\alpha := \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, $\beta := \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}$, $G_1 := \langle \alpha \rangle$, and $G_2 := \langle \beta \rangle$. Let $X_1 := \{(x, y) \in \mathbb{R}^2 | \; |x| \geq \frac{a}{2}|y|\}$ and $X_2 := \{(x, y) \in \mathbb{R}^2 | \; |x| \leq \frac{a}{2}|y|\}$.

   a) Consider the natural linear action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2$. Prove that
   $$(G_1 \setminus \{I\}) \cdot X_2 \subseteq X_1 \quad \text{and} \quad (G_2 \setminus \{I\}) \cdot X_1 \subseteq X_2.$$

   b) Prove that α and β freely generate a free subgroup of $\mathrm{SL}_2(\mathbb{R})$.

7. Let $\alpha := \pm \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$ and $\beta := \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ be two elements of the group $\mathrm{PSL}_2(\mathbb{Z}) := \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$.

   a) Use Exercise 2 to show that $\mathrm{PSL}_2(\mathbb{Z}) = \langle \alpha, \beta \rangle$ and deduce that there is a surjective group homomorphism from $\mathbb{Z}/3\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$ to $\mathrm{PSL}_2(\mathbb{Z})$.

   b) Consider the Möbius group action of $\mathrm{PSL}_2(\mathbb{R})$ on $\mathbb{C} \cup \{\infty\}$; that means
   $$\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az + b}{cz + d}$$
   (justify that it is a group action). Notice that
   $$\alpha \cdot z = -1 - \frac{1}{z}, \quad \alpha^{-1} \cdot z = -\frac{1}{z + 1}, \quad \text{and} \quad \beta \cdot z = -\frac{1}{z}.$$
   Let X_1 be the set of all positive irrational real numbers and X_2 be the set of all the negative irrational real numbers. Show that
   $$(\langle \alpha \rangle \setminus \{I\}) \cdot X_1 \subseteq X_2 \quad \text{and} \quad (\langle \beta \rangle \setminus \{I\}) \cdot X_2 \subseteq X_1.$$

   c) Prove that there is an isomorphism $\mathrm{PSL}_2(\mathbb{Z}) \simeq \mathbb{Z}/3\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$ which factors through $\langle \alpha \rangle * \langle \beta \rangle$.

8. Prove that $\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}$ generate a subgroup of infinite index in $\mathrm{SL}_2(\mathbb{Z})$. (Hint: Use Exercise 7.)

Chapter 2

Random-walks on a graph and expanders

2.1 Basics of random-walks on a finite graph

A *random walk* on a graph $\mathscr{G}$ is a sequence of random variables $\{X_i\}_{i=0}^{\infty}$ with values on the set of vertices $V_{\mathscr{G}}$ of $\mathscr{G}$ such that, for very non-negative integer i, X_{i+1} is chosen independently at random from the neighbors of X_i. For every vertex v,

$$\mathbb{P}(X_{i+1} = v) = \sum_{w \in V_{\mathscr{G}}} \mathbb{P}(X_i = w)\mathbb{P}(w \to v). \tag{2.1}$$

Here for every $w \in V_{\mathscr{G}}$, $\mathbb{P}(w \to v) = \frac{1}{d_w}[\{w, v\} \in E_{\mathscr{G}}]$ where $[\{w, v\} \in E_{\mathscr{G}}] = 1$ if w is connected to v in $\mathscr{G}$ and it is zero otherwise and d_w is the degree of the vertex w; that means the number of edges that have w as one of their vertices. Let μ_i be the distribution of X_i; that means

$$\mu_i : V \to [0, 1], \quad \mu_i(v) = \mathbb{P}(X_i = v).$$

Suppose the set of vertices $V := V_{\mathscr{G}}$ is $\{v_1, \ldots, v_n\}$. Then $\mathfrak{B} := \{\delta_{v_1}, \ldots, \delta_{v_n}\}$ is an orthonormal basis of $L^2(V)$. For every function $f \in L^2(V)$, $\langle f|$ denotes the row matrix $\begin{pmatrix} f(v_1) & \cdots & f(v_n) \end{pmatrix}$ and $|f\rangle$ denotes the transpose of $\langle f|$. Notice that $f = \sum_{i=1}^{n} f(v_i)\delta_{v_i}$, and so $\langle f|$ is simply the matrix representation of f with respect to the basis $\mathfrak{B}$.

Let T be the *transition matrix* of the random-walk; that means the (i, j)-entry of T is equal to

$$\mathbb{P}(v_i \to v_j) = \frac{1}{d_{v_i}}[\{v_i, v_j\} \in E_{\mathscr{G}}].$$

Then by (2.1), we have

$$\langle \mu_{i+1}| = \langle \mu_i|T,$$

and so the probability law after l steps random-walk is given by $\langle \mu_l| = \langle \mu_0|T^l$. We can understand and compute powers of a matrix the best if it is diagonal or at least diagonalizable. We know that a symmetric matrix is diagonalizable. We notice that in general the transition matrix is, however, not necessarily symmetric. In the case of random-walk on a finite graph the transition matrix is symmetric if and only if all the vertices have the same degree; that means when the graph is regular. In general, we have

$$T = D_{\mathscr{G}}^{-1} A_{\mathscr{G}},$$

where $D_{\mathscr{G}}$ is the diagonal matrix $\operatorname{diag}(d_{v_1}, \ldots, d_{v_n})$ and $A_{\mathscr{G}}$ is the adjacency matrix of the graph; that means the (i,j) entry is 1 if v_i is connected to v_j and 0 otherwise. Hence for every integer l, we have

$$T^l = D_{\mathscr{G}}^{-1} A_{\mathscr{G}} D_{\mathscr{G}}^{-1} A_{\mathscr{G}} \cdots D_{\mathscr{G}}^{-1} A_{\mathscr{G}}.$$

Therefore

$$T^l = D_{\mathscr{G}}^{-1/2} M_{\mathscr{G}}^l D_{\mathscr{G}}^{1/2}, \tag{2.2}$$

where

$$M_{\mathscr{G}} = D_{\mathscr{G}}^{-1/2} A_{\mathscr{G}} D_{\mathscr{G}}^{-1/2}.$$

We notice that $M_{\mathscr{G}}$ is a real symmetric. Hence it has a right orthonormal basis $\{|\phi_1\rangle, \ldots, |\phi_n\rangle\}$ with real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Since $M_{\mathscr{G}}$ is symmetric, $\{\langle\phi_1|, \ldots, \langle\phi_n|\}$ is left eigenbasis. By (2.2), we deduce T is diagonalizable with a right eigenbasis $\{D_{\mathscr{G}}^{-1/2}|\phi_1\rangle, \ldots, D_{\mathscr{G}}^{-1/2}|\phi_n\rangle\}$, a left eigenbasis $\{\langle\phi_1|D_{\mathscr{G}}^{1/2}, \ldots, \langle\phi_n|D_{\mathscr{G}}^{1/2}\}$, and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. If $\langle\mu_0| = \sum_{i=1}^n c_i \langle\phi_i|D_{\mathscr{G}}^{1/2}$, then by (2.2) we obtain that

$$\langle\mu_l| = \sum_{i=1}^n \lambda_i^l c_i \langle\phi_i|D_{\mathscr{G}}^{1/2}. \tag{2.3}$$

So it is crucial to gain a better understanding of λ_i's. This is achieved by looking at the multiplication by T from left:

$$\mathscr{T} : L^2(V) \to L^2(V) \quad |\mathscr{T}(f)\rangle = T|f\rangle.$$

Notice that for every $v \in V$, $\mathscr{T}(f)(v) = \sum_w \mathbb{P}(v \to w) f(w)$ is the average of the values of f at the neighbors of v. Based on the fact that $\mathscr{T}$ is an averaging operator and the maximum modulus principle, we can gain some basic information on λ_i's.

For every i, let $\bar{\phi}_i : V \to \mathbb{R}$, $\bar{\phi}_i(v) := d_v^{-1/2}\phi_i(v)$. Then $\mathscr{T}(\bar{\phi}_i) = \lambda_i \bar{\phi}_i$ for every i. After replacing $\bar{\phi}_i$ with $-\bar{\phi}_i$, if needed, we can and will assume that for some $w_0^{(i)} \in V$, $\bar{\phi}_i(w_0^{(i)}) = \|\bar{\phi}_i\|_\infty = \max_{v\in V}\{|\bar{\phi}_i(v)|\}$. Hence

$$|\lambda_i||\bar{\phi}_i(w^{(i)})| = |\mathscr{T}(\bar{\phi}_i)(w_0^{(i)})| \leq \sum_v \mathbb{P}(w^{(i)} \to v)|\bar{\phi}_i(v)| \leq \|\bar{\phi}_i\|_\infty. \tag{2.4}$$

By (2.4), we obtain that $|\lambda_i| \leq 1$ for every i.

Lemma 5. *Every eigenvalue of $T_{\mathscr{G}}$ is in the interval $[-1, 1]$.*

Next, we investigate the extreme possible values. Notice that $\mathscr{T}\mathbb{1}_V = \mathbb{1}_V$ where $\mathbb{1}_V$ is the constant function 1; we can observe this based on the fact that $\mathscr{T}f(v)$ is the average of the values of f at the neighbors of v. Hence 1 is definitely an eigenvalue of $\mathscr{T}$, and so by Lemma 5, $\lambda_1 = 1$.

A function in the kernel of $\mathscr{T} - I$ is called *harmonic*. Suppose f is a non-zero real harmonic function. After replacing f with $-f$, if needed, we can and will assume that $f(w_0) = \|f\|_\infty$ for some $w_0 \in V$. Then

$$f(w_0) = |f(w_0)| \le \sum_v \mathbb{P}(w_0 \to v)|f(v)| \le \|f\|_\infty,$$

which implies that for every $v \in V$, either $\mathbb{P}(w_0 \to v) = 0$ or $f(v) = f(w_0)$. This means $f(v) = f(w_0)$ for v that is connected to w_0. Repeating this argument, we obtain that $f(v) = f(w_0)$ for every v in the *connected component* of w_0 in $\mathscr{G}$. Conversely, characteristic functions of connected components of $\mathscr{G}$ are harmonic functions. Altogether, we have proved the following statement.

Lemma 6. *The dimension of the operator $\mathscr{T} - I$ is equal to the number of connected components of $\mathscr{G}$. In particular, $\mathscr{G}$ is connected if and only if $\lambda_2 < 1$.*

Suppose $\mathscr{T}$ has eigenvalue -1 and $\mathscr{T} f = -f$ for a nonzero function f. Replacing f with $-f$, if needed, we can and will assume that $f(w_0) = \|f\|_\infty$ for some $w_0 \in V$. Hence

$$0 = \sum_v \mathbb{P}(w_0 \to v)(f(w_0) + f(v)) \quad \text{and} \quad f(w_0) + f(v) \ge 0, \text{ for every } v.$$

Therefore for every neighbor v of w_0, we have $f(v) = -f(w_0)$. Repeating this argument, we see that the value of f at every neighbor of a neighbor of w_0 is again $f(w_0)$. We deduce that the connected component of w_0 is a bipartite graph; this means the vertices of this connected component can be partitioned into two sets A and B, and every edge has an element in A and an element in B.

Lemma 7. *In the above setting $\lambda_n = -1$ if and only if $\mathscr{G}$ has a bipartite connected component.*

Proof. We have already proved that if $\lambda_n = -1$, then $\mathscr{G}$ has a bipartite connected component. For the converse look at Exercise 1. □

By (2.3), and Lemmas 6 and 7, we obtain the following result on the rate of convergence of random-walks on a finite connected non-bipartite regular graph.

Proposition 8. *Suppose $\{X_i\}_{i=0}^\infty$ is a random-walk on a finite connected non-bipartite k-regular graph $\mathscr{G}$ (k-regular means that the degree of all the vertices are k). Suppose μ_i is the distribution of X_i. Suppose $\lambda_1 \ge \ldots \ge \lambda_n$ are as before the eigenvalues of the transition matrix. Let $\lambda_{\mathscr{G}} := \max\{|\lambda_2|, |\lambda_n|\}$. Then the following statements hold.*

1. *(L^2-convergence) For every $f \in L^2(V)$ and every positive integer l,*

$$\left\| \mathscr{T}^l f - \frac{\langle f, \mathbb{1}_V \rangle}{|V|} \mathbb{1}_V \right\|_2 \le \lambda_{\mathscr{G}}^l \|f\|_2$$

where $\mathscr{T}$ is as before.

2. *(L^1-convergence) For every $f \in L^1(V)$ and every positive integer l,*

$$\left| \mathbb{E}[f(X_l)] - \frac{\sum_{v\in V} f(v)}{|V|} \right| \le \lambda_{\mathscr{G}}^l \|f\|_2;$$

in particular, for every $A \subseteq V$,

$$\left| \mathbb{P}(X_l \in A) - \frac{|A|}{|V|} \right| \le \lambda_{\mathscr{G}}^l \sqrt{|A|}.$$

3. *(Mixing) For every $f, g \in L^2(V)$,*

$$\left| \langle f, \mathscr{T}^l g\rangle - \left(\sum_{v\in V} f(v)\right) \frac{\sum_{v\in V} g(v)}{|V|} \right| \le \lambda_{\mathscr{G}}^l \|f\|_2 \|g\|_2.$$

Proof. Suppose $\{\phi_1, \ldots, \phi_n\}$ is as before an orthonormal basis of $M_{\mathscr{G}}$. Notice that since $\mathscr{G}$ is k-regular, $M_{\mathscr{G}} = T_{\mathscr{G}}$. Also notice that $\phi_1 = \frac{1}{\sqrt{|V|}} \mathbb{1}_V$, and for every $f \in L^2(V)$, the orthogonal projection of f to the space of constant functions is

$$\frac{\langle f, \mathbb{1}_V\rangle}{|V|} \mathbb{1}_V. \tag{2.5}$$

For $f \in L^2(V)$, suppose $f = \sum_{i=1}^n c_i\phi_i$. Then $\|f\|_2^2 = \sum_{i=1}^n |c_i|^2$ and by (2.5), we have

$$\mathscr{T}^l f - \frac{\langle f, \mathbb{1}_V\rangle}{|V|} \mathbb{1}_V = \sum_{i=2}^n \lambda_i^l c_i \phi_i.$$

This implies that

$$\left\| \mathscr{T}^l f - \frac{\langle f, \mathbb{1}_V\rangle}{|V|} \mathbb{1}_V \right\|_2^2 = \sum_{i=2}^n |\lambda_i|^{2l} |c_i|^2 \le \lambda_{\mathscr{G}}^{2l} \sum_{i=2}^n |c_i|^2 \le \lambda_{\mathscr{G}}^l \|f\|_2^2.$$

This completes the proof of the first part.

Assuming that the first L^1-convergence inequality is proved, we let f be the characteristic function $\mathbb{1}_A$ of A. The desired inequality follows from the fact that $\mathbb{E}[\mathbb{1}_A(X_l)] = \mathbb{P}(X_l \in A)$. For the rest of the inequalities look at the exercise 2. □

We refer to $\lambda_{\mathscr{G}}$ as the *spectral gap* of this random walk. Notice that

$$\lambda_{\mathscr{G}} = \| \mathscr{T}_{\mathscr{G}}|_{L^2(V)^\circ} \|_{\mathrm{op}}$$

where $L^2(V)^\circ := \{f \in L^2(V) | \ \sum_{v\in V} f(v)\}$ is the space of functions that are orthogonal to the space of constant functions.

It is intuitive that a random-walk on a well-connected regular graph should quickly converge to equidistribution. This means having a lower bound for the Cheeger constant $h(\mathscr{G})$ of a finite k-regular graph $\mathscr{G}$ should give us an upper bound for $\lambda_{\mathscr{G}}$. In the rest of this chapter, we prove a variant of this result.

2.2 Discrete Laplacian

Suppose $\mathscr{G}$ is a finite k-regular graph. Pick an orientation for the edges. For every edge $e \in E_{\mathscr{G}}$, let e^- be the initial vertex and e^+ be the terminal vertex of the oriented version. Thinking about a function $f : V_{\mathscr{G}} \to \mathbb{R}$ as the amount of charge on nodes, $df(e) := f(e^+) - f(e^-)$ measures the amount of the resistance times the current on that edge. We can also think of the vertices as 0-cells, the edges as 1-cells, and

$$d : L^2(V) \to L^2(E), \quad df(e) := f(e^+) - f(e^-)$$

as the boundary map. We notice that

$$d^* : L^2(E) \to L^2(V), \quad d^*g(v) = \sum_{v=e^+} g(e) - \sum_{v=e^-} g(e)$$

(See Exercise 3). Thinking about a function g on the edges as the amount of a flow going through that edge, we can think about $d^*g(v)$ as the amount of the flow that *sinks* in v. For every $f \in L^2(V)$, we have

$$\begin{aligned} d^*df(v) &= \sum_{v=e^+} df(e) - \sum_{v=e^-} df(e) \\ &= \sum_{v=e^+} (f(e^+) - f(e^-)) - \sum_{v=e^-} (f(e^+) - f(e^-)) \\ &= d_v f(v) - \sum_{w \sim v} f(w) \\ &= k((I - \mathscr{T}_{\mathscr{G}})(f))(v), \end{aligned}$$

where $w \sim v$ means $\{w, v\}$ is an edge in $\mathscr{G}$. Hence

$$\mathscr{L}_{\mathscr{G}} := I - \mathscr{T}_{\mathscr{G}} = \frac{1}{k} d^* d, \tag{2.6}$$

and it is called the *discrete Laplacian* of the k-regular graph $\mathscr{G}$. Assuming that $\{\phi_1, \ldots, \phi_n\}$ is an orthonormal basis of $\mathscr{T}$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, by (2.6) we obtain that

$$\mathscr{L}_{\mathscr{G}}(\phi_i) = (1 - \lambda_i)\phi_i$$

for every i. Hence assuming $\mathscr{G}$ is connected, eigenvalues of $\mathscr{L}_{\mathscr{G}}$ are

$$0 = 1 - \lambda_1 < 1 - \lambda_2 \leq \cdots \leq 1 - \lambda_n \leq 2.$$

For $f \in L^2(V)$, suppose $f = \sum_{i=1}^n c_i \phi_i$. Then $\|f\|_2 = \sum_{i=1}^n |c_i|^2$ and

$$\begin{aligned} \|df\|_2^2 &= \langle df, df \rangle = \langle f, d^* df \rangle = k \langle f, \mathscr{L} f \rangle \\ &= k \sum_{i,j} \overline{c}_i c_j \langle \phi_i, \mathscr{L} \phi_j \rangle = k \sum_{i,j} (1 - \lambda_j) \overline{c}_i c_j \langle \phi_i, \phi_j \rangle \\ &= k \sum_{i=1}^n (1 - \lambda_i) |c_i|^2 = k \sum_{i=2}^n (1 - \lambda_i) |c_i|^2. \end{aligned} \tag{2.7}$$

By (2.7), we obtain the following description of $1 - \lambda_2$.

Lemma 9. *In the previous setting,*

$$1-\lambda_2 = \frac{1}{k}\min\left\{ \frac{\|df\|_2^2}{\|f\|_2^2} \,\middle|\, f\in L^2(V)^\circ\setminus\{0\}\right\}.$$

Proof. By (2.7), we have $\|df\|_2^2 \geq k(1-\lambda_2)\sum_{i=2}^n |c_i|^2$ where $f=\sum_{i=1}^n c_i\phi_i$. Notice that $c_i=\langle f,\phi_i\rangle$ for every i; in particular $c_1=0$ as ϕ_1 is constant and $f\in L^2(V)^\circ$. Therefore $\|f\|_2^2=\sum_{i=2}^n|c_i|^2$. Altogether, we have $\|df\|_2^2\geq k(1-\lambda_2)\|f\|_2^2$ for every $f\in L^2(V)^\circ$. Therefore

$$1-\lambda_2 \leq \frac{1}{k}\min\left\{ \frac{\|df\|_2^2}{\|f\|_2^2} \,\middle|\, f\in L^2(V)^\circ\setminus\{0\}\right\}. \tag{2.8}$$

We also notice that $\|d\phi_2\|_2^2 = k\langle\phi_2,\mathscr{L}\phi_2\rangle = k(1-\lambda_2)$, and this shows that the equality in (2.8) holds. This completes the proof. $\square$

Sometimes it is useful to notice that

$$\|df\|_2^2=\sum_{e\in E}|f(e^+)-f(e^-)|^2=\sum_{w\sim v}|f(v)-f(w)|^2,$$

and so

$$k(1-\lambda_2)\leq \frac{\sum_{w\sim v}|f(v)-f(w)|^2}{\sum_v|f(v)|^2}$$

if $\sum_v f(v)=0$ and $f\neq 0$.

Next we show that the Cheeger constant can be described based on an L^1-version of Lemma 9. This is done based on finding various *good* cuts.

2.3 Finding good cuts

In a society, communities shape based on certain *features*. Inspired by this, in social medias, we try to find certain *features* that can *distinguish* various communities. A basic such example is finding a *feature* that can split people into two communities; this means finding a *good cut* in the underlying graph. In mathematical language, a *feature* is simply a function f on the set of the vertices of the given (social media) graph, and after picking a *critical value* c_0, we split the vertices based on whether the value of f at the given vertex is more or less than c_0.

Suppose $\mathscr{G}$ is a finite graph with the set of vertices V and set of edges E. For $f:V\to\mathbb{R}$ and $c\in\mathbb{R}$, let

$$V_{f,c}^- := \{v\in V|\ f(v)<c\},$$

and

$$h_{\mathscr{G}}(f):=\inf_c \frac{|E(V_{f,c}^-,V\setminus V_{f,c}^-)|}{\min\{|V_{f,c}^-|,|V\setminus V_{f,c}^-|\}}.$$

This means $h_{\mathscr{G}}(f)$ quantifies how good of a cut we can get using f. We can view f as a projection of the graph $\mathscr{G}$ to a line. Starting with a measure μ on $\mathbb{R}$, using the projection given by f, we can put a weight on each edge. For an edge $e = \{v, w\}$, let I_e be the interval with the end points $f(v)$ and $f(w)$. Then the weight of e corresponding to μ and f is $\mu(I_e)$.

In this section, we use a probabilistic method to find upper bounds for $h_{\mathscr{G}}(f)$.

Lemma 10. *Suppose μ is a measure on $\mathbb{R}$, and $f : V \to \mathbb{R}$. For every edge $e = \{v, w\}$, let I_e be the interval with end points $f(v)$ and $f(w)$. Then*

$$\int |E(V^-_{f,c}, V \setminus V^-_{f,c})| d\mu(c) = \sum_e \mu(I_e).$$

Proof. Notice that $e \in E(V^-_{f,c}, V \setminus V^-_{f,c})$ if and only if $c \in I_e$. Hence

$$\int |E(V^-_{f,c}, V \setminus V^-_{f,c})| d\mu(c) = \sum_e \int [e \in E(V^-_{f,c}, V \setminus V^-_{f,c})] d\mu(c) = \sum_e \mu(I_e).$$

This completes the proof. $\square$

Lemma 11. *Suppose μ is a measure on $\mathbb{R}$ such that $\mu(\{c\}) = 0$ for every $c \in \mathbb{R}$, and $f : V \to \mathbb{R}$. Let c_0 be the median of $f(v)$'s as v ranges in V. For every $v \in V$, let I_v be the interval with the end points c_0 and $f(v)$. Then*

$$\int \min\{|V^-_{f,c}|, |V \setminus V^-_{f,c}|\} d\mu(c) = \sum_v \mu(I_v).$$

Proof. Notice that $\min\{|V^-_{f,c}|, |V \setminus V^-_{f,c}|\} = |V^-_{f,c}|$ if and only if $c \leq c_0$. Hence

$$\begin{aligned} \int \min\{|V^-_{f,c}|, |V \setminus V^-_{f,c}|\} d\mu(c) &= \int_{c \leq c_0} |V^-_{f,c}| d\mu(c) + \int_{c > c_0} |V \setminus V^-_{f,c}| d\mu(c) \\ &= \sum_v [f(v) < c_0] \mu(f(v), c_0] + [f(v) \geq c_0] \mu(c_0, f(v)] \\ &= \sum_v \mu(I_v). \end{aligned}$$

This completes the proof. $\square$

Theorem 12. *Suppose μ is a measure on $\mathbb{R}$ such that $\mu(\{c\}) = 0$ for every $c \in \mathbb{R}$. Suppose $f : V \to \mathbb{R}$ is a function such that $\mu([\min_v f(v), \max_v f(v)]) \neq 0$. Let c_0 be the median of $f(v)$'s as v ranges in V. For $v \in V$, let I_v be the interval with the end points c_0 and $f(v)$, and for $e = \{v, w\} \in E$, let I_e be the interval with the end points $f(v)$ and $f(w)$. Then*

$$h_{\mathscr{G}}(f) \leq \frac{\sum_{e \in E} \mu(I_e)}{\sum_{v \in V} \mu(I_v)}.$$

Proof. By Lemmas 10 and 11, we have

$$\int (\sum_v \mu(I_v)) |E(V^-_{f,c}, V \setminus V^-_{f,c})| - (\sum_{e \in E} \mu(I_e)) \min\{|V^-_{f,c}|, |V \setminus V^-_{f,c}|\} d\mu(c) = 0.$$

Therefore for some c we have that

$$(\sum_v \mu(I_v))|E(V^-_{f,c}, V \setminus V^-_{f,c})| - \sum_{e\in E} \mu(I_e) \min\{|V^-_{f,c}|, |V \setminus V^-_{f,c}|\} \leq 0,$$

and so

$$h_{\mathscr{G}}(f) \leq \frac{|E(V^-_{f,c}, V \setminus V^-_{f,c})|}{\min\{|V^-_{f,c}|, |V \setminus V^-_{f,c}|\}} \leq \frac{\sum_{e\in E} \mu(I_e)}{\sum_v \mu(I_v)}.$$

This finishes the proof. □

Special cases of μ give us interesting results. For instance the case when μ is the Lebesgue measure implies the following Theorem.

Theorem 13. *Suppose $\mathscr{G}$ is a finite graph with the set of vertices V. Let $h(\mathscr{G})$ be the Cheeger constant of $\mathscr{G}$. Then*

$$h(\mathscr{G}) = \inf\left\{ \frac{\|df\|_1}{\|f\|_1} \,\middle|\, f \in L^1(V) \setminus \{0\}, \mathrm{Med}(f) = 0 \right\},$$

where $\mathrm{Med}(f)$ is the median of $f(v)$'s as v ranges in V.

Notice that for the L^2-norm in the denominator we took a shift of f which minimized the L^2-norm, and here for the L^1-norm we are taking a shift of f which minimizes the L^1-norm! It worths pointing out that df does not change as we shift f by a constant.

Proof of Theorem 13. Applying Theorem 12 for the case when μ is the Lebesgue measure ℓ, we obtain that

$$h(\mathscr{G}) \leq \frac{\sum_{e\in E} \ell(I_e)}{\sum_{v\in V} \ell(I_v)},$$

where E is the set of edges of $\mathscr{G}$, for $e = \{v, w\}$, I_e is an interval with the end points $f(v), f(w)$, and for $v \in V$, I_v is an interval with the end points $\mathrm{Med}(f) = 0$ and $f(v)$. Hence

$$\ell(I_e) = |df(e)| \quad \text{and} \quad \ell(I_v) = |f(v)|.$$

Therefore $h(\mathscr{G}) \leq \frac{\|df\|_1}{\|f\|_1}$ if $\mathrm{Med}(f) = 0$ and $f \neq 0$.

Suppose $h(\mathscr{G}) = \frac{|E(A,A^c)|}{|A|}$ for some $A \subseteq V$ with $|A| \leq |V|/2$. Let $f = \mathbb{1}_A$ be the characteristic function of A. Since $|A| \leq |V|/2$, $\mathrm{Med}(f) = 0$. Notice that $\|df\|_1 = |E(A, A^c)|$ and $\|f\|_1 = |A|$, and so $h(\mathscr{G}) = \frac{\|df\|_1}{\|f\|_1}$. This completes the proof. □

2.4 Discrete isoperimetric inequalities

In this section, we use the L^2-optimization description of $1 - \lambda_2$ (see Lemma 9) and the bounds that we have found for $h(\mathscr{G})$ using Theorem 12 , and prove isoperimetric inequalities.

Using Theorem 12 for the case when μ is the Lebesgue measure, we found an L^1-optimization description of $h(\mathscr{G})$. Thinking about edges as wires laying on a surface, the Lebesgue measure more or less ends up giving us the total weight on top of a point. Next, we roughly think about this graph balanced about the median of f and measure the *torque* of each edge. This means we assume that 0 is the median and consider the measure μ given by the density function

$$d\mu(t) := |t|dt.$$

Notice that $\mu([a,b]) = \int_a^b |t|dt = \frac{b|b|-a|a|}{2}$. Hence

$$\sum_v \mu(I_v) = \frac{\|f\|_2^2}{2}, \tag{2.9}$$

where I_v is the interval with the endpoints $\mathrm{Med}(f) = 0$ and $f(v)$. We also have

$$\sum_e \mu(I_e) = \frac{1}{2}\sum_e (f(e^+)|f(e^+)| - f(e^-)|f(e^-)|) \tag{2.10}$$

Notice that for every $a, b \in \mathbb{R}$, we have

$$b|b| - a|a| \leq |b-a|(|b|+|a|). \tag{2.11}$$

By (2.10) and (2.11), we obtain that

$$\sum_e \mu(I_e) \leq \sum_e |df(e)|\left(\frac{|f(e^+)| + |f(e^-)|}{2}\right),$$

and so by the Cauchy-Schwarz inequality, we have

$$\sum_e \mu(I_e) \leq \|df\|_2 \sqrt{\sum_e \left(\frac{|f(e^+)| + |f(e^-)|}{2}\right)^2}. \tag{2.12}$$

Because $(\frac{a+b}{2})^2 \leq \frac{a^2+b^2}{2}$, by (2.12), we obtain

$$\sum_e \mu(I_e) \leq \|df\|_2 \sqrt{\sum_e \frac{|f(e^+)|^2 + |f(e^-)|^2}{2}} = \sqrt{\frac{k}{2}}\,\|df\|_2\|f\|_2, \tag{2.13}$$

if $\mathscr{G}$ is a k-regular graph.

By (2.9), (2.13), and Theorem 12, we deduce the following result.

Lemma 14. *Suppose $\mathscr{G}$ is a k-regular graph and $f : V \to \mathbb{R}$ is a function whose median is 0. Then*

$$h(\mathscr{G}) \leq \sqrt{2k}\frac{\|df\|_2}{\|f\|_2}.$$

Now we are ready to prove the discrete isoperimetric inequalities.

Theorem 15. *Suppose $\mathscr{G}$ is a k-regular graph. Then in the previous setting, we have*

$$\frac{1-\lambda_2}{2} \leq \frac{h(\mathscr{G})}{k} \leq \sqrt{2(1-\lambda_2)}.$$

Proof. Applying Lemma 14 to $f = \phi_2$, the second eigenfunction of the discrete Laplacian, we deduce that

$$h(\mathscr{G}) \leq \sqrt{2k}\frac{\|d\phi_2\|_2}{\|\phi_2\|_2} = \sqrt{2k}\sqrt{k(1-\lambda_2)} = k\sqrt{2(1-\lambda_2)}. \tag{2.14}$$

To obtain a lower bound for the Cheeger constant, we start with a subset A which gives us the Cheeger constant; that means we assume

$$h(\mathscr{G}) = \frac{|E(A, A^c)|}{|A|} \quad \text{and} \quad |A| \leq \frac{1}{2}|V|.$$

Let f be the orthogonal projection of the characteristic function $\mathbb{1}_A$ of A to $L^2(V)^\circ$; that means

$$f := \mathbb{1}_A - \frac{|A|}{|V|}\mathbb{1}_V.$$

Then by Lemma 9, we have

$$k(1-\lambda_2) \leq \frac{\|df\|_2^2}{\|f\|_2^2}. \tag{2.15}$$

Notice that $df = d\mathbb{1}_A$, and for every edge e, we have

$$|d\mathbb{1}_A(e)| = |\mathbb{1}_A(e^+) - \mathbb{1}_A(e^-)| = \mathbb{1}_{E(A,A^c)}(e). \tag{2.16}$$

By (2.15) and (2.16), we deduce that

$$k(1-\lambda_2) \leq \frac{\|\mathbb{1}_{E(A,A^c)}\|_2^2}{\|f\|_2^2} = \frac{|E(A,A^c)|}{\|\mathbb{1}_A\|_2^2 - \left(\frac{|A|}{|V|}\right)^2\|\mathbb{1}_V\|_2^2} = \frac{|E(A,A^c)|}{|A| - \frac{|A|^2}{|V|}}. \tag{2.17}$$

The term in the denominator is $|A|\left(\frac{|A|^c}{|V|}\right)$, and $|A^c| \geq |V|/2$. Hence the denominator in (2.17), is at least $|A|/2$. Therefore, we deduce that

$$k(1-\lambda_2) \leq 2\frac{|E(A,A^c)|}{|A|} = 2h(\mathscr{G}). \tag{2.18}$$

By (2.14) and (2.18), we obtain that

$$\frac{1-\lambda_2}{2} \leq \frac{h(\mathscr{G})}{k} \leq \sqrt{2(1-\lambda_2)},$$

which finishes the proof. □

Based on these isoperimetric inequalities, we get a spectral description of a family of expander graphs.

Theorem 16. *An infinite family $\{\mathscr{G}_i\}_{i\in I}$ of k-regular graphs is a family of expanders if and only if $\sup_i \lambda_2(\mathscr{G}_i) < 1$. In particular, a family $\{\mathscr{G}_i\}_{i\in I}$ of k-regular graphs is a family expanders if $\sup_i \lambda_{\mathscr{G}_i} < 1$.*

Proof. If $\{\mathscr{G}_i\}_i$ is a family of expanders, then there is a positive number c_0 such that $h(\mathscr{G}_i) \geq c_0$ for every i. By Theorem 15, we deduce that

$$\sqrt{2(1-\lambda_2(\mathscr{G}_i))} \geq \frac{c_0}{k}, \quad \text{and so} \quad \lambda_2(\mathscr{G}_i) \leq 1 - \frac{1}{2}\left(\frac{c_0}{k}\right)^2$$

for every i. Hence $\sup_i \lambda_2(\mathscr{G}_i) < 1$.

For the converse, suppose $\sup_i \lambda_2(\mathscr{G}_i) = c_0'$ for some $c_0' < 1$. Then by Theorem 15, for every i, we have

$$\frac{k}{2}(1-c_0') \leq h(\mathscr{G}_i),$$

and so $\{\mathscr{G}_i\}_i$ is a family of expanders.

The final claim follows from the fact that $\lambda_{\mathscr{G}_i} = \max\{|\lambda_2(\mathscr{G}_i)|, |\lambda_n(\mathscr{G}_i)|\}$. □

In some texts, having a non-trivial bound for $\lambda_2(\mathscr{G}_i)$'s is called *one-sided expander* and having a non-trivial bound for $\lambda_{\mathscr{G}_i}$'s is called *two-sided expander*.

2.5 Exercises

1. Suppose $\mathscr{G}$ is a finite graph that has a bipartite connected component. Prove that $\mathscr{T}_{\mathscr{G}}$ has eigenvalue -1.

   (Hint: Suppose V has two disjoint subsets A and B such that if an edge e intersects $A \cup B$, then $|e \cap A| = |e \cap B| = 1$. Let $f := \mathbb{1}_A - \mathbb{1}_B$ where for a subset Y of V, $\mathbb{1}_Y$ is the characteristic function of Y. Prove that $\mathscr{T} f = -f$.)

2. Prove the L^1-convergence and the mixing property of a random-walk in a finite regular graph given in Proposition 8.

   (Hint. For the mixing, use the Cauchy-Schwarz inequality and obtain

   $$\left|\left\langle f, \mathscr{T}^l g - \frac{\langle g, \mathbb{1}_V\rangle}{|V|}\mathbb{1}_V \right\rangle\right| \leq \|f\|_2 \left\|\mathscr{T}^l g - \frac{\langle g, \mathbb{1}_V\rangle}{|V|}\mathbb{1}_V\right\|_2 \leq \lambda_{\mathscr{G}}^l \|f\|_2 \|g\|_2,$$

   and finish the proof. For the L^1-convergence, use the mixing inequality for g equals to the initial distribution μ_0. Notice that

   $$\langle f, \mathscr{T}^l \mu_0\rangle = \mathbb{E}[f(X_l)] \quad \text{and} \quad \sum_{v\in V} \mu_0(v) = 1.)$$

3. Suppose $\mathscr{G} = (V, E)$ is a directed graph. Let

   $$d : L^2(V) \to L^2(E), df(e) := f(e^+) - f(e^-).$$

   Prove that

   $$d^* g(v) = \sum_{v=e^+} g(e) - \sum_{v=e^-} g(e).$$

(Hint. Notice that

$$\begin{aligned}\langle df, g\rangle &= \sum_{e\in E} \overline{df(e)}g(e) = \sum_{e\in E}(\overline{f(e^+)} - \overline{f(e^-)})g(e)\\ &= \sum_{v\in V}\overline{f(v)}(\sum_{v=e^+} g(e) - \sum_{v=e^-} g(e)).)\end{aligned}$$

Chapter 3

Fourier analysis and equidistribution

Based on the spectral description of expanders (see Theorem 16), we can say that super-approximation is about the study of the rate of convergence of random-walks to equidistribution on Cayley graphs of congruence quotients of finitely generated subgroups of $\mathrm{GL}_n(\mathbb{Z}[1/q_0])$.

One of the classical tools for proving convergence to equidistribution is Fourier analysis. In this chapter, we start by reviewing how Fourier analysis can help us prove the equidistribution of irrational rotations. Then the basics of Fourier analysis on finite groups is reviewed. We finish this chapter by defining and proving basic properties of *quasi-random groups*; a concept introduced by Gowers. A *mixing inequality* and a *product result for large subsets* will be proved for a quasi-random group.

3.1 Equidistribution of irrational rotations

We identify $\mathbb{R}/\mathbb{Z}$ with a circle. This way a rotation can be identified with addition in $\mathbb{R}/\mathbb{Z}$,

$$x+\mathbb{Z}\mapsto \alpha+x+\mathbb{Z}.$$

We say a sequence $\{a_i+\mathbb{Z}\}_{i=1}^{\infty}$ of points in $\mathbb{R}/\mathbb{Z}$ is *equidistributed* if for every smooth function $f:\mathbb{R}/\mathbb{Z}\to\mathbb{C}$, we have

$$\lim_{N\to\infty}\frac{\sum_{i=1}^{N} f(a_i+\mathbb{Z})}{N}=\int_{\mathbb{R}/\mathbb{Z}} f(t)dt. \tag{3.1}$$

The following is a classical result, and we give a sketch of its proof based on Fourier analysis on the compact abelian group $\mathbb{R}/\mathbb{Z}$.

Theorem 17. *Suppose α is an irrational number. Then $\{n\alpha+\mathbb{Z}\}_{n=1}^{\infty}$ is equidistributed in $\mathbb{R}/\mathbb{Z}$.*

Notice that if a sequence $\{a_i+\mathbb{Z}\}_{i=1}^{\infty}$ of points is equidistributed, then their shift $\{t+a_i+\mathbb{Z}\}_{i=1}^{\infty}$ by t is also equidistributed for every t. Hence $\{a_i+\mathbb{Z}\}_{i=1}^{\infty}$ is equidistributed if and only if $\{T_N(f)\}_{N=1}^{\infty}$ converges to the constant function $\langle f,\mathbb{1}\rangle\mathbb{1}$ pointwise where

$$T_N(f)=\frac{\sum_{i=1}^{N} f(t+a_i+\mathbb{Z})}{N}.$$

It is often hard to prove *pointwise convergence* of a sequence of functions. So first we start with proving the *weak convergence*.

Lemma 18. *Suppose α is an irrational number, and $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is a smooth function. For a positive integer N, let*

$$T_N(f)(t) = \frac{\sum_{n=1}^{N} f(t + n\alpha + \mathbb{Z})}{N}.$$

Then for every $g \in L^2(\mathbb{R}/\mathbb{Z})$,

$$\lim_{N\to\infty} \langle T_N(f), g\rangle = \langle f, \mathbb{1}\rangle\langle \mathbb{1}, g\rangle. \tag{3.2}$$

Proof. From Fourier analysis on $\mathbb{R}/\mathbb{Z}$, we know that $\{e_n(t)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R}/\mathbb{Z})$ where

$$e_n(t + \mathbb{Z}) := e^{2\pi i n t}.$$

Hence we have

$$g = \sum_{n\in\mathbb{Z}} \hat{g}(n)e_n \quad \text{in the } L^2\text{-norm, where } \hat{g}(n) := \langle e_n, g\rangle.$$

This means $\lim_{N\to\infty} \|g - \sum_{|n|\le N} \hat{g}(n)e_n\|_2 = 0$. Hence by the Cauchy-Schwarz inequality,

$$\langle T_N(f), g\rangle = \sum_{n\in\mathbb{Z}} \hat{g}(n)\langle T_N(f), e_n\rangle \tag{3.3}$$

By (3.3), for every $\varepsilon > 0$, there is M_ε such that

$$|\langle T_N(f), g\rangle - \sum_{|n|\le M_\varepsilon} \hat{g}(n)\langle T_N(f), e_n\rangle| \le \varepsilon. \tag{3.4}$$

(So to speak, we are focusing on the *low frequencies*.) Suppose we have proved that (3.2) holds for $g = e_n$ for every integer n. Then for every $\varepsilon > 0$, there exists N_ε such that

$$|\langle T_N(f), e_n\rangle - \langle f, \mathbb{1}\rangle\langle \mathbb{1}, e_n\rangle| \le \frac{\varepsilon}{M_\varepsilon \max\{|\hat{g}(n)|| \, |n| \le M_\varepsilon\}} \tag{3.5}$$

for every $n \le M_\varepsilon$ and $N \ge N_\varepsilon$. Notice that $\mathbb{1} = e_0$, and so $\langle \mathbb{1}, e_n\rangle = [n = 0]$. Therefore by (3.4) and (3.5), we obtain that

$$|\langle T_N(f), g\rangle - \langle f, \mathbb{1}\rangle\hat{g}(0)| \le 2\varepsilon$$

for every $N \ge N_\varepsilon$. Notice that $\hat{g}(0) = \langle \mathbb{1}, g\rangle$, and so (3.2) follows. A closer look at the above argument shows that in order to prove a weak convergence it is enough to show it for *test functions* from an orthonormal basis.

Next we show that

$$\lim_{N\to\infty} \langle T_N(f), e_m\rangle = \langle f, \mathbb{1}\rangle[m = 0]. \tag{3.6}$$

To this end, we notice that

$$T_N(f) = \frac{\sum_{n=1}^N \lambda(-n\alpha)(f)}{N} \quad \text{where} \quad (\lambda(a)(f))(t) := f(-a+t),$$

and

$$\begin{aligned}\widehat{\lambda(a)(f)}(m) &= \int_{\mathbb{R}/\mathbb{Z}} (\lambda(a)(f))(t)\overline{e_m(t)}dt = \int_{\mathbb{R}/\mathbb{Z}} f(-a+t)\overline{e_m(t)}dt \\ &= \int_{\mathbb{R}/\mathbb{Z}} f(s)\overline{e_m(a+s)}ds = e_m(a)\int_{\mathbb{R}/\mathbb{Z}} f(s)\overline{e_m(s)}ds = \quad e_m(a)\hat{f}(m).\end{aligned}$$

Hence

$$\widehat{T_N(f)}(m) = \hat{f}(m)\left(\frac{\sum_{n=1}^N e_m(-n\alpha)}{N}\right). \tag{3.7}$$

By (3.7), we see that (3.6) holds if and only if there is a non-trivial *cancellation* in the *exponential sum* $\sum_{n=1}^N e_m(-n\alpha)$. Here this cancellation comes for free as the given exponential sum is simply a geometric series. This technique (due to H. Weyl), however, is quite general (see Exercise 2).

Since α is irrational, if $m \neq 0$, then

$$|\sum_{n=1}^N e_m(-n\alpha)| = \left| e(-m\alpha)\frac{1-e(-Nm\alpha)}{1-e(-m\alpha)}\right| \le \frac{2}{|1-e(-m\alpha)|}.$$

Hence by (3.7), we obtain that

$$|\widehat{T_N(f)}(m)| \le \frac{2|\hat{f}(m)|}{N|1-e(-m\alpha)|}.$$

This implies that $\lim_{N\to\infty} \widehat{T_N(f)}(m) = 0$ if $m \neq 0$. For $m = 0$, it is clear that $\widehat{T_N(f)}(0) = \hat{f}(0) = \langle \mathbb{1}, f\rangle$. Altogether (3.5) follows, which in turn completes the proof. □

Next we show how one can go from a weak convergence to a pointwise convergence.

Proof of Theorem 17. Suppose $\{\psi_\varepsilon\}$ is a family of smooth functions on $\mathbb{R}/\mathbb{Z}$ converging to the dirac mass at $x_0 \in \mathbb{R}/\mathbb{Z}$; that means the following properties hold:

1. $\psi_\varepsilon \ge 0$, $\langle \psi_\varepsilon, \mathbb{1}\rangle = 1$.
2. The support of ψ_ε is a subset of the ε-neighborhood of x_0.
3. For every smooth function $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$, $\lim_{\varepsilon\to 0}\langle \psi_\varepsilon, f\rangle = f(x_0)$.

Since $\mathbb{R}/\mathbb{Z}$ is compact, every continuous function f on $\mathbb{R}/\mathbb{Z}$ is uniformly continuous. Hence for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $x, y \in \mathbb{R}/\mathbb{Z}$ that are δ-close, we have

$$|f(x) - f(y)| \le \varepsilon.$$

Hence for every positive integer N, we obtain

$$|T_N(f)(x) - T_N(f)(y)| \le \frac{1}{N}\sum_{n=1}^{N} |f(n\alpha + x) - f(n\alpha + y)| \le \varepsilon \tag{3.8}$$

if x and y are δ-close. Therefore by (3.8), for every y in the support of ψ_δ, we have $|T_N(f)(y) - T_N(f)(x_0)| \le \varepsilon$. Thus

$$|\langle \psi_\delta, T_N(f)\rangle - T_N(f)(x_0)| \le \varepsilon, \tag{3.9}$$

for every δ that is small enough depending on ε and f. On the other hand, by Lemma 18, for a fixed δ, if N is large enough depending on ε and δ, we have

$$|\langle \psi_\delta, T_N(f)\rangle - \langle f, \mathbb{1}\rangle| \le \varepsilon. \tag{3.10}$$

By (3.9) and (3.10), we obtain that

$$|\langle \mathbb{1}, f\rangle - T_N(f)(x_0)| \le 2\varepsilon,$$

for every N which is large enough depending on ε. This implies that

$$\lim_{N\to\infty} T_N(f)(x_0) = \int_{\mathbb{R}/\mathbb{Z}} f(t)dt,$$

which finishes the proof. □

3.2 Fourier analysis on finite groups

The Fourier analysis on a compact group G is the study of its (unitary) representations as a way of analyzing its (complex valued) functions. In the proof of the equidistribution of irrational rotations, we saw the importance of having an orthonormal basis of $L^2(\mathbb{R}/\mathbb{Z})$ that consists of eigenfunctions of the (left) translation by $\mathbb{R}/\mathbb{Z}$. When G is a non-abelian compact group, $L^2(G)$, however, does not have an orthonormal eigenbasis for the left translation action of G. Instead, we write $L^2(G)$ as a direct sum of *irreducible representations* of G, and use that to find a semi-canonical orthonormal basis for $L^2(G)$. This in turn can help us prove the Plancherel theorem.

In this section, we study the Fourier analysis only on *finite groups*, but as much as possible, we treat it in a way that can be extended to compact groups.

A group homomorphism $\pi : G \to \mathrm{GL}_n(\mathbb{C})$ is called a *representation* of G (for a compact group, we further assume that π is continuous). We say π is a representation of *degree* n and write $\deg \pi = n$ if it is a group homomorphism from G to $\mathrm{GL}_n(\mathbb{C})$. Notice that a group representation of degree n induces a G-group action on $\mathbb{C}^n$; that means for every $v \in \mathbb{C}^n$ and $x \in G$, $x \cdot v := \pi(x)(v)$ is a group action. To emphasize that we are looking at $\mathbb{C}^n$ together with the G-action induced by π, we write V_π instead of $\mathbb{C}^n$; and so $\deg \pi = \dim V_\pi$. Since the G-group action on V_π is linear, we can view V_π as a $\mathbb{C}G$-module where

$$\mathbb{C}G := \left\{ \sum_{x\in G} f(x)x \,|\; f(x) \ne 0 \text{ only for finitely many terms}\right\}$$

is the group ring of G. This point of view works the best when G is a finite group.

We say two representation π_1 and π_2 are *equivalent* if there is a $\mathbb{C}$-isomorphism $T : V_{\pi_1} \to V_{\pi_2}$ such that the following is a commuting diagram

$$\begin{tikzcd} V_{\pi_1} \arrow[r, "\pi_1(x)"] \arrow[d, "T"] & V_{\pi_1} \arrow[d, "T"] \\ V_{\pi_2} \arrow[r, "\pi_2(x)"] & V_{\pi_2} \end{tikzcd}$$

for every $x \in G$. This means for every $x \in G$, we have

$$\pi_1(x) = T^{-1}\pi_2(x)T.$$

We say a representation is *unitary* if V_π has an inner product which is invariant under G. Notice that for every inner product $\langle \cdot, \cdot \rangle'$ on $\mathbb{C}^n$, there is a basis $\{u_1, \ldots, u_n\}$ of $\mathbb{C}^n$ such that $\langle u_i, u_j \rangle' = [i = j]$. Suppose $T : \mathbb{C}^n \to \mathbb{C}^n$ is such that $Te_i = u_i$ for every i. Then for every $v, w \in \mathbb{C}^n$, we have

$$\langle v, w \rangle = \langle Tv, Tw \rangle',$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product. If $\langle \cdot, \cdot \rangle'$ is G-invariant, we obtain that for every $x \in G$ and $v, w \in \mathbb{C}^n$ the following holds

$$\langle v, w \rangle = \langle Tv, Tw \rangle' = \langle \pi(x)Tv, \pi(x)Tw \rangle' = \langle T^{-1}\pi(x)Tv, T^{-1}\pi(x)Tw \rangle.$$

Hence π is equivalent with a representation $\pi' : G \to \mathrm{U}_n(\mathbb{C})$ where

$$\mathrm{U}_n(\mathbb{C}) := \{y \in \mathrm{GL}_n(\mathbb{C}) |\ y^*y = I\}$$

is the group of n-by-n matrices that preserve the standard inner product of $\mathbb{C}^n$.

The set of fixed points of every group action is of special interest in almost every example. When the action has additional geometric properties, one can use an *averaging* technique to find a *projection* to the set of fixed points; for instance consider an affine action on a convex set, isometries on a non-positive curvature symmetric space or more generally isometries on a CAT(0) space. For the purposes of understanding representations of a finite group, we use the averaging technique for the following actions.

1. (Space of inner products) Let P_n^+ be the set of n-by-n positive definite Hermitian forms:

$$P_n^+ := \{a \in \mathrm{GL}_n(\mathbb{C}) |\ a = a^*, \forall v \in \mathbb{C}^n \setminus \{0\}, \langle \overline{v} | a | v \rangle > 0\}.$$

Notice that for every $a \in P_n^+$, the following defines an inner product on $\mathbb{C}^n$:

$$\langle v, w \rangle_a := \langle \overline{v} | a | w \rangle.$$

Vice versa, if $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{C}^n$, then $[\langle e_i, e_j \rangle]$ is in P_n^+ (why?). For every $x \in \mathrm{GL}_n(\mathbb{C})$ and inner product $\langle \cdot, \cdot \rangle$, the following is another inner product

$$\langle xv, xw \rangle.$$

This gives us a group action of $\mathrm{GL}_n(\mathbb{C})$ on the set of inner products. Working with elements of P_n^+, we can concretely see this action: the following gives us an affine action of $\mathrm{GL}_n(\mathbb{C})$ on P_n^+

$$x \cdot a := x^* a x$$

(see Exercise 5). Suppose G is a compact subgroup of $\mathrm{GL}_n(\mathbb{C})$. Let $\int_G f(x)dx$ be the integration with respect to the unique Borel G-invariant probability measure on G (this measure exists and it is called the *Haar measure* of G). For instance, for a finite group G,

$$\int_G f(x)dx = \frac{\sum_{x\in G} f(x)}{|G|}.$$

Let $A_G : P_n^+ \to \mathrm{M}_n(\mathbb{C})$, $A_G(a) := \int_G x \cdot a\; da$; this means we are integrating each entry separately. Since P_n^+ is a convex, one can show that the image of A_G is a subset of P_n^+. Because the considered measure is G-invariant, one can show that the image of A_G is exactly the set of G-fixed points of P_n^+.

2. (Intertwining operators) Suppose $\pi_1 : G \to \mathrm{GL}(V_1)$ and $\pi_2 : G \to \mathrm{GL}(V_2)$ are two representations of G. Let $\mathrm{Hom}_{\mathbb{C}}(V_1, V_2)$ be the set of $\mathbb{C}$-linear homomorphisms from V_1 to V_2. For every $x \in G$ and $a \in \mathrm{Hom}_{\mathbb{C}}(V_1, V_2)$, let

$$x \cdot a : V_1 \to V_2, \quad (x \cdot a)(v_1) := \pi_2(x)(a(\pi_1(x^{-1})(v_1)));$$

alternatively, we can write

$$x \cdot a := \pi_2(x) \circ a \circ \pi_1(x)^{-1}.$$

Notice that $\cdot$ defines a linear action of G on $\mathrm{Hom}_{\mathbb{C}}(V_1, V_2)$. Viewing V_i's as $\mathbb{C}G$-modules, $a \in \mathrm{Hom}_{\mathbb{C}}(V_1, V_2)$ is a G-fixed point precisely when $a : V_1 \to V_2$ is a $\mathbb{C}G$-module homomorphism. Such a map sometimes called an *intertwining operator*, and the set of all intertwining operators is denoted by $I(\pi_1, \pi_2)$. So

$$I(\pi_1, \pi_2) = \mathrm{Hom}_{\mathbb{C}G}(V_1, V_2) = \mathrm{Hom}_{\mathbb{C}}(V_1, V_2)^G.$$

As before, for a compact group G and two of its representations π_1, π_2, let

$$A_G : \mathrm{Hom}_{\mathbb{C}}(V_1, V_2) \to \mathrm{Hom}_{\mathbb{C}G}(V_1, V_2), \quad A_G(a) := \int_G x \cdot a\; dx,$$

and notice since the action is linear and the measure is G-invariant, A_G is well-defined.

Let's make the above statements more precise by proving a more general result about affine actions of finite groups. Let's recall that we say that a subset X of $\mathbb{C}^n$ is called a *convex* subset if for every $x_1, x_2 \in X$, the segment connecting them is a subset of X; that means

$$\{p_1 x_1 + p_2 x_2 |\ 0 \le p_1, p_2 \text{ and } p_1 + p_2 = 1\} \subseteq X.$$

Suppose X is a convex subset of $\mathbb{C}^n$ and G acts on X. We say this is an *affine action* if for every $x_1, x_2 \in X$, non-negative p_1 and p_2 such that $p_1 + p_2 = 1$, and $g \in G$, we have

$$g \cdot (p_1 x_1 + p_2 x_2) = p_1\ g \cdot x_1 + p_2\ g \cdot x_2.$$

Notice that if X is a convex set and G acts on X by affine transformations, then for every $x_1, \ldots, x_n \in X$, non-negative numbers $p_1, \ldots, p_n$ that add up to 1, and $g \in G$, we have that

$$\sum_{i=1}^{m} p_i x_i \in X \quad \text{and} \quad g \cdot (\sum_{i=1}^{n} p_i x_i) = \sum_{i=1}^{n} p_i\ g \cdot x_i,$$

(see Exercise 6).

Lemma 19 (Averaging trick)**.** *Suppose G is a finite group, X is a convex subset of $\mathbb{C}^n$, and G acts by affine transformations on X. Then the set X^G of the G-fixed points is non-empty, the map*

$$A_G : X \to X^G, \quad A_G(x) := \int_G g \cdot x\, dg$$

where $\int_G f(g)\, dg = \frac{1}{|G|} \sum_{g \in G} f(g)$ is well-defined, and $A_G \circ A_G = A_G$.

Proof. For every $x \in X$, the G-orbit of x is a subset of X, and so by the previous remark (see Exercise 6),

$$\frac{1}{|G|} \sum_{g \in G} g \cdot x \in X \text{ and so } A_G(x) \in X.$$

Next we show that $A_G(x)$ is a G-fixed point and deduce that X^G is not empty. For every $g \in G$, by Exercise 6, we have

$$\begin{aligned} g \cdot A_G(x) =& g \cdot \left(\frac{1}{|G|} \sum_{g' \in G} g' \cdot x \right) = \frac{1}{|G|} \sum_{g' \in G} g \cdot (g' \cdot x) \\ =& \frac{1}{|G|} \sum_{g' \in G} (gg') \cdot x = \frac{1}{|G|} \sum_{g' \in G} g' \cdot x = A_G(x). \end{aligned}$$

Hence $A_G(x) \in X^G$. Finally we notice that if $x \in X^G$, then $A_G(x) = x$. Therefore $A_G \circ A_G = A_G$, which finishes the proof. □

Next we see how the mentioned actions on the space of inner products and linear maps between spaces of two representations together with Lemma 19 help us prove *complete reducibility* of every (finite dimensional) representation of a finite group, and prove *Schur's orthogonality relations*.

Lemma 20 (Unitarization)**.** *Suppose G is a compact group, and $\pi : G \to \mathrm{GL}_n(\mathbb{C})$ is a finite dimensional representation. Then there exists a unitary representation $\pi' : G \to \mathrm{U}_n(\mathbb{C})$ which is equivalent to π.*

Proof. Consider the affine action of G on P_n^+ via π, and notice that P_n^+ is a convex subset of $\mathrm{M}_n(\mathbb{C})$. Hence by Lemma 19, there is a G-fixed point a. This implies that the inner product $\langle\cdot,\cdot\rangle_a$ is G-invariant, and so π is equivalent to a unitary representation. $\square$

Lemma 21 (Complete reducibility)**.** *Suppose G is a finite group, and $\pi : G \to \mathrm{GL}(V)$ is a complex representation of G. Then V is a completely reducible $\mathbb{C}G$-module.*

We also say that the representation $\pi : G \to \mathrm{GL}(V)$ is a completely reducible representation if V is a completely reducible $\mathbb{C}G$-module.

Proof. By Lemma 20, we can and will assume that π is a unitary representation. We proceed by induction on $\dim V$ to prove that V is a completely reducible $\mathbb{C}G$-module. The base of induction is clear. If V does not have any non-trivial $\mathbb{C}G$-module, then V is a simple $\mathbb{C}G$-module, and there is nothing to prove. Next assume that W is a non-trivial $\mathbb{C}G$-module. We claim that

$$W^\perp := \{v \in V | h\forall w \in W, \langle v, w\rangle = 0\}$$

is a $\mathbb{C}G$-submodule. For every $g \in G$, $v, w \in V$, we have

$$\langle \pi(g)(v), w\rangle = \langle v, \pi(g^{-1})(w)\rangle \tag{3.11}$$

as $\pi(g)$ is unitary. Notice that for $w \in W$, $\pi(g^{-1})(w) \in W$ as W is G-invariant. Hence by (3.11), we obtain that for every $v \in W^\perp$ and $w \in W$ the following holds:

$$\langle \pi(g)(v), w\rangle = 0;$$

and so $\pi(g)(v) \in W^\perp$. This implies that $W^\perp$ is a G-invariant subspace of V. Hence $W^\perp$ is a $\mathbb{C}G$-submodule of V. Notice that since $\langle\cdot,\cdot\rangle$ is an inner product, $V = W \oplus W^\perp$. Since W and $W^\perp$ are $\mathbb{C}G$-modules of dimension smaller than $\dim V$, by the induction hypothesis, they are completely reducible. Thus V is completely reducible. $\square$

Using Lemma 20, we can and will focus on studying only *unitary representations* of a finite group. One of most important unitary representations of a finite group G is its action on $L^2(G)$ given by *left translations* (or right translations)

$$\lambda_0 : G \to \mathrm{GL}(L^2(G)), \quad (\lambda_0(x)(f))(y) := f(x^{-1}y),$$

and

$$\rho_0 : G \to \mathrm{GL}(L^2(G)), \quad (\rho_0(x)(f))(y) := f(yx).$$

In fact, for us the study of representations of compact groups is a mean to analyze the space of (L^2-integrable) functions on G. Notice that for a finite group G, every complex function on G is in $L^2(G)$, and the only reason we emphasize on this perspective is because we want to remind ourselves of the natural inner product on this space which is G-invariant (for both left and right translations):

$$\langle f, g\rangle := \frac{1}{|G|}\sum_{x\in G} \overline{f(x)}g(x).$$

Sometimes we work with the *counting measure* on G, instead of the *probability counting measure*; that means we let

$$\langle f, g\rangle := \sum_{x\in G} \overline{f(x)}g(x).$$

In order to distinguish these spaces, we denote the former with $\widetilde{L^2}(G)$ and the latter with $L^2(G)$.

To every unitary representation $\pi : G \to \mathrm{U}_n(\mathbb{C})$, we associate a subspace H_π of $L^2(G)$. We refer to H_π as the space of all the *matrix coefficients* of π. It is defined as the span of $\{e_{ij} \circ \pi |\ 1 \le i, j \le n\}$ where $e_{ij}(x)$ is the (i, j)-entry of x. Here are some of the basic properties of the space of matrix coefficients of a unitary representation.

Lemma 22 (Space of matrix coefficients)**.** *1. Suppose π and π' are two equivalent representations of G. Then they have the same space of matrix coefficients; that means $H_\pi = H_{\pi'}$.*

2. Suppose π is a unitary representation. Then for every $v, w \in V_\pi$,

$$f_{v,w}(x) := \langle \pi(x)v, w\rangle \in H_\pi.$$

3. For a representation π, H_π is a $G \times G$-invariant subspace of $L^2(G)$, where the action of $G \times G$ on $L^2(G)$ is given by

$$((x_1, x_2) \cdot f)(y) := f(x_1^{-1}yx_2).$$

4. If $\pi, \pi_1, \ldots, \pi_m$ are a representation of G and $V_\pi = \sum_{j=1}^m V_{\pi_j}$, then

$$H_\pi = \sum_{j=1}^{m} H_{\pi_j}.$$

Proof. Since π and π' are equivalent, $\deg\pi = \deg\pi' = n$ for some positive integer n, and there is $g \in \mathrm{GL}_n(\mathbb{C})$ such that for every $x \in G$,

$$\pi_2(x) = g\pi_1(x)g^{-1}. \tag{3.12}$$

Hence for every $1 \le i, j \le n$, we have

$$e_{i,j} \circ \pi' = \sum_{l,k} e_{i,l}(g)e_{k,j}(g^{-1})e_{l,k} \circ \pi,$$

and so $H_{\pi'} \subseteq H_\pi$. Similarly we have that $H_\pi \subseteq H_{\pi'}$, which implies the first part.

To show the second part, we notice that

$$f_{v,w}(x) = \sum_{i,j} \overline{v}_j w_i \langle \pi(x)(e_j), e_i\rangle = \sum_{i,j} \overline{v}_j w_i e_{ij} \circ \pi(x)$$

for every $v, w \in V_\pi$. This implies that

$$f_{v,w} = \sum_{i,j} \overline{v}_j w_i \, e_{ij} \circ \pi \in H_\pi.$$

The third part is obtained from the following equation which implies that the set $\{f_{v,w}\}_{v,w\in V_\pi}$ is invariant under the action of $G \times G$:

$$\begin{aligned}((x_1, x_2) \cdot f_{v,w})(y) =& f_{v,w}(x_1^{-1} y x_2)\\ =& \langle \pi(x_1^{-1} y x_2)v, w\rangle\\ =& \langle \pi(y)(\pi(x_2)v), \pi(x_1)(w)\rangle = f_{\pi(x_2)(v),\pi(x_1)(w)}(y).\end{aligned}$$

Notice that since $f_{e_j,e_i} = e_{ij} \circ \pi$, the span of $f_{v,w}$'s is H_π.

The last part is left as an exercise. $\square$

The main algebraic lemma which helps us make use of an *irreducibility* assumption is the following which is known as Schur's lemma.

Lemma 23 (Schur's lemma)**.** *Suppose G is a finite group.*

1. *If π_1 and π_2 are two irreducible representations of G and $f \in I(\pi_1, \pi_2)$ is an intertwining operator, then either $f = 0$ or f is a bijection and π_1 and π_2 are equivalent. In particular, if π_1 and π_2 are not equivalent, $I(\pi_1, \pi_2) = \{0\}$.*

2. *Suppose π is an irreducible representation of G. Then the set $I(\pi, \pi)$ of intertwining operators between π and itself consists of scalars; that means*

$$I(\pi, \pi) = \{c \, \mathrm{id}_{V_\pi} \mid c \in \mathbb{C}\}.$$

Proof. Since f is an intertwining operator, $\ker f$ and $\operatorname{im} f$ are G-invariant. Because both π_1 and π_2 are irreducible, we deduce that $\ker f$ and $\operatorname{im} f$ are trivial submodules of V_{π_1} and V_{π_2}, respectively. If $\ker f = V_{\pi_1}$ or $\operatorname{im} f = \{0\}$, then $f = 0$. Otherwise, $\ker f = \{0\}$ and $\operatorname{im} f = V_{\pi_2}$. The former implies that f is injective, and latter implies that f is surjective. Hence f is a bijection if it is not zero. Existence of a bijective intertwining operator implies that π_1 and π_2 are equivalent. The first part follows.

Suppose $f \in I(\pi, \pi)$, and let $c \in \mathbb{C}$ be an eigenvalue of f. Then $f - c \, \mathrm{id}_{V_\pi}$ is a non-bijective element of $I(\pi, \pi)$. Hence by the first part, it is zero. This means $f = c \, \mathrm{id}_{V_\pi}$, which finishes the proof. $\square$

Using Schur's lemma (Lemma 23) and the averaging trick (see Lemma 19), we prove Schur's orthogonality relations which play a central role in the Fourier analysis on finite groups.

Theorem 24 (Schur's orthogonality relations)**.** *Suppose G is a finite group.*

1. *If π_1 and π_2 are two non-equivalent unitary irreducible representations of G, then H_{π_1} is perpendicular to H_{π_2}.*

2. *If π is a unitary irreducible representation of G, then for every $a \in \operatorname{Hom}(V_\pi, V_\pi)$*

$$\frac{1}{|G|}\sum_{x\in G}\pi(x)\circ a\circ\pi(x)^{-1} = \frac{\operatorname{tr} a}{\deg \pi}\operatorname{id}_{V_\pi}.$$

3. *If π is a unitary irreducible representation of G, then $\{\sqrt{\deg \pi}\, e_{ij}\circ\pi\}_{i,j}$ is an orthonormal basis of $H_\pi \subseteq \widetilde{L^2}(G)$.*

Proof. By Lemma 19, for every $a \in \operatorname{Hom}(V_{\pi_1}, V_{\pi_2})$,

$$\frac{1}{|G|}\sum_{g\in G}\pi_2(x)\circ a\circ\pi_1(x)^{-1} \in I(\pi_1,\pi_2).$$

By Schur's lemma, $I(\pi_1,\pi_2)=\{0\}$. Hence

$$\frac{1}{|G|}\sum_{g\in G}\pi_2(x)\circ a\circ\pi_1(x)^{-1} = 0 \tag{3.13}$$

for every $a \in \operatorname{Hom}(V_{\pi_1}, V_{\pi_2})$. Since π_1 is unitary, $\pi_1(x)^{-1} = \pi_1(x)^*$; and so by (3.13), we obtain

$$\begin{aligned}
0 =& \frac{1}{|G|}\sum_{x\in G} e_{i,j}(\pi_2(x)\circ a\circ\pi_1(x)^{-1})\\
=& \frac{1}{|G|}\sum_{x\in G}\sum_{k,l} e_{i,k}(\pi_2(x))e_{k,l}(a)e_{l,j}(\pi_1(x)^{-1})\\
=& \sum_{k,l} e_{k,l}(a)\frac{1}{|G|}\sum_{x\in G} e_{i,k}\circ\pi_2(x)\overline{e_{j,l}\circ\pi_1(x)}\\
=& \sum_{k,l} e_{k,l}(a)\langle e_{j,l}\circ\pi_1, e_{i,k}\circ\pi_2\rangle
\end{aligned} \tag{3.14}$$

Applying (3.14) for a with 1 in its (k,l)-entry and 0 everywhere else, we obtain that

$$e_{j,l}\circ\pi_1 \perp e_{i,k}\circ\pi_2$$

for every i, j, k, and l. This implies the first part.

For the second part, we again use Lemma 19 and Schur's lemma to obtain that

$$\frac{1}{|G|}\sum_{x\in G}\pi(x)\circ a\circ\pi(x)^{-1} = c\operatorname{id}_{V_\pi} \tag{3.15}$$

for some $c \in \mathbb{C}$. Taking the trace of both sides of (3.15), we deduce that

$$\operatorname{tr} a = c \deg \pi.$$

Hence $c = \frac{\operatorname{tr} a}{\deg \pi}$, and so by (3.15), the second part follows.

Using the second part, similar to a computation done in (3.13), we obtain

$$\frac{\operatorname{tr} a}{\deg \pi}[i=j] = \sum_{k,l} e_{k,l}(a)\langle e_{j,l}\circ \pi_1, e_{i,k}\circ \pi_2\rangle \tag{3.16}$$

for every i, j. Applying (3.16) for a with 1 in its (k,l)-entry and 0 everywhere else, we obtain that

$$\frac{[k=l][i=j]}{\deg \pi} = \langle e_{j,l}\circ \pi_1, e_{i,k}\circ \pi_2\rangle.$$

The third part follows, which finishes the proof. □

One of the implications of the third part of Theorem 24 is that $\dim H_\pi = (\deg \pi)^2$ for every unitary irreducible representation π of G. Next we describe the $\mathbb{C}G$-module structure of H_π.

Lemma 25. *Suppose G is a finite group and π is a unitary irreducible representation of G. Then $H_\pi \simeq V_\pi^{\deg \pi}$ as a $\mathbb{C}G$-module where H_π is the space of matrix coefficients of π and is considered as a $\mathbb{C}G$-module via the right-translations action.*

Proof. Suppose $\mathfrak{B} := \{e_1, \ldots, e_n\}$ is an orthonormal basis of V_π. Using the basis $\mathfrak{B}$, identify V_π with $\mathbb{C}^n$ and π with a group homomorphism from G to $\mathrm{U}_n(\mathbb{C})$.

Now notice that for every $1 \le i, j \le n$ and $x \in G$, we have

$$(\rho_0(x)(e_{i,j}\circ\pi))(y) = e_{i,j}\circ \pi(yx) = \sum_{k=1}^{n} e_{i,k}\circ\pi(y) e_{k,j}\circ \pi(x).$$

Hence

$$\rho_0(x)(e_{i,j}\circ \pi) = \sum_{k=1}^{n} e_{k,j}\circ\pi(x)\; e_{i,k}\circ \pi. \tag{3.17}$$

Equation 3.17 implies that

$$H_{\pi,i} := \operatorname{span}\{e_{i,j}\circ\pi \mid 1\le i\le n\}$$

is a G-invariant subspace of H_π (under the right-translations action). Moreover by the third part of Theorem 24, we have that $\{e_{i,j}\circ\pi\}_i$ is an orthogonal basis of $H_{\pi,j}$, $H_{\pi,i}$'s are pairwise orthogonal to each other, and $H_\pi = H_{\pi,1}\oplus\cdots\oplus H_{\pi,n}$.

We also notice that for every j

$$\pi(x)(e_j) = \sum_{k=1}^{n} e_{k,j}\circ \pi(x)\; e_k. \tag{3.18}$$

And so $T_i x : V_\pi \to H_{\pi,i}, \quad T_i(e_j) := e_{i,j}\circ \pi$ is a $\mathbb{C}G$-module isomorphism. Altogether, we obtain that

$$H_\pi = H_{\pi,1}\oplus\cdots\oplus H_{\pi,n} \simeq \underbrace{V_\pi\oplus\cdots\oplus V_\pi}_{n \text{ times}},$$

which finishes the proof. □

For two unitary representations π and ρ of G, we say π is a *subrepresentation* of ρ if there is an embedding $\phi : V_\pi \to V_\rho$ such that for every $v, w \in V_\pi$

$$\langle \phi(v), \phi(w) \rangle = \langle v, w \rangle$$

where the former inner product is taken in V_ρ and the latter in V_π, and for every $x \in G$ and $v \in V_\pi$,

$$\phi(\pi(x)(v)) = \rho(x)(\phi(v)).$$

If π is a subrepresentation of ρ, we write $\pi \le \rho$.

Corollary 26. *Suppose G is a finite group. Suppose $\pi_1, \ldots, \pi_m$ are all the non-equivalent irreducible subrepresentations of the regular representation λ_0. Then*

$$L^2(G) = H_{\pi_1} \oplus \cdots \oplus H_{\pi_m}.$$

Proof. For $f \in L^2(G)$, let M_f be the submodule of $L^2(G)$ generated by f. Let $e_1, \ldots, e_k$ be an orthonormal basis of M_f. Let

$$f_{ij}(x) := \langle \lambda_0(x) e_i, e_j \rangle.$$

Since M_f is G-invariant, for every x, $\lambda_0(x)e_i$ is in the span of e_j's, and so

$$\lambda_0(x)e_i = \sum_{j=1}^m \langle e_j, \lambda_0(x)e_i \rangle e_j = \sum_{j=1}^m \langle \lambda_0(x^{-1})e_j, e_i \rangle e_j = \sum_{j=1}^m f_{ji}(x^{-1})e_j. \quad (3.19)$$

Evaluating both sides of (3.19) at 1, we obtain

$$\lambda_0(x)e_i(1) = \sum_{j=1}^m f_{ji}(x^{-1})e_j(1),$$

and so

$$e_i(x^{-1}) = \sum_{j=1}^m e_j(1) f_{ji}(x^{-1}) \in H_{M_f}.$$

Hence $f \in H_{M_f}$. Therefore by the last part of Lemma 25, we deduce that

$$f \in \sum_{j=1}^m H_{\pi_j}.$$

By the first part of Theorem 24 (Schur's orthogonality relations), we have that H_{π_j}'s are pairwise orthogonal to each other, and for every unitary irreducible subrepresentation π of λ_0, $H_\pi = H_{\pi_j}$ for some j. Altogether, we conclude that

$$L^2(G) = H_{\pi_1} \oplus \cdots \oplus H_{\pi_m}.$$

This finishes the proof. $\square$

For a finite group G, let $\widehat{G}$ be the set of all the non-equivalent irreducible subrepresentations of λ_0.

Corollary 27. *Suppose G is a finite group and π is a unitary irreducible representation of G. Then π is equivalent to an element of $\widehat{G}$.*

Proof. If π is not equivalent to elements of $\widehat{G}$, then by the first part of Theorem 24,

$$H_\pi \perp \sum_{\pi' \in \widehat{G}} H_{\pi'}$$

which contradicts Corollary 26. □

Theorem 28 (Fourier analysis: an orthonormal basis)**.** *Suppose G is a finite group, and $\widehat{G}$ is the set of all the non-equivalent irreducible subrepresentations of the regular representation λ_0. Then*

$$\{\sqrt{\deg \pi}\; e_{i,j} \circ \pi |\; \pi \in \widehat{G}, 1 \le i, j \le \deg \pi\}$$

is an orthonormal basis of $L^{\tilde{2}}(G)$.

Proof. This is an immediate consequence of Corollary 26 and the third part of Theorem 24. □

By Theorem 28, we have that for every $f \in L^2(G)$

$$f = \sum_{\pi \in \widehat{G}} \deg \pi \sum_{i,j} \langle e_{i,j} \circ \pi, f\rangle \; e_{i,j} \circ \pi, \tag{3.20}$$

and so we are interested in

$$\langle e_{i,j} \circ \pi, f\rangle = \int_G \overline{e_{i,j} \circ \pi(x)} f(x) dx = \int f(x) e_{j,i}(\pi(x)^*) dx. \tag{3.21}$$

This brings us to the definition of the Fourier inverse of a function in $L^2(G)$ where G is a finite group. The *Fourier inverse* $\widehat{f}$ of a function $f \in L^2(G)$ is a function on $\widehat{G}$. For every $\pi \in \widehat{G}$, $\widehat{f}(\pi)$ is an element of $\mathrm{Hom}(V_\pi, V_\pi)$ that is the *average* of $\pi(x)^*$'s with the weights given by f; that means

$$\widehat{f}(\pi) := \int_G f(x) \pi(x)^* dx.$$

By (3.20) and (3.21), the following result follows.

Theorem 29 (Fourier expansion)**.** *In the setting of Theorem 28, for every $f \in L^2(G)$, we have*

$$f = \sum_{\pi \in \widehat{G}} \deg \pi \sum_{i,j} e_{j,i}(\widehat{f}(\pi)) \; e_{i,j} \circ \pi;$$

and so for every $x \in G$,

$$f(x) = \sum_{\pi \in \widehat{G}} \deg \pi \;\; \mathrm{tr}(\widehat{f}(\pi)\pi(x)).$$

In the next proposition, we see the basic properties of the Fourier inversion with respect to the regular representations and convolution.

Proposition 30 (Basic properties of Fourier inverse)**.** *Suppose G is a finite group. Then for every $f, g \in L^2(G)$, $\pi \in \widehat{G}$, and $x \in G$, we have*

$$\widehat{\lambda_0(x)f}(\pi) = \widehat{f}(\pi)\pi(x)^*, \widehat{\rho_0(x)f}(\pi) = \pi(x)\widehat{f}(\pi), \text{ and } \widehat{f * g}(\pi) = |G|\ \widehat{g}(\pi)\widehat{f}(\pi).$$

Proof. All the statements are easy to show. Here we only discuss the last one and leave the rest as an exercise. By the definition of convolution and Fourier inverse, we have

$$\begin{aligned}\widehat{f * g}(\pi) = & \int_G (f * g)(x)\pi(x)^* dx = \int_G \sum_{y \in G} f(y)g(y^{-1}x)\pi(x)^* dx \\ = & \sum_{y \in G} f(y)(\int_G \lambda_0(y)(g)(x)\pi(x)^* dx) = \sum_{y \in G} f(y)\widehat{\lambda_0(y)g}(\pi) \\ = & |G| \int_G f(y)\widehat{g}(\pi)\pi(y)^* dy = |G|\ \widehat{g}(\pi)\widehat{f}(\pi).\end{aligned}$$

□

It is worth to point out that if convolution is defined using the *probability* counting measure instead of the counting measure, then the Fourier inverse of the convolution of f and g is simply the product of the Fourier inverse of g and the Fourier inverse of f.

We finish our discussion of Fourier analysis on finite groups by proving the Plancherel Theorem.

Theorem 31 (Plancherel's theorem)**.** *In the setting of Theorem 28, for every $f \in L^{\widetilde{2}}(G)$, we have*

$$\|f\|_{\widetilde{2}}^2 = \sum_{\pi \in \widehat{G}} \deg \pi\ \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2,$$

where $\|a\|_{\mathrm{HS}}^2 = \sum_{i,j} |a_{ij}|^2$ for $a \in \mathrm{M}_n(\mathbb{C})$.

Proof. By Theorems 28 and Theorem 29, we have

$$\|f\|_{\widetilde{2}}^2 = \sum_{\pi \in \widehat{G}} \deg \pi \sum_{i,j} |e_{j,i}(\widehat{f}(\pi))|^2 = \sum_{\pi \in \widehat{G}} \deg \pi\ \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2.$$

This finishes the proof. □

Notice that by Theorem 31 in $L^2(G)$, we have

$$\|f\|_2^2 = |G| \sum_{\pi \in \widehat{G}} \deg \pi\ \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2.$$

3.3 Quasi-randomness, a mixing and a product theorem

As we have seen in the example of equidistribution of irrational rotations, we can use the Fourier inverse of a smooth function on the *circle* $\mathbb{R}/\mathbb{Z}$ and approximate it with a finite sum of the form $\sum_{|n|\le N}\widehat{f}(n)e_n$. Notice that as x ranges in $\mathbb{R}/\mathbb{Z}$ once, $e_n(x)$ goes around the circle n times. The real (or the imaginary) part of $e_n(x)$ oscillates with frequency n. In other words, by approximating based on the Fourier inverse of a smooth function, we ignore *high frequency* terms. This idea has been heavily used in *data compression* algorithms. It is not surprising that such algorithms started with *audio data compression* as we can only hear sounds at certain frequencies. Here we employ the same philosophy to study random-walks on (certain) finite groups.

Notice that if f is a *more smooth* function on $\mathbb{R}/\mathbb{Z}$, then it should have less correlation with high frequency oscillations. This means the *tail* of the Fourier expansion should be *quite small*. The extreme case is when f is *constant*; in this case, only the first term is non-zero. For a non-abelian finite group G, we can use the degree of a representation instead of the frequency in order to measure its *complexity*. We take a threshold D, and collect the *high frequency* terms in the Plancherel formula of the L^2-norm of f and call it $H(f;D)$; that means

$$H(f;D) := \sum_{\pi\in\widehat{G},\deg\pi\ge D} \deg\pi\ \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2.$$

The smaller $H(f;D)$ (for a smaller threshold D) the *smoother* the function f.

The other general concept is that the convolution of two functions is *smoother* than the original functions. It is worth pointing out that the same principle is used in softwares that process photos in order to make the pictures more smooth: we essentially consider the convolution of the original function with another function which is called a *filter*. In order to find out, how much more $f*g$ is *smooth* compared to f and g, we look at $H(f*g;D)$.

Theorem 32 (Gowers's mixing theorem: version 1)**.** *Suppose G is a finite group and D is a positive number. Then for every $f,g\in L^2(G)$,*

$$H(f*g;D) \le \frac{|G|^2}{D} H(f;D)H(g;D).$$

Proof. We have $H(f*g;D)=\sum_{\pi\in\widehat{G},\deg\pi\ge D}\deg\pi\ \|\widehat{f*g}(\pi)\|_{\mathrm{HS}}^2$. And so by Proposition 30, we obtain

$$H(f*g;D) = |G|^2 \sum_{\pi\in\widehat{G},\deg\pi\ge D} \deg\pi\ \|\widehat{g}(\pi)\widehat{f}(\pi)\|_{\mathrm{HS}}^2.$$

Since $\|AB\|_{\mathrm{HS}}\le\|A\|_{\mathrm{HS}}\|B\|_{\mathrm{HS}}$, we deduce that

$$H(f*g;D) \le |G|^2 \sum_{\pi\in\widehat{G},\deg\pi\ge D} \deg\pi\ \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2\|\widehat{g}(\pi)\|_{\mathrm{HS}}^2. \tag{3.22}$$

Notice that the right hand side of the inequality given in (3.22) is at most

$$\frac{|G|^2}{D}\underbrace{\left(\sum_{\pi\in\widehat{G},\deg\pi\geq D}\deg\pi\ \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2\right)}_{H(f;D)}\underbrace{\left(\sum_{\pi\in\widehat{G},\deg\pi\geq D}\deg\pi\ \|\widehat{g}(\pi)\|_{\mathrm{HS}}^2\right)}_{H(g;D)},$$

which finishes the proof. □

It is worth mentioning that the factor $|G|^2$ is in some sense an unnatural factor which appears only because we defined the convolution of two functions using the counting measure instead of the *probability* counting measure; that means if we define $f*g(x)$ as $\int_G f(y)g(y^{-1}x)dy$ instead of our convention $f*g(x) := \sum_{y\in G} f(y)g(y^{-1}x)$, the factor $|G|^2$ disappears.

Theorem 32 is particularly strong when D is chosen to be

$$D_G := \min\{\deg\pi |\ \pi\in\widehat{G}\setminus\{1\}\}, \tag{3.23}$$

where 1 denotes the trivial representation of G, and D_G is *large*.

Theorem 33 (Gowers's mixing theorem: version 2)**.** *Suppose G is a finite group and D_G is as in (3.23). Then for every $f\in L^2(G)^\circ$ and $g\in L^2(G)$, we have*

$$\|f*g\|_2 \leq \sqrt{\frac{|G|}{D_G}}\|f\|_2\|g\|_2.$$

Proof. Notice that $L^2(G)^\circ$ is an ideal of $(L^1(G),+,*)$; that means if $f\in L^2(G)^\circ$ and $g\in L^2(G)$, then $f*g\in L^2(G)^\circ$. Next, we mention that for a function $h\in L^2(G)^\circ$,

$$\widehat{h}(1) = \int_G h(x)dx = 0;$$

and so by the Plancherel theorem

$$\|h\|_2^2 = |G|\ H(h;D_G).$$

Hence by Theorem 32, we obtain

$$\begin{aligned}\|f*g\|_2^2 =& |G|H(f*g;D_G) \leq \frac{|G|^3}{D_G}H(f;D_G)H(g;D_G)\\ \leq& \frac{|G|}{D_G}\|f\|_2^2\|g\|_2^2,\end{aligned}$$

which finishes the proof. □

Similar to the case of the equidistribution of irrational rotations, we are interested in a *pointwise* estimate instead of an L^2-estimate.

Lemma 34. *Suppose G is a finite group, and $f, g \in L^2(G)$. Then for every $x \in G$, we have*

$$f * g(x) = \langle \lambda_0(x)\check{g}, f\rangle,$$

where $\check{g}(x) := \overline{g(x^{-1})}$, and

$$\|f * g\|_\infty \leq \|f\|_2 \|g\|_2.$$

Proof. For every $x \in G$,

$$f * g(x) = \sum_{y \in G} f(y) g(y^{-1}x) = \sum_{y \in G} f(y)\overline{\lambda_0(x)(\check{g})(y)} = \langle \lambda_0(x)\check{g}, f\rangle.$$

This together with the Cauchy-Schwarz inequality implies that

$$|f * g(x)| = |\langle \lambda_0(x)\check{g}, f\rangle| \leq \|\lambda_0(x)\check{g}\|_2 \|f\|_2 = \|f\|_2 \|g\|_2,$$

which finishes the proof. □

Using induction, Theorem 33, and Lemma 34, we obtain the following result.

Theorem 35 (Gowers's mixing theorem: version 3)**.** *Suppose G is a finite group and $f \in L^2(G)^\circ$, $g_1, \ldots, g_n \in L^2(G)$. Then for $n \geq 2$*

$$\|f * g_1 * \cdots * g_{n-1}\|_2 \leq \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-1} \|f\|_2 \prod_{i=1}^{n-1} \|g_i\|_2, \tag{3.24}$$

$$\|f * g_1 * \cdots * g_n\|_\infty \leq \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-1} \|f\|_2 \prod_{i=1}^{n-1} \|g_i\|_2, \tag{3.25}$$

and if $\mu_1, \ldots, \mu_n$ are probability measures on G, then

$$\|\mu_1 * \cdots * \mu_n - \mathscr{P}_G\|_\infty \leq \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-2} \prod_{i=1}^{n} \|\mu_i\|_2, \tag{3.26}$$

where $\mathscr{P}_G$ is the probability counting measure on G.

Proof. We proceed by induction on n to prove (3.24). The base case of $n = 2$ follows from Theorem 33. Let $h := f * g_1 * \cdots * g_{n-2}$. Then $h \in L^2(G)^\circ$, and so by Theorem 33, we deduce that

$$\|h * f_{n-1}\|_2 \leq \sqrt{\frac{|G|}{D_G}} \|h\|_2 \|f_{n-1}\|_2. \tag{3.27}$$

By the induction hypothesis, we have

$$\|h\|_2 \leq \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-2} \|f\|_2 \prod_{i=1}^{n-2} \|f_i\|_2. \tag{3.28}$$

By (3.27) and (3.28), (3.24) follows.

The inequality (3.25) follows from (3.24) and Lemma 34.

To prove the last inequality, it is enough to observe that $\mu_1 - \mathscr{P}_G \in L^2(G)^\circ$ and $\mu_i * \mathscr{P}_G = \mathscr{P}_G$. These equalities imply the following

$$\mu_1 * \cdots * \mu_n - \mathscr{P}_G = (\mu_1 - \mathscr{P}_G) * \mu_2 * \cdots * \mu_n.$$

Hence by (3.25), we obtain

$$\|\mu_1 * \cdots * \mu_n - \mathscr{P}_G\|_\infty \le \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-2} \prod_{i=1}^n \|\mu_i\|_2.$$

□

The following product theorem is an important corollary of Theorem 35.

Theorem 36. *Suppose G is a finite group and $D_G := \min\{\deg \pi | \ \pi \in \widehat{G} \setminus \{1\}\}$. Suppose $A_1, \ldots, A_n$ are subsets of G such that*

$$\prod_{i=1}^n |A_i| > \frac{|G|^n}{D_G}.$$

Then $A_1 \cdots A_n = G$.

Proof. Let $\mu_i := \mathscr{P}_{A_i}$ be the probability counting measure on A_i. Then by Theorem 35, we obtain that for every $x \in G$

$$\begin{aligned}(\mathscr{P}_{A_1} * \cdots * \mathscr{P}_{A_n})(x) \ge& \mathscr{P}_G(x) - \|\mathscr{P}_{A_1} * \cdots * \mathscr{P}_{A_n} - \mathscr{P}_G\|_\infty \\ \ge& \frac{1}{|G|} - \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-2} \prod_{i=1}^n \|\mathscr{P}_{A_i}\|_2 \\ =& \frac{1}{|G|} - \left(\sqrt{\frac{|G|}{D_G}}\right)^{n-2} \prod_{i=1}^n \frac{1}{\sqrt{|A_i|}} \\ =& \frac{1}{|G|} - \sqrt{\frac{|G|^{n-2}}{D_G \prod_{i=1}^n |A_i|}} > 0.\end{aligned}$$

This means every $x \in G$ can be written as a product of elements of A_i's. This finishes the proof. □

The mixing theorems and the product theorems are effective when D_G is *large*. We say G is *c-quasi-random* if

$$D_G \ge |G|^c;$$

this means $\deg \pi \ge |G|^c$ for every $\pi \in \widehat{G} \setminus \{1\}$. We finish this section by proving that $\mathrm{SL}_n(\mathbb{F}_p)$'s are c_n-quasi-random for some c_n which only depends on n.

Theorem 37. *Suppose p is a prime number, and $G := \mathrm{SL}_2(\mathbb{F}_p)$. Then $D_G \geq \frac{p-1}{2}$, and G is $1/4$-quasi-random if $p \geq 7$.*

Proof. The usual technique of studying representations of a linear group is starting with its restriction to the *Borel* subgroup. In the case of $\mathrm{SL}_2(\mathbb{F}_p)$, this means upper triangular matrices. Notice that since $\mathrm{SL}_2(\mathbb{F}_p)$ is generated by $a := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $b := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and a and b are conjugate of each other, neither a nor b can be in the kernel of a non-trivial representation π. Hence $\pi(a)$ is an element of order p in $\mathrm{U}_d(\mathbb{C})$ where $d := \deg \pi$. Therefore after conjugation, if needed, we can and will assume that $\pi(a)$ is a diagonal matrix $\mathrm{diag}(\zeta_1, \ldots, \zeta_d)$ and ζ_i's are p-th roots of unity and $\zeta_1 \neq 1$.

So far we understood, the restriction of π to the *unipotent radical* of the Borel subgroup. Next, we investigate the diagonal matrices in $\mathrm{SL}_2(\mathbb{F}_p)$, and the key point is that the diagonal matrices normalize the unipotent radical. For every $\alpha \in \mathbb{F}_p^\times$, we have

$$\pi \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \pi \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \pi \begin{pmatrix} \alpha^{-1} & 0 \\ 0 & \alpha \end{pmatrix} = \pi \begin{pmatrix} 1 & \alpha^2 \\ 0 & 1 \end{pmatrix} = \pi(a)^{\alpha^2}.$$

Hence $\pi(a)^{\alpha^2}$ is a conjugate of $\pi(a)$. Hence $\zeta_1^{\alpha^2}$ is an eigenvalue of $\pi(a)$ for every $\alpha \in \mathbb{F}_p^\times$. Since the multiplicative order of ζ_1 is p and there are $\frac{p-1}{2}$ perfect residues in $\mathbb{F}_p^\times$, we deduce that $\pi(a)$ has at least $\frac{p-1}{2}$ distinct eigenvalues. Hence $d \geq \frac{p-1}{2}$, which implies that $D_G \geq \frac{p-1}{2}$.

The last part follows from the fact that $|\mathrm{SL}_2(\mathbb{F}_p)| = p(p-1)(p+1)$ and

$$\frac{p-1}{2} \geq (p(p-1)(p+1))^{1/4}$$

for $p \geq 7$. $\square$

3.4 Exercises

1. (Uniform convergence of Fourier series) Suppose $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is a smooth function. Prove the uniform convergence of the Fourier series of f; that means for every $\varepsilon > 0$ there is N_ε such that for every $x \in \mathbb{R}/\mathbb{Z}$, we have

$$|f(x) - \sum_{|n| \leq N_\varepsilon} \hat{f}(n) e_n(x)| \leq \varepsilon$$

(Hint. Use the same scheme of argument as in the proof of Theorem 17. For $\varepsilon > 0$, choose M_ε such that

$$\|f - \sum_{|n| \leq M_\varepsilon} \hat{f}(n) e_n\|_2 \leq \varepsilon.$$

Let $\{\psi_\varepsilon\}$ be a family of smooth test functions approximating the Dirac mass at x_0. Choose δ small enough so that

$$|e_n(x) - e_n(x_0)| \leq \frac{\varepsilon}{M_\varepsilon \max\{|\hat{f}(n)| \mid |n| \leq M_\varepsilon\}}$$

for every x in the δ-neighborhood of x_0 and $|n| \leq M_\varepsilon$. Then show that for small enough δ (depending on ε and f), we have

$$|f(x_0) - \langle f, \psi_\delta\rangle| \leq \varepsilon, \quad |\langle f, \psi_\delta\rangle - \sum_{|n|\leq M_\varepsilon} \hat{f}(n)\langle e_n, \psi_\delta\rangle| \leq \varepsilon,$$

and

$$\sum_{|n|\leq M_\varepsilon} |\hat{f}(n)||\langle e_n, \psi_\delta\rangle - e_n(x_0)| \leq \varepsilon.$$

Finish the proof.)

2. (Weyl's equidistribution criterion) A sequence $\{a_n\}_{n=1}^\infty$ of points in $\mathbb{R}/\mathbb{Z}$ is equidistributed if and only if for every non-zero integer m, we have the following *exponential sum cancellation*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^N e_m(a_n)}{N} = 0. \tag{3.29}$$

(Hint. Follow the same scheme of argument as in the proof of Theorem 17. For every smooth function f, let

$$T_N(f)(t) := \frac{\sum_{n=1}^N f(t+a_n)}{N}.$$

Notice that (3.29) is equivalent to saying $\lim_{N\to\infty} T_N(e_m)(0) = 0$ and equidistribution is equivalent to saying $\lim_{N\to\infty} T_N(f)(0) = \langle f, \mathbb{1}\rangle$. Use the uniform convergence of the Fourier series to show the existence of N_ε such that for every x,

$$|f(x) - \sum_{|m|\leq N_\varepsilon} \hat{f}(m)e_m(x)| \leq \varepsilon.$$

Deduce that

$$|f(x+a_n) - \sum_{|m|\leq N_\varepsilon} e_m(a_n)\hat{f}(m)e_m(x)| \leq \varepsilon,$$

and so

$$\left| T_N(f)(x) - \sum_{|m|\leq N_\varepsilon} \left(\frac{\sum_{|n|\leq N} e_m(a_n)}{N}\right)\hat{f}(m)e_m(x)\right| \leq \varepsilon.$$

Choosing N large enough, deduce that

$$|T_N(f) - \hat{f}(0)| \leq 2\varepsilon,$$

and complete the proof.)

3. (The van der Corput trick) Suppose $\{a_n\}_{n=1}^{\infty}$ is a sequence of points in $\mathbb{R}/\mathbb{Z}$. Suppose for every positive integer h, the sequence $\{a_{n+h}-a_n\}_{n=1}^{\infty}$ is equidistributed. Prove that $\{a_n\}_{n=1}^{\infty}$ is equidistributed.

   (Hint. By Weyl's equidistribution criterion, it is enough to prove that

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} e_m(a_n)}{N} = 0$$

   for every non-zero integer m. We pick an *intermediate range* H, and till that range, we use the *trivial* bound. To go beyond the intermediate range, we use the *exponential sum cancellation* given by the assumption.

   For every $1 \le h \le H$, we have

$$\left| \frac{\sum_{n=1}^{N} e_m(a_n)}{N} - \frac{\sum_{n=1}^{N} e_m(a_{n+h})}{N} \right| = O\left(\frac{H}{N}\right).$$

   Averaging over $1 \le h \le H$, deduce that

$$\left| \frac{\sum_{n=1}^{N} e_m(a_n)}{N} \right| = \left| \frac{1}{N} \sum_{n=1}^{N} \left(\frac{1}{H} \sum_{h=1}^{H} e_m(a_{n+h}) \right) \right| + O\left(\frac{H}{N}\right).$$

   Using the Cauchy-Schwarz inequality, obtain

$$\left| \frac{\sum_{n=1}^{N} e_m(a_n)}{N} \right| \le \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left| \frac{1}{H} \sum_{h=1}^{H} e_m(a_{n+h}) \right|^2} + O\left(\frac{H}{N}\right).$$

   Notice that

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^{N} \left| \frac{1}{H} \sum_{h=1}^{H} e_m(a_{n+h}) \right|^2 &= \frac{1}{H^2} \sum_{1\le h_1, h_2 \le H} \left(\frac{1}{N} \sum_{n=1}^{N} e_m(a_{n+h_1} - a_{n+h_2}) \right) \\ &= \frac{1}{H^2} \left(1 + \sum_{1 \le h_1 \ne h_2 \le H} \left(\frac{1}{N} \sum_{n=1}^{N} e_m(a_{n+h_1} - a_{n+h_2}) \right) \right). \end{aligned}$$

   For a fixed H, let N go to infinity and use the assumption, and deduce that

$$\left| \frac{\sum_{n=1}^{N} e_m(a_n)}{N} \right| \le \frac{1}{H}$$

   for every H. Finish the proof.)

4. (Weyl's theorem) Suppose $p(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{R}[x]$, $n > 0$, and there exists $i_0 > 0$ such that $a_{i_0} \notin \mathbb{Q}$. Prove that

$$\{p(n) + \mathbb{Z}\}_{n=1}^{\infty}$$

   is equidistributed.

   (Hint. Use the van der Corput trick and proceed by induction on the degree of p.)

5. (Space of inner products) Let P_n^+ be the set of n-by-n positive definite Hermitian matrices. For $x \in \mathrm{GL}_n(\mathbb{C})$ and $a \in P_n^+$, let $x \cdot a := x^* a x$.

   a) Prove that P_n^+ is a convex set.

   b) Prove that $\cdot$ defines an affine group action.

   c) Prove that $\mathrm{GL}_n(\mathbb{C})$ acts transitively on P_n^+.

   d) Prove that if G is a compact subgroup of $\mathrm{GL}_n(\mathbb{C})$, then there is $x \in \mathrm{GL}_n(\mathbb{C})$ such that $G \subseteq x\,\mathrm{U}_n(\mathbb{C}) x^{-1}$.

   e) Prove that the set of fixed points of $\mathrm{U}_n(\mathbb{C})$ under the action induced by $\cdot$ is $\{cI \mid c \in \mathbb{C} \setminus \{0\}\}$.

   f) Prove that $\mathrm{U}_n(\mathbb{C})$ is a maximal compact subgroup of $\mathrm{GL}_n(\mathbb{C})$ and every maximal compact subgroup of $\mathrm{GL}_n(\mathbb{C})$ is a conjugate of $\mathrm{U}_n(\mathbb{C})$.

6. (Weighted center of mass) Suppose $X \subseteq \mathbb{C}^n$ is a convex set and G acts on X by affine transformations, then for every $x_1, \ldots, x_n \in X$, non-negative numbers $p_1, \ldots, p_n$ that add up to 1, and $g \in G$, we have that

$$\sum_{i=1}^{m} p_i x_i \in X \quad \text{and} \quad g \cdot (\sum_{i=1}^{n} p_i x_i) = \sum_{i=1}^{n} p_i \; g \cdot x_i,$$

   (Hint. Use induction on n.)

7. Suppose G is a finite group, $\pi, \pi_1, \ldots, \pi_m$ are finite dimensional unitary representations of G. Suppose $V_\pi = \sum_{j=1}^m V_{\pi_j}$. Prove that

$$H_\pi = \sum_{j=1}^{m} H_{\pi_j},$$

   where H_π and H_{π_j} are spaces of matrix coefficients.

8. (Space of matrix coefficients: $G \times G$-representation and Wedderburn) Suppose G is a finite group and π is a degree n unitary irreducible representation of G. Then the following statements hold.

   a) The space H_π of matrix coefficients of π is invariant under the left-right translations action, and $H_\pi \simeq \mathrm{M}_n(\mathbb{C})$ as a $\mathbb{C}(G \times G)$-module where for every $(x_1, x_2) \in G$ and $a \in \mathrm{M}_n(\mathbb{C})$,

$$(x_1, x_2) \cdot a := \pi(x_2)\; a \; \pi(x_1)^{-1}.$$

   b) Let $\mathbb{C}[\pi(G)] \subseteq \mathrm{M}_n(\mathbb{C})$ be the $\mathbb{C}$-span of $\pi(G)$. Prove that

$$\mathbb{C}[\pi(G)] = \mathrm{M}_n(\mathbb{C}).$$

   c) Prove that H_π is an irreducible $G \times G$-representation.

d) Let $\widehat{G} := \{\pi_1, \ldots, \pi_m\}$ be a set of unitary irreducible representations of G which are pairwise non-equivalent and every unitary irreducible representation is equivalent to one of the elements of $\widehat{G}$. Prove that

$$L^2(G) = H_{\pi_1} \oplus \cdots \oplus H_{\pi_m}.$$

e) The space H_π is an ideal of the group ring $\mathbb{C}G$, after identifying $\mathbb{C}G$ with $(L^1(G), +, *)$.

f) Suppose $\widehat{G}$ is as in part (8d). Prove that $\mathbb{C}G \simeq \prod_{\pi \in \widehat{G}} \mathrm{M}_{\deg \pi}(\mathbb{C})$ as a ring.

(Hint. For part (a), notice that

$$((x_1, x_2) \cdot e_{i,j} \circ \pi)(y) = e_{i,j}(\pi(x_1^{-1} y x_2));$$

and so

$$(x_1, x_2) \cdot (e_{i,j} \circ \pi) = \sum_{k,l} e_{i,k}(\pi(x_1)^{-1}) e_{l,j}(\pi(x_2))\ e_{k,l} \circ \pi.$$

Let $T : H_\pi \to \mathrm{M}_n(\mathbb{C})$ be a linear map such that $T(e_{i,j} \circ \pi) := E_{j,i}$ where $E_{j,i}$ is the n-by-n matrix that has 1 in its (j, i)-entry and 0 everywhere else. Then

$$\begin{aligned} T\Big((x_1, x_2) \cdot (\sum_{i,j} a_{j,i}\ e_{i,j} \circ \pi)\Big) &= \sum_{i,j} a_{j,i} \sum_{k,l} e_{i,k}(\pi(x_1)^{-1}) e_{l,j}(\pi(x_2))\ E_{l,k} \\ &= \sum_{k,l} \Big(\sum_{i,j} e_{l,j}(\pi(x_2)) a_{j,i} e_{i,k}(\pi(x_1)^{-1}) \Big) E_{l,k} \\ &= \sum_{k,l} e_{l,k}(\pi(x_2) a \pi(x_1)^{-1}) E_{l,k} \\ &= \pi(x_2) a \pi(x_1)^{-1}. \end{aligned}$$

For part (b), suppose to the contrary that $\mathbb{C}[\pi(G)]$ is a proper subspace and deduce that there exists $a := (a_{ij}) \in \mathrm{M}_n(\mathbb{C}) \setminus \{0\}$ such that $\mathrm{tr}(a^* \mathbb{C}[\pi(G)]) = 0$. Let $f_a := \sum_{i,j} \overline{a}_{ji} e_{i,j} \circ \pi \in H_\pi$. Show that $f_a(x) = 0$ for every $x \in G$. Deduce that $a = 0$ which is a contradiction.

For part (d), use Corollary 26 and Corollary 27.

For part (e), use the following equation

$$\begin{aligned} (g * f_{v,w})(x) &= \sum_y g(y) \langle \pi(y^{-1} x) v, w \rangle \\ &= \Big\langle \pi(x) v, \sum_y g(y) \pi(y) w \Big\rangle = f_{v, \sum_y g(y)\pi(y)(w)}(x).) \end{aligned}$$

9. Suppose G is a finite group. Then for every $f, g \in L^2(G)$, $\pi \in \widehat{G}$, and $x \in G$, we have

$$\widehat{\lambda_0(x) f}(\pi) = \widehat{f}(\pi) \pi(x)^*, \text{ and } \widehat{\rho_0(x) f}(\pi) = \pi(x) \widehat{f}(\pi).$$

10. (Finite simple groups of Lie type are quasi-random) Suppose p is a prime number and $q = p^n$ for some positive integer n.

    a) Prove that $D_{\mathrm{SL}_2(\mathbb{F}_q)} \geq \frac{q-1}{2}$.
    b) Suppose H is a subgroup of G and the normal closure of H is G; that means the smallest normal subgroup of G which contains H is G. Prove that $D_G \geq D_H$.
    c) Prove that $D_{\mathrm{SL}_n(\mathbb{F}_q)} \geq \frac{q-1}{2}$, and deduce that $\mathrm{SL}_n(\mathbb{F}_q)$ is c_n-quasi-random for a positive number c_n which only depends on n.

    (Hint. Similar to the case of $q = p$, start with the restriction of π to the subgroup of unipotent upper triangular matrices. Argue why you can assume that

    $$\pi \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} = \mathrm{diag}(\phi_1(a), \ldots, \phi_d(a))$$

    where $\phi_j : \mathbb{F}_q \to \{\zeta \in \mathbb{C} |\ \zeta^p = 1\}$ is a group homomorphism for every j. Using conjugation by diagonal matrices in $\mathrm{SL}_2(\mathbb{F}_q)$, argue that for every, $b \in \mathbb{F}_q^\times$ and $a \in \mathbb{F}_p$,

    $$(\phi_1(b^2 a), \ldots, \phi_d(b^2 a))$$

    is a permutation of

    $$(\phi_1(a), \ldots, \phi_d(a)).$$

    This implies that

    $$\{b^{-2} \cdot \phi_1 |\ b \in \mathbb{F}_q^\times\} \subseteq \{\phi_1, \ldots, \phi_d\}.$$

    as functions on $\mathbb{F}_q$. Argue that if ϕ_1 is not trivial, then $b^{-2} \cdot \phi_1 = \phi_1$ if and only if $b^{-2} = 1$. Deduce the first part.

    For the last part, show that $\mathrm{SL}_n(\mathbb{F}_q)$ is generated by conjugates of a copy of $\mathrm{SL}_2(\mathbb{F}_q)$. Using this idea one can show that apart from Suzuki groups every finite simple group of Lie type of rank r is $c(r)$-quasi-random for some positive number $c(r)$ which only depends on r.

    This bound is not the optimal bound. The optimal bound is found by Landazuri and Seitz.)