MARKOV CHAINS ON FINITE GROUPS: CHARACTERS AND REPRESENTATIONS

SHANSHAN DING

1. Introduction

Let $(G, \circ)$ be a finite group. Unless otherwise specified, we will let $n = |G|$ and omit the group operation $\circ$ from our notation. Given a probability distribution μ on G, the transition probabilities $P(g, hg) := \mu(h)$ define a Markov chain. In words, the chain moves via left multiplication by a random element of G selected according to μ. The measure μ is called the *increment distribution* on G.

For a familiar example, consider the cyclic group $\mathbb{Z}_n = \{1, 2, \dots, n-1\}$, with μ assigning $1/2$ to the elements 1 and $-1 \equiv n-1 \bmod n$. This gives the simple random walk on the n-cycle.

The objective of these notes is to investigate the mixing properties of Markov chains on finite groups. To do so, we will develop the theory of harmonic analysis on finite groups. The basic idea is that Fourier transforms give us a particularly nice way to study convolution of probability measures. Since Fourier transforms relate functions on a domain with functions on the dual of the domain, we will need the appropriate notions of the group duals. When G is abelian, duality is defined via *characters*, and we generalize to the non-abelian case by studying *representations*.

Before bringing in the heavy machinery, let us get a few easy facts out of the way:

Proposition 1.1. *For a Markov chain P on G with increment distribution μ,*

(i) The stationary distribution of P is the uniform distribution U on G.

(ii) Let H be the subgroup of G generated by supp $\mu := \{g \in G : \mu(g) > 0\}$, then the irreducible subsets of P are precisely the right cosets of H in G, i.e. sets of the form Hg with $g \in G$. In particular, P is irreducible iff supp μ generates G.

(iii) Let e be the identity element of G. If $e \in$ supp μ, then P is aperiodic.

Proof. (i) For any $g \in G$,

$$\sum_h U(h)P(h,g) = \frac{1}{n}\sum_h P(h,g) = \frac{1}{n}\sum_h \mu(gh^{-1}) = \frac{1}{n} = U(g), \tag{1.1}$$

where the second to last equality uses the observation that the operation $h \to gh^{-1}$ re-indexes G.

(ii) A chain which starts at g_0 has a positive probability of arriving at each state of the form $g_n \dots g_1 g_0$, where n is arbitrary and each g_i is in supp μ. Thus it suffices to show that H coincides with the set $H' = \{g_l \dots g_1 : l \in \mathbb{N}, g_1, \dots, g_l \in \text{supp } \mu\}$. Since H and H' have the same generators, we just need to show that H' contains e and is closed under taking inverses. This follows from the finiteness of G: each $g \in G$ has finite order, so if $g \in$ supp μ, then $g^k = e$ for some $k \in \mathbb{N}$, and thus e and $g^{k-1} = g^{-1}$ are both in H'.

(iii) If $\mu(e) > 0$, then $P(g, g) > 0$ for all $g \in G$, so that all states have period 1. □

Note that as a result of (i), the distance to stationarity does not depend on the initial state: a chain which starts at g is simply a translation by g of a walking starting at e, and the uniform distribution is translation-invariant.

2. Markov chains on finite abelian groups

2.1. **Characters.** Let G be a finite abelian group, and let Γ be the multiplicative group of all complex numbers of modulus 1.

Definition 2.1. *A* character[1] χ *on* G *is a group homomorphism from* G *to* Γ. *That is,* $\chi(gh) = \chi(g)\chi(h)$ *for all* $g, h \in G$.

In particular, the map which sends every $g \in G$ to 1 is the *trivial character* χ_{triv} on G. It is straightforward to verify that the collection of all characters on G, $\hat{G}$, is a group under the operation of pointwise multiplication, with χ_{triv} as the identity.

Example. Consider the cyclic group $\mathbb{Z}_n$, with the group operation being addition mod n. It is clear that $\chi_j : g \to \exp(2\pi ijg/n)$ is a character for each $0 \leq j \leq n-1$, and that the map $j \to \chi_j$ is an injective homomorphism. For an arbitrary character χ on G, since $\chi(0) = 1$, $\chi(1)$ must be an n-th root of unity, i.e. $\chi(1) = \exp(2\pi ij/n)$ for some j. Thus $\chi = \chi_j$, and the map $j \to \chi_j$ is onto as well. This shows that in fact G and $\hat{G}$ are isomorphic when G is cyclic.

Example. Let G be the hypercube $\{0,1\}^l$, with the group operation being componentwise addition mod 2. Since every $g \in G$ has order 2, each character on G can only take values in $\{-1,1\}$. It is natural, then, to consider the homomorphisms

$$\chi_{\epsilon_1,\epsilon_2,\ldots,\epsilon_l} : (r_1, r_2, \ldots, r_l) \to \epsilon_1^{r_1}\epsilon_2^{r_2}\cdots\epsilon_l^{r_l}, \tag{2.1}$$

where $(\epsilon_1, \epsilon_2, \ldots, \epsilon_l) \in \{-1,1\}^l$ and $(r_1, r_2, \ldots, r_l) \in \{0,1\}^l$. Each $\chi_{\epsilon_1,\epsilon_2,\ldots,\epsilon_l}$ is clearly a character. That these give all the characters will follow from the next proposition.

Theorem 2.2. *Any finite abelian group,* G*, is isomorphic to its character group,* $\hat{G}$.

Proof. By the fundamental theorem of finite abelian groups, G is isomorphic to a direct sum of cyclic groups: $G = G_1 \oplus G_2 \oplus \cdots \oplus G_k$, where each G_i is of the form $\mathbb{Z}_{n_i}$. Recall that G_i is isomorphic to $\hat{G}_i$ for each i. It is easy to see that if $\chi_i \in \hat{G}_i$, then

$$(\chi_1, \chi_2, \ldots, \chi_k) \to ((g_1, g_2, \ldots, g_k) \to \chi_1(g_1)\chi_2(g_2)\cdots\chi_n(g_n)) \tag{2.2}$$

gives an isomorphism from $\hat{G}_1 \oplus \hat{G}_2 \oplus \cdots \oplus \hat{G}_n$ to $\hat{G}$. Hence $G \cong \hat{G}$. □

Proposition 2.3. *If* $\chi \neq \chi_{triv}$*, then* $\sum_g \chi(g) = 0$.

Proof. For any $\chi \neq \chi_{triv}$, there exists $g_0 \in G$ such that $\chi(g_0) \neq 1$. Observe that since $\{g_0 g : g \in G\} = G$,

$$\sum_g \chi(g) = \sum_g \chi(g_0 g) = \chi(g_0)\sum_g \chi(g). \tag{2.3}$$

As $\chi(g_0) \neq 1$, we must have that $\sum_g \chi(g) = 0$. □

[1] For an arbitrary group, a character is defined as the trace of a representation (see Section 3.4). The two definitions are equivalent when G is abelian.

For a group G with n elements, the set X_G of all complex-valued functions on G is an n-dimensional vector space over $\mathbb{C}$. Furthermore, we can turn X_G into a Hilbert space by defining the inner product

$$\langle f_1, f_2 \rangle_G = \frac{1}{n} \sum_g f_1(g) \overline{f_2(g)}. \tag{2.4}$$

Theorem 2.4. *$\hat{G}$ is an orthonormal basis of X_G.*

Proof. For $\chi_1, \chi_2 \in \hat{G}$, if $\chi_1 = \chi_2$, then $\chi_1 \overline{\chi_2} = \chi_1 \chi_2^{-1} = \chi_{triv}$, otherwise $\chi_1 \overline{\chi_2} \neq \chi_{triv}$. Thus by Proposition 2.3,

$$\langle \chi_1, \chi_2 \rangle_G = \frac{1}{n} \sum_g (\chi_1 \overline{\chi_2})(g) = \begin{cases} 1 & \text{if } \chi_1 = \chi_2 \\ 0 & \text{if } \chi_1 \neq \chi_2. \end{cases}$$

This shows that $\hat{G}$ is an orthonormal set in X_G, so that the χ are all linearly independent. By Theorem 2.2, $\hat{G}$ has exactly n elements, which proves that $\hat{G}$ is in fact an orthonormal basis of X_G. □

Corollary 2.5. *(i) For any $f \in X_G$, $f = \sum_\chi \langle f, \chi \rangle_G \chi$. In words, f can be expressed as a unique linear combination of the characters.*

(ii) $\hat{G}$ separates the points of G, i.e. for distinct $g, h \in G$, there exists $\chi \in \hat{G}$ such that $\chi(g) \neq \chi(h)$.

Proof. (i) is a direct consequence of Theorem 2.4 and properties of inner products. By considering an $f \in X_G$ with $f(g) \neq f(h)$,(ii) follows from (i). □

2.2. Fourier transform of functions.

Definition 2.6. *For complex-valued function f on G, the* Fourier transform *$\hat{f}$ of f is the complex-valued funcion on $\hat{G}$ given by $\hat{f}(\chi) = \frac{1}{n} \sum_g f(g)\chi(g)$.*

Note that the Fourier transform $f \to \hat{f}$ is a linear map from X_G to $X_{\hat{G}}$. By Theorem 2.4, the Fourier transform of a character χ is the indicator function on $\{\bar{\chi}\}$, i.e. $\hat{\chi}(\chi')$ is 1 if $\chi' = \bar{\chi}$ and 0 otherwise. Since there are n different $\bar{\chi}$, the range of X_G under the Fourier transform is n-dimensional, and as a surjective linear map between n-dimensional vector spaces, the Fourier transform is also injective.

Theorem 2.7 (Fourier inversion). *For any $f \in X_G$ and $g \in G$,*

$$f(g) = \sum_\chi \hat{f}(\chi) \overline{\chi(g)} \tag{2.5}$$

Proof. Since $f \to \hat{f}$ is a linear map, it suffices to check (2.5) on the basis $\hat{G}$ of X_G. By the preceding discussion, for $f = \chi_0$, $\hat{\chi_0}(\chi)$ is 1 if $\chi = \overline{\chi_0}$ and 0 otherwise, so that

$$\sum_\chi \hat{f}(\chi) \overline{\chi(g)} = \overline{\overline{\chi_0}(g)} = \chi_0(g), \tag{2.6}$$

as desired. □

Not only is the Fourier transform a linear bijection from X_G to $X_{\hat{G}}$, is is an isometry under a suitably normalized L^2-norm:

Theorem 2.8 (Plancherel formula). *Let $f_1, f_2 \in X_G$, then*

$$\langle f_1, f_2\rangle_G = \sum_\chi \hat{f}_1(\chi)\overline{\hat{f}_2(\chi)}. \tag{2.7}$$

In particular,

$$\frac{1}{n}\sum_g |f(g)|^2 = \langle f, f\rangle_G = \sum_\chi |\hat{f}(\chi)|^2 \tag{2.8}$$

for any $f \in X_G$.

Proof. First, suppose $f_1 = \chi_1$ and $f_2 = \chi_2$ are characters. By Theorem 2.4, the LHS of (2.7) is 1 when $\chi_1 = \chi_2$ and 0 otherwise. On the RHS, $\hat{\chi}_1$ and $\hat{\chi}_2$ are indicator functions on $\{\bar{\chi}\}$, so at least one is 0 at each χ if $\chi_1 \neq \chi_2$, in which case the sum is 0. If $\chi_1 = \chi_2$, then one of the summands is 1 and the others are 0, so the RHS is 1.

For the general case, recall that inner products are linear in the first argument and antilinear in the second, so (2.7) follows if we write f_1 and f_2 as linear combinations of characters. □

2.3. Fourier transform of measures.

Definition 2.9. *For a probability measure μ on G, the* Fourier transform *of μ is the integral of χ with respect to μ, i.e. $\hat{\mu}(\chi) = \sum_g \mu(\{g\})\chi(g)$.*

Remark. Observe that $\hat{\mu}$ is not quite the Fourier transform of the function $g \to \mu(\{g\})$, as there is a factor of $\frac{1}{n}$ missing on the RHS. However, if we define $f_\mu(g)$ as the *density of with respect to the uniform distribution*, namely $f_\mu(g) = n\mu(\{g\})$, then the Fourier transform of the measure μ is indeed the Fourier transform of this density. The reason for this scaling is to allow the Fourier transform of any probability measure to equal 1 on χ_{triv}. For convenience, we will suppress the set notation and use $\mu(g)$ to mean $\mu(\{g\})$.

Example. For a Bernoulli trial, $G = \mathbb{Z}_2$, with $\mu(1) = p$ and $\mu(0) = 1 - p$. The character group $\hat{G}$ consists of χ_{triv} and the character $\chi(g) = \exp(\pi i g) = (-1)^g$, and $\hat{\mu}$ maps χ_{triv} to $1(p) + 1(1-p) = 1$ and χ to $-1(p) + 1(1-p) = 1 - 2p$.

Example. More generally, for a distribution μ on $\mathbb{Z}_n$, the characters are of the form $\chi_j : g \to \exp(2\pi i j g/n)$, so that

$$\hat{\mu}(\chi_j) = \sum_{g=0}^{n-1} \mu(g)\exp(2\pi i j g/n). \tag{2.9}$$

Here the χ_j are mapped to convex combinations of n-th roots of unity. Furthermore, suppose that μ is uniform, so that $\hat{\mu}(\chi_j) = \frac{1}{n}\sum_{g=0}^{n-1}\exp(2\pi i j g/n)$. Letting $\xi = \exp(2\pi i j/n)$, we can rewrite the sum as $\frac{1}{n}(1 + \xi + \cdots + \xi^{n-1})$, and when $j \neq 0$,

$$\hat{\mu}(\chi_j) = \frac{1}{n}(1 + \xi + \cdots + \xi^{n-1}) = \frac{1}{n}\left(\frac{1-\xi^n}{1-\xi}\right) = 0. \tag{2.10}$$

Of course, when $j = 0$, χ_j is trivial, and $\hat{\mu}(\chi_{triv}) = 1$. This will turn out to be a special case of Proposition 2.10

Example. On the hypercube $\{0,1\}^l$, consider the increment distribution

$$\mu(r_1, r_2, \ldots, r_l) = \begin{cases} 1/(l+1) & \text{if } r_1 + r_2 + \ldots + r_l \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Recall that characters on the hypercube are of the form

$$\chi_{\epsilon_1,\epsilon_2,\ldots,\epsilon_l}(r_1, r_2, \ldots, r_l) = \epsilon_1^{r_1}\epsilon_2^{r_2}\cdots\epsilon_l^{r_l}, \tag{2.11}$$

where $(\epsilon_1, \epsilon_2, \ldots, \epsilon_l) \in \{-1,1\}^l$. Thus

$$\begin{aligned} \hat{\mu}(\chi_{\epsilon_1,\ldots,\epsilon_l}) &= \frac{1}{l+1}\sum_{r_1+\ldots+r_l\leq 1} \epsilon_1^{r_1}\cdots\epsilon_l^{r_1} \\ &= \frac{1}{l+1}(1+\epsilon_1+\cdots+\epsilon_l) = 1 - \frac{2\omega(\epsilon_1,\ldots,\epsilon_l)}{l+1}, \end{aligned} \tag{2.12}$$

where ω counts the number of -1 in $(\epsilon_1, \ldots, \epsilon_l)$.

Proposition 2.10. *Let μ, ν be probability measures on G, with U denoting the uniform distribution.*

(i) $\mu = U$ if and only if $\hat{\mu}$ is 1 on χ_{triv} and 0 on all other χ.

(ii) Taking $\frac{1}{2}\sum_g|\mu(g) - \nu(g)|$ as the definition of the total variation distance $\|\mu - \nu\|_{TV}$ between μ and ν, we have

$$\|\mu - \nu\|_{TV} \leq \frac{1}{2}\left(\sum_\chi |\hat{\mu}(\chi) - \hat{\nu}(\chi)|^2\right)^{1/2}. \tag{2.13}$$

In particular,

$$\|\mu - U\|_{TV} \leq \frac{1}{2}\left(\sum_{\chi\neq\chi_{triv}} |\hat{\mu}(\chi)|^2\right)^{1/2}. \tag{2.14}$$

(iii) $|\hat{\mu} - \hat{\nu}|_{max} := \max_\chi|\hat{\mu}(\chi) - \hat{\nu}(\chi)| \leq 2\|\mu - \nu\|_{TV}$.

Proof. (i) Since $\hat{U}(\chi) = \sum_g \chi(g)U(g) = \frac{1}{n}\sum_g \chi(g)$, by Proposition 2.3, $\hat{U}(\chi)$ is 1 if $\chi = \chi_{triv}$ and 0 otherwise. Since the Fourier transform is injective, only the uniform distribution can have this property.

(ii) Applying Cauchy-Schwarz to vectors $(1, \ldots, 1)$ and $(a_1, \ldots, a_n)$ gives the inequality $(\sum_{i=1}^n a_i)^2 \leq n\sum_{i=1}^n a_i^2$, so that

$$4\|\mu - \nu\|_{TV}^2 = \left(\sum_g |\mu(g) - \nu(g)|\right)^2 \leq n\sum_g |\mu(g) - \nu(g)|^2. \tag{2.15}$$

By the Plancherel formula (2.8) and linearity,

$$\frac{1}{n}\sum_g n^2|\mu(g) - \nu(g)|^2 = \sum_\chi |\hat{\mu}(\chi) - \hat{\nu}(\chi)|^2. \tag{2.16}$$

Putting the two together gives (2.13). (2.14) follows from (2.13) with the additional observations that the Fourier transform of any probability measure is 1 on χ_{triv} and that $\hat{U}(\chi) = 0$ whenever $\chi \neq \chi_{triv}$.

(iii) For any χ,

$$\begin{aligned}|\hat{\mu}(\chi)-\hat{\nu}(\chi)| &= |\sum_g \chi(g)(\mu(g)-\nu(g))| \\ &\leq \sum_g |\chi(g)(\mu(g)-\nu(g))| = \sum_g |\mu(g)-\nu(g)| = 2\|\mu-\nu\|_{TV},\end{aligned} \tag{2.17}$$

as desired. □

2.4. **Convolutions.** The previous proposition is helpful because it gives us bounds on the total variation distance between measures. (2.14) is especially useful, as it translates the question of "how close is μ to uniformity (the stationary distribution)" to the question of "how small is $\hat{\mu}$ on the nontrivial characters". In order to apply to our Markov chain on G, it remains to investigate the Fourier transforms of k-step transition measures.

Suppose, given the increment distribution μ on G, we want to know the probability of moving from g to hg in two steps. It is easy to see that this is equal to $\sum_{h_2h_1=h}\mu(h_1)\mu(h_2)$. Since, for every $h_1 \in G$, there exists a unique $h_2 \in G$ such that $h_2h_1 = h$, we can rewrite the sum as $\sum_{h_1}\mu(h_1)\mu(hh_1^{-1})$. This is the motivation for the following definition:

Definition 2.11. *For probability measures μ,ν on G, we define the* convolution *of μ and ν as the measure*

$$(\nu * \mu)(h) = \sum_{h_1}\mu(h_1)\nu(hh_1^{-1}). \tag{2.18}$$

In particular, we use the notation $\mu^{(2)}$ for $\mu*\mu$ and define the k-fold convolution of μ inductively as $\mu^{(k)} = \mu * \mu^{(k-1)}$.

It is straightforward to check that the convolution of probability measures is again a probability measure. Of course, the special case of $\mu^{(k)}$ is exactly the k-step transition distribution on G, which is indeed a probability measure. Now, to compute how fast a Markov chain with increment distribution μ converges to stationarity, it suffices to compute how fast $\mu^{(k)}$ tends to the uniform distribution. The Fourier transform is an especially useful tool, as it transforms convolutions into pointwise products:

Theorem 2.12. *For probability measures μ,ν on G, $\widehat{\nu*\mu} = \hat{\nu}\hat{\mu}$.*

Proof. For any χ,

$$\begin{aligned}\widehat{\nu*\mu}(\chi) &= \sum_g((\nu*\mu)(g))\chi(g) = \sum_g \chi(g)\sum_{g_1}\mu(g_1)\nu(gg_1^{-1}) \\ &= \sum_{g,g_1}\chi(gg_1^{-1}g_1)\mu(g_1)\nu(gg_1^{-1}) = \sum_{g_1,g_2}\chi(g_2)\chi(g_1)\mu(g_1)\nu(g_2) \\ &= \left(\sum_{g_2}\nu(g_2)\chi(g_2)\right)\left(\sum_{g_1}\mu(g_1)\chi(g_1)\right) = \hat{\nu}(\chi)\hat{\mu}(\chi).\end{aligned} \tag{2.19}$$

as was to be shown. □

Combining Theorem 2.12 with (2.14) gives the following:

Corollary 2.13. $\|\mu^{(k)} - U\|_{TV} \leq \frac{1}{2}(\sum_{\chi\neq\chi_{triv}}|\hat{\mu}(\chi)|^{2k})^{1/2}$.

Recall that $\hat{\mu}(\chi) = \sum_g \mu(g)\chi(g)$, so that each $\hat{\mu}(\chi)$ is a convex combination of points on the unit circle, and that as such, it lies in the unit disk. The further it is inside the boundary, the faster $|\hat{\mu}(\chi)|^k$ tends to 0. Intuitively, it makes sense geometrically that the closer μ already is to U, the small the moduli of the $\hat{\mu}(\chi)$ are, and the faster $\mu^{(k)}$ converges to U.

Corollary 2.14. *The sequence $\mu^{(k)}$ converges to U with respect to the total variation norm if and only if $|\hat{\mu}(\chi)| < 1$ for every nontrivial χ.*

Proof. This follows directly from Theorem 2.12 and (2.16). □

Example. For a distribution μ on $\mathbb{Z}_n$, each $\hat{\mu}(\chi_j)$ is a convex combination of the roots of unity $\exp(2\pi ijg/n)$, $g = 0, \dots, n-1$, which lies on the unit circle if and only if all weights are concentrated on the same number. Thus $|\hat{\mu}(\chi_j)| = 1$ if and only if supp μ is contained in a set of the form $\{a + lb : l = 0, 1, \dots, n-1\}$ with $\gcd(b, n) = n/j$, i.e. a nontrivial arithmetic progression in $\mathbb{Z}_n$.

2.5. **Application to mixing.** We end the section with some examples of applications of Corollary 2.13.

Example. Consider μ on $\mathbb{Z}_n$ with supp $\mu = \{0, 1\}$. What is the optimal choice of $p := \mu(1)$ for mixing? Intuitively, we don't want p to be close to either 0 or 1: the chain moves too slowly in the former case and behaves nearly deterministically in the latter. Formally, we want to minimize the moduli of

$$\hat{\mu}(\chi_j) = (1-p) + p\exp(2\pi ij/n), 1 \le j \le n-1 \tag{2.20}$$

which are points on the line segment between 1 and $\exp(2\pi ij/n)$. It is not difficult to see that the point closest to the origin on such a line segment is the midpoint, so that $\hat{\mu}(\chi_j)$ attains minimum modulus at $p = 1/2$ for each j. Thus $p = 1/2$ is indeed the optimal choice for mixing.

Example. Now consider μ on $\mathbb{Z}_n$ with $\mu(1) = \mu(n-1) = 1/2$. This is the simple random walk on the cycle. To avoid periodicity, we will assume that n is odd. Since

$$\hat{\mu}(\chi_j) = \frac{1}{2}(\exp(2\pi ij/n) + \exp(-2\pi ij/n)) = \cos(2\pi j/n), \tag{2.21}$$

we see that

$$\|\mu^{(k)} - U\|_{TV}^2 \le \frac{1}{4}\sum_{j=1}^{n-1}(\cos(2\pi j/n))^{2k}. \tag{2.22}$$

For this bound to be useful, we need to put it into a more effective form. Recalling the cosine symmetries $\cos(-x) = \cos(x)$ and $\cos(\pi - x) = -\cos x$, we can rewrite

$$\frac{1}{4}\sum_{j=1}^{n-1}(\cos(2\pi j/n))^{2k} = \frac{1}{2}\sum_{j=1}^{(n-1)/2}(\cos(\pi j/n))^{2k}, \tag{2.23}$$

to which we apply the inequality $\cos x \leq e^{-x^2/2}$ $(0 \leq x \leq \pi/2)$ to derive that

$$\begin{aligned} \|\mu^{(k)} - U\|_{TV}^2 &\leq \frac{1}{2} \sum_{j=1}^{(n-1)/2} \exp(-\pi^2 j^2 k/n^2) \\ &= \frac{1}{2} \exp(-\pi^2 k/n^2) \sum_{j=1}^{n-1} \exp(-\pi^2(j^2-1)k/n^2) \\ &\leq \frac{1}{2} \exp(-\pi^2 k/n^2) \sum_{j=1}^{\infty} \exp(-\pi^2(j^2-1)k/n^2) \\ &\leq \frac{1}{2} \exp(-\pi^2 k/n^2) \sum_{j=1}^{\infty} \exp(-3(j-1)\pi^2 k/n^2) \\ &= \frac{1}{2} \exp(-\pi^2 k/n^2) \sum_{j=0}^{\infty} \exp(-3j\pi^2 k/n^2) \\ &= \frac{\exp(-\pi^2 k/n^2)}{2(1-\exp(-3\pi^2 k/n^2))}, \end{aligned} \tag{2.24}$$

where the last inequality comes from the fact that $j^2 - 1 \geq 3(j-1)$ for all $j \geq 1$, and the final step is just the summation formula for infinite geometric series. When $k \geq n^2$, we have that $2(1-\exp(-3\pi^2 k/n^2)) \geq 1$, and hence

$$\|\mu^{(k)} - U\|_{TV} \leq \exp(-\pi^2 k/2n^2). \tag{2.25}$$

This assures that the chain will be close to stationarity after $O(n^2)$ steps.

Example. Recall the hypercube example from Section 2.3, where

$$\hat{\mu}(\chi_{\epsilon_1,\ldots,\epsilon_l}) = 1 - \frac{2\omega(\epsilon_1,\ldots,\epsilon_l)}{l+1}. \tag{2.26}$$

Here χ_{triv} corresponds to the $(\epsilon_1,\ldots,\epsilon_l)$ of all 1's, i.e. $\omega(\epsilon_1,\ldots,\epsilon_l) = 0$. For $1 \leq s \leq l$, there are exactly $\binom{l}{s}$ vectors $(\epsilon_1,\ldots,\epsilon_l)$ such that $\omega(\epsilon_1,\ldots,\epsilon_l) = s$. Thus Corollary 2.13 gives us that

$$\|\mu^{(k)} - U\|_{TV} \leq \frac{1}{2} \left(\sum_{s=1}^{l} \binom{l}{s} \left(1 - \frac{2s}{l+1}\right)^{2k} \right)^{1/2}. \tag{2.27}$$

Note that $(1 - \frac{2s}{l+1})^{2k}$ is largest when $s = 1$ or l, in which cases $(1 - \frac{2s}{l+1})^{2k} = (\frac{l-1}{l+1})^{2k}$. It follows that

$$\|\mu^{(k)} - U\|_{TV} \leq C_l \left(\frac{l-1}{l+1}\right)^k, \tag{2.28}$$

where C_l is a constant for each fixed l (and increases as l increases).

3. Markov chains on arbitrary finite groups

3.1. **Representations and Schur's lemma.** Now suppose that G is a non-abelian group. For $g, h \in G$ with $gh \neq hg$, any homomorphism ϕ from G to an abelian group has to satisfy

$$\phi(gh) = \phi(g)\phi(h) = \phi(h)\phi(g) = \phi(hg), \tag{3.1}$$

which fails to preserve the non-abelian structure of G. Thus in order to generalize the results of the previous section to arbitrary finite groups, we need to generalize the notion of characters by replacing Γ with a suitable non-abelian group. Fortunately, there is a natural choice for this.

Recall that a d-by-d matrix $M = (a_{ij})$ is *unitary* if $MM^* = M^*M = I_d$, where $M^* = (\overline{a_{ji}})$ is the adjoint of M. Let $\mathcal{U}_d$ denote the set of d-by-d unitary matrices.

Definition 3.1. *A d-dimensional* (unitary) representation ρ *of* G *is a group homomorphism from* G *to* $\mathcal{U}_d$. *The 1-dimensional representation which sends every* $g \in G$ *to 1 is the* trivial representation ρ_{triv}.

Remark. When G is abelian, the collection of characters on G is exactly the collection of 1-dimensional representations of G.

Example. Consider the symmetric group S_l. Every $\tau \in S_l$ can be written as a product of transpositions, and $\rho : S_l \to \{\pm 1\}$ defined by

$$\rho(\tau) = \begin{cases} 1 & \text{if } \tau \text{ is the product of an even number of transpositions} \\ -1 & \text{if } \tau \text{ is the product of an odd number of transpositions} \end{cases}$$

is the *sign representation* ρ_{sign} of S_l.

Another noteworthy representation of S_l is the l-dimensional representation ρ where

$$(\rho(\tau))_{ij} = \begin{cases} 1 & \text{if } \tau(j) = i \\ 0 & \text{otherwise.} \end{cases}$$

This is the *defining representation* of S_l. Let us work it out for S_3:

$$\rho(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho(1,2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho(1,3) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\rho(2,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \rho(1,2,3) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \rho(1,3,2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Observe that $\rho(\tau)$ is a permutation matrix for each τ.

As in Section 2, we use X_G to denote the Hilbert space of complex-valued functions on G, endowed with the inner product $\langle f_1, f_2 \rangle_G = \frac{1}{n} \sum_g f_1(g)\overline{f_2(g)}$. Every $g \in G$ induces a map $T_g : X_G \to X_G$ by sending $f \in X_G$ to the function $h \to f(g^{-1}h)$. T_g is clearly linear, surjective (for each $f \in X_g$, T_g maps the function $h \to f(gh)$ to f), and satisfies

$T_{g_1g_2} = T_{g_1}T_{g_2}$ (where the operation on the RHS is function composition). Furthermore, T_g is an isometry:

$$\begin{aligned}\langle T_g(f_1), T_g(f_2)\rangle_G &= \frac{1}{n}\sum_h f_1(g^{-1}h)\overline{f_2(g^{-1}h)} \\ &= \frac{1}{n}\sum_{h'} f_1(h')\overline{f_2(h')} = \langle f_1, f_2\rangle_G.\end{aligned} \tag{3.2}$$

Thus T_g is a unitary operator on X_G, so that if we fix an orthonormal basis $\{f_1, f_2, \ldots, f_n\}$ of X_G, then every T_g corresponds to the unitary matrix M_g whose ij-th entry is given by $\langle T_g(f_j), f_i\rangle_G$. This gives us the n-dimensional representation $\rho(g) = M_g$, which is called the *left regular representation of* G and denoted as $\rho_{regular}$.

New representations can be built from existing representations. One way to do this is via a change of basis. More precisely, starting with a d-dimensional represention ρ, then $\rho_M : g \to M\rho(g)M^{-1}$ is a d-dimensional representation as well. Another way to create new representations from existing ones is via direct sums: if ρ_1 and ρ_2 are d_1- and d_2-dimensional representations respectively, then

$$\rho_1 \oplus \rho_2 : g \to \begin{pmatrix} \rho_1(g) & 0 \\ 0 & \rho_2(g) \end{pmatrix} \tag{3.3}$$

is a $(d_1 + d_2)$-dimensional representation.

It is natural to ask the question, then, of what are the most "essential" representations, i.e. the smallest collection $\mathcal{C}$ of representations which fully preserves the structure of G. Representations which are constructed from other representations in manners described in the previous paragraph are clearly not good candidates for inclusion in $\mathcal{C}$. So what are the good candidates? This will be answered by the Peter-Weyl theorem, but to get there we need some preparations.

Definition 3.2. *(i) Representations ρ_1 and ρ_2 of the same dimension d are* equivalent *if there exists $M \in \mathcal{U}_d$ such that $\rho_2(g) = M\rho_1(g)M^{-1}$ for all $g \in G$.*

(ii) ρ is irreducible *if it is not equivalent to a representation of the form $\rho_1 \oplus \rho_2$ (where ρ_1 and ρ_2 need not be of the same dimension).*

Remark. 1-dimensional representations are trivially irreducible. Additionally, different 1-dimensional representations are obviously not equivalent.

Theorem 3.3 (Schur's lemma). *Let ρ be a d-dimensional representation of G. The following are equivalent:*

(i) ρ is irreducible.

(ii) If A is any d-by-d matrix such that $A\rho(g) = \rho(g)A$ for all g. Then A is a scalar multiple of I_d.

(iii) Suppose that V is a subspace of $\mathbb{C}^d$ that is invariant under $\rho(g)$ for all $g \in G$, i.e. $\rho(g)v \in V$ for all $v \in V$ (here V is said to reduce *ρ). Then either $V = \{0\}$ or $V = \mathbb{C}^d$.*

Proof. We will prove the contrapositive of each implication. (i) ⇒ (ii): Suppose that $A \notin \mathbb{C}I_d$ commutes with all $\rho(g)$. Since the $\rho(g)$ is unitary, $(\rho(g))^* = (\rho(g))^{-1} = \rho(g^{-1})$, and

$$(\rho(g)A^*)^* = A(\rho(g))^* = A\rho(g^{-1}) = \rho(g^{-1})A = (\rho(g))^*A = (A^*\rho(g))^*, \tag{3.4}$$

hence A^* also commutes with all $\rho(g)$. This means that $\rho(g)$ commutes with the Hermitian matrices $A_1 = \frac{1}{2}(A + A^*)$ and $A_2 = \frac{1}{2i}(A - A^*)$, and since $A = A_1 + iA_2$, at least one of A_1, A_2 is not in $\mathbb{C}I_d$.

Assume without loss of generality that $A_1 \notin \mathbb{C}I_d$. Because it is Hermitian, there exists $M \in \mathcal{U}_d$ such that $B = MAM^{-1}$ is diagonal. Let $\lambda_1, \dots, \lambda_d$ be the diagonal entries of B. Since $A_1 \notin \mathbb{C}_d$, there exists $d' < d$ such that $\lambda_1 = \cdots = \lambda_{d'}$ and that $\lambda_1 \neq \lambda_j$ for any $j > d'$. As B commutes with $\rho_M(g) = M\rho(g)M^{-1}$, so does $P(B)$ for any polynomial P. In particular, we may choose P so that $C = P(B)$ with $C_{11} = \cdots = C_{d'd'} = 1$ and $C_{ij} = 0$ for all other entries (i, j). But because $C\rho_M(g) = \rho_M(g)C$, each $\rho_M(g)$ can have a nonzero entry at (j, k) only if $1 \leq i, j \leq d'$ or $d' < i, j \leq d$, so that each $\rho_M(g)$ is the direct sum of two square matrices (of dimensions d'-by-d' and $(d - d')$-by-$(d - d')$). Thus $\rho_M = \rho_1 \oplus \rho_2$ for some ρ_1 and ρ_2.

(ii) ⇒ (iii): Let V be a reducing subspace for ρ. We will prove that there is a matrix not in $\mathbb{C}I_d$ which commutes with all $\rho(g)$. First assume that V is of the form

$$W_{d'} = \{(v_1, \dots, v_d) \in \mathbb{C}^d : v_{d'+1}, \dots, v_d = 0\} \tag{3.5}$$

for some $1 \leq d' < d$. Since $\rho(g)v \in V$ for all $v \in V$, we must have that $(\rho(g))_{ij} = 0$ for all (i, j) such that $d' + 1 \leq i \leq d$ and $1 \leq j \leq d'$. Since $\rho(g) = (\rho(g^{-1}))^*$, $(\rho(g^{-1}))_{ij} = 0$ for all (i, j) with $1 \leq i \leq d$ and $d' + 1 \leq j \leq d$, and because $\rho(g)\rho(g^{-1}) = I_d$, this is true for $\rho(g)$ as well. Thus $\rho(g)$ is a direct sum of a d'-by-d' matrix and a $(d - d')$-by-$(d - d')$ matrix, and as such commutes with the matrix C from the preceding part of the proof.

In the general case, if V is a proper nontrivial subspace of $\mathbb{C}^d$, it can be put into the form $W_{d'}$ $(1 \leq d' < d)$ via a change of basis, i.e. there exists $M \in \mathcal{U}_d$ such that $W_{d'} = \{Mv : v \in V\}$. Thus C commutes with $\rho_M(g) = M\rho(g)M^{-1}$ for all g, and therefore $M^{-1}CM$ commutes with $\rho(g)$ for all g.

(iii) ⇒ (i): Suppose $\rho = (\rho_1 \oplus \rho_2)_M$, with $\dim(\rho_1) = d'$. Then $V = \{v \in \mathbb{C}^d : Mv \in W_{d'}\}$ is a d'-dimensional reducing subspace for ρ. □

Corollary 3.4. *If G is abelian, then every irreducible representation is 1-dimensional, and thus the collection of irreducible representations of G can be identified with $\hat{G}$.*

Proof. Suppose that ρ is irreducible. Since G is abelian, $\rho(g)\rho(h) = \rho(h)\rho(g)$ for all $g, h \in G$, and therefor every $\rho(g)$ is of the form $a_g Id$ by Schur's lemma. But every matrix commutes with $a_g Id$, so if $\dim(\rho) > 1$, then there is a plethora of matrices A such that $A\rho(g) = \rho(g)A$ for every g, a contradiction. □

Definition 3.5. *Let ρ_1 and ρ_2 be irreducible representations of an arbitrary G, with dimensions d_1 and d_2 respectively. We say that a d_2-by-d_1 matrix A* connects *ρ_1 and ρ_2 if $A\rho_1(g) = \rho_2(g)A$ for all $g \in G$.*

Corollary 3.6. *(i) If A connects ρ_1 and ρ_2, then either $A = 0$ or $d_1 = d_2$, in which case $A = aM$ for some positive a and unitary matrix M.*

(ii) ρ_1 is equivalent to ρ_2 if and only if there exists a nonzero matrix A which connects ρ_1 and ρ_2.

Proof. (i) Let $V = \{v \in \mathbb{C}^{d_1} : Av = 0\}$. Since $A\rho_1(g) = \rho_2(g)A$, $\rho_1(g)v \in V$ for all $v \in V$, which by Schur's lemma implies that either $V = \{0\}$ or $V = \mathbb{C}^{d_1}$. Similarly, by considering $W = \{Av : v \in \mathbb{C}^{d_1}\}$ with ρ_2, we see that either $W = \{0\}$ or $W = \mathbb{C}^{d_2}$. If A is not the zero matrix, then $V = \{0\}$ and $W = \mathbb{C}^{d_2}$, so that A has independent rows and the column rank (and thereby the row rank) of d_2. This is possible only if $d_1 = d_2$.

It remains to show that, in this case, A is a positive multiple of a unitary matrix. Since $A\rho_1(g) = \rho_2(g)A$ for all g, taking adjoints give us that $p_1(g^{-1})A^* = A^*\rho_2(g^{-1})$ for all g, so that $p_1(g)A^* = A^*\rho_2(g)$ for all g. Therefore $A^*A\rho_1(g) = A^*\rho_2(g)A = \rho_1 A^*A$ for all g, and by Schur's lemma, A^*A is of the form cI_{d_1}. Observe that the inner product $\langle v_1, v_2\rangle := v_2^* v_1$ on complex vector spaces has the property that $\langle Av, Av\rangle = \langle A^*Av, v\rangle = c\langle v, v\rangle \geq 0$, and since A is not the zero matrix, $c\langle v, v\rangle > 0$ whenever $v \neq 0$. Hence c must be real and positive, and $A = \sqrt{c}M$, where M is unitary.
(ii) If ρ_1 and ρ_2 are equivalent, then $\rho_2 = M\rho_1 M^{-1}$ for some unitary matrix M, so that M connects ρ_1 and ρ_2. Conversely, if a nonzero A connects ρ_1 and ρ_2, then by part (i) A is of the form aM for some unitary M, so that $\rho_2 = M\rho_1 M^{-1}$. □

The next proposition will tell us how to construct connecting matrices:

Proposition 3.7. *Let ρ_1 and ρ_2 be irreducible representations of G with dimensions d_1 and d_2 respectively. Then, for any d_2-by-d_1 matrix A, the matrix*

$$\tilde{A} = \frac{1}{n}\sum_g \rho_2(g^{-1})A\rho_1(g) \tag{3.6}$$

connects ρ_1 and ρ_2. In particular, $\tilde{A}$ is a positive multiple of a unitary matrix if ρ_1 is equivalent to ρ_2 and the zero matrix otherwise.

In the case that $\rho_1 = \rho_2$, $\tilde{A}$ is of the form cI_{d_1}, where $c =$tr$(A)/d_1$.

Proof. Fix any $g \in G$. Since

$$\sum_h \rho_2(h^{-1})A\rho_1(hg) = \sum_{hg} \rho_2(g(hg)^{-1})A\rho_1(hg) = \sum_h \rho_2(gh^{-1})A\rho_1(h), \tag{3.7}$$

we see that $\tilde{A}\rho_1(g) = \rho_2(g)\tilde{A}$. If $\rho_1 = \rho_2$, then $\tilde{A}\rho_1(g) = \rho_1(g)\tilde{A}$ for all g, so by Schur's lemma $\tilde{A} = cI_{d_1}$, and

$$c = \frac{\mathrm{tr}(\tilde{A})}{d_1} = \frac{1}{d_1 n}\sum_g \mathrm{tr}(\rho_1(g^{-1})A\rho_1(g)) = \frac{1}{d_1 n}\sum_g \mathrm{tr}(A) = \frac{\mathrm{tr}(A)}{d_1}, \tag{3.8}$$

where we have used trace property that $\mathrm{tr}(PQ) = \mathrm{tr}(QP)$ for arbitrary matrices P and Q (so long that PQ and QP are both defined). □

3.2. **Duals and the Peter-Weyl theorem.**

For an irreducible representation ρ of G, we define the functions $f_{ij}^\rho : G \to \mathbb{C}$ by assigning to $f_{ij}^\rho(g)$ the ij-th entry of $\rho(g)$. The f_{ij}^ρ are called the *coordinate functions* of ρ.

Theorem 3.8. *As elements of X_G, the coordinate functions of irreducible representations satisfy the following orthogonality relations:*

(i) If ρ_1 and ρ_2 are not equivalent, then $\langle f_{ij}^{\rho_1}, f_{lm}^{\rho_2}\rangle_G = 0$ for any i, j, l, m.
(ii) $\langle f_{ij}^\rho, f_{lm}^\rho\rangle_G$ is $1/\dim(\rho)$ if $(i, j) = (l, m)$ and 0 otherwise.

Proof. (i) Let $d_1 = \dim(\rho_1)$ and $d_2 = \dim(\rho_2)$. Define the d_2-by-d_1 matrix A where $A_{li} = 1$ and all other entires are 0. With respect to the usual inner product on $\mathbb{C}^{d_2}$, we have that

$$\langle A\rho_1(g)e_j^{d_1}, \rho_2(g)e_m^{d_2}\rangle = (\rho_2(g)e_m^{d_2})^* A\rho_1(g)e_j^{d_1} = f_{ij}^{\rho_1}(g)\overline{f_{lm}^{\rho_2}(g)}, \tag{3.9}$$

where $e_j^{d_1}$ is the j-th standard unit vector of $\mathbb{C}^{d_1}$ and $e_m^{d_2}$ is the m-th standard unit vector of $\mathbb{C}^{d_2}$. Since $(\rho_2(g))^* = \rho_2(g^{-1})$, $\langle A\rho_1(g)e_j^{d_1}, \rho_2(g)e_m^{d_2}\rangle = \langle \rho_2(g^{-1})A\rho_1(g)e_j^{d_1}, e_m^{d_2}\rangle$, so that

$$\begin{aligned}\langle \tilde{A}e_j^{d_1}, e_m^{d_2}\rangle &= e_m^{d_2}\tilde{A}e_j^{d_1} = \frac{1}{n}\sum_g e_m^{d_2}\rho_2(g^{-1})A\rho_1(g)e_j^{d_1} \\ &= \frac{1}{n}\sum_g \langle \rho_2(g^{-1})A\rho_1(g)e_j^{d_1}, e_m^{d_2}\rangle = \frac{1}{n}\sum_g f_{ij}^{\rho_1}(g)\overline{f_{lm}^{\rho_2}(g)} = \langle f_{ij}^{\rho_1}, f_{lm}^{\rho_2}\rangle_G.\end{aligned} \tag{3.10}$$

When ρ_1 and ρ_2 are not equivalent, $\tilde{A}$ is the zero matrix by Corollary 3.6, and hence $\langle f_{ij}^{\rho_1}, f_{lm}^{\rho_2}\rangle_G = 0$.

(ii) We apply part (i) with $\rho = \rho_1 = \rho_2$ and $d = \dim(\rho)$. If $i \neq l$, then A has all 0's on main diagonal and therefore $\mathrm{tr}(A) = 0$, which by Corollary 3.6 implies that $\tilde{A}$ is the zero matrix, so that $\langle f_{ij}^{\rho}, f_{lm}^{\rho}\rangle_G = 0$. If $i = l$, then $\mathrm{tr}(A) = 1$ and $\tilde{A} = \frac{1}{d}I_d$. It follows from (3.10) that

$$\langle f_{ij}^{\rho}, f_{im}^{\rho}\rangle_G = \langle \tilde{A}e_j^d, e_m^d\rangle = \begin{cases} \frac{1}{d} & \text{if } j = m \\ 0 & \text{otherwise,} \end{cases}$$

as was to be shown. □

Part (i) of Theorem 3.8 tells us that there cannot be "too many" non-equivalent irreducible representations: since each pair of $(f_{ij}^{\rho}, f_{lm}^{\rho'})$ are orthogonal, each ρ gives rise to d_ρ^2 functions which are orthogonal to all other $f_{lm}^{\rho'}$. As are at most $\dim(X_G) = n$ such functions, $\sum_\rho d_\rho^2 \leq n$. But are there "too few" non-equivalent irreducible representations?

Definition 3.9. *A* dual *of G, denoted as $\hat{G}$, is a collection of representations that contains precisely one representative from each equivalence class of irreducible representations of G.*

Remark. In general, duals are not unique. In the abelian case, however, G has only one dual, namely the character group of G.

The Peter-Weyl theorem will tell us that in fact, the number of non-equivalent irreducible representations is just right. More precisely, it states that the normalized coordinate functions associated with a dual $\hat{G}$ form an orthonormal basis of X_G. This will allow us to generalize the Fourier analysis of the previous section, so let us continue with preparations for the Peter-Weyl theorem.

Proposition 3.10. *Let ρ be a d-dimensional representation of G. If we define $\bar{\rho} : G \to \mathcal{U}_d$ by $\bar{\rho}(g) = \overline{\rho(g)}$, then $\bar{\rho}$ is also a d-dimensional representation of G. Furthermore, if ρ is irreducible, then so is $\bar{\rho}$.*

Remark. Note that this generalizes the fact about abelian groups that the complex conjugate of a character is again a character.

Proof. That $\bar{\rho}$ is a representation follows from the fact that $\overline{M}$ is unitary for a unitary matrix M. If, in addition, ρ is irreducible, then for any matrix A such that $A\bar{\rho} = \bar{\rho}A$, $\bar{A}$ must be in $\mathbb{C}I_d$ by Schur's lemma, which means that A is as well. □

Proposition 3.11. *Any representation is either irreducible or equivalent to a representation $\rho_1 \oplus \rho_2$ where ρ_1 is irreducible.*

Proof. The proof is via strong induction on d_ρ. The base case of $d_\rho = 1$ is trivial. For $d_\rho = k$, if ρ is not irreducible, then $\rho_M = \rho_1 \oplus \rho_2$ for some $M \in \mathcal{U}_k$. Let $k_1 = \dim(\rho_1)$. If ρ_1 is irreducible, we are done, otherwise by the inductive hypothesis, $M_1\rho_1M_1^{-1} = \rho' \oplus \rho''$ for irreducible representation ρ', representation ρ'', and $M_1 \in \mathcal{U}_{k_1}$. Then

$$\begin{aligned}
&((M_1 \oplus I_{k-k_1})M)\rho((M_1 \oplus I_{k-k_1})M)^{-1}\\
&= ((M_1 \oplus I_{k-k_1})M)(M^{-1}\rho_1 \oplus \rho_2 M)((M_1 \oplus I_{k-k_1})M)^{-1}\\
&= (M_1 \oplus I_{k-k_1})(\rho_1 \oplus \rho_2)(M_1 \oplus I_{k-k+1})^{-1}\\
&= M_1\rho_1M_1^{-1} \oplus \rho_2 = \rho' \oplus (\rho'' \oplus \rho_2),
\end{aligned} \tag{3.11}$$

and hence ρ is equivalent to the direct sum of ρ' and $\rho'' \oplus \rho_2$, where ρ' is irreducible. □

Recall that each $g \in G$ induces a map $T_g : X_G \to X_G$ by sending f to the function $h \to f(g^{-1}h)$. Suppose that W is a subspace of X_G that is invariant under all T_g, i.e. for each $g \in G$, if $f \in W$, then $T_g(f) \in W$. Fix any orthonormal basis $\{f_1, \dots, f_d\}$ of W, and define ρ_g as the d-by-d matrix whose ij-th entry is given by $\langle T_g(f_j), f_i\rangle_G$. Then ρ is a representation. If $W = X_G$, this is the left regular representation which we introduced in Section 3.1, otherwise we say that ρ is the representation induced by $\rho_{regular}$ on W. The next proposition plays a crucial role in the proof of the Peter-Weyl theorem:

Proposition 3.12. *Suppose that ρ, as defined above, is irreducible. Let ρ' be an irreducible representation which is equivalent with $\bar{\rho}$. Then not all coordinate functions $f_{ij}^{\rho'}$ of ρ' are in $W^\perp$.*

Proof. If all $f_{ij}^{\rho'}$ are in $W^\perp$, then since the $f_{ij}^{\bar{\rho}}$ are linear combinations of the $f_{ij}^{\rho'}$, we have that $\langle f_l, f_{ij}^{\bar{\rho}}\rangle_G$ for all $1 \le i, j, l \le d$. Therefore,

$$0 = \sum_g f_l(g)\overline{f_{ij}^{\bar{\rho}}(g)} = \sum_g f_l(g) f_{ij}^{\rho}(g) = \sum_g f_l(g)\langle T_g(f_j), f_i\rangle_G. \tag{3.12}$$

Moreover, since $\{f_1, \dots, f_d\}$ is an orthonormal basis of W,

$$T_g(f_j) = \sum_{f_i}\langle T_g(f_j), f_i\rangle_G f_i. \tag{3.13}$$

Evaluating at each g,

$$\sum_{f_i}\langle T_g(f_j), f_i\rangle_G f_i(g) = (T_g(f_j))(g) = f_j(g^{-1}g) = f_k(e), \tag{3.14}$$

and thus

$$f_j(e) = \frac{1}{n}\sum_{g,f_i}\langle T_g(f_j), f_i\rangle_G f_i(g). \tag{3.15}$$

By (3.12), the RHS is 0, so we have that $f_j(e) = 0$ for each j. Then by (3.13) evaluated at e, $f_j(g^{-1}) = 0$ for all $g \in G$, which means that $f_j(g) = 0$ for all j and g. This is clearly absurd, since $\{f_1, \dots, f_d\}$ is an orthonormal set. □

Theorem 3.13 (Peter-Weyl). *Let $\hat{G}$ be a dual of G. Then $\{\sqrt{d_\rho} f_{ij}^{\rho} : \rho \in \hat{G}, 1 \le i, j \le d_\rho\}$ is an orthonormal basis of X_G.*

Proof. Let V be the linear span of the f^{ρ}_{jk}. By the orthogonality properties proved in Theorem 3.8, it suffices to show that $V = X_G$, or equivalently, that $W = V^{\perp}$ is empty.

We claim that W is invariant under T_g for all $g \in G$. To show this, we first show that V is invariant. Since $\rho(g^{-1}h) = \rho(g^{-1})\rho(h)$ for all h, every translate $T_g(f^{\rho}_{ij})$ of a coordinate function is a linear combination of coordinate functions and therefore lies in V. This verifies that V is invariant under all T_g. For W, we need to show that $\langle T_g(f_1), (f_2)\rangle_G = 0$ whenever $f_1 \in W$ and $f_2 \in V$. Since V is invariant, if $f_2 \in V$, then $T_{g^{-1}}(f_2) \in V$, so that $\langle f_1, T_{g^{-1}}(f_2)\rangle_G = 0$. But then since T_g is an isometry,

$$\langle T_g(f_1), f_2\rangle = \langle T_g(f_1), T_g(T_{g^{-1}}(f_2))\rangle_G = \langle f_1, T_{g^{-1}}(f_2)\rangle_G = 0, \tag{3.16}$$

which proves the claim.

Now we put can finally everything together. Suppose that W is nonempty, and consider the representation ρ induced by $\rho_{regular}$ on W. It may not be irreducible, but by Proposition 3.11, there exists nonempty orthonormal set $f_1, \dots, f_d \in W$ with linear span W_1 such that the representation ρ_1 induced by $\rho_{regular}$ on W_1 is irreducible. Let ρ' be the element of $\hat{G}$ that is equivalent to $\bar{\rho_1}$. Since $W_1 \subseteq W$, all coordinate functions $f^{\rho'}_{ij}$, which are in $W^{\perp}$, are in $W_1^{\perp}$ as well. But this is impossible by Proposition 3.12, which means that W must be empty. □

Corollary 3.14. *(i)* $\sum_{\rho\in\hat{G}} d_\rho^2 = n$.
(ii) For every $f \in X_G$, $f = \sum_{\rho,i,j} d_\rho \langle f, f^{\rho}_{ij}\rangle_G f^{\rho}_{ij}$.
(iii) If $g \neq h$, *then there exists* $\rho \in \hat{G}$ *such that* $\rho(g) \neq \rho(h)$.
(iv) G *is commutative if and only if all irreducible representations are 1-dimensional.*

Proof. Parts (i) and (ii) are immediate consequences of the Peter-Weyl theorem. For (iii), it suffices to find f^{ρ}_{ij} with $\rho \in \hat{G}$ such that $f^{\rho}_{ij}(g) \neq f^{\rho}_{ij}(h)$, and as in Corollary 2.5, this follows from the existence of an $f \in X_G$ such that $f(g) \neq f(h)$. The only if direction of (iv) has already been proved in Corollary 3.4. For the other direction, if all irreducible representations of G are 1-dimensional, then $\rho(gh) = \rho(g)\rho(h) = \rho(h)\rho(g) = \rho(hg)$ for all g, h, and ρ. By (iii), this means that $gh = hg$ for all $g, h \in G$. □

In general, the task of finding a dual is daunting. The following converse of Peter-Weyl is helpful:

Proposition 3.15. *Suppose that* $\mathcal{C}$ *is a collection of representations of* G *whose normalized coordinate functions* $\{\sqrt{d_\rho} f^{\rho}_{ij} : \rho \in \mathcal{C}, 1 \le i, j \le d_\rho\}$ *form an orthonormal basis of* X_G. *Then* $\mathcal{C}$ *is a dual of* G.

Proof. Let ρ be a d-dimensional representation of G whose coordinate functions are orthogonal and have norm $1/\sqrt{d}$. Also, let $M \in \mathcal{U}_d$. Since $f^{\rho_M}_{ij} = \sum_{l,m} M_{im}\overline{M_{jl}} f^{\rho}_{ml}$, the unitarity of M gives us that

$$\begin{aligned}\langle f^{\rho_M}_{ij}, f^{\rho_M}_{i'j'}\rangle_G &= \sum_{l,m,l',m'} \langle f^{\rho}_{ml}, f^{\rho}_{m'l'}\rangle_G M_{im}\overline{M_{jl}} M_{j'l'}\overline{M_{i'm'}} \\ &= \frac{1}{d}\sum_{l,m} M_{im}\overline{M_{jl}} M_{j'l}\overline{M_{i'm}} \\ &= \frac{1}{d}\left(\sum_{l} \overline{M_{jl}} M_{j'l}\right)\left(\sum_{m} M_{im}\overline{M_{i'm}}\right) = \frac{1}{d}\delta_{jj'}\delta_{ii'},\end{aligned} \tag{3.17}$$

so that the coordinate functions of ρ_M also have norm $1/\sqrt{d}$.

We claim that every $\rho \in \mathcal{C}$ must be irreducible. Otherwise, there exists $M \in \mathcal{U}_d$ such that $\rho_M = \rho_1 \oplus \rho_2$, where ρ_1 is irreducible and $0 < d_{\rho_1} < d$. By (3.17), the coordinate functions of ρ_M, and therefore those of ρ_1, have norm $1/\sqrt{d}$. However, by Theorem 3.8, the coordinate functions of ρ_1 have norm $1/\sqrt{d_1}$. This is a contradiction, and hence $\mathcal{C}$ contains only irreducible representations.

Now we show that no two representations $\rho_1, \rho_2 \in \mathcal{C}$ are equivalent. Suppose that $\rho_2 = (\rho_1)_M$ for some unitary M. By assumption, the coordinate functions of ρ_1 and ρ_2 are orthogonal, but since the coordinate functions of ρ_2 are just linear combinations of the coordinate functions of ρ_1, this cannot happen. Therefore the representations contained in $\mathcal{C}$ are pairwise non-equivalent.

It remains to argue that $\mathcal{C}$ contains a representative from each equivalent class of irreducible representations. If this were not the case, i.e. there exists irreducible ρ that is not equivalent to any representation in $\mathcal{C}$, then Theorem 3.8 stipulates that the coordinate functions of ρ are orthogonal to the coordinate functions of the representations in $\mathcal{C}$. But this contradicts the condition that the coordinate functions of the representations in $\mathcal{C}$ form a basis of X_G, so we are done. □

Example. Let us find a dual of the symmetric group S_3. Since S_3 is not abelian, it has at least one representation of degree greater than 1. On the other hand, $\sum_\rho d_\rho^2 = 6$, so the only option is that S_3 has two representations of dimension 1 and one representation of dimension 2. (By the way, this tells us that the defining representation of S_3 is not irreducible.)

The two 1-dimensional representations are clearly ρ_{triv} and ρ_{sign}. The natural choice for the 2-dimensional representation ρ arises from the correspondence between elements of S_3 and motions that fix the origin while permuting the vertices of an equilateral triangle, which in turn corresponds to 2-by-2 unitary matrices:

$$\rho(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \rho(1,2) = \begin{pmatrix} \frac{1}{2} & \sqrt{\frac{3}{4}} \\ \sqrt{\frac{3}{4}} & -\frac{1}{2} \end{pmatrix}, \quad \rho(1,3) = \begin{pmatrix} \frac{1}{2} & -\sqrt{\frac{3}{4}} \\ -\sqrt{\frac{3}{4}} & -\frac{1}{2} \end{pmatrix},$$

$$\rho(2,3) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \ \rho(1,2,3) = \begin{pmatrix} -\frac{1}{2} & -\sqrt{\frac{3}{4}} \\ \sqrt{\frac{3}{4}} & -\frac{1}{2} \end{pmatrix}, \ \rho(1,3,2) = \begin{pmatrix} -\frac{1}{2} & \sqrt{\frac{3}{4}} \\ -\sqrt{\frac{3}{4}} & -\frac{1}{2} \end{pmatrix}.$$

We now check that $\{\rho_{triv}, \rho_{sign}, \rho\}$ is indeed a dual of S_3 using Proposition 3.15. The following table gives the associated coordinate functions:

g	e	(1, 2)	(1, 3)	(2, 3)	(1, 2, 3)	(1, 3, 2)
ρ_{triv}	1	1	1	1	1	1
ρ_{sign}	1	-1	-1	-1	1	1
f_{11}^ρ	1	$1/2$	$1/2$	-1	$-1/2$	$-1/2$
f_{12}^ρ	0	$\sqrt{3/4}$	$-\sqrt{3/4}$	0	$-\sqrt{3/4}$	$\sqrt{3/4}$
f_{21}^ρ	0	$\sqrt{3/4}$	$-\sqrt{3/4}$	0	$\sqrt{3/4}$	$-\sqrt{3/4}$
f_{22}^ρ	1	$-1/2$	$-1/2$	1	$-1/2$	$-1/2$

It is straightforward to verify that the six coordinate functions are an orthonormal set, and thus $\widehat{S_3} = \{\rho_{triv}, \rho_{sign}, \rho\}$.

3.3. **Fourier transforms of functions and measures.** We proceed to generalize the Fourier analysis of the previous section. Let f be a complex-valued function and μ be a probability measure, both on G. Recall that when G is abelian, the Fourier transforms of f and μ are defined by

$$\hat{f}(\chi) = \frac{1}{n}\sum_g f(g)\chi(g) \text{ and } \hat{\mu}(\chi) = \sum_g \mu(g)\chi(g). \tag{3.18}$$

Naturally, as dimension increases, complexity expands.

Definition 3.16. *When G is an arbitrary group, the* Fourier transforms *of f and μ are the matrix-valued maps defined by*

$$\hat{f}(\rho) = \frac{1}{n}\sum_g f(g)\rho(g) = \left(\left(\frac{1}{n}\sum_g f(g)f^\rho_{ij}(g)\right)_{i,j=1,\dots,d_\rho}\right) \tag{3.19}$$

and $\hat{\mu}(\rho) = \sum_g \mu(g)\rho(g)$, where $\rho \in \hat{G}$.

As before, Fourier transforms are linear maps. Formally, $\hat{f}$ and $\hat{\mu}$ can be identified with elements of the n-dimensional linear space $\prod_{\rho\in\hat{G}}\mathcal{M}_{d_\rho}$, where $\mathcal{M}_d$ is the set of d-by-d complex matrices. $\prod_{\rho\in\hat{G}}\mathcal{M}_{d_\rho}$ can be viewed as a space of functions defined on $\hat{G}$, analogous to $X_{\hat{G}}$ from the abelian case.

Example. Define $f : S_3 \to \mathbb{C}$ by $f(\tau) = \tau(1)$. Then

$$\begin{aligned}
\hat{f}(\rho_{triv}) &= \frac{1}{6}(1+2+3+1+2+3) = 2,\\
\hat{f}(\rho_{sign}) &= \frac{1}{6}(1-2-3-1+2+3) = 0, \text{ and}\\
\hat{f}(\rho) &= \frac{1}{6}\left(1\begin{pmatrix} 1 & 0\\ 0 & 1\end{pmatrix} + \cdots + 3\begin{pmatrix} -\frac{1}{2} & \sqrt{\frac{3}{4}}\\ -\sqrt{\frac{3}{4}} & -\frac{1}{2}\end{pmatrix}\right) = \begin{pmatrix} 0 & 0\\ -\sqrt{\frac{1}{12}} & -\frac{1}{2}\end{pmatrix}.
\end{aligned} \tag{3.20}$$

Moreover, let μ be the measure on G that assigns mass $1/3$ to e, $(1,2)$, and $(1,3)$. Then

$$\begin{aligned}
\hat{\mu}(\rho_{triv}) &= \frac{1}{3}(1+1+1) = 1,\\
\hat{\mu}(\rho_{sign}) &= \frac{1}{3}(1-1-1) = -\frac{1}{3}, \text{ and}\\
\hat{\mu}(\rho) &= \frac{1}{3}\left(\begin{pmatrix} 1 & 0\\ 0 & 1\end{pmatrix} + \begin{pmatrix} \frac{1}{2} & \sqrt{\frac{3}{4}}\\ \sqrt{\frac{3}{4}} & -\frac{1}{2}\end{pmatrix} + \begin{pmatrix} \frac{1}{2} & -\sqrt{\frac{3}{4}}\\ -\sqrt{\frac{3}{4}} & -\frac{1}{2}\end{pmatrix}\right) = \begin{pmatrix} \frac{2}{3} & 0\\ 0 & 0\end{pmatrix}.
\end{aligned} \tag{3.21}$$

Proposition 3.17. *(i) $\widehat{f^\rho_{ij}}(\rho_0)$ is the zero matrix whenever $\rho_0 \neq \rho$, otherwise it is E^ρ_{ij}, the d_ρ-by-d_ρ matrix with $1/d_\rho$ at the ij-th entry and 0 everywhere else.*

(ii) The Fourier transform of the uniform distribution is 1 at ρ_{triv} and vanishes at all other $\rho \in \hat{G}$.

Proof. (i) is merely a restatement of Theorem 3.8. The proof for (ii) is essentially the same as that of Proposition 2.3. Clearly $\hat{U}(\rho_{triv}) = 1$. If $\rho \neq \rho_{triv}$, then there exists $g_0 \in G$ such that $\rho(g_0) \neq I_{d_\rho}$, and

$$\hat{U}(\rho) = \frac{1}{n}\sum_g \rho(g) = \frac{1}{n}\sum_g \rho(g_0 g) = \rho(g_0)\frac{1}{n}\sum_g \rho(g). \tag{3.22}$$

Because $\rho(g_0) \neq I_{d_\rho}$, $\frac{1}{n}\sum_g \rho(g) = 0$ must be the zero matrix. □

Theorem 3.18 (Fourier inversion)**.** *For any $f \in X_G$ and $g \in G$,*

$$f(g) = \sum_{\rho \in \hat{G}} d_\rho \mathrm{tr}(\rho(g^{-1})\hat{f}(\rho)). \tag{3.23}$$

It follows that for any measure μ on G,

$$\mu(g) = \frac{1}{n}\sum_{\rho \in \hat{G}} d_\rho \mathrm{tr}(\rho(g^{-1})\hat{\mu}(\rho)). \tag{3.24}$$

Proof. First consider the case when $f = \overline{f_{ij}^\rho}$. Since $\rho_0(g^{-1}) = (\rho_0(g))^*$, by Proposition 3.17, $\rho_0(g^{-1})\hat{f}(\rho_0)$ is the zero matrix when $\rho_0 \neq \rho$ and the matrix A/d_ρ, where A is the matrix with $(\overline{f_{i1}^\rho}(g), \overline{f_{i2}^\rho}(g), \ldots, \overline{f_{i,d_{\rho_0}}^\rho}(g))$ in the j-th column and 0 everywhere else, when $\rho_0 = \rho$. Thus $\mathrm{tr}(\rho(g^{-1})\hat{\mu}(\rho)) = \overline{f_{ij}^\rho}/d_\rho$. The general case follows from linearity. □

Theorem 3.19 (Plancherel formula)**.** *For any $f_1, f_2 \in X_G$,*

$$\langle f_1, f_2 \rangle_G = \sum_{\rho \in \hat{G}} d_\rho \mathrm{tr}(\hat{f}_1(\rho)(\hat{f}_2(\rho))^*). \tag{3.25}$$

In particular,

$$\frac{1}{n}\sum_g |f(g)|^2 = \sum_{\rho \in \hat{G}} d_\rho \mathrm{tr}(\hat{f}(\rho)(\hat{f}(\rho))^*) \tag{3.26}$$

for any $f \in X_G$.

Proof. Again, it suffices to consider only the complex conjugates of the coordinate functions. Let $f_1 = \overline{f_{ij}^{\rho'}}$ and $f_2 = \overline{f_{lm}^{\rho''}}$. If $\rho' = \rho''$ and $(i,j) = (l,m)$, then the LHS of (3.25) is $1/d_{\rho'}$; otherwise it is 0. On the RHS, $\hat{f}_1(\rho')$ is the matrix $E_{ij}^{\rho'}$ if $\rho = \rho' = \rho''$ and the zero matrix otherwise. Therefore if $\rho' = \rho''$ and $(i,j) = (l,m)$, then $\hat{f}_1(\rho)(\hat{f}_2(\rho))^*$ is the $d_{\rho'}$-by-$d_{\rho'}$ matrix with $1/(d_{\rho'})^2$ at the ii-th entry and 0 everywhere else; otherwise it is the zero matrix. This gives the desired results. □

Corollary 3.20. *$f \to \hat{f}$ is a bijection between X_G and $\prod_{\rho \in \hat{G}} \mathcal{M}_{d_\rho}$.*

Proof. The inversion formula shows that the Fourier transform is injective. Surjectivity follows from the fact that both X_G and $\prod_{\rho \in \hat{G}} \mathcal{M}_{d_\rho}$ are of the same dimension n. □

Proposition 3.21. *(i) $\mu = U$ if and only if $\hat{\mu}$ is 1 on ρ_{triv} and 0 on all other ρ.*

(ii)

$$\|\mu - U\|_{TV} \le \frac{1}{2}\left(\sum_{\rho\neq\rho_{triv}\in\hat{G}} d_\rho \mathrm{tr}(\hat{\mu}(\rho)(\hat{\mu}(\rho))^*)\right)^{1/2}. \tag{3.27}$$

Proof. (i) The only if direction has already been noted in part (ii) of Proposition 3.17. The if direction is due to the injectivity of the Fourier transform.
(ii) By the Plancherel formula (3.26), linearity, and part (i),

$$\sum_g \left(\mu(g) - \frac{1}{n}\right)^2 = \frac{1}{n}\sum_{\rho\neq\rho_{triv}\in\hat{G}} d_\rho \mathrm{tr}(\hat{\mu}(\rho)(\hat{\mu}(\rho))^*). \tag{3.28}$$

The result now follows from (2.15). □

Theorem 2.12 on the Fourier transform of convolutions still applies when G is non-abelian (note that we were very careful in the proof with the order of operations), so we finally arrive at the following analogue of Corollary 2.13:

Corollary 3.22.

$$\|\mu^{(k)} - U\|_{TV} \le \frac{1}{2}\left(\sum_{\rho\neq\rho_{triv}\in\hat{G}} d_\rho \mathrm{tr}((\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k)\right)^{1/2}. \tag{3.29}$$

Corollary 3.23. *The sequence $\mu^{(k)}$ converges to U in the total variation norm if and only if $\|(\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k\| \to 0$, where $\|\cdot\|$ is any matrix norm, for all $\rho \neq \rho_{triv}$ in $\hat{G}$.*

Proof. The matrices $(\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k$ are positive semidefinite, so their diagonal entries are non-negative, and thus $\mathrm{tr}((\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k) \to 0$ if and only if $(\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k$ converges to the zero matrix in some matrix norm. Hence the corollary follows from Theorem 2.12 and (3.28). □

3.4. **Convergence in matrix norm.** Unfortunately, the question of whether and how fast $(\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k$ converges to the zero matrix is significantly more difficult than the analogous question in the abelian case, because we are now dealing with matrices instead of numbers.

Define the *operator norm* of a d-by-d matrix A as

$$\|Av\|_{op} = \sup\{\|Av\| : v \in \mathbb{C}^d \text{ with } \|v\| = 1\}. \tag{3.30}$$

The operator norm is sub-multiplicative, i.e. it satisfies $\|AB\|_{op} \le \|A\|_{op}\|B\|_{op}$, so that $\|(\hat{\mu}(\rho))^k((\hat{\mu}(\rho))^*)^k\|_{op} \le \|\hat{\mu}(\rho)\|_{op}^k\|(\hat{\mu}(\rho))^*\|_{op}^k$. However, in general this is a rather weak bound.

Let λ_0 be the eigenvalue of AA^* with the largest modulus. It can be shown that $\|A\|_{op} = \sqrt{|\lambda_0|}$. Suppose that A is Hermitian. Then $\|A^k(A^*)^k\|_{op} = \|A^{2k}\|_{op} = \sqrt{|\lambda_1|}$, where λ_1 is the eigenvalue of $A^{2k}(A^{2k})^* = A^{4k}$ with the largest modulus. Since the eigenvalues of matrix powers are exactly the powers of the eigenvalues, $\lambda_1 = \lambda^{4k}$, where λ is the eigenvalue of A with the largest modulus. Hence $\|A^k(A^*)^k\|_{op} = |\lambda|^{2k}$. (We can drop the absolute value sign, as the eigenvalues of a Hermitian matrix is always real.) This suggests that those measures μ for which the matrices $\hat{\mu}(\rho)$ are all Hermitian may be nicer to work with. Fortunately, we have a simple criterion for this.

Proposition 3.24. *If μ is a* symmetric *probability measure on G, i.e. $\mu(g) = \mu(g^{-1})$ for all $g \in G$, then $\hat{\mu}(\rho)$ is Hermitian for all $\rho \in \hat{G}$.*

Proof. From the unitarity of $\rho(g)$,

$$\begin{aligned}(\hat{\mu}(\rho))^* &= \left(\sum_g \mu(g)\rho(g)\right)^* = \sum_g \mu(g)(\rho(g))^* = \sum_g \mu(g)\rho(g^{-1}) \\ &= \sum_g \mu(g^{-1})\rho(g^{-1}) = \sum_g \mu(g)\rho(g) = \hat{\mu}(\rho),\end{aligned} \tag{3.31}$$

as desired. □

If G is non-abelian, a certain type of measures on G could mimic measures on abelian groups.

Definition 3.25. *A probability measure μ that is constant on the conjugacy classes of G, i.e. $\mu(g) = \mu(hgh^{-1})$ for all $g, h \in G$, is called a* class measure *on G.*

Observe that if μ is a class measure, then $\mu(gh) = \mu(hghh^{-1}) = \mu(hg)$ for all $g, h \in G$.

Theorem 3.26. *Let μ be a class measure. Then, for every $\rho \in \hat{G}$, we have that*

$$\hat{\mu}(\rho) = \left(\frac{1}{d_\rho}\sum_g \mu(g)\chi_\rho(g)\right) I_{d_\rho}, \tag{3.32}$$

where $\chi_\rho(g) = \operatorname{tr}(\rho(g))$ is the character associated with ρ *at g.*

Proof. For any $h \in G$,

$$\begin{aligned}\rho(h)\hat{\mu}(\rho) &= \rho(h)\sum_g \mu(g)\rho(g) = \sum_g \mu(g)\rho(hg) \\ &= \sum_g \mu(h^{-1}hg)\rho(hg) \overset{g'=hg}{=} \sum_{g'} \mu(h^{-1}g')\rho(g') \\ &= \sum_{g'} \mu(g'h^{-1})\rho(g') \overset{g''=g'h^{-1}}{=} \sum_{g''} \mu(g'')\rho(g''h) \\ &= \left(\sum_{g''} \mu(g'')\rho(g'')\right)\rho(h) = \hat{\mu}(\rho)\rho(h).\end{aligned} \tag{3.33}$$

Thus by Schur's lemma, $\hat{\mu}(\rho) = cI_{d_\rho}$ for some c, so that $\operatorname{tr}(\hat{\mu}(\rho)) = cd_\rho$. On the other hand, $\operatorname{tr}(\hat{\mu}(\rho)) = \sum_g \mu(g)\chi_\rho(g)$, and hence $c = \frac{1}{d_\rho}\sum_g \mu(g)\chi_\rho(g)$. □

Example. Consider the *random transposition shuffle* of a deck of l cards, where we select i, j uniformly from $\{1, 2, \ldots, l\}$ and exchange the i-th and j-th cards. This corresponds to the measure μ on S_l that assigns mass $1/l$ to the identity and $2/l^2$ to each of the $l(l-1)/2$ transpositions. The set of all transpositions is a conjugacy class in S_l, so μ is a class measure. Clearly $\chi_\rho(e) = d_\rho$, and since traces are similarity-invariant, $\chi_\rho(\tau)$ is the same for any transposition τ. Thus by Theorem 3.26,

$$\hat{\mu}(\rho) = \left(\frac{1}{l} + \frac{(l-1)\chi_\rho(\tau)}{ld_\rho}\right) I_{d_\rho}. \tag{3.34}$$

Furthermore, each transposition is its own inverse, so μ is also symmetric, and by Proposition 3.24, $\hat{\mu}(\rho) = (\hat{\mu}(\rho))^*$. Hence Corollary 3.22 gives us that

$$\|\mu^{(k)} - U\|_{TV} \leq \frac{1}{2} \left(\sum_{\rho \neq \rho_{triv} \in \hat{G}} d_\rho^2 \left(\frac{1}{l} + \frac{(l-1)\chi_\rho(\tau)}{l d_\rho} \right)^{2k} \right)^{1/2} . \tag{3.35}$$

To convert this into a more functional bound, we would need to delve deeply into the properties of $\chi_\rho(\tau)$. In [2], Diaconis did so and proved that after $O(l \log l)$ shuffles, all permutations of the cards are approximately equally likely.

References

[1] E. Behrends, *Introduction to Markov Chains (with Special Emphasis on Rapid Mixing)*, Vieweg Verlag, Braunschweig/Wiesbaden, 2000.

[2] P. Diaconis, *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, CA, 1988.

[3] D.A. Levin, Y. Peres, and E.L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, Providence, RI, 2008.

[4] B. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, Springer-Verlag, New York, 2010.

E-mail address: shanshand@math.upenn.edu