Random Walks on Groups

Eshaan Nichani

December 10, 2019

Abstract

This paper focuses on applications of representation theory to the study of random walks on finite groups [1, 2]. We are interested in analyzing the *mixing time* of a random walk - given a group and a random walk on this group, how many steps are required until our final distribution is "close" to uniform? I summarize the general technique used to bound the mixing time of a random walk, and show that the mixing time of the transposition random walk on the symmetric group S_n is exactly $\frac{1}{2}n \log n$.

1 Introduction

Random walks on graphs show up in a variety of fields in mathematics and science, arising from numerous physical processes, or used in many computational algorithms. It is known that under some simple conditions all random walks, or Markov Chains, have a stationary distribution; one interesting question is thus to calculate the number of steps needed to get within ϵ of the stationary distribution, also known as the *mixing time*. General techniques bound the mixing time of a Markov chain by analyzing the spectral gap of the transition matrix, which is a function of the second largest eigenvalue.

However, we can obtain much tighter bounds on the mixing when we restrict random walks to be on groups. For example, consider the question of how many shuffles are needed to fully shuffle a deck of cards. This can be viewed as a random walk on the symmetric group S_{52}. The study of random walks on the symmetric group is also important in the study of Markov-Chain Monte-Carlo (MCMC) algorithms, where physical processes can be simulated or samples from a probability distribution can be generated based on a sequence of random transpositions. Now that our walks are restricted to groups, we can rely on results in representation theory to bound the mixing time.

In this paper, I begin by formalizing the notion of a random walk, as well as what it means for two distributions to be "close". I then review some representation theory of finite groups, and discuss the Fourier Transform. I next prove the Upper Bound Lemma, which shows how to apply representation theory to obtain bounds on mixing times, and present a simple example of a random walk on the group $\mathbb{Z}/p\mathbb{Z}$. Finally, I analyze the transposition random walk on the symmetric group S_n. By investigating the characters of representations on S_n, we will prove that exactly $\frac{1}{2}n \log n$ shuffles are needed [3]. This presentation follows that of Diaconis in [1].

2 Preliminaries

We must first define what it means to have a random walk on a group. Let G be a finite group.

Definition. *A function $P : G \to \mathbb{R}^{\geq 0}$ is a* **probability distribution** *if $\sum_{g \in G} P(g) = 1$.*

Assume that our random walk starts at the identity 1. Let $g^{(t)}$ be the position of our walk at time t. We advance to $g^{(t+1)}$ by sampling some g from our probability P, and letting $g^{(t+1)} = gg^{(t)}$.

Definition. *The* **convolution** *between two probabilities P, Q, denoted by $P * Q$, is given by*

$$(P * Q)(g) = \sum_{h \in G} P(gh^{-1})Q(h)$$

We see that $P*Q$ is also a probability distribution. If we define P^{*k} to be the distribution formed by taking the convolution between P and itself k times, we see that P^{*k} is simply the distribution of the random walk at time k. It is known via standard Markov Chain techniques that the uniform distribution $U(g) = \frac{1}{|G|}$ is a stationary distribution of a random walk on a group. Thus we want P^{*k} to be close to the uniform distribution.

Definition. *The* **total variation distance** *between distributions P and Q is defined as*

$$\|P - Q\| = \max_{A \subset G} |P(A) - Q(A)|.$$

The exact choice of a distance metric isn't very important, since many distances can be related to the total variation distance.

Proposition 1. *Define the L_1 distance between P and Q to be $\|P - Q\|_1 = \sum_{g \in G} |P(g) - Q(g)|$. Then $\|P - Q\| = \frac{1}{2}\|P - Q\|_1$.*

Proof. $|P(A) - Q(A)|$ is maximized when we choose $A = \{g : P(g) \geq Q(g)\}$. We then see that

$$\begin{aligned}
\|P - Q\|_1 &= \sum_{g \in G} |P(g) - Q(g)| \\
&= \sum_{g \in A} P(g) - Q(g) + \sum_{g \notin A} Q(g) - P(g) \\
&= \sum_{g \in A} P(g) - Q(g) + (1 - \sum_{g \in A} Q(g)) - (1 - \sum_{g \in A} P(g) \\
&= 2\sum_{g \in A} P(g) - Q(g) = 2\|P - Q\|.
\end{aligned}$$

□

Now we can ask ourselves the following question: Given some $\epsilon > 0$, for what value of k do we have

$$\|P^{*k} - U\| \leq \epsilon?$$

3 Representation Theory of Finite Groups

We next examine some classical results in the representation theory of finite groups, which will prove useful in our analysis. Let $\rho : G \to GL(V)$ be a representation with dimension d_ρ.

Since G is finite, we can characterize the irreducible representations of G.

Definition. *The* **Regular Representation** *of G is constructed as follows. Let V be a $|G|$-dimensional vector space with basis indexed by the elements of G, i.e. $\{e_g\}$ for $g \in G$. Then ρ_{reg} is defined on the basis vectors as $\rho(h)(e_g) = e_{hg}$.*

Proposition 2. *a) The character of the regular representation χ_{reg} satisfies $\chi_{reg}(1) = |G|$ and $\chi_{reg}(g) = 0$ for $g \neq 1$.*

b) Each irreducible representation ρ is contained in the regular representation with multiplicity equal to d_ρ.

Proof. For $g \neq 1$, $\rho(g)(e_h) = e_{gh} \neq e_h$, so $\chi_{reg}(g) = 0$. This proves a). To prove b), recall that we can find the multiplicity of ρ in ρ_{reg} by taking the inner product of the characters:

$$\langle \chi_{reg}, \chi \rangle = \frac{1}{|G|} \sum_{g \in G} = \chi_{reg}(g)\chi(g^{-1}) = \chi(1) = d_\rho.$$

□

One corollary is that there are finitely many irreducible representations of G, and if $\mathrm{Irr}(G)$ is the set of all irreducible representations, then $\sum_{\rho \in \mathrm{Irr}(G)} d_\rho^2 = |G|$.

Definition. *Given a probability P, its* **Fourier transform** *$\widehat{P}$ is given by $\widehat{P}(\rho) = \sum_{g \in G} P(g)\rho(g)$.*

Proposition 3. $\widehat{P * Q}(\rho) = \widehat{P}(\rho)\widehat{Q}(\rho)$

Proof. A straightforward computation:

$$\begin{aligned}
\widehat{P * Q}(\rho) &= \sum_g P * Q(g)\rho(g) \\
&= \sum_g \sum_h P(gh^{-1})Q(h)\rho(g) \\
&= \sum_{g'} \sum_{h'} P(g')Q(h')\rho(g'h') \\
&= \sum_{g'} \sum_{h'} P(g')Q(h')\rho(g)\rho('h') \\
&= \widehat{P}(\rho)\widehat{Q}(\rho)
\end{aligned}$$

□

The Fourier transform is useful as it allows us to turn convolutions into products. This requires us to work in the space of representations, and necessitates some way to go from representations back to a probability. This can be done using the Fourier Inversion formula:

Theorem 1. (Fourier Inversion) *Any function $f : G \to \mathbb{C}$ satisfies*

$$f(g) = \frac{1}{|G|} \sum_{\rho \in Irr(G)} d_\rho \operatorname{Tr}(\widehat{f}(\rho)\rho(g^{-1})).$$

Proof. Since the Fourier transform is linear, this equality is linear in f, and thus it suffices to prove the claim for functions of the form δ_h, where $\delta_h(g) = 1$ iff $h = g$ and $\delta_h(g)$ otherwise. We first see that:

$$\widehat{\delta_h}(\rho) = \sum_g \delta_h(g)\rho(g) = \rho(h).$$

Therefore

$$\begin{aligned}\frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \operatorname{Tr}(\widehat{\delta_h}(\rho)\rho(g^{-1})) &= \frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \operatorname{Tr}(\rho(h)\rho(g^{-1})) \\ &= \frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \chi(hg^{-1}) \\ &= \frac{1}{|G|} \chi_{reg}(hg^{-1}) = \delta_h(g),\end{aligned}$$

so the claim is true for δ_h and thus all f, as desired. □

We also have a theorem which relates inner products between functions and their Fourier transforms.

Theorem 2. (Plancherel's Formula) *For $f, h : G \to \mathbb{C}$,*

$$\sum_{g \in G} f(g)\overline{h(g)} = \frac{1}{|G|} \sum_{\rho \in Irr(G)} d_\rho \operatorname{Tr}(\widehat{f}(\rho)\widehat{h}(\rho)^*),$$

where A^* is the conjugate transpose of a matrix A.

Proof. The formula is linear in f, so it suffices to prove the claim for $f(g) = \delta_s(g)$ for some $s \in G$. Also, note that with appropriate choice of a basis we can make our representations unitary, namely that $\rho(g)^* = \rho(g^{-1})$ for all g. The LHS is just $\overline{h(s)}$, which by the Fourier transform equals

$$\begin{aligned}\overline{h(s)} &= \overline{\frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \operatorname{Tr}(\widehat{h}(\rho)\rho(s^{-1}))} \\ &= \frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \overline{\operatorname{Tr}(\widehat{h}(\rho)\rho(s^{-1}))} \\ &= \frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \operatorname{Tr}(\rho(s^{-1})^*\widehat{h}(\rho)^*) \\ &= \frac{1}{|G|} \sum_{\rho \in \mathrm{Irr}(G)} d_\rho \operatorname{Tr}(\rho(s)\widehat{h}(\rho)^*),\end{aligned}$$

which is the RHS, as desired. □

We can use the Plancherel Formula to bound the L_2 norm of f based on the norms of its Fourier coefficients:

Corollary.

$$\|f\|_2^2 = \frac{1}{|G|} \sum_{\rho \in Irr(G)} d_\rho \operatorname{Tr}(\hat{f}(\rho)\hat{f}(\rho)^*).$$

4 The Upper Bound Lemma

Now that we have developed some theory on the representations of finite groups, let us return to our original question. Given a probability P, how large should k be such that $\|P^{*k} - U\| \leq \epsilon$?

Theorem 3. (Upper Bound Lemma) *Let Q be a probability distribution, and let ρ_0 be the trivial representation. Then*

$$\|Q - U\|^2 \leq \frac{1}{4} \sum_{\rho \in Irr(G) \setminus \{\rho_0\}} d_\rho \operatorname{Tr}(\widehat{Q}(\rho)\widehat{Q}(\rho)^*)$$

Proof. Recall that the total variation distance is half the L_1 distance, which yields

$$4\|Q - U\|^2 = (\sum_{g \in G} |Q(g) - U(g)|)^2 \leq |G| \sum_{g \in G} (Q(g) - U(g))^2 = |G| \cdot \|Q - U\|_2^2,$$

where the inequality is due to Cauchy-Schwarz. By the corollary to the Plancherel Theorem, the RHS equals

$$\sum_{\rho \in \mathrm{Irr}} d_\rho \operatorname{Tr}\left((\widehat{Q - U}(\rho))(\widehat{Q - U}(\rho))^*\right) = \sum_{\rho \in \mathrm{Irr}} d_\rho \operatorname{Tr}\left((\widehat{Q}(\rho) - \widehat{U}(\rho))(\widehat{Q}(\rho)^* - \widehat{U}(\rho)^*)\right)$$

If $\rho = \rho_0$, then $\widehat{Q}(\rho_0) = \widehat{U}(\rho_0) = 1$. Otherwise, $\widehat{U}(\rho) = 0$. Therefore the sum becomes

$$\sum_{\rho \in \mathrm{Irr}(G) \setminus \{\rho_0\}} d_\rho \operatorname{Tr}(\widehat{Q}(\rho)\widehat{Q}(\rho)^*),$$

as desired. □

This means that we can use the upper bound lemma to give us the following bound:

$$\|P^{*k} - U\| \leq \frac{1}{4} \sum_{\rho \in \mathrm{Irr}(G) \setminus \{\rho_0\}} d_\rho \operatorname{Tr}\left(\widehat{P^{*k}}(\rho)\widehat{P^{*k}}(\rho)^*\right) = \frac{1}{4} \sum_{\rho \in \mathrm{Irr}(G) \setminus \{\rho_0\}} d_\rho \operatorname{Tr}\left(\widehat{P}(\rho)^k(\widehat{P}(\rho)^k)^*\right).$$

This gives us a strategy for showing when P^{*k} is close to U. We simply need to analyze when the Fourier transform of non-trivial representations is small.

5 Random walk on $\mathbb{Z}/p\mathbb{Z}$

We will first see this technique in action for a simple random walk on $\mathbb{Z}/p\mathbb{Z}$, the group of integers mod p for a prime p. Let P be the probability distribution such that $P(1) = P(-1) = \frac{1}{2}$ and $P(g) = 0$ elsewhere.

Theorem 4. *Let $k = cp^2, c > 1, p > 7$. Then*

$$\|P^{*k} - U\| \leq e^{-c\pi^2/2}.$$

This tells us that we need $k = \Theta(p^2)$ steps to approach the uniform distribution.

Proof. $\mathbb{Z}/p\mathbb{Z}$ is an Abelian group, which means that it has one-dimensional representations $\rho_j(x) = \exp(\frac{2\pi ixj}{p})$ for $0 \leq j \leq p-1$. The Fourier transform of ρ_j is then

$$\widehat{P}(\rho_j) = \frac{1}{2}\left(\exp\left(\frac{2\pi ij}{p}\right) + \exp\left(\frac{-2\pi ij}{p}\right)\right) = \cos\left(\frac{2\pi j}{p}\right).$$

Thus by the upper bound lemma:

$$\begin{aligned}
\|P^{*k} - U\|^2 &\leq \frac{1}{4}\sum_{j=1}^{p-1} d_{\rho_j} \operatorname{Tr}\left(\widehat{P}(\rho_j)^k(\widehat{P}(\rho_j)^k)^*\right) \\
&= \frac{1}{4}\sum_{j=1}^{p-1} \cos\left(\frac{2\pi j}{p}\right)^{2k} \\
&= \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}} \cos\left(\frac{\pi j}{p}\right)^{2k}.
\end{aligned}$$

Bounding this quantity is now simple calculus. We can use the fact that $\cos(x) \leq \exp(-\frac{x^2}{2})$ on the interval $[0, \pi]$. This gives us

$$\begin{aligned}
\|P^{*k} - U\|^2 &\leq \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}} \exp\left(-\frac{\pi^2 j^2 k}{p^2}\right) \\
&\leq \frac{1}{2}\exp\left(-\frac{\pi^2 k}{p^2}\right)\sum_{j=1}^{\infty}\left(-\frac{\pi^2(j^2-1)k}{p^2}\right) \\
&\leq \frac{1}{2}\exp\left(-\frac{\pi^2 k}{p^2}\right)\sum_{j=1}^{\infty}\left(-\frac{3\pi^2(j-1)k}{p^2}\right) \\
&= \frac{1}{2}\cdot\frac{\exp\left(-\frac{\pi^2 k}{p^2}\right)}{1-\exp\left(-\frac{3\pi^2 k}{p^2}\right)},
\end{aligned}$$

where we used that $j^2 - 1 \geq 3(j-1)$ for integers $j \geq 1$. Now if $k \geq p^2$, we can check that $2(1 - \exp\left(-\frac{3\pi^2 k}{p^2}\right)) > 1$, giving us $\|P^{*k} - U\|^2 \leq \exp(-\pi^2 c)$, as desired. □

We can use a similar method to lower bound the distance as well, using the following fact.

Definition. *Let $\mathbb{E}_P(f)$ denote the expectation of a function f under the probability distribution P, so*

$$\mathbb{E}_P(f) = \sum_{g \in G} P(g) f(g).$$

Proposition 4. *The total variation distance can be written as:*

$$\|P - Q\| = \frac{1}{2} \max_{\|f\|_\infty \leq 1} |\mathbb{E}_P(f) - \mathbb{E}_Q(f)|,$$

where the maximum is taken over all functions satisfying $\|f\|_\infty = \sup_{g \in G} |f(g)| \leq 1$.

Theorem 5. *If $p \geq 7, k = cp^2$, there is a lower bound of*

$$\|P^{*k} - U\| \geq e^{-c\pi^2/2 + ac/p^2}$$

for some constant a.

Proof. Let $f(g) = \cos \frac{\pi(p-1)g}{p}$. The expected value of f under U is 0, while the expected value under P^{*k} is

$$\sum_{g \in G} P^{*k}(g) \cos\left(\frac{\pi(p-1)g}{p}\right) = P^{*k}(\rho_{\frac{p-1}{2}}) = \cos^k\left(\frac{\pi(p-1)}{p}\right) = (-1)^k \cos^k\left(\frac{\pi}{p}\right).$$

Therefore

$$\|P^{*k} - U\| = \frac{1}{2}|\mathbb{E}_{P^{*k}}(f) - \mathbb{E}_U(f)| = \frac{1}{2}\left|\cos^k\left(\frac{\pi}{p}\right)\right| \geq \frac{1}{2}e^{-k\pi^2/2p^2 - O(k/p^4)},$$

via a Taylor series expansion. □

There are a couple interesting takeaways from this analysis. First, note that that P^{*k} becomes close to U very sharply around $k = n^2$. This is known as the *cutoff phenomenon*, where P^{*k} rapidly goes from being very far from the uniform distribution to very close to it. In general, upper bounds can be formed by showing that all non-trivial representations are small, while lower bounds can be formed by finding a function, usually corresponding to the slowest decreasing representation, such that P^{*k} and U differ in expected value when k is small. We can think of such representations as being "close" to the trivial representation.

Another important note is that since we're working in an Abelian group, all our representations are one-dimensional. In this case, our representation theory approach is simply an application of Fourier Analysis. This connects rather nicely to spectral graph theory, where the Fourier coefficients are simply the eigenvalues of the adjancency matrix of the Cayley graph. However, to fully unlock the power of representation theory, we must consider an example of a walk on a non-abelian group.

6 Random walk on S_n

One natural such example is shuffling a deck of n cards, which can be viewed as a random walk on S_n. One way to shuffle a deck of cards is through random transpositions, where we select two cards independently at random and swap the cards. This is known as the transposition random walk. We see that this walk has probability P satisfying:

$$P(g) = \begin{cases} 1 & \text{if } g = \mathrm{id} \\ \frac{2}{n^2} & \text{if } g = \tau \text{ is a transposition} \\ 0 & \text{otherwise} \end{cases}$$

The following theorem bounds the mixing time of this random walk:

Theorem 6. *Let $k = \frac{1}{2} n \log n + cn$ for some $c > 0$. Then there exists a constant a such that*

$$\|P^{*k} - U\| \leq ae^{-2c}.$$

In order to use the upper bound to prove this theorem, we need to first calculate the Fourier transform $\widehat{P}$ of a representation ρ. Since the transpositions form a conjugacy class of S_n, note that our probability distribution is constant on conjugacy classes (i.e. that $P(\eta\pi\eta^{-1}) = P(\pi)$ for any $\pi, \eta \in S_n$). We then see that

$$\begin{aligned} \rho(h^{-1})\widehat{P}(\rho)\rho(h) &= \sum_{g \in S_n} \rho(h^{-1})\rho(g)\rho(h)P(g) \\ &= \sum_{g} \rho(h^{-1}gh)P(g) \\ &= \sum_{g} \rho(h^{-1}gh)P(h^{-1}gh) = \widehat{P}(\rho) \end{aligned}$$

Therefore $\rho(h)\widehat{P}(\rho) = \widehat{P}(\rho)\rho(h)$ for all $h \in S_n$. Thus $\widehat{P}(\rho)$ is a G-invariant homomorphism, and since ρ is irreducible by Schur's Lemma we have that $\hat{P}(\rho) = cI$ for some constant c. We can calculate c by taking traces:

$$cd_\rho = \mathrm{tr}(\hat{P}(\rho)) = \sum_{g \in G} P(g)\chi_\rho(g) = \frac{1}{n}d_\rho + \frac{\binom{n}{2}}{n^2/2}\chi_\rho(\tau) = \frac{1}{n}d_\rho + \frac{n-1}{n}\chi_\rho(\tau),$$

where $\chi_\rho(1) = d_\rho$ and $\chi_\rho(\tau)$ is the character of a transposition. Therefore

$$c = \frac{1}{n} + \frac{n-1}{n} \cdot \frac{\chi_\rho(\tau)}{d_\rho}.$$

Definition. *Let $r(\rho)$ be the ratio $\frac{\chi_\rho(\tau)}{d_\rho}$.*

The upper bound lemma then yields

$$\begin{aligned}\|P^{*k}-U\|^2 &\leq \frac{1}{4}\sum_{\rho\in\mathrm{Irr}(G)\setminus\{\rho_0\}} d_\rho\|\widehat{P}(\rho)^k\|^2\\ &= \frac{1}{4}\sum_{\rho\in\mathrm{Irr}(G)\setminus\{\rho_0\}} d_\rho\cdot d_\rho c^{2k}\\ &= \frac{1}{4}\sum_{\rho\in\mathrm{Irr}(G)\setminus\{\rho_0\}} d_\rho^2\left(\frac{1}{n}+\frac{n-1}{n}r(\rho)\right)^{2k}\end{aligned}$$

This gives us a path to bounding the distance between P^{*k} and the uniform distribution. We just need to show that $r(\rho)$ is smaller than 1 for all non-trivial representations. To show this we must develop some theory on representations of the symmetric group.

7 Representations of the Symmetric Group

Many of the following facts are stated without justification, in the same order as presented in [1]. Some proofs are mentioned there, while others are discussed in [4].

Definition. *A* **partition** *of n is a tuple $\lambda = (\lambda_1, \dots, \lambda_m)$ of nonnegative integers such that $\lambda_1 \geq \cdots \geq \lambda_m$ and $\lambda_1 + \cdots + \lambda_m = n$.*

Definition. *A* **tableux** *associated with a partition λ is a diagram of squares such that the ith row has λ_i squares, and each square in the diagram is given a unique number from 1 to n. Two tableux are* **equivalent** *if they have the same row sets; an equivalence class of tableux is known as a* **tabloid**.

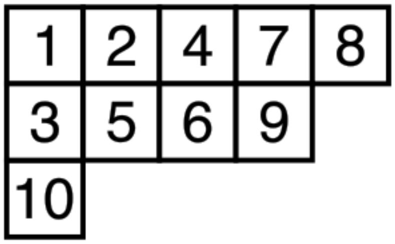

Figure 1: A tableux for $n = 10$.

Given a partition λ, there are $\frac{n!}{\lambda_1!\cdots\lambda_m!}$ tabloids corresponding to λ.

Definition. *We can define the representation ρ on the vector space M^λ as follows. Consider a basis $\{e_t\}$ where t runs over all tabloids of shape λ. Then ρ acts on M^λ as $\rho(\pi)(e_t) = e_{\pi(t)}$ for $\pi \in S_n$, where π acts on the tabloid by permuting the numbers in the squares.*

M^λ is not an irreducible representation, however, it has a unique irreducible subspace S^λ. Here are a few facts involving S^λ, as well as corollaries derived from these facts.

Fact 1. *There is a one-to-one correspondence between partitions of n and irreducible representatinos of S_n, where the partition λ corresponds to the representation on the subspace S^λ. Furthermore, the dimension of S^λ, denoted by d_λ, is the number of ways to place $1, \dots, n$ into a tableux corresponding to λ such that all rows and columns are increasing.*

Corollary. *Let $\lambda^* = (\lambda_2, \dots, \lambda_m)$ be a partition of $n - \lambda_1$. Then $d_\lambda \le \binom{n}{\lambda_1} d_{\lambda^*}$.*

Proof. There are $\binom{n}{\lambda_1}$ ways to choose the numbers in the first row, which must be in increasing order. There are at most d_{λ^*} ways to arrange the remaining numbers. □

Now we can use the values of the partition to bound the ratio r. Let ρ_λ be the representation corresponding to λ.

Fact 2.

$$r(\rho_\lambda) = \frac{1}{n(n-1)} \sum_{j=1}^{m} \left(\lambda_j^2 - (2j-1)\lambda_j\right)$$

Proof. The proof is quite involved and thus omitted. See [4] for a full proof. □

Corollary. (Monotonicity) *We say that, for partitions λ, λ', $\lambda \ge \lambda'$ if the tableux corresponding to λ can be acheived from the tableux corresponding to λ' by moving blocks up and to the right. Then, if $\lambda \ge \lambda'$, $r(\rho_\lambda) \ge r(\rho_{\lambda'})$.*

Proof. This can be deduced from fact 2 via simple algebra in the case where a single block is moved from one row to another. □

Fact 3. *Let λ^t be the partition corresponding to flipping the corresponding tableux across the down-right diagonal. Then $\chi_{\rho_\lambda} = -\chi_{\rho_{\lambda^t}}$.*

8 Proof of the Upper Bound

We can now use the tools in the previous section to simplify the upper bound.

Proposition 5. *Let λ be a partition such that $r(\rho_\lambda) \ge 0$ Then*

$$\left| \frac{1}{n} + \frac{n-1}{n} r(\rho_\lambda) \right| \le \begin{cases} 1 - \frac{2(\lambda_1+1)(n-\lambda_1)}{n^2} & \text{if } \lambda_1 \ge \frac{n}{2} \\ \frac{\lambda_1}{n} & \text{otherwise} \end{cases}$$

Proof. For the first case, see that $\lambda \le (\lambda_1, n - \lambda_1)$. Thus by monotonicity, we have that

$$\begin{aligned} r(\rho_\lambda) &\le r(\rho_{(\lambda_1, n-\lambda_1)}) \\ &= \frac{1}{n(n-1)} \left[\lambda_1^2 - \lambda_1 + (n-\lambda_1)^2 - 3(n-\lambda_1)\right] \\ &= 1 - \frac{2(\lambda_1+1)(n-\lambda_1)}{n(n-1)}, \end{aligned}$$

so

$$\left| \frac{1}{n} + \frac{n-1}{n} r(\rho_\lambda) \right| \le 1 - \frac{2(\lambda_1+1)(n-\lambda_1)}{n^2}.$$

Otherwise, we can bound $r(\rho_\lambda)$ directly:

$$\begin{aligned} r(\rho_\lambda) &= \frac{1}{n(n-1)}\sum_{j=1}^{m}(\lambda_j^2-(2j-1)\lambda_j) \\ &\le \frac{1}{n(n-1)}\sum_{j=1}^{m}\lambda_j(\lambda_j-1) \\ &\le \frac{\lambda_1-1}{n(n-1)}\sum_{j=1}^{m}\lambda_j = \frac{\lambda_j-1}{n-1}, \end{aligned}$$

so $\left|\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right| \le \frac{\lambda_1}{n}$, as desired. □

We want to find the sum

$$\sum_\lambda d_\lambda^2\left(\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right)^{2k},$$

where the sum is over all nontrivial partitions λ. We decompose this sum as

$$\begin{aligned} \sum_\lambda d_\lambda^2\left(\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right)^{2k} &\le \sum_{\lambda:r(\rho_\lambda)\ge 0} d_\lambda^2\left(\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right)^{2k} + \sum_{\lambda:r(\rho_\lambda)<0} d_\lambda^2\left(\frac{1}{n}-\frac{n-1}{n}r(\rho_\lambda^t)\right)^{2k} \\ &\le 2\sum_{\lambda:r(\rho_\lambda)\ge 0} d_\lambda^2\left|\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right|^{2k} \end{aligned}$$

Let $0<\alpha<\frac{1}{4}$ by a constant. We can split up our sum as

$$\le 2\sum_{\lambda:r(\rho_\lambda)\ge 0} d_\lambda^2\left|\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right|^{2k} = \sum_{j=1}^{\alpha n}\sum_{\lambda:r(\rho_\lambda)\ge 0,\lambda_1=n-j} d_\lambda^2\left|\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right|^{2k} \qquad (1)$$

$$+\sum_{j>\alpha n}\sum_{\lambda:r(\rho_\lambda)\ge 0,\lambda_1=n-j} d_\lambda^2\left|\frac{1}{n}+\frac{n-1}{n}r(\rho_\lambda)\right|^{2k} \qquad (2)$$

We first bound part (1), using the above corollary, and get that

$$(1) \le \sum_{j=1}^{\alpha n}\left(1-\frac{2j(n-j+1)}{n^2}\right)^{2k}\sum_{\lambda:\lambda_1=n-j} d_\lambda^2.$$

Now by corollary 2, $d_\lambda \le \binom{n}{\lambda_1}d_{\lambda^*}$, so

$$\sum_{\lambda:\lambda_1=l} d_\lambda^2 = \binom{n}{l}^2\sum_{\lambda^*} d_{\lambda^*}^2,$$

where this sum is over all partitions of $n-l$ (and thus irreducible representations of S_{n-l}). Since the sum of the dimensions squared is the order of the group, we have that the sum equals $\binom{n}{l}^2(n-l)! = \binom{n}{l}\frac{n!}{l!}$. Putting it together yields

$$(1) \leq \sum_{j=1}^{\alpha n} \left(1 - \frac{2j(n-j+1)}{n^2}\right)^{2k} \binom{n}{j} \frac{n!}{(n-j)!}.$$

Similarly, using corollary 4 we get that

$$(2) \leq \sum_{j>\alpha n} \left(1 - \frac{j}{n}\right)^{2k} \sum_{\lambda:\lambda_1 = n-j} d_\lambda^2 \leq \sum_{j>\alpha n} \left(1 - \frac{j}{n}\right)^{2k} \binom{n}{j} \frac{n!}{(n-j)!}.$$

We have now reduced our problem to bounding two (relatively) tractable sums, and it can be shown that by picking α to be slightly less than $\frac{1}{4}$, if $k = \frac{1}{2}n\log n + cn$, each quantity is $O(e^{-4c})$. ■

It turns out that this bound is tight; namely, that if $k = \frac{1}{2}n\log n - cn$ for some $c > 0$, then

$$\|P^{*k} - U\| \geq e^{-1} - e^{-\Omega(c)}.$$

This can be proven via standard probabilistic techniques by showing that for small enough k, P^{*k} assigns much more probability to a particular set A than expected by the uniform distribution. The set A is the set of permutations with at least one fixed point. Under the uniform distribution $U(A) \approx e^{-1}$, but for k much smaller than $\frac{1}{2}n\log n$, the probability of there being a fixed point after k steps of the random walk is exponentially close to 1.

References

[1] Persi Diaconis. *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics Lecture Notes - Monograph Series, 1988.

[2] Persi Diaconis. Random walks on groups: Characters and geometry. *Groups St. Andrews*, 2003.

[3] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(2):159–179, Jun 1981.

[4] G. D. James. *Representation Theory of the Symmetric Group.* Springer Lecture Notes in Mathematics 682, Springer-Verlag: New York, 1978.