

## Group Representations and Mixing Times

*Li-Chung Chen and Nicholas Eriksson*

The analysis of the mixing time of a Markov chain can sometimes be simplified if the underlying state space is a finite group and if the transition matrix has a special structure. In such cases, Fourier analysis can be used to bound the mixing time.

First we fix some notation. Given a finite group  $G$ , a *representation* of  $G$  is a group homomorphism  $\rho: G \rightarrow GL(V)$ , where  $V$  is a finite dimensional vector space over  $\mathbb{C}$ . The dimension of  $\rho$ , denoted as  $d_\rho$ , is the dimension of  $V$ . The *character* of  $\rho$  is defined as  $\chi_\rho(g) = \text{Tr } \rho(g)$  for  $g \in G$ . Below  $\sum_\rho^*$  denotes a sum over all nontrivial irreducible representations  $\rho$  of  $G$ .

We write  $U$  for the uniform probability distribution on  $G$ . If  $P$  is a probability on  $G$ , define the *Fourier transform* at a representation  $\rho$  by  $\widehat{P}(\rho) = \sum_{s \in G} P(s)\rho(s)$ . By the convolution  $P * Q$ , we mean the probability  $P * Q(s) = \sum_{t \in G} P(st^{-1})Q(t) = \sum_{t, u: tu=s} P(t)Q(u)$ . Note that  $\widehat{P * Q}(\rho) = \widehat{P}(\rho)\widehat{Q}(\rho)$ . Let  $P^{*k}$  denote the  $k$ -th convolution power.

The space of functions  $f: G \rightarrow \mathbb{C}$  has an inner product defined by  $\langle \phi | \psi \rangle = \frac{1}{|G|} \sum_{g \in G} \phi(g)\psi(g)^*$ . We will use the Plancherel formula, which says roughly that the inner product of two functions equals the “inner product” of their transforms.

**Proposition 1 (Plancherel Formula)** *If  $f, h$  are real valued functions on  $G$ , then*

$$\sum_{s \in G} f(s)h(s) = \frac{1}{|G|} \sum_{\rho} d_{\rho} \text{Tr}(\widehat{h}(\rho)\widehat{f}(\rho)^*).$$

For a probability distribution  $P$  on  $G$ , define an associated Markov chain on  $G$  in which we transition from  $s$  to  $t$  with probability  $P(s^{-1}t)$ . In other words, if we are at state  $s$ , pick a random  $g$  according to  $P$  and move to state  $sg$ . This Markov chain is doubly stochastic, so if it is ergodic then the stationary distribution is uniform. Observe that if we start at the identity  $e \in G$ , then the distribution of the Markov chain after  $k$  steps is exactly  $P^{*k}$ . By symmetry we can assume that we start at the identity. Hence good estimates on the variation distance  $\|P^{*k} - U\|$  will allow us to estimate the mixing time. The following elegant bound of Diaconis and Shahshahani has led to many novel results.

**Theorem 2 (Upper Bound Theorem)** *Let  $P$  be a probability distribution on  $G$ . Then*

$$\|P^{*k} - U\|^2 \leq \frac{1}{4} \sum_{\rho}^* d_{\rho} \text{Tr}(\widehat{P}(\rho)^k(\widehat{P}(\rho)^k)^*).$$

**Sketch of proof:** Note that  $4\|P^{*k} - U\|^2 = (\sum_s |P^{*k}(s) - U(s)|)^2 \leq |G| \sum_s |P^{*k}(s) - U(s)|^2$  by Cauchy-Schwartz. Then we use the Plancherel formula to show that the last term equals  $\sum_{\rho}^* d_{\rho} \text{Tr}(\widehat{P^{*k}}(\rho)\widehat{P^{*k}}(\rho)^*)$ . Now use the fact that  $\widehat{P^{*k}}(\rho) = \widehat{P}(\rho)^k$ . ■

If  $P$  is constant on conjugacy classes of  $G$ , then the above bound simplifies dramatically.

**Corollary 3** *Suppose  $P$  is a probability distribution that is constant on conjugacy classes of  $G$ . On the  $i$ th conjugacy class, let  $P_i$  be the value of  $P$ ,  $n_i$  be the size of this conjugacy class, and  $\chi_\rho^i$  be the value of the character  $\chi_\rho$ . Let  $C_\rho = \frac{1}{d_\rho} \sum_i P_i n_i \chi_\rho^i$ . Then*

$$\|P^{*k} - U\|^2 \leq \frac{1}{4} \sum_{\rho}^* d_{\rho}^2 C_{\rho}^{2k}.$$

**Sketch of proof:** We can show that  $\widehat{P}(\rho) = C_{\rho} I_{d_{\rho}}$  using Schur's lemma. Thus  $\widehat{P}(\rho)^k (\widehat{P}(\rho)^k)^* = C_{\rho}^{2k} I_{d_{\rho}}$  and has trace  $d_{\rho} C_{\rho}^{2k}$ . Now substitute this into the above theorem. ■

To assess the usefulness of these theorems, we apply them to two Markov chains discussed in lecture.

**Example 1: Random Walk on the Hypercube.** The hypercube is the group  $\mathbb{Z}_2^n$ , the group of length- $d$  0-1 strings with coordinate-wise mod 2 addition. Define  $P$  to be the distribution with value  $\frac{1}{n+1}$  at  $0 \dots 0, 10 \dots 0, 010 \dots 0, \dots, 0 \dots 01$  and with value 0 everywhere else. The random walk generated by  $P$  corresponds to staying put or moving to one of  $n$  nearest neighbors, each with probability  $\frac{1}{n+1}$ .

**Proposition 4** *For the above random walk on  $\mathbb{Z}_2^n$  with  $k = \frac{1}{4}(n+1)(\log n + c)$ , we have*

$$\|P^{*k} - U\|^2 \leq \frac{1}{2}(e^{e^{-c}} - 1).$$

Thus  $\tau_{mix} \leq \frac{1}{4}(n+1)(\log n + c)$  for some  $c$ . In fact  $\tau_{mix} \sim \frac{1}{4}n \log n$ .

**Sketch of proof:** The irreducible representations of  $\mathbb{Z}_2^n$  are one-dimensional. They are indexed by  $\mathbb{Z}_2^n$  and given by  $\rho_x(y) = (-1)^{x \cdot y}$ , where  $x \cdot y \in \mathbb{Z}_2$  is the usual dot product. Thus for  $x \in \mathbb{Z}_2^n$ ,  $\widehat{P}(x) = \sum_y (-1)^{x \cdot y} P(y) = 1 - \frac{2w(x)}{n+1}$ , where  $w(x)$  is the number of ones (or weight) of  $x$ . The trivial representation is  $\rho_{0 \dots 0}$ , so Theorem 2 gives  $\|P^{*k} - U\| \leq \frac{1}{4} \sum_{x \neq 0 \dots 0} \widehat{P}(x)^{2k} = \frac{1}{4} \sum_{j=1}^n \binom{n}{j} \left(1 - \frac{2j}{n+1}\right)^{2k}$ , which can be shown to be at most  $\frac{1}{2}(e^{e^{-c}} - 1)$ . ■

**Example 2: Generating a Random Permutation by Random Transpositions.** Define  $P$  on  $S_n$  to be the distribution with  $P(\text{id}) = \frac{1}{n}$ ,  $P(\tau) = \frac{2}{n^2}$  for transpositions  $\tau$ , and  $P(\pi) = 0$  otherwise. If  $S_n$  consists of orderings of  $n$  cards in a deck, then a step of the random walk is to choose  $s, t \in \{1, \dots, n\}$  u.a.r. independently and then swap the  $s$ th and  $t$ th card.

**Proposition 5** *For the above random walk on  $S_n$ , let  $k = \frac{1}{2}n \log n + cn$ . For  $c > 0$  and a fixed constant  $a$ ,*

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

Thus  $\tau_{mix} \leq \frac{1}{2}n \log n + cn$  for some  $c$ . In fact  $\tau_{mix} \sim \frac{1}{2}n \log n$ .

**Sketch of proof:** Since  $P$  is constant on conjugacy classes, we apply Corollary 3 to get the bound  $\|P^{*k} - U\|^2 \leq \frac{1}{4} \sum_{\rho}^* d_{\rho}^2 \left(\frac{1}{n} + \frac{n-1}{n} r(\rho)\right)^{2k}$ , where  $r(\rho) = \chi_{\rho}(\tau)/d_{\rho}$ .

It is well-known that irreducible representations of  $S_n$  are indexed by *partitions* of  $n$ , where a partition  $\lambda = (\lambda_1, \dots, \lambda_m)$  is a nonincreasing sequence of positive integers whose sum is  $n$ . For shorthand, let  $\lambda$  also denote its corresponding irreducible representation. The Frobenius character formula implies that  $r(\lambda) = \frac{1}{n(n-1)} \sum_j (\lambda_j^2 - (2j-1)\lambda_j)$ . This implies that  $r(\lambda)$  is at most  $1 - \frac{2(\lambda_1+1)(n-\lambda_1)}{n^2}$  if  $\lambda_1 \geq n/2$  and is at most  $\frac{\lambda_1}{n}$  for all  $\lambda$ .

For the proof, first we show that  $\sum_{\lambda}^* d_{\rho}^2 \left(\frac{1}{n} + \frac{n-1}{n} r(\lambda)\right)^{2k} \leq 2 \sum_{\lambda: r(\lambda) \geq 0}^* d_{\rho}^2 \left(\frac{1}{n} + \frac{n-1}{n} r(\lambda)\right)^{2k}$ . Then we split this sum into two parts, according to whether  $\lambda_1$  is small or large. To each part we apply the above bounds for  $r(\lambda)$  and also the inequality  $\sum_{\lambda: \lambda_1 = \ell} d_{\lambda}^2 \leq \binom{n}{\ell} \frac{n!}{\ell!}$ . After more analysis, we can show the desired bound on  $\|P^{*k} - U\|$ . ■

The upper bound theorem gives the exact asymptotic bound for the mixing time in these two examples. For the hypercube, the bound given by the coupling argument in lecture is off by a factor of 2. For  $S_n$ , the bound given by the simple coupling argument in lecture is  $O(n^2)$  and a better coupling argument is not known. Thus the representation theory method gives better bounds than coupling for the mixing times in these two examples, although at the cost of a complex mathematical analysis for  $S_n$ .

We now interpret Corollary 3 in terms of eigenvalues of the transition matrix, thus giving us some intuition about why the group representation method gives such good bounds for the mixing times.

Suppose  $P$  is a probability distribution that is constant on conjugacy classes of  $G$ . Let  $\mathbb{C}(G)$  be the group algebra of  $G$ . Let  $M: \mathbb{C}(G) \rightarrow \mathbb{C}(G)$  be the linear map whose matrix with respect to the standard basis,  $\{1 \cdot g : g \in G\}$ , is the transition matrix of the random walk determined by  $P$ . Let  $\mathbb{C}(G) = \bigoplus_{\rho} V_{\rho}$  be the isotypic decomposition of the left regular representation of  $\mathbb{C}(G)$ .

**Theorem 6** *Let  $C_{\rho}$  be as defined in Corollary 3. Then  $M$  maps each  $V_{\rho}$  into itself. Furthermore,  $M|_{V_{\rho}} = C_{\rho} I_{d_{\rho}}$ . Therefore,  $M$  has eigenvalue  $C_{\rho}$  with multiplicity  $\dim(V_{\rho}) = d_{\rho}^2$ , and  $M$  has an orthonormal basis of eigenvectors.*

**Sketch of proof:** It can be shown that  $M$  is just left multiplication by  $\sum_{s \in G} P(s)s$  in  $\mathbb{C}(G)$ . Because left multiplication by  $s \in G$  maps  $V_{\rho}$  to itself, so does  $M$ . Now  $V_{\rho}$  is a direct sum of  $d_{\rho}$  copies of  $\rho$ , and  $M$  acts as multiplication by  $\widehat{P}(\rho)$  (with respect to some basis) in each copy. By Schur's lemma,  $\widehat{P}(\rho) = C_{\rho} I_{d_{\rho}}$ , so  $M$  acts as multiplication by  $C_{\rho}$  on  $V_{\rho}$ . Because the isotypic subspaces  $V_{\rho}$  are pairwise orthogonal, we can find an orthonormal basis of eigenvectors. ■

Therefore, the bound given by Corollary 3 is actually a linear combination of power of the eigenvalues. Since all eigenvalues are used, we might expect to get a fairly tight bound that is better than the one given by the eigenvalue gap, which only uses the second largest eigenvalue.

Because of the tight bounds in the above examples, we are led to hope that there are many problems where we can use the upper bound theorem. Diaconis [D88] gives several other examples, including random walks on cyclic and affine groups with applications to pseudo random number generators.

Unfortunately, the setting of a random walk on a group is fairly special, basically requiring a high level of symmetry in the Markov chain. If the group is non-abelian, the simplifying assumption that  $P$  is constant on conjugacy classes is even more special.

However, these methods can be generalized to many other situations with some care. Infinite groups generally have a more complicated representation theory, but there are elegant results for many classes of groups that allow a generalization of the upper bound theorem. For example, if  $G$  is a compact Lie group, many random walks converge to Haar measure in total variation distance and the representation theory is very well developed. For example, Rosenthal [R94] has analyzed methods of producing random rotations in  $SO(n, \mathbb{R})$  using the Weyl character formula. For an overview of several other uses, see [D86].

The upper bound lemma can also be generalized to spaces with a transitive group action and a probability which is induced by one on the group, see [D88, Chapter 3F].

We have shown that the upper bound theorem and related Fourier analysis techniques give a very powerful method of understanding random walks on those finite groups whose representation theory takes a particularly nice form (e.g., abelian and symmetric groups). When the probability distribution is concentrated on a union of conjugacy classes, we have seen that the formulas become particularly nice. We conclude that the group representation method is a viable way to estimate the mixing time in a limited but substantial number of Markov chains.

## References

- [D86] P. DIACONIS, “The cutoff phenomenon in finite Markov chains,” *Proc. Nat. Acad. Sci. U.S.A.* **93**, pp. 1659-1664.
- [D88] P. DIACONIS, *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes, Monograph Series 11, 1988.
- [DS81] P. DIACONIS and M. SHAHSHAHANI, “Generating a Random Permutation with Random Transpositions,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **57**, pp. 159-179.
- [R94] J.S. ROSENTHAL, “Random Rotations: Characters and Random Walks on  $SO(N)$ ”, *Ann. Prob.* **22**, pp. 398-423.