RANDOM WALKS IN GROUPS

KEVIN XU

Abstract. This paper goes over random walks over groups. We first start with definitions of groups and other related material, and then transition into convolutions of distributions and Fourier transforms. Lastly, we use this to prove a nice upper bound on the mixing time, which we then illustrate in an example of the lazy random walk on Z/nZ. The paper mostly follows [Dia87], though hopefully is easier to follow. Huge thanks to Simon, whose help otherwise would not have made this paper possible.

1. Preliminaries

A paper on random walks in groups would be incomplete if it didn't start out with a good introduction. Here we define the basics of a group.

Definition 1.1. A *group* G is a set closed under a binary operation satisfying

(1) $\exists e \in G \mid eg = ge = g \quad \forall g \in G$
(2) $\exists h \in G \mid gh = hg = e \quad \forall g \in G$
(3) $f(gh) = (fg)h \quad \forall f, g, h \in G$

.

These are the *identity*, *inverse*, and *associative* properties. Subsets of groups that are also groups called *subgroups*. Because subgroups are closed under their operation, we can split a group into *cosets* of a subgroup H, where each coset is gH for elements $g \in G$.

In fact, gH is actually a left coset, and Hg would be a right coset. However, we usually talk about cosets of *normal subgroups*, which is when gH and Hg are the same set. Note that this is equivalent to H being closed under *conjugation* (i.e subgroup $H \in G$ is normal iff $ghg^{-1} \in H$ for all $g \in G, h \in H$). The set of cosets of a normal subgroup forms a group called the *quotient group*.

Groups also contain a *generating set*, which is the smallest subset such that every element can be expressed as the combination of finitely many elements in the set and their inverses. Groups are usually represented by their generating set.

Since groups are essentially objects, we can define functions that send elements in one group to another. But we only want functions that will preserve the group structure, such as sending the identity to the identity and the generating set to the generating set.

Definition 1.2. For groups G, H, a *homomorphism* is a function $f : G \to H$ satisfying

$$f(g_1 g_2) = f(g_1) f(g_2) \quad \forall g_1, g_2 \in G.$$

When we usually think of probability, we think of events such as rolling a die or flipping a coin, and associate a value from 0 to 1 to that event. We can do the same with groups, by using a function to associate a probability to each element of a group.

Definition 1.3. A *probability distribution* is a function $P\colon G \to [0,1]$ such that

$$\sum_{g\in G} P(g) = 1.$$

For example, rolling a fair die is a distribution on $\mathbb{Z}/6\mathbb{Z}$, with equal probabilities for each element. A game where you can move forward or backward at every step with equal probability is best represented by $\mathbb{Z}^+$ with distribution $P(1) = P(-1) = \frac{1}{2}$ and $P(X) = 0$ otherwise. In these non-uniform distributions, the elements that have probability zero are not important, so in many cases we want to work with the *support* of P.

Definition 1.4. The *support* of a distribution P is the set of elements of G with nonzero probability.

Now just like how we can construct a sequence of coin flips or dice rolls, we can construct a sequence by starting from the identity and then every step multiplying by an element g with probability $P(g)$.

Definition 1.5. A *random walk* on a group G along with its associated probability distribution P is a sequence of elements starting from e such that the probability of g moving to gh is $P(h)$ for $g, h \in G$.

A random walk is *irreducible* if the probability of getting to h from g is positive for all $g, h \in G$, and it is *aperiodic* if for time $t > T$ where T is fixed the probability of reaching a state is positive. If both conditions are satisfied, the random walk is *ergodic*.

Example. Shuffling cards is a random walk on S_n, the symmetric group. The probability distribution depends on what method is used to shuffle the cards.

One easy albeit inefficient way to shuffle a deck is by transposition, or switching the positions of two not necessarily distinct cards in the deck. This can be represented by the distribution

$$P(\pi) = \begin{cases} \frac{1}{n} & \pi = e \\ \frac{2}{n^2} & \pi \text{ is a transposition} \\ 0 & \text{otherwise.} \end{cases}$$

In standard probability, as long as we repeat for a large amount of times, our sequence will "converge" to some value. For example, a fair coin will always converge to a 50-50 distribution after a large number of tries no matter what the first few flips are.

In the same way, as long as we make a large amount of steps in any random walk, we also approach a "stationary distribution" consisting of the probabilities of landing on each element. But finding the probability of being at element g after a long time k is quite hard in practice.

Definition 1.6. The *convolutions* of a distribution P are defined recursively by

$$P \star P(g) = \sum_h P(gh^{-1})P(h)$$

and

$$P^k(g) = P^{k-1}(g) \star P(g).$$

Note that $P^2(g)$ is the summing over all instances of first rolling element h and then rolling gh^{-1}. The

In many cases we want to find what the convolutions converge to. It turns out that ergodic random walks always have a unique stationary distribution. This distribution is uniform if the P is *symmetric*, i.e. $P(g) = P(g^{-1})\ \forall g \in G$. But when are random walks ergodic?

Theorem 1.7. *For a group G and its probability distribution P, the random walk will be ergodic iff the support of P cannot be contained by a proper subgroup or a coset of a proper normal subgroup of G.*

Proof. Let the support of P be Σ. If the random walk is irreducible, then starting from e we can arrive at any element g, so Σ generates G. But then Σ cannot be contained by a proper subgroup of G, as then due to closure the set generated by Σ must be contained within the subgroup. Conversely, if Σ cannot be contained within a proper subgroup H of G, then since the set generated by Σ is a subgroup, it must be G. Then we can always get from $g \to h$ by multiplying through the corresponding generators in $g^{-1}h$, so the random walk is irreducible.

Assume that the random walk is irreducible. Now suppose that Σ is contained by some coset gH. Then after the nth step we will arrive at some $(gh_1)(gh_2)\cdots(gh_n) = g^n(h_1\cdots h_n) \in g^nH$. Thus we can pick some element $g' \in G\backslash g^nH$, we can never get to g' in n steps, so the random walk is periodic. Conversely, suppose the random walk is periodic with period $m > 1$. In formal terms, this is

$$m = \gcd\{k \mid \xi_1\cdots\xi_k = e,\ \xi_i \in \Sigma\}.$$

Now define

$$\Sigma_i = \{\xi_1\cdots\xi_k \mid \xi_j \in \Sigma,\ k \equiv j \pmod m\}.$$

We prove that these sets do not overlap. Note that if $x \in \Sigma_i,\ y \in \Sigma_j$, then $xy \in \Sigma_{i+j}$, which also implies $x^{-1} \in \Sigma_{-i}$. Suppose we have $\xi_1\cdots\xi_p = \xi_1'\cdots\xi_q'$, and thus

$$e = \xi_1'\cdots\xi_q'\xi_p^{-1}\cdots\xi_1 - 1 \in \Sigma_{p(m-1)+q}.$$

Therefore $p(m-1)+q \equiv 0 \pmod m$ and $p \equiv q$. Finally, note that Σ_0 is a normal subgroup of G, so because $\Sigma \subset \Sigma_1$ it is contained in a coset of Σ_0. □

Mixing Times

Because mathematicians are greedy, we also want to find how many steps we need to get a close approximation of the stationary distribution.

Definition 1.8. *Total variation* measures how much a distribution deviates from its stationary distribution u after a certain amount of steps, and is calculated by

$$||P^k - u||_{TV} = \frac{1}{2}\sum_g \left|P^k(g) - \frac{1}{|G|}\right|.$$

When the total variation is small, commonly measured with $\frac{1}{4}$ or other small number, we say that the random walk is sufficiently *mixed*. The *mixing time*, is the minimum time k for which a random walk is mixed.

It is always nice to find the mixing time or find better bounds on it in a random walk in order to have more efficient processes. For example, Diaconis and Shahshahani proved in [DS80] that it takes 270 transpositions on a 52-card deck to ensure the total variation was

less than 0.01. In comparison, the popular riffle shuffle takes about 6 shuffles to ensure the same "randomness."

The mixing time is often very hard to compute, so we rely on bounds to get a rough estimate. One such bound uses Fourier analysis.

For a finite group G, we can define a *representation* ρ as a homomorphism $\rho\colon G \to \mathrm{GL}(V)$ that sends group elements to a matrix. Here, $\mathrm{GL}(V)$ is the general linear group, or the group of invertible matrices with dimension $\dim(V)$. Now we introduce *Fourier transforms*:

Definition 1.9. The *Fourier transform* of a distribution P on a finite group G at ρ is

$$\widehat{P}(\rho) = \sum_{g \in G} P(g)\rho(g).$$

For the uniform distribution, we have the following nice property:

$$\widehat{P}(\rho) = \begin{cases} 1 & \rho \text{ is trivial} \\ 0 & \text{otherwise} \end{cases}.$$

Here 1 and 0 refer to the identity and zero matrices respectively, and the trivial representation is the map that sends everything to the identity matrix. The proof for trivial ρ is trivial, while the other part relies on $\rho(g^{-1})\rho(g)$ being the identity.

They also play nicely with convolutions:

Proposition 1.10. *Fourier transforms convert convolutions to products:*

$$\widehat{P^n(\rho)} = \left(\widehat{P}(\rho)\right)^n.$$

Proof. In the case of $n = 2$, we have

$$\begin{aligned}
\widehat{P^2(\rho)} &= \sum_{g \in G} \left(\rho(g) \sum_{h \in G} P(gh^{-1})P(h) \right) \\
&= \sum_{h \in G} \sum_{g \in G} P(gh^{-1})\rho(gh^{-1})P(h)\rho(h) \\
&= \sum_{h \in G} \left(P(h)\rho(h) \sum_{gh^{-1} \in G} P(gh^{-1})\rho(gh^{-1}) \right) \\
&= \left(\sum_{h \in G} P(h)\rho(h) \right) \left(\sum_{gh^{-1} \in G} P(gh^{-1})\rho(gh^{-1}) \right) \\
&= \left(\widehat{P}(\rho)\right)^2.
\end{aligned}$$

The proof then follows inductively. □

Now for a matrix A we define A^* to be its conjugate transpose, or the matrix obtained by taking the transpose and then replacing each entry with its complex conjugate. For a real-valued Matrix, A^* will just be the transpose.

Note that A^*A must be positive semi-definite, as for any nonzero vector z we have $z^*A^*Az = (Az)^*(Az) = |Az|^2 \geq 0$. Thus we can let B be a matrix such that $BB = A^*A$, or in bad notation $B = \sqrt{A^*A}$. We say the trace of B is the *trace norm* of A, denoted as $||A||$.

Theorem 1.11. *There exists a function $f\colon G \to \mathbb{C}$ satisfying*

$$1)\ f(g) = \frac{1}{|G|}\sum_{\rho} \dim(V)\,\mathrm{Tr}\left(\widehat{f}(\rho)\rho(g^{-1})\right)$$

$$2)\ ||f||^2 = \frac{1}{|G|}\sum_{\rho} \dim(V)||\widehat{f}(\rho)||^2.$$

Proof. I claim that we can find this function. Let

$$f(g) = \begin{cases} 1 & g = h \\ 0 & \text{otherwise} \end{cases}$$

for some $h \in G$. Taking the Fourier transform, we have $\widehat{f}(g) = \rho(h)$, so

$$f(g) = \frac{1}{|G|}\sum_{p} d_p\,\mathrm{Tr}\left(\rho(hg^{-1})\right).$$

Since the trace of the regular representation is $|G|$ and 0 when hg^{-1} is the identity or not, the sum evaluates to 1 when $g = h$ and 0 otherwise, as desired.

We can think of f as a matrix corresponding to $f(g)$ for all $g \in G$. Therefore the trace norm is the trace of f^*f, or just the sum of the products between elements and conjugates. In other words,

$$||f(g)||^2 = \sum f_1(g)\overline{f_2}(g)$$

where f_1, f_2 are permutations on f. Now fix $f_2 = f$ as defined above, so we want to prove

$$f_1(h) = \frac{1}{|G|}\sum_{\rho} d_\rho\,\mathrm{Tr}\left(\widehat{f_1}(\rho)\rho^*(h)\right)$$

which follows from the above equation. □

Using this theorem, we then prove an upper bound for the mixing time:

Corollary 1.12. *For a probability distribution P on a finite group G,*

$$4||P^k - u||^2 \leq \sum_{\rho} \dim(V)||\widehat{P}(\rho)||^{2k}.$$

Proof. From our definition and then Cauchy-Schwartz, we have

$$4||P^k - u||^2 = \left(\sum_g \left|P^k(g) - u(g)\right|\right)^2 \leq |G|\sum_g \left|P^k(g) - u(g)\right|^2.$$

We can then substitute the above theorem and then apply Plancherel's Theorem (integral is taken over all elements of G) which gives us

$$4||P^k - u||^2 \leq \sum_{\rho\neq 1} d_p\,\mathrm{Tr}\left(\widehat{P}(\rho)^k\widehat{P}(\rho)^{k*}\right).$$

The proof then follows from noticing that $||AB|| \leq ||A|| \cdot ||B||$. □

We can now compute the mixing times of some Markov chains using this theorem. Suppose we have the lazy random walk on $\mathbb{Z}/n\mathbb{Z}$ where $P(0) = \frac{1}{2}$ and $P(k) = P(-k) = \frac{1}{4}$ for k, n relatively prime. The irreducible representations are one-dimensional and must be given by $\rho_a(g) = e^{2\pi iag/n}\rho(e)$, so then the Fourier transform is

$$\widehat{P}(\rho_a) = \sum_g P(g)\rho_a(g) = \frac{1}{2} + \frac{1}{4}e^{2\pi iak/n} + \frac{1}{4}e^{-2\pi iak/n} = \frac{1}{2} + \frac{1}{2}\cos(2\pi ak/n).$$

Plugging in the above corollary, we have

$$4||P^l - u|| \leq 2^{-2l}\sum_{a=1}^{n-1}(1 + \cos(2\pi ak/n))^{2l} = 2^{-2l}\sum_{a=1}^{n-1}\cos^{4l}(\pi ak/n).$$

Approximating with $\cos x \leq e^{-x^2/2}$ (which holds as long as $k \leq \frac{n}{2}$), we arrive at

$$\begin{aligned}
4||P^l - u|| &\leq 2^{-2l}\sum_{a=1}^{\infty} e^{-2\pi^2a^2k^2l/n^2} = 2^{-2l}e^{-2\pi^2k^2l/n^2}\sum_{a=1}^{\infty} e^{-2\pi^2(a^2-1)k^2l/n^2} \\
&\leq 2^{-2l}e^{-2\pi^2k^2l/n^2}\sum_{a=1}^{\infty} e^{-6\pi^2ak^2l/n^2} \\
&= 2^{-2l}e^{-2\pi^2k^2l/n^2}\cdot\frac{e^{-6\pi^2k^2l/n^2}}{1 - e^{-6\pi^2k^2l/n^2}} = \frac{2^{-2l}e^{-8\pi^2k^2l/n^2}}{1 - e^{-6\pi^2k^2l/n^2}}.
\end{aligned}$$

If $l \geq \frac{n^2}{k^2}$, then clearly this expression is less than 1. Therefore the walk becomes well-mixed after at least $\frac{n^2}{k^2}$ steps.

References

[Dia87] Persi Diaconis. *Random Walks on Groups: Characters and Geometry*. Stanford University, 1987.

[Dia12] Persi Diaconis. *Group Representations in Probability and Statistics*. Harvard University, 2012.

[DS80] Persi Diaconis and Mehrdad Shahshahani. *Generating a random permutation with random transpositions*. Stanford University, 1980.

[Out15] Nolan Outlaw. *Markov Chains, Random Walks, and Card Shuffling*. Illinois Institute of Technology, 2015.