17. Random walks on finite groups I

April 17.

In the next several sections, we will discuss applications of projection theory to different areas. First we will discuss random walks on finite groups. Then we will discuss the distribution of orbits in homogeneous dynamics.

We here apply projection theory to studying the behavior of random walks on a finite group. Let G be a finite group and μ be a probability measure on G. A random walk starting at g_0 is defined as a sequence of random variables $(g_n)_{n\geq 0}$ such that $g_{n+1} = g_n g$ with probability $\mu(g)$. Essentially, at every step, a random element is chosen from G using μ, and then the current state is right multiplied by the chosen element. The guiding question is how evenly distributed the random walk is after K steps. We now develop several formal definitions to phrase this question more precisely. First, define a convolution of functions $f_1, f_2 : G \to \mathbb{C}$ in the standard way:

$$f_1 * f_2(g) = \sum_{g_1, g_2 \in G : g_1 g_2 = g} f_1(g_1) f_2(g_2) \tag{31}$$

We now view the random walk as a Markov chain with transitions given by a linear operator T_μ defined as

$$T_\mu f = f * \mu \tag{32}$$

When f is viewed as a probabilty distribution of a state g_n, $T_\mu f$ gives the probability distribution of g_{n+1}. When the random walk starts at a state g_0, that is equivalent to starting with initial probability distribution δ_{g_0}. Then after one step the probability distribution is $T_\mu \delta_{g_0}$, so the probability of state $g_0 h$ is

$$T_\mu \delta_{g_0}(g_0 h) = \sum_{g_1 g_2 = g_0 h} \delta_{g_0}(g_1) \mu(g_2) \tag{33}$$

$$= \delta_{g_0}(g_0) \mu(h) = \mu(h) \tag{34}$$

After K steps, the probability distribution of the random walk position g_K is $T_\mu^K \delta_{g_0}$. This leads to the first question, which is to estimate the L^2 norm

$$||T_\mu^K \delta_{g_0} - \frac{1}{|G|}||_{L^2} \tag{35}$$

or alternatively, other L^p norms. The $1/|G|$ term is the average value of the distribution over all of G, so the L^p norms are measures of the regularity of the distribution. Since T_μ is a linear operator, we can approach this by examining the singular values of T_μ. The squares of the singular values are the eigenvalues of the matrix $T_\mu^T T_\mu$ where T_μ^T is the transpose. When T is symmetric, the singular values are the same as the eigenvalues, but in general they are different.

We first show the following lemma:

Lemma 17.1.

$$||T_\mu f||_{L^2} \leq ||f||_{L^2} \tag{36}$$

Proof.

$$\begin{aligned} T_\mu f(g) =& f * \mu(g) & (37)\\ =& \sum_{g_1 g_2 = g} f(g_1)\mu(g_2) & (38)\\ =& \sum_{g_2} f(g g_2^{-1})\mu(g_2) & (39)\end{aligned}$$

We then define the right multiplication operator R_g so that $R_g f(h) = f(hg^{-1})$ Then applying the triangle inequality and the translation invariance of the L^2 norm,

$$\begin{aligned} ||T_\mu f||_{L^2} =& ||\sum_{g_2} \mu(g_2) R_{g_2} f||_{L^2} & (40)\\ \leq& \sum_{g_2} \mu(g_2) ||R_{g_2} f||_{L^2} & (41)\\ \leq& ||f||_{L^2} & (42)\end{aligned}$$

□

Then since $T_\mu 1 = 1$, 1 is the largest singular value of T_μ. We now define the subspace

$$L^2(G)_0 = \{f \in L^2(G) : \langle f, 1\rangle = 0\}$$

where $\langle , \rangle$ is the standard inner product with the counting measure on G. We can then analyze the restriction of T_μ

$$T_\mu : L^2(G)_0 \to L^2(G)_0$$

This restriction quotients out the trivial singular value 1 and allows us to examine the next singular value, which governs the decay rate of the L^p norms. Denote $\sigma_1(T_\mu)$ as the largest singular value of T_μ restricted to $L^2(G)_0$. Then we can quantitatively express the decay of the L^2 norm in terms of the following proposition:

Proposition 17.2.

$$||T_\mu^K \delta_{g_0} - \frac{1}{|G|}||_{L^2} \leq |\sigma_1(T_\mu)|^K$$

Proof. Note that

$$\langle \delta_{g_0} - \frac{1}{|G|}, 1\rangle = 0$$

so

$$\delta_{g_0} - \frac{1}{|G|} \in L^2(G)_0$$

Then T_μ maps $\delta_{g_0} - \frac{1}{|G|}$ to $L^2(G)_0$, so the claim follows from the fact that the largest singular value of a linear operator is also its operator norm. □

This proposition leads to the second guiding question, which is to estimate $\sigma_1(T_\mu)$. The proposition shows that an estimate on $\sigma_1(T_\mu)$ is sufficient to give an estiamte on the decay of the L^2 norm. We additionally remark that since we are using the counting measure, the L^∞ norm is bounded by the L^2 norm, so this gives an estimate of the L^∞ norm as well.

We now examine the group $G = SL_2(\mathbb{F}_p)$ where p is prime. The case where μ is the uniform measure on a subset A of G was studied by Selberg. For convenience, define $T_A \equiv T_{\mu_A}$ to be the operator corresponding to the measure on A. In particular, Selberg studied the particular set

$$A = \left\{ \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm 1 & 1 \end{pmatrix} \right\}$$

which has four elements. Selberg essentially proved the following theorem about this case:

Theorem 17.3. *There exists a universal constant $c > 0$ so that for every prime p, then*

$$\sigma_1(T_A) \leq 1 - c \tag{43}$$

The theorem that Selberg actually proved is about the smallest eigenvalue of the Laplacian on a hyperbolic surface X_p whose geometry is closely related to $SL_2(\mathbb{F}_p)$ with the generating set A above. Using modern techniques such as Cheeger's inequality, it is not difficult to translate between Selberg's eigenvalue bound and the mixing bound in Theorem 17.3.

Before discussing the proof of Selberg's theorem, we recall the connection between mixing estimates and isoperimetric inequalities on graphs. For a finite group G and a subset $A \subset G$ define a graph $C(G, A) = (V, E)$ with set of vertices V indexed by G and an edge $(g_1, g_2) \in E$ if $g_1^{-1}g_2 \in A$, or equivalently there exists $a \in A$ such that $g_2 = g_1 a$. Therefore the nodes that are connected by edges are the nodes that can be connected by a single step of the random walk. Now for two subsets $S, T \subset V$, define

$$E(S, T) \equiv \{(g_1, g_2) \in S \times T : (g_1, g_2) \in E\}$$

or equivalently, $E(S, T) = E \cap S \times T$. We now consider the following proposition:

Proposition 17.4. *If S is a subset of G, then*

$$|E(S, S^c)| \geq (1 - \sigma_1(T_A))\frac{|A||S||S^c|}{|G|} \tag{44}$$

Proof. We first prove that

$$E(S, S^c)| = |A|\langle T_A 1_S, 1_{S^c}\rangle \tag{45}$$

which follows from the following computation:

$$T_A 1_S(g) = \frac{1}{|A|}\sum_{a \in A} 1_S(ga^{-1})$$

$$\langle T_A 1_S(g), 1_{S^c}\rangle = \sum_{g \in G}\frac{1}{|A|}\sum_{a \in A} 1_S(ga^{-1})1_{S^c}(g)$$

Note that $1_S(ga^{-1})1_{S^c}(g) = 1$ if $ga^{-1} \in S$ and $g \in S^c$, which is equivalent to the statement $(ga^{-1}, g) \in E(S, S^c)$, which shows equation 45. We then decompose 1_S into a constant and mean zero part as

$$1_S = \frac{|S|}{|G|} + 1(_{S-}\frac{|S|}{|G|})$$

Applying this decomposition to 1_{S^c} as well gives

$$\begin{aligned}\langle T_A 1_S, 1_{S^c}\rangle =& \langle T_A\left(\frac{|S|}{|G|} + 1_S - \frac{|S|}{|G|}\right), 1 - \frac{|S|}{|G|} + 1_{S^c} - (1 - \frac{|S|}{|G|})\rangle \\ =& \langle\frac{|S|}{|G|} + T_A\left(1_S - \frac{|S|}{|G|}\right), 1 - \frac{|S|}{|G|} + 1_{S^c} - (1 - \frac{|S|}{|G|})\rangle\end{aligned}$$

This then decomposes into the inner products of the constant and the non-constant terms. The inner product of the constant terms is

$$|G|\frac{|S|}{|G|}(1-\frac{|S|}{|G|}) = \frac{|S||S^c|}{|G|}$$

The inner product of the non-constant terms is

$$\langle T_A\left(1_S - \frac{|S|}{|G|}\right), 1_{S^c} - (1-\frac{|S|}{|G|})\rangle$$

Applying Proposition 17.2 and Cauchy Schwartz gives the upper bound

$$\sigma_1(T_A)||1_S - \frac{|S|}{|G|}||_{L^2}||1_{S^c} - (1-\frac{|S|}{|G|})||_{L^2} = \sigma_1(T_A)\frac{|S||S^c|}{|G|}$$

Then combining the terms from the constant and nonconstant parts gives

$$\langle T_A 1_S, 1_{S^c}\rangle \geq (1-\sigma(T_A))\frac{|S||S^c|}{|G|}$$

Multiplying by A and applying equation 45 then gives the desired result:

$$E(S,S^c) \geq (1-\sigma(T_A))\frac{|A||S||S^c|}{|G|}$$

□

Without loss of generality S can be chosen so that $|S| \leq |G|/2$. Then if $\sigma(T_A) \leq 1$

$$E(S,S^c) \gtrsim |S|$$

where the implicit constants depend on A. This property of a subset of vertices and its complement sharing a large number of edges is known as an expander graph. Note that when A is a subset of a proper subgroup H of G, then the set of elements generated by A is at most H. The distribution will therefore never become uniform after repeatedly applying T_A, which implies that $\sigma_1(T_A) = 1$.

The original proof of Selberg's theorem was difficult and relied on the Riemann hypothesis for curves over a finite field. Around 1990, Sarnak and Xue gave a more elementary proof (with slightly weaker bounds on the constants). We will discuss some of the ideas in that proof. The first idea has to do with the representation theory of the group $SL_2(\mathbb{F}_p)$/ Consider the following proposition:

Proposition 17.5. *If $\rho : SL_2(\mathbb{F}_p) \to U(d)$ is a nontrivial representation of $SL_2(\mathbb{F}_p)$ mapping to the unitary group with d dimensions, then $d \geq \frac{p+1}{2}$*

Proof. This proof relies on the existence of the elements

$$u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$v = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

These elements generate $SL_2(\mathbb{F}_p)$, and are tranposes of each other, so without loss of generality we assume that $\rho(u) \neq e$. u and v have the property that they are conjugate to powers of themselves. In particular:

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix} = \begin{pmatrix} 1 & a^2 \\ 0 & 1 \end{pmatrix} = u^{a^2}$$

The conjugates of v similarly are powers of v. Then because representations preserve conjugacy classes, $\rho(u)$ must be conjugate to $\rho(u)^{a^2}$. Since conjugate matrices have the same eigenvalues, then $\rho(u)$ and $\rho(u)^{a^2}$ must have the same set of eigenvalues. $\rho(u)$ has order p, so its eigenvalues must be roots of unity of order p, or equivalently of the form $e^{2\pi i n/p}$ for integer n. Then the eigenvalues of $\rho(u)^{a^2}$, and equivalently of $\rho(u)$, are of the form $e^{2\pi i a^2 n/p}$. Since this is true for arbitrary a, a single nontrivial eigenvalue $e^{2\pi i n/p}$ generates all eigenvalues corresponding to $a^2 n$ mod p. $\rho(u)$ is by hypothesis not the identity, so must have at least one eigenvalue not equal to 1. Since there are $\frac{p-1}{2}$ distinct nonzero quadratic residues (and 1 is an eigenvalue of $\rho(u)$), then $\rho(u)$ has at least $\frac{p+1}{2}$ distinct eigenvalues, and so has dimension at least $\frac{p+1}{2}$. This completes the proof. □

We now apply this proposition to prove a further proposition.

Proposition 17.6. *Let μ be a measure on $SL_2(\mathbb{F}_p)$. Then*

$$\sigma_1(T_\mu)^2 \frac{p+1}{2} \leq |SL_2(\mathbb{F}_p)| \; ||u||_{L^2}^2$$

In particular, since $|SL_2(\mathbb{F}_p)| \sim p^3$, this implies

$$\sigma_1(T_\mu) \lesssim p||u||_{L^2}$$

Proof. Note that $\sigma_i(T_\mu)^2$ is the ith eigenvalue of $T_\mu T_\mu^*$. Since $T_\mu T_\mu^*$ is a right action, its eigenspaces have a left G action $L_g f(h) = f(g^{-1}h)$, which is nontrivial except for the constant functions. Each action on an eigenspace induces a representation of $SL_2(\mathbb{F}_p)$, which is unitary because

$$\begin{aligned}\langle L_g f, h\rangle =& \sum_{\ell \in G} f(g^{-1}\ell)h(\ell)\\ =& \sum_{\ell \in G} f(\ell)h(g\ell)\\ =& \langle f, L_{g^{-1}}h\rangle\end{aligned}$$

Therefore the representation must have dimension at least $\frac{p+1}{2}$, so the singular values must value multiplicity at least $\frac{p+1}{2}$. Then because the Frobenius is invariant under unitary operations, and since $T_\mu T_\mu^*$ is symmetric it is diagonalizable by a unitary transformation:

$$\begin{aligned}\frac{p+1}{2}\sigma_1(T_\mu)^2 \leq & \sum_{i \text{ with multiplicity}} \sigma_i(T_\mu)^2\\ =& \sum_{g_1, g_2} |(T_\mu)_{g_1,g_2}|^2\\ =& \sum \mu(g_1 g_2^{-1})^2\\ =& |SL_2(\mathbb{F}_p)| \sum_g \mu(g)^2\\ =& |SL_2(\mathbb{F}_p)| \; ||\mu||_{L^2}^2\end{aligned}$$

□

Then returning to the case that $\mu = \mu_A$ for a subset A

$$||\mu_A||_{L^2}^2 = \frac{1}{|A|^2}|A| = \frac{1}{|A|}$$

This together with proposition 17.6 implies the following corollary:

Corollary 17.7.

$$\sigma_1(T_A)^2 \lesssim \frac{p^2}{|A|}$$

This bound is only nontrivial when $|A| \gtrsim p^2$. The bound is tight in the sense that there are sets A with $|A| \sim p^2$ and with $\sigma_1(T_A) = 1$. Indeed, if A is a proper subgroup of $SL_2(\mathbb{F}_p)$, then $\sigma_1(T_A) = 1$. The subgroup of upper triangular matrices in $SL_2(\mathbb{F}_p)$ has cardinality $\sim p^2$.

Therefore, this estimate implies that every proper subgroup of $SL_2(\mathbb{F}_p)$ has cardinality $\lesssim p^2$. We state this result as a corollary.

Corollary 17.8. *If H is a proper subgroup of $SL_2(\mathbb{F}_p)$ then $|H| \lesssim p^2$.*

Proof. If H is a proper subgroup, then $\sigma(T_H) = 1$, which implies that $p^2/|H| \gtrsim 1$. Multiplying both sides by $|H|$ gives the desired result. □

(Note that the order of $SL_2(\mathbb{F})_p$ is $p(p-1)(p+1)$, so this corollary is not a consequence of Lagrange's theorem.)

To get further bounds for $\sigma_1(T_A)$ we will need to take account of other features of A besides just the cardinality of A. We will explore how to do in the next lecture.

18.156 Projection Theory
Spring 2025