Aspects of Markov chains

Yuval Peres
Microsoft Research

Outline

Discrete isoperimetric inequality

For a graph $G = (V, E)$, define the *boundary* ∂A of a set $A \subset V$ as the set of edges with one end in A and the other in $V - A$.

Theorem 1.1 (Discrete isoperimetric inequality)

Let $A \subset \mathbf{Z}^d$ be a finite set, then

$$|\partial A| \geq 2d|A|^{\frac{d-1}{d}} .$$

Remark 1.2

Observe that the 2d constant in the inequality is the best possible as the example of the d-dimensional cube shows: If $A = [0, n)^d \cap \mathbf{Z}^d$, then $|A| = n^d$ and $|\partial A| = 2dn^{d-1}$.

Discrete Loomis and Whitney inequality

For every $1 \leq i \leq d$, define the projection $\mathcal{P}_i : \mathbf{Z}^d \to \mathbf{Z}^{d-1}$ simply as the function dropping the ith coordinate, i.e.,

$$\mathcal{P}_i(x_1, \ldots, x_d) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) .$$

Theorem 1.1 follows easily from the following lemma.

Lemma 1.3 (Discrete Loomis and Whitney inequality, 1949)

For any finite $A \subset \mathbf{Z}^d$,

$$|A|^{d-1} \leq \prod_{i=1}^{d} |\mathcal{P}_i(A)| .$$

Proof of Theorem 1.1

Proof of Theorem 1.1.

The important observation is that $|\partial A| \geq 2\sum_{i=1}^{d} |\mathcal{P}_i A|$. To see this, observe that any vertex in $\mathcal{P}_i(A)$ matches to a straight line in the ith coordinate direction which "stabs" A. Thus, since A is finite, to any vertex in $\mathcal{P}_i(A)$ you can always match two distinct edges in ∂A: the first and last edges on the straight line which intersects A. Using this and the arithmetic-geometric mean inequality we get

$$|A|^{d-1} \leq \prod_{i=1}^{d} |\mathcal{P}_i(A)| \leq \left(\frac{1}{d}\sum_{i=1}^{d} |\mathcal{P}_i(A)|\right)^d \leq \left(\frac{|\partial A|}{2d}\right)^d ,$$

as required. □

Entropy and conditional entropy

To prove Lemma 1.3, we introduce some notions of entropy.

Definition 1.4 (Entropy and conditional entropy)

Let X and Y be random variables that take values $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$, respectively. Denote $p(x) := \mathbb{P}[X = x]$ and $p(x, y) := \mathbb{P}[X = x, Y = y]$. Then,

- *The entropy of X is defined to be*

$$H(X) = \sum_{i=1}^{n} p(x_i) \log(1/p(x_i)) .$$

- *The conditional entropy $H(X \mid Y)$ of X given Y is defined as*

$$H(X \mid Y) = H(X, Y) - H(Y) = \sum_{x_i, y_j} p(x_i, y_j) \log(1/p(x_i, y_j)) - \sum_{y_j} p(y_j) \log(1/p(y_j)) .$$

Simple properties of entropy

Proposition 1.5

(i) *If X takes n values, then $H(X) \leq \log n$.*

(ii) *$H(X \mid Y) \leq H(X)$ and $H(X \mid Y, Z) \leq H(X \mid Z)$.*

Proof.

Note that $H(X) - \log n = \sum_{i=1}^{n} p(x_i) \log \frac{1}{np(x_i)}$. Plugging in the inequality $\log t \leq t - 1$, valid for all $t > 0$, we get

$$H(X) - \log n \leq \sum_{i=1}^{n} p(x_i) \left(\frac{1}{np(x_i)} - 1 \right) = 0 .$$

Part (ii) of the Proposition is proved similarly and is left as an exercise for the reader. □

Han-Shearer inequality

Our next step in proving Lemma 1.3 is a theorem of J. Shearer (see Han 1978 and Chung, Frankl, Graham and Shearer 1986).

Theorem 1.6 (Han-Shearer inequality)

Let $X_1, \ldots, X_d$ be random variables taking finitely many values, and let $S_1, \ldots, S_l \subset \{1, \ldots, d\}$ be sets with the property that any $j = 1, \ldots, d$ is a member of precisely r of the S_i's. Then

$$rH(X_1, \ldots, X_d) \leq \sum_{i=1}^{l} H(\{X_j : j \in S_i\}) .$$

Proof of Han-Shearer inequality

Proof of Theorem 1.6.

For any i, by definition, we have the telescoping sum

$$H(X_j : j \in S_i) = \sum_{j \in S_i} H(X_j \mid X_u : u \in S_i, u < j) \geq \sum_{j \in S_i} H(X_j \mid X_u : u < j) ,$$

where the last inequality follows from part (ii) of Proposition 1.5. Sum over i and recall that each j appears in precisely r of the S_i's to get

$$\sum_{i=1}^{l} H(X_j : j \in S_i) \geq \sum_{i=1}^{l} \sum_{j \in S_i} H(X_j \mid X_u : u < j)$$
$$= \sum_{j=1}^{d} rH(X_j \mid X_u : u < j) = rH(X_1, \ldots, X_d) ,$$

where the last equality follows from the definition of conditional entropy. □

Proof of discrete Loomis and Whitney inequality

Proof of Lemma 1.3.

Take random variables $X_1, \ldots, X_d$ such that $(X_1, \ldots, X_d)$ is distributed uniformly on A. It is clear that $H(X_1, \ldots, X_d) = \log |A|$, and by Proposition 1.5, $H(X_1, \ldots X_{i-1}, X_{i+1}, \ldots, X_d) \leq \log |\mathcal{P}_i(A)|$. Now use Theorem 1.6 on $X_1, \ldots, X_d$ and $S_i = \{1, \ldots, d\} - \{i\}$ to find that

$$(d-1) \log |A| \leq \sum_{i=1}^{d} \log |\mathcal{P}_i(A)| ,$$

as required. □

A lower bound on isoperimetric profile of Cayley graphs

We now present an argument which gives a lower bound to the isoperimetric profile of a Cayley graph based only on the rate of growth of the graph. This proof is attributed to Coulhon and Saloff-Coste. Denote by $B(n)$ the ball of radius n surrounding the identity and by $V(n)$ the number of vertices in $B(n)$.

Theorem 1.7

Let $G = (V, E)$ be a Cayley graph of degree d. Let $\phi(\ell) = \inf\{n : V(n) > \ell\}$, then for all finite $K \subset V$ such that $|K| \leq |V|/2$, we have

$$\frac{|\partial K|}{|K|} \geq \frac{1}{2\phi(2|K|)} .$$

Proof of Theorem 1.7

Proof.

Let $n = \phi(2|K|)$ and take the ball such that $V(n) \geq 2|K|$. Fix $x \in K$, then a uniform random $g \in B(n)$ has probability at least 1/2 to have $gx \notin K$. So the expected value of $|\{x \in K : gx \notin K\}|$ is at least $|K|/2$, hence there is some $g \in B(n)$ that achieves this. Now notice that if s is a generator, the number of vertices that leave K due to s acting on them is at most $|\partial K|$, i.e., $|\{x \in K : sx \notin K\}| \leq |\partial K|$. By the same token, if g is at distance at most n from the origin, then $|\{x \in K : gx \notin K\}| \leq n|\partial K|$, so we get

$$|K|/2 \leq n|\partial K| ,$$

which yields the result. □

Exercises

1. Prove part (ii) of Proposition 1.5.
2. Show that there exists a constant $C_d > 0$ such that if A is a subgraph of the box $\{0, \ldots, n-1\}^d$ and $|A| < \frac{n^d}{2}$, then

$$|\partial A| \geq C_d |A|^{\frac{d-1}{d}} .$$

Outline

Definition of Markov type 2

Recall that a Markov chain $\{Z_t\}_{t=0}^{\infty}$ with transition probabilities $p_{ij} := \mathbb{P}[Z_{t+1} = j \mid Z_t = i]$ on the state space $\{1, \ldots, n\}$ is *stationary* if $\pi_i := \mathbb{P}[Z_t = i]$ does not depend on t, and it is (*time*) *reversible* if $\pi_i p_{ij} = \pi_j p_{ji}$ for every $i, j \in \{1, \ldots, n\}$.

Definition 2.1 (Ball 1992)

Given a metric space (X, d) we say that X has Markov type 2 if there exists a constant $M > 0$ such that for every stationary reversible Markov chain $\{Z_t\}_{t=0}^{\infty}$ on $\{1, \ldots, n\}$, every mapping $f : \{1, \ldots, n\} \to X$ and every time $t \in \mathbf{N}$,

$$\mathbb{E} d(f(Z_t), f(Z_0))^2 \leq M^2 t \mathbb{E} d(f(Z_1), f(Z_0))^2 .$$

Real line has Markov type 2

Theorem 2.2

$\mathbf{R}$ *has Markov type* 2.

Trees have Markov type 2

Recall that any tree with edge weights defines a metric on the vertices: the distance between two vertices in a weighted tree is the sum of the weights along the unique path between the vertices.

Theorem 2.3 (Naor, P., Schramm and Sheffield 2004)

Weighted trees have uniform Markov type 2.

Some elementary facts on transition matrix

Let $P = (p_{ij})$ be the transition matrix of the Markov chain. Observe that time reversibility is equivalent to the assertion that P is a self-adjoint operator in $L^2(\pi)$. This is because

$$\langle Pf, g \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \pi(i) f(j) g(i) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ji} \pi(j) f(j) g(i) = \langle f, Pg \rangle .$$

This in turn implies that $L^2(\pi)$ has an orthogonal basis of eigenfunctions of P with real eigenvalues. Since P is a stochastic matrix,

$$\max_i \sum_{j=1}^{n} p_{ij} |f(j)| \leq \max_i |f(i)| ,$$

which implies $\|Pf\|_\infty \leq \|f\|_\infty$, and thus if λ is an eigenvalue of P then $|\lambda| \leq 1$.

Proof of Theorem 2.2 I

Proof. We prove the statement with constant $M = 1$. Note that

$$\mathbb{E}d(f(Z_t), f(Z_0))^2 = \sum_{i,j} \pi_i p_{ij}^{(t)} [f(i) - f(j)]^2 = 2\langle (I - P^t)f, f\rangle ,$$

and similarly

$$\mathbb{E}d(f(Z_1), f(Z_0))^2 = 2\langle (I - P)f, f\rangle .$$

Thus it suffices to prove that

$$\langle (I - P^t)f, f\rangle \le t\langle (I - P)f, f\rangle .$$

Proof of Theorem 2.2 II

Indeed, if f is an eigenfunction with eigenvalue λ, this reduces to proving $(1 - \lambda^t) \le t(1 - \lambda)$. Since $|\lambda| \le 1$, this reduces to

$$1 + \lambda + \cdots + \lambda^{t-1} \le t ,$$

which is obviously true.
The claim follows for any other f by taking $f = \sum_{j=1}^n a_j f_j$ where $\{f_j\}$ is an orthonormal basis of eigenfunctions:

$$\langle (I-P^t)f, f\rangle = \sum_{j=1}^n a_j^2 \langle (I-P^t)f_j, f_j\rangle \le \sum_{j=1}^n a_j^2 t\langle (I-P)f_j, f_j\rangle = t\langle (I-P)f, f\rangle .$$

□

A maximal inequality for Markov chain

To prove Theorem 2.3 we first prove a lemma.

Lemma 2.4

Let $\{Z_t\}_{t=0}^\infty$ be a stationary time reversible Markov chain on $\{1, \ldots, n\}$ and $f : \{1, \ldots, n\} \to \mathbf{R}$. Then, for every time $t > 0$,

$$\mathbb{E} \max_{0\le s\le t} [f(Z_s) - f(Z_0)]^2 \le 15t\mathbb{E}[f(Z_1) - f(Z_0)]^2 .$$

Proof of Lemma 2.4 I

Proof. Let π be the stationary distribution. Define $P : L^2(\pi) \to L^2(\pi)$ by $(Pf)(i) = \mathbb{E}[f(Z_{s+1})|Z_s = i] = \sum_{j=1}^n p_{ij} f(j)$. For any $s \in \{0, \ldots, t-1\}$, let

$$D_s = f(Z_{s+1}) - (Pf)(Z_s) .$$

Since $\mathbb{E}[D_s|Z_1, \ldots, Z_s] = \mathbb{E}[D_s|Z_s] = 0$, the D_s are martingale differences with respect to the natural filtration of $Z_1, \ldots, Z_t$. Also, because of time-reversibility,

$$\tilde{D}_s = f(Z_{s-1}) - (Pf)(Z_s)$$

are martingale differences with respect to the natural filtration on $Z_t, \ldots, Z_1$. Note that $D_s - \tilde{D}_s = f(Z_{s+1}) - f(Z_{s-1})$, which implies that for any m,

$$f(Z_{2m}) - f(Z_0) = \sum_{k=1}^m D_{2k-1} - \sum_{k=1}^m \tilde{D}_{2k-1} .$$

Proof of Lemma 2.4 II

So,

$$\max_{0\le s\le t} f(Z_s) - f(Z_0)$$
$$\le \max_{m\le t/2} \sum_{k=1}^m D_{2k-1} + \max_{m\le t/2} \sum_{k=1}^m -\tilde{D}_{2k-1} + \max_{\ell\le t/2} |f(Z_{2\ell+1}) - f(Z_{2\ell})| .$$

Take squares and use the fact $(a + b + c)^2 \le 3(a^2 + b^2 + c^2)$, which is implied by the Cauchy-Schwarz inequality, to get

$$\max_{0\le s\le t} |f(Z_s) - f(Z_0)|^2 \le 3 \max_{m\le t/2} \Big|\sum_{k=1}^m D_{2k-1}\Big|^2 + 3 \max_{m\le t/2} \Big|\sum_{k=1}^m \tilde{D}_{2k-1}\Big|^2 + 3 \sum_{\ell\le t/2} \Big|f(Z_{2\ell+1}) - f(Z_{2\ell})\Big|^2 .$$

Proof of Lemma 2.4 III

We will use Doob's L^2 maximum inequality for martingales (see, e.g., Durrett 1996)

$$\mathbb{E} \max_{0\le s\le t} M_s^2 \le 4\, \mathbb{E}|M_t|^2 .$$

Consider

$$M_{s+1} = \sum_{j\le s, j \text{ odd}} D_j .$$

Since M_s is still a martingale, we have

$$\mathbb{E} \max_{0\le s\le t} |f(Z_s) - f(Z_0)|^2 \le 12\mathbb{E}\Big| \sum_{k=1}^{\lfloor t/2\rfloor} D_{2k-1}\Big|^2 + 12\mathbb{E}\Big| \sum_{k=1}^{\lfloor t/2\rfloor} \tilde{D}_{2k-1}\Big|^2 + 3 \sum_{0\le\ell\le t/2} \mathbb{E}\Big|f(Z_{2\ell+1}) - f(Z_{2\ell})\Big|^2 .$$

Proof of Lemma 2.4 IV

Denote $V = \mathbb{E}[|f(Z_1) - f(Z_0)|^2]$, and notice that

$$D_0 = f(Z_1) - f(Z_0) - \mathbb{E}[f(Z_1) - f(Z_0) \mid Z_0]\,,$$

which implies that D_0 is orthogonal to $\mathbb{E}[f(Z_1) - f(Z_0) \mid Z_0]$ in $L^2(\pi)$. So, by the Pythagorian law, for any s we have $\mathbb{E}[D_s^2] = \mathbb{E}[D_0^2] \le V$. Summing everything up gives

$$\mathbb{E} \max_{0 \le s \le t} |f(Z_s) - f(Z_0)|^2 \le 6tV + 6tV + 3(t/2+1)V \le 15tV\,,$$

which concludes our proof. □

Proof of Theorem 2.3 I

Proof. Let T be a weighted tree, $\{Z_j\}$ be a reversible Markov chain on $\{1,\ldots,n\}$ and $F : \{1,\ldots,n\} \to T$. Choose an arbitrary root and set for any vertex v, $\psi(v) = d(root, v)$. If $v_0,\ldots,v_t$ is a path in the tree, then

$$d(v_0, v_t) \le \max_{0 \le j \le t} \left(|\psi(v_0) - \psi(v_j)| + |\psi(v_t) - \psi(v_j)|\right)\,,$$

since choosing the closest vertex to the root on the path yields equality.

Proof of Theorem 2.3 II

Let $X_j = F(Z_j)$. Connect X_i to X_{i+1} by the shortest path for any $0 \le i \le t-1$ to get a path between X_0 and X_t. Since now the closest vertex to the root can be on any of the shortest paths between X_j and X_{j+1}, we get

$$d(X_0, X_t) \le \max_{0 \le j < t} \Big(|\psi(X_0) - \psi(X_j)| + |\psi(X_t) - \psi(X_j)| + 2d(X_j, X_{j+1})\Big)\,.$$

Square, and use Cauchy-Schwarz again,

$$d(X_0, X_t)^2$$
$$\le 3 \max_{0 \le j \le t} \Big(|\psi(X_0) - \psi(X_j)|^2 + |\psi(X_t) - \psi(X_j)|^2\Big) + 12 \sum_{0 \le j < t} d^2(X_j, X_{j+1})\,.$$

Proof of Theorem 2.3 III

By Lemma 2.4 with $f = \psi F$ we get,

$$\mathbb{E}d(X_0, X_t)^2 \le 90t\mathbb{E}|\psi(X_0) - \psi(X_1)|^2 + 6 \sum_{0 \le j \le t} \mathbb{E}d^2(X_j, X_{j+1})\,.$$

Since in any metric space $|\psi(X_1) - \psi(X_0)| \le d(X_0, X_1)$ and since the Markov chain is stationary we have $\mathbb{E}d(X_0, X_1) = \mathbb{E}d(X_j, X_{j+1})$ for any j. So

$$\mathbb{E}d(X_0, X_t)^2 \le 96t\mathbb{E}d(X_0, X_1)^2\,,$$

which concludes our proof. □

Exercise

1. Prove that the n-dimensional hypercubes do not have uniform Markov type 2.

Outline

Capture convergence rate via spectral gap

In this section we investigate the speed with which a finite, reversible Markov chain converges to the stationary distribution. Let P be the transition matrix of the chain, and let π be the stationary distribution. For any two states x and y, consider the distance

$$\frac{P^t(x,y)-\pi(y)}{\pi(y)}.$$

Since P is reversible, as we have seen before, it has only real eigenvalues, all in $[-1,1]$, denoted $\lambda_n \leq \ldots \leq \lambda_1 = 1$. We will show that under some conditions, when P exhibits what is known as a *spectral gap*, i.e., the second largest eigenvalue of P is strictly smaller than 1, the chain converges to the stationary distribution in an exponential rate in t.

Some preliminary definitions

Note that in this section all inner products are with respect to the stationary measure π, i.e.,

$$\langle f,g\rangle = \sum_{i=1}^{n} \pi(i)f(i)g(i).$$

Definition 3.1

We call P irreducible if

$$\forall x,y \quad \exists k \quad P^k(x,y) > 0.$$

We call P irreducible a-periodic if

$$\exists k \quad \forall x,y \quad P^k(x,y) > 0.$$

Strictly positive spectral gap implies convergence at exponential rate

Lemma 3.2

Let $\lambda_* = \max_{i\geq 2}|\lambda_i|$ be the second largest (in absolute value) eigenvalue of a finite, irreducible, a-periodic, reversible Markov chain with transition matrix P, then $\lambda_* < 1$.

For a proof of this lemma see exercise 2.

Theorem 3.3

Under the conditions of Lemma 3.2, let π be the stationary distribution of P and $\pi_* = \min_x \pi(x)$ and denote $g = 1-\lambda_*$. Then, for any states x and y,

$$\left|\frac{P^t(x,y)-\pi(y)}{\pi(y)}\right| \leq \frac{e^{-gt}}{\pi_*}.$$

Definition of Cheeger constant

Intuitively, if a Markov chain has "bottlenecks", it will take it more time to mix. To formulate this intuition, we define what is known as the *Cheeger constant*.
For any two states in a stationary Markov chain a and b, let $q(a,b) = \pi(a)p(a,b)$, and for any two subsets of states A and B, let $Q(A,B) = \sum_{a\in A, b\in B} q(a,b)$.

Definition 3.4

The Cheeger constant of a stationary Markov chain is

$$\Phi_* = \min_{S:\pi(S)\leq 1/2} \Phi_S,$$

where

$$\Phi_S = \frac{Q(S,S^c)}{\pi(S)}.$$

Upper and lower bounds on spectral gap via Cheeger constant

The following theorem (see Alon, 1986, Jerrum and Sinclair, 1989 and Lawler and Sokal, 1988), together with the previous one, connects the isoperimetric properties of a Markov chain with its mixing time. We assume that the chain is “lazy”, i.e., for any state x we have that $p(x,x) \geq \frac{1}{2}$. In that case, $P = \frac{I+\tilde{P}}{2}$ where $\tilde{P}$ is another stochastic matrix, and hence all the eigenvalues of P are in $[0,1]$, and $\lambda_* = \lambda_2$.

Theorem 3.5

Let λ_2 be the second eigenvalue of a reversible and lazy Markov chain, and $g = 1-\lambda_2$. Then,

$$\frac{\Phi_*^2}{2} \leq g \leq 2\Phi_*.$$

Proof of Theorem 3.5 - upper bound I

Proof. Recall that

$$\lambda_2 = \max_{f\perp \mathbf{1}} \frac{\langle Pf,f\rangle}{\langle f,f\rangle},$$

so

$$g = \min_{f\perp \mathbf{1}} \frac{\langle (I-P)f,f\rangle}{\langle f,f\rangle} \text{ ..cheegergap}$$

As we have seen before, expanding the nominator gives

$$\langle (I-P)f,f\rangle = \frac{1}{2}\sum_{x,y}\pi(x)p(x,y)[f(y)-f(x)]^2,$$

thus

$$g = \min_{f\perp \mathbf{1}} \frac{\sum_{x,y}\pi(x)p(x,y)[f(y)-f(x)]^2}{\sum_{x,y}\pi(x)\pi(y)[f(x)-f(y)]^2}.$$

Proof of Theorem 3.5 - upper bound II

To get $g \le 2\Phi_*$, for any S with $\pi(S) \le 1/2$ define a function f by $f(x) = \pi(S^c)$ for $x \in S$ and $f(x) = \pi(S)$ for $x \notin S$. $\mathbb{E}f = 0$, so $f \perp \mathbf{1}$. Using this f we get that

$$g \le \frac{2Q(S,S^c)}{2\pi(S)\pi(S^c)} \le \frac{2Q(S,S^c)}{\pi(S)} \le 2\Phi_S\,,$$

and so $g \le 2\Phi_*$. □

A useful lemma to prove the lower bound

Lemma 3.6

Given a function $\psi \ge 0$ on a state space V of a stationary Markov chain, order V such that ψ is monotone decreasing and assume $\pi\{\psi > 0\} \le 1/2$, then

$$\|\psi\|_{L^1(\pi)} \le \Phi_*^{-1} \sum_{x<y} [\psi(x) - \psi(y)] q(x,y)\,.$$

Proof. Let $t > 0$, then $\Phi_* \le \Phi_S$ with $S = \{\psi > t\}$, so we have

$$\pi\{\psi > t\} \le \Phi_*^{-1} \textstyle\sum_{x,y} q(x,y) \mathbf{1}_{\{\psi(x) > t \ge \psi(y)\}}\,.$$

Now, integrating on t gives

$$\|\psi\|_{L^1(\pi)} \le \Phi_*^{-1} \textstyle\sum_{x<y} [\psi(x) - \psi(y)] q(x,y)\,,$$

as $\int_0^\infty \pi\{\psi > t\} dt = \|\psi\|_{L^1(\pi)}$ and $\int_0^\infty \mathbf{1}_{\{\psi(x) > t \ge \psi(y)\}} = \psi(x) - \psi(y)$.

Proof of Theorem 3.5 - lower bound I

Proof. Now take an eigenfunction f_2 such that $Pf_2 = \lambda_2 f_2$ and $\pi\{f_2 > 0\} \le 1/2$ (if this does not hold, take $-f_2$). Define a new function $f = \max(f_2, 0)$. Observe that,

$$[(I-P)f](x) \le gf(x) \qquad \forall x\,.$$

This is because if $f(x) = 0$, this translates to $-(Pf)(x) \le 0$ which is true since $f \ge 0$, and if $f(x) > 0$, then, since $f \ge f_2$, we have $[(I-P)f](x) \le [(I-P)f_2](x) \le gf_2(x) = gf(x)$. Since $f \ge 0$, we get

$$\langle (I-P)f, f\rangle \le g\langle f, f\rangle\,,$$

or equivalently,

$$g \ge \frac{\langle (I-P)f, f\rangle}{\langle f, f\rangle}\,.$$

Proof of Theorem 3.5 - lower bound II

Note that this looks like a contradiction to 30.cheegergap/, but it is not since f is not orthogonal to $\mathbf{1}$. Denote $R = \frac{\langle (I-P)f, f\rangle}{\langle f, f\rangle}$. Now, apply the lemma with $\psi = f^2$ to get

$$\langle f, f\rangle^2 \le \Phi_*^{-2} \left[\sum_{x<y} (f^2(x) - f^2(y)) q(x,y)\right]^2\,.$$

By the Cauchy-Schwarz inequality we get

$$\langle f, f\rangle^2 \le \Phi_*^{-2} \left[\sum_{x<y} (f(x) - f(y))^2 q(x,y)\right] \left[\sum_{x<y} (f(x) + f(y))^2 q(x,y)\right]\,.$$

Recall that $\langle (I-P)f, f\rangle = \sum_{x<y} (f(x) - f(y))^2 q(x,y)$ and use the fact that $(f(x) + f(y))^2 = 2f^2(x) + 2f^2(y) - (f(x) - f(y))^2$ to get that

$$\langle f, f\rangle^2 \le \Phi_*^{-2} \langle (I-P)f, f\rangle [2\langle f, f\rangle - \langle (I-P)f, f\rangle]\,.$$

Proof of Theorem 3.5 - lower bound III

Divide by $\langle f, f\rangle^2$ to get

$$\Phi_*^2 \le R(2-R)\,,$$

so

$$1 - \Phi_*^2 \ge 1 - 2R + R^2 = (1-R)^2 \ge (1-g)^2\,.$$

One additional computation,

$$\left(1 - \frac{\Phi_*^2}{2}\right)^2 \ge 1 - \Phi_*^2 \ge (1-g)^2\,,$$

yields that $g \ge \frac{\Phi_*^2}{2}$, as required.

Definition of expander

Definition 3.7

A family of graphs $\{G_n\}$ is said to be an (n, d, λ) expander family if all of the following three conditions hold for all n:

- **(i)** $|V(G_n)| = n$.
- **(ii)** *G_n is d-regular.*
- **(iii)** *The Cheeger constant of the simple random walk on the graph satisfies $\Phi_*(G_n) \ge \lambda$.*

Exercises

1. Prove that there exists $\delta > 0$ such that

$$\sum_{k=1}^{\frac{n}{2}} \frac{\binom{n}{\delta k}\binom{(1+\delta)k}{\delta k}^2}{\binom{n}{k}} < 1\,.$$

2. Let $\lambda_* = \max_{i\geq 2}|\lambda_i|$ be the second largest (in absolute value) eigenvalue of a finite, irreducible, a-periodic, reversible Markov chain with transition matrix P, and stationary distribution π. Prove that $\lambda_* < 1$.

Outline

Definition of harmonic function and coupling

Definition 4.1

Let P be a Markov operator and let f be a real function on the state space. f is called harmonic if $Pf = f$, i.e. if

$$\sum_{y\sim x} P(x,y)f(y) = f(x)\,.$$

Definition 4.2

Let $\mathcal{S}$ be a measure space. A coupling of two $\mathcal{S}$-valued random variables X and X', or of their distributions μ and μ', is a probability measure on $\mathcal{S}\times\mathcal{S}$ having marginals μ and μ', i.e., for every event $A \subset \mathcal{S}$

$$\mathbb{P}\Big[\{(x,x') : x \in A\}\Big] = \mu(A)$$

and

$$\mathbb{P}\Big[\{(x,x') : x' \in A\}\Big] = \mu'(A)\,.$$

Theorem 4.3

Let P be a transition matrix for a Markov chain on a countable state space V. If for any $x, y \in V$ there exists a coupling (X_n, Y_n) such that $\{X_n\}$ and $\{Y_n\}$ are distributed according to the chain, $(X_0, Y_0) = (x, y)$ and

$$\lim_{n\to\infty} \mathbb{P}[X_n \neq Y_n] = 0\,,$$

then every bounded harmonic function $f : V \to \mathbf{R}$ is constant.

Proof of Theorem 4.3

Proof.

Let $x, y \in V$ and let (X_n, Y_n) be such a coupling. Let $f : V \to \mathbf{R}$ be a bounded harmonic function. Observe that since f is harmonic, for any n we have

$$\mathbb{E}[f(X_{n+1}) - f(X_n) \mid X_n = x] = \sum_{y\sim x} P(x,y)(f(y) - f(x)) = 0\,,$$

and by taking expectation over x we get that $\mathbb{E}f(X_{n+1}) = \mathbb{E}f(X_n)$. So for any n we have that $|f(x) - f(y)| = |\mathbb{E}f(X_n) - \mathbb{E}f(Y_n)|$, and if $|f|$ is bounded by M, then

$$|f(x) - f(y)| = |\mathbb{E}f(X_n) - \mathbb{E}f(Y_n)| \leq 2M\mathbb{P}[X_n \neq Y_n] \to 0\,.$$

That is, $f(x) = f(y)$ for every $x, y \in V$. □

Bounded harmonic functions on $\mathbf{Z}^d$ are constant I

Theorem 4.4 (Blackwell, 1955)

All bounded harmonic functions on $\mathbf{Z}^d$ are constant.

Proof. By Theorem 4.3 it is enough to find a Markov chain and a coupling (X_n, Y_n) such that $(X_0, Y_0) = (x, y)$ and $P(X_n \neq Y_n) \to 0$. We take $\{X_n\}$ and $\{Y_n\}$ to be lazy simple random walks on $\mathbf{Z}^d$ starting at x and y respectively, i.e. for any $x \in \mathbf{Z}^d$ we have $p(x,x) = 1/2$, and the transition probability to each of the $2d$ neighbors is $1/(4d)$.

Bounded harmonic functions on $\mathbf{Z}^d$ are constant II

We couple $\{X_n\}$ and $\{Y_n\}$ coordinate-wise. Draw a direction $i \in \{1, \ldots, d\}$ at random. If $X_n(i) = Y_n(i)$, then with probability $1/2$ leave both at the same position and with probability $1/2$ move them together in direction i. If $X_n(i) \neq Y_n(i)$, choose either X_n or Y_n with probability $1/2$ and move it in direction i. Observe that both walks are distributed as lazy simple random walks. Also, in any coordinate the difference between the walks is a lazy simple one-dimensional random walk with 0 as an absorbing state. This implies that with probability 1 there exists some N such that for all $n > N$ we have $X_n = Y_n$, and thus $P(X_n \neq Y_n) \to 0$, as required.

Example of non-constant harmonic function on trees

On infinite regular trees the situation is quite different.

Proposition 4.5

For any $d \geq 3$, let T be the infinite d-regular tree, and let o be its root. Choose one neighbor of o and let $A \subset G$ be all the descendants of that vertex. Let f be the real function on T defined by

$$f(x) := P_x[X_n \in A \text{ for all but finitely many } n] .$$

Then f is a non-constant bounded harmonic function.

Proof of Proposition 4.5.

Denote by B the event $\{X_n \in A \text{ for all but finitely many } n\}$. Observe that for any x,

$$f(x) = P_x(B) = \sum_y P_x(X_1 = y)P_y(B) = \sum_y P_x(X_1 = y)f(y) ,$$

implying that f is a bounded harmonic function.

To show f is not constant, consider $x, y \in A$ such that y is a child of x. Note that the probability of the simple random walk starting at y never to hit x is the same as the probability that a random walk on $\mathbf{Z}$, with probability $1/d$ going left and $1 - 1/d$ going right, starting at 1, never hits 0. If p denotes the above probability, then clearly $f(y) = f(x) + p$, and since $d > 2$, it is known that $p > 0$, and so $f(y) > f(x)$. □

Escape rate and non-constant bounded harmonic functions

Our aim is to formulate a connection between the rate of escape of a Markov chain on a graph on the one hand and the existence of non-constant bounded harmonic functions on the graph on the other hand.

Proposition 4.6

If $\{X_n\}$ is a simple random walk on a vertex transitive graph, then

$$\lim_{n\to\infty} \frac{\mathbb{E}d(X_0, X_n)}{n}$$

exists.

Proof of Proposition 4.6

Proof.

The sequence $\mathbb{E}d(X_0, X_n)$ is sub-additive, since using the triangle inequality and the transitivity, we get

$$\mathbb{E}d(X_0, X_{n+m}) \leq \mathbb{E}d(X_0, X_n) + \mathbb{E}d(X_n, X_{n+m}) = \mathbb{E}d(X_0, X_n) + \mathbb{E}d(X_0, X_m) ,$$

and so we conclude the proof by the following lemma:

Lemma 4.7 (Fekete)

If $\{a_n\}$ sub-additive, i.e. $a_{n+m} \leq a_n + a_m$ for all n, m, then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n} .$$

□

Definition of lamplighter

Definition 4.8

The lamplighter group over $\mathbf{Z}$ is a group with elements being all pairs (f, x), where $x \in Z$ and $f : \mathbf{Z} \to \mathbf{Z}_2$, with $f(i) = 0$ for all but finitely many i. Traditionally, x represents the lamplighter's location and f the states of the lamps. The identity element of the group is (f, x) where $x = 0$ and $f(i) = 0$ for all i. We take three generators of the group which yield the following moves:

- **(i)** *Lamplighter moves one step left.*
- **(ii)** *Lamplighter moves one step right.*
- **(iii)** *Lamplighter changes the current state of the lamp at her position.*

Equivalently, the multiplication rule of the group is $(f, x)(g, y) = (h, x + y)$ where $h(i) = f(i) + g(i - x)$.

Remark 4.9

A natural generalization of this group is the lamplighter group over $\mathbf{Z}^d$, in which the lamplighter moves in d dimensions.

Speed of random walk on lamplighter group

We are interested in the speed of the simple random walk on the Cayley graph of the lamplighter group over $\mathbf{Z}^d$ with the generators specified above.

Theorem 4.10

Let $d > 2$ and let $\{X_n\}$ be a random walk on the lamplighter group with the specified generators. Assume that X_n starts at the identity, and that it has the "lazy" property (i.e., with probability $1/4$, we have $X_n = X_{n-1}$, with probability $1/4$, the lamplighter changes the lamp's state, and with probability $1/(4d)$ the lamplighter walks to one of the $2d$ neighbors). Then

$$\lim_{n\to\infty} \frac{\mathbb{E}d(X_n, id)}{n} > 0\,.$$

In other words, the walk has positive speed.

To prove Theorem 4.10, we first provide a lemma.

Lemma 4.11

Let p denote the probability that a simple random walk on a vertex-transitive graph, starting at the origin, never returns to the origin. Let R_n denote the number of distinct sites visited by the walk up to time n. Then

$$\lim_{n\to\infty} \frac{\mathbb{E}R_n}{n} = p\,.$$

Proof of Lemma 4.11

Proof.

Let I_i^n be the indicator random variable indicating whether X_i is a state that is not visited again between time $i+1$ and n, or equivalently

$$I_i^n = \begin{cases} 1 & \text{if } X_i \notin \{X_{i+1}, \ldots, X_n\}, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $R_n = \sum_{i=0}^n I_i^n$, and for any fixed i we have $\mathbb{E}I_i^n \to p$ as $n \to \infty$, since the graph is vertex-transitive. $\mathbb{E}I_i^n$ is decreasing in n, and increasing in i, hence, for any $\epsilon > 0$ there exists m such that for all i and $n > i + m$ we have $p \le \mathbb{E}I_i^n \le p + \epsilon$. So

$$p \le \frac{\mathbb{E}R_n}{n} = \frac{\sum_{i=1}^n \mathbb{E}I_i^n}{n} \le \frac{(p+\epsilon)(n-m)+m}{n} \to p + \epsilon\,.$$

□

Proof of Theorem 4.10

Proof.

Let p be the probability that the lazy simple random walk on $\mathbf{Z}^d$, starting at the origin, never returns to the origin. Since $d > 2$, we have $p > 0$. Notice that since all the lamps in the identity element are turned off, $d(X_n, id)$ is at least the number of lit lamps. Note that each lamp that the lamplighter visited has probability at least $1/4$ of being lit after the lamplighter leaves, so by Lemma 4.11,

$$\frac{\mathbb{E}d(X_n, id)}{n} \ge \frac{\mathbb{E}\#\text{ lamps lit after } n \text{ steps}}{n} \ge \frac{\mathbb{E}\#\text{ lamps visited}}{4n} \to p/4 > 0\,.$$

□

Harmonic functions on lamplighter groups on $\mathbf{Z}^d$

Theorem 4.12

All bounded harmonic functions on the lamplighter group on $\mathbf{Z}^d$ are constant for $d = 1, 2$. This is not true for $d > 2$.

μ-harmonic functions on Abelian groups

Definition 4.13

Suppose μ is a probability measure on a group G such that the set $\{g : \mu(g) > 0\}$ generates G. A function f on G is μ-harmonic if $\int_G f(xy)d\mu(y) = f(x)$ for all $x \in G$.

The following is an extension of Theorem 4.4.

Theorem 4.14 (Choquet and Deny (1960))

Let h be a bounded μ-harmonic function on an abelian group G. Then for all $x \in G$, for μ-almost every $y \in G$, we have $h(x + y) = h(x)$ (i.e., h is constant).

Exercises

1. Suppose μ is a probability measure on an Abelian group G such that $\{g : \mu(g) > 0\}$ generates G. Show that any bounded μ-harmonic function on G is constant, using a coupling argument.
2. Given a transition matrix P and a positive harmonic function h on the state space, check that the matrix $\tilde{P}$ defined by

$$\tilde{P}(x, y) = P(x, y)h(y)/h(x)$$

is a transition matrix (this is known as the Doob h-transform).
3. Show that on $\mathbf{Z}^d$, any harmonic function with sublinear growth is constant.
4. Show that on $\mathbf{Z}^d$, any positive harmonic function is constant.
5. Calculate the function f defined in Proposition 4.5.

Outline

C-embedding and distortion

Definition 5.1

An invertible mapping $f : X \to Y$, where (X, d_X) and (Y, d_Y) are metric spaces, is a C-embedding if there exists a number $r > 0$ such that for all $x, y \in X$

$$rd_X(x, y) \le d_Y(f(x), f(y)) \le Crd_X(x, y)\,.$$

The infimum of numbers C such that f is a C-embedding is called the distortion of f and is denoted by $dist(f)$. Equivalently, $dist(f) = \|f\|_{\mathrm{Lip}}\|f^{-1}\|_{\mathrm{Lip}}$, where

$$\|f\|_{\mathrm{Lip}} = \sup\left\{\frac{d_Y(f(x), f(y))}{d_X(x, y)} : x, y \in X, x \neq y\right\}.$$

A dimension reduction lemma

Lemma 5.2 (Johnson and Lindenstrauss (1984))

For any $0 < \epsilon < 1/2$ and $v_1, \ldots, v_n \in \mathbf{R}^n$ with Euclidean metric, there exists a linear map $A : \mathbf{R}^n \to \mathbf{R}^k$ where $k = O(\log n/\epsilon^2)$, with distortion at most $1 + \epsilon$ on the n point space $\{v_1, \ldots, v_n\}$.

Embedding of general metric spaces

Theorem 5.3 (Bourgain, 1985)

Every n-point metric space (X, d) can be embedded in an $O(\log n)$ -dimensional Euclidean space with an $O(\log n)$ distortion.

Embedding the hypercube to Hilbert space I

Proposition 5.4 (Enflo, 1969)

There exists $c > 0$ such that any embedding of the hypercube $\{0, 1\}^k$ in Hilbert space has distortion at least $c\sqrt{k}$.

Proof. Recall that in Exercise 1 of Chapter 3 we proved that if $\{X_j\}$ is a simple random walk in the hypercube, then

$$\mathbb{E}d(X_0, X_j) \ge \frac{j}{2} \qquad \forall j \le k/4\,.$$

Take $j = \frac{k}{4}$. By Jensen's inequality, $\mathbb{E}d^2(X_0, X_{k/4}) \ge k^2/64$. Now let $f : \{0, 1\}^k \to L^2$ be a map. Assume without loss of generality that f is a non-expanding, i.e., $\|f\|_{\mathrm{Lip}} = 1$ (otherwise, take $f/\|f\|_{\mathrm{Lip}}$).

Embedding the hypercube to Hilbert space II

By Theorem 2.2 it follows that L^2 has Markov type 2 with constant $M = 1$, so,

$$\mathbb{E}d^2\Big(f(X_0), f(X_{k/4})\Big) \le k\,.$$

We conclude

$$\|f^{-1}\|_{\mathrm{Lip}}^2 k \ge \|f^{-1}\|_{\mathrm{Lip}}^2 \mathbb{E}d^2(f(X_0), f(X_{k/4})) \ge \mathbb{E}d^2(X_0, X_{k/4}) \ge k^2/64\,,$$

hence $\|f^{-1}\|_{\mathrm{Lip}} \ge \frac{\sqrt{k}}{8}$, which implies the result. □

Remark 5.5

Enflo's original proof gives $c = 1$. See Exercise 1 for the proof of this fact.

Exercises I

1. (Enflo, 1969) Let $\Omega_d = \{-1, 1\}^d$ be the d-dimensional hypercube with ℓ_1 metric. Show that any $f : \Omega_d \to L^2$ has distortion at least $\sqrt{d}$ and give an example for an f with $\sqrt{d}$ distortion.
<u>*Solution*</u> As an example, take f to be the identity embedding $\Omega_d \to \ell_2^d$. Clearly $d_2(f(x), f(y)) \le d(x, y)$, and by Cauchy-Schwarz we have

$$d(x, y) = 2\sum_{i=1}^{d} \mathbf{1}_{(x_i \ne y_i)} \le d_2(x, y)\sqrt{d}\,.$$

Exercises II

For the lower bound, a simple computation gives that for any $x_1, x_2, x_3, x_4 \in L^2$ we have that the sum of squares of the diagonals is smaller than the sum of squared edges lengths

$$d(x_1, x_4)^2 + d(x_2, x_3)^2 \le d(x_1, x_2)^2 + d(x_2, x_4)^2 + d(x_3, x_4)^2 + d(x_1, x_3)^2\,.$$

This can be easily extended by induction on d to

$$\sum_{x\in\Omega_d} \|f(x) - f(-x)\|^2 \le \sum_{x\sim y} \|f(x) - f(y)\|^2\,.$$

Assume now that $L > 0$ satisfies

$$d_1(x, y) \le \|f(x) - f(y)\|_2 \le L d_1(x, y)\,.$$

Exercises III

Then

$$\sum_{x\sim y} \|f(x) - f(y)\|^2 \le 4L^2 d 2^d\,,$$

and also

$$\sum_{x\in\Omega_d} \|f(x) - f(-x)\|^2 \ge 4d^2 2^d\,,$$

which, together with the previous argument, imply that $L \ge \sqrt{d}$, as required. □

Exercises IV

2. Use the proof that **R** has Markov type 2 to show that for an (n, d, λ)-expander family, any invertible mapping f of the vertices to a Hilbert space have

$$\|f\|_{\mathrm{Lip}} \|f^{-1}\|_{\mathrm{Lip}} \ge C_{d,\lambda} \log n\,,$$

where $C_{d,\lambda}$ is constant.
<u>*Solution*</u> Let $\{X_j\}$ be the simple random walk on the expander, with transition matrix P. Since the family is d-regular, the random walk has uniform stationary distribution. By Theorem 3.3 it follows that for any $x, y \in V$

$$P^t(x, y) \le \pi(y) + e^{-gt}\,,$$

where $g = 1 - \lambda$. Take $\alpha > 0$ such that $g\alpha < 1$, and take $t = \alpha \log n$.

Exercises V

Then

$$P^t(x, y) \le \frac{1}{n} + e^{-g\alpha \log n} \le 2e^{-g\alpha \log n}\,.$$

Fix $\gamma > 0$ small enough such that $d^\gamma e^{-g\alpha} < 1$. We wish to show that up to time $t = \alpha \log n$, the random walk on the expander has positive speed. Indeed, for any $x \in V$, since the ball $B(x, \gamma \log n)$ of radius $\gamma \log n$ around x has at most $d^{\gamma \log n}$ vertices, it follows that

$$\mathbb{P}_x[X_t \in B(x, \gamma \log n)] \le d^{\gamma \log n} 2e^{-g\alpha \log n} \to 0\,.$$

Exercises VI

This in turn implies that for large enough n

$$\mathbb{E}d^2(X_0, X_t) > \frac{\gamma^2 \log^2 n}{2}\,.$$

Let $f : V \to L^2$, and assume without loss of generality that $\|f\|_{\mathrm{Lip}} = 1$ (otherwise take $f/\|f\|_{\mathrm{Lip}}$). Recall that in the proof of Theorem 2.2 we actual-ly proved that

$$\mathbb{E}d(f(X_t), f(X_0))^2 \le (1 + \lambda + \lambda^2 + \cdots + \lambda^{t-1})\mathbb{E}d(f(Z_1), f(Z_0))^2\,.$$

This immediately implies that

$$\mathbb{E}d^2(f(X_0), f(X_t)) \le \frac{1}{g}\,.$$

Now, similarly to the proof of Proposition 5.4 we conclude that $\|f^{-1}\|_{\mathrm{Lip}} \ge \sqrt{g}\gamma \log n/\sqrt{2}$. □