The Annals of Applied Probability
1991, Vol. 1, No. 1, 36–61

GEOMETRIC BOUNDS FOR EIGENVALUES OF MARKOV CHAINS

BY PERSI DIACONIS[1] AND DANIEL STROOCK[2]

Harvard University and Massachusetts Institute of Technology

We develop bounds for the second largest eigenvalue and spectral gap of a reversible Markov chain. The bounds depend on geometric quantities such as the maximum degree, diameter and covering number of associated graphs. The bounds compare well with exact answers for a variety of simple chains and seem better than bounds derived through Cheeger-like inequalities. They offer improved rates of convergence for the random walk associated to approximate computation of the permanent.

1. Introduction.

A. *Basic notation*. Let X be a finite set and $P(x, y)$ the transition probability for an irreducible Markov chain. We assume throughout that $P(x, y)$ is reversible relative to the probability distribution π. That is,

$$(1.1) \qquad Q(x,y) \equiv \pi(x)P(x,y) = \pi(y)P(y,x) \quad \text{for all } x, y \in X.$$

This means that π is a stationary distribution for $P(x, y)$ and so (because of irreducibility) π charges every point. Equivalently, the operator P given by

$$[P\phi](x) = \sum_{y \in X} P(x,y)\phi(y), \qquad x \in X,$$

is a self-adjoint contraction on $L^2(\pi)$. This P has largest eigenvalue 1 and (again because of irreducibility) the constant functions are the only eigenfunctions with eigenvalue 1. The eigenvalues are denoted

$$1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{m-1} \geq -1 \quad \text{where } m = |X|.$$

The chain is aperiodic precisely when $\beta_{m-1} > -1$.

This paper develops methods for bounding β_1, β_{m-1} and $\beta_* = \max(\beta_1, |\beta_{m-1}|)$. Bounds for rates of convergence to stationarity in variation distance in terms of eigenvalues are given at the end of this introduction.

There are technical advantages in considering the Laplacian $L = I - P$ instead of P. Obviously the spectrum of L consists of the numbers $\lambda_i = 1 - \beta_i$, $0 \leq i \leq m - 1$. The usual minimax characterization of eigenvalues [see, e.g.,

---

Received December 1989; revised July 1990.

[1]Partially supported by NSF Grant DMS-86-00235 and by a grant from Teledyne Electronics.

[2]Partially supported by NSF Grant DMS-86-11487 and U.S. Army Research Office Grant DAAL 03-86-K-0171.

AMS 1980 *subject classifications*. 60J10, 60C05.

Key words and phrases. Markov chains, eigenvalues, random walk.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Applied Probability*.
www.jstor.org

Horn and Johnson (1985), page 176] gives

$$\lambda_1 = \inf\left\{\frac{\mathscr{E}(\phi,\phi)}{\operatorname{Var}(\phi)} : \phi \text{ is nonconstant}\right\}, \tag{1.2}$$

where $\operatorname{Var}(\phi)$ denotes the variance of ϕ relative to π and

$$\mathscr{E}(\phi,\phi) \equiv \tfrac{1}{2}\sum_{x,y} (\phi(y) - \phi(x))^2 Q(x,y) \tag{1.3}$$

is easily identified [use symmetry of $Q(x,y)$] as the quadratic form

$$\phi \mapsto (\phi, L\phi)_{L^2(\pi)} \quad \text{determined by } L \text{ on } L^2(\pi).$$

REMARK. For historical reasons, quadratic forms determined by operators which, like L, generate a Markov semigroup, are called Dirichlet forms. The classical Dirichlet form is

$$\phi \mapsto \int_r |\nabla\phi(x)|^2\,dx,$$

where r is a region in $\mathbb{R}^n$. The quadratic form $\mathscr{E}(\phi,\phi)$ is nothing but a discrete version of this classical form.

B. *Poincaré inequalities.* The next ingredient is a graph with vertex set X and $\{x,y\}$ an edge iff $Q(x,y) > 0$. For each pair of distinct points $x, y \in X$, choose a path γ_{xy} from x to y. Paths may have repeated vertices but a given edge appears at most once in a given path. Let Γ denote the collection of paths (one for each ordered pair x, y). Irreducibility guarantees that such paths exist. However, as will become obvious, the quality of our estimate depends on making a judicious selection of Γ.

For $\gamma_{xy} \in \Gamma$ define the path length by

$$|\gamma_{xy}|_Q = \sum_{e \in \gamma_{xy}} Q(e)^{-1}, \tag{1.4}$$

where the sum is over edges in the path and $Q(e) = Q(z,w)$ if $e = \{z,w\}$.

The *geometric* quantity that appears in our estimate is

$$\kappa = \kappa(\Gamma) = \max_e \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_Q \pi(x)\pi(y), \tag{1.5}$$

where the maximum is over directed edges in the graph and the sum is over all paths γ which traverse e. As will emerge from the examples in Section 2, κ is a measure of bottlenecks. It will be small if it is possible to choose paths which do not traverse any one edge too often. As will also emerge, κ can be effectively bounded in examples of interest.

With this notation, a first form of our estimate can be stated.

PROPOSITION 1 (Poincaré inequality). *For an irreducible Markov chain P the second largest eigenvalue satisfies*

$$\beta_1 \le 1 - \frac{1}{\kappa}$$

with κ defined by (1.5).

PROOF. Write

$$\begin{aligned}\operatorname{Var}(\phi) &= \frac{1}{2}\sum_{x,y\in X}(\phi(x)-\phi(y))^2\pi(x)\pi(y)\\ &= \frac{1}{2}\sum_{x,y\in X}\left(\sum_{e\in\gamma_{xy}}\left(\frac{Q(e)}{Q(e)}\right)^{1/2}\phi(e)\right)^2\pi(x)\pi(y)\\ &\le \frac{1}{2}\sum_{x,y}|\gamma_{xy}|_Q\pi(x)\pi(y)\sum_{e\in\gamma_{xy}}Q(e)\phi(e)^2\\ &= \frac{1}{2}\sum_{e}Q(e)\phi(e)^2\sum_{\gamma_{xy}\ni e}|\gamma_{xy}|_Q\pi(x)\pi(y).\end{aligned}$$

Here $\phi(e) = \phi(e^+) - \phi(e^-)$ where e is the oriented edge in a path from e^- to e^+, the inequality is Cauchy–Schwarz and the final sum is over all oriented edges in the graph. Bounding the final inner sum by κ, we arrive at

$$\operatorname{Var}(\phi) \le \kappa\mathscr{E}(\phi,\phi),$$

so the result follows from the variational characterization (1.2). □

REMARK. Proposition 1 is a discrete analog of the classical method of Poincaré for estimating the spectral gap of the Laplacian on a domain [see, e.g., Bandle (1980)]. Related ideas were used by Landau and Odlyzko (1981), by Holley and Stroock (1988) and by Mohar (1989a, b). Section 1E gives further discussion and a comparison with other techniques such as ergodic coefficients. Our own realization of just how much can be gained by careful choice of paths in Γ came from reading Sinclair and Jerrum's (1989) lovely solution to a problem from computer science.

Finally, it should be clear that we have made only one of many possible choices for estimating $(\sum_{e\in\gamma_{xy}}\phi(e))^2$ in terms of $\sum_e Q(e)\phi(e)^2$. For example, Sinclair (1990) has suggested the one leading to

PROPOSITION 1′. *With notation as in Proposition* 1,

$$\beta_1 \le 1 - \frac{1}{K},$$

where

$$K = \max_e Q(e)^{-1}\sum_{\gamma_{xy}\ni e}|\gamma_{xy}|\pi(x)\pi(y)$$

and $|\gamma_{xy}|$ *denotes the number of edges in the path* γ_{xy}.

This bound is often easier to use and examples presented in the next section show it can be more effective than the bound involving K. For random walks on graphs, explained below, the bounds involving κ and K coincide.

Sinclair (1990) has used these Poincaré inequalities to get bounds in the approach to equilibrium in the Metropolis algorithm for simulating Ising models and in several complex Markov chains for solving problems in computer science. Ingrassia (1990) has used the techniques of the present paper to get bounds on the rate of convergence in simulated annealing.

In Section 1D it is shown how bounds on β_1 translate into bounds on rates of convergence for chains run in continuous time. In Section 1C it is shown how to use similar ideas to bound the smallest eigenvalue and so the spectral gap, $1 - \beta^*$.

An important special case occurs when the Markov chain is the random walk on a graph. That is, let $G = (X, E)$ be an undirected graph with vertex set X and edge set E. We assume that G is connected and simple, i.e., that G has no loops or multiple edges. A random walk begins at an initial vertex x_0 and thereafter proceeds by choosing a neighboring vertex with uniform probability. Thus, if $d(x)$ is the degree of x, then

$$(1.6) \qquad P(x,y) = \begin{cases} \dfrac{1}{d(x)} & \text{if } \{x,y\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Since the graph is connected, the chain is irreducible. It is clear that the chain is reversible with respect to

$$\pi(x) = \frac{d(x)}{2|E|}, \qquad x \in X.$$

Hence in this case

$$Q(x,y) = \begin{cases} \dfrac{1}{2|E|} & \text{if } \{x,y\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

The estimate of Proposition 1 specializes to the following.

Corollary 1. *Let (X, E) be a connected graph. Then, for any choice of Γ,*

$$\beta_1 \le 1 - \frac{2|E|}{d_*^2 \gamma_* b}$$

with $d_* = \max d(x)$, γ is the maximum number of edges in any $\gamma \in \Gamma$ (γ_* is the diameter of G if paths are chosen as geodesics) and

$$b = \max_e \{\gamma \in \Gamma : e \in \gamma\}.$$

Proof. $\kappa(\Gamma)$ of (1.5) is bounded above by

$$\left(\frac{d_*}{2|E|}\right)^2 2|E| \gamma_* b.$$

Thus the corollary follows immediately from Proposition 1. □

REMARK. The geometric aspects of the bound come out clearly in the graph case. The quantity b is exactly a measure of bottlenecks. If paths can be chosen with small overlap, the bound is good.

C. *Smallest eigenvalue.* The techniques introduced above can be applied to get lower bounds on the smallest eigenvalue β_{m-1}. Toward this end introduce a graph with vertex set X, an edge from x to y if $Q(x,y) > 0$ and a self loop from x to x if $Q(x,x) > 0$. The chain $P(x,y)$ is aperiodic if and only if this graph is not bipartite: The set of vertices cannot be partitioned into disjoint sets such that edges only go from one set to the other. In particular, a connected graph cannot be bipartite if $P(x,x) > 0$ for some x. As is well known, the chain is aperiodic if and only if $\beta_{m-1} > -1$.

Let σ_x be a path from x to x *with an odd number of edges*. Such paths always exist for irreducible aperiodic chains. Let Σ be the collection of paths (one for each x). Define the path length $|\sigma_x|_Q$ by analogy with (1.4). The geometric quantity that appears now is

$$\iota = \iota(\Sigma) = \max_{e} \sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x). \tag{1.7}$$

PROPOSITION 2. *For an irreducible aperiodic Markov chain P the smallest eigenvalue $\beta_{\min} = \beta_{m-1}$ satisfies*

$$\beta_{\min} \geq -1 + \frac{2}{\iota},$$

with ι defined in (1.7).

PROOF. The following simple identity will be used:

$$\tfrac{1}{2}\sum_{x,y} (\phi(x) + \phi(y))^2 Q(x,y) = E(\phi^2) + \langle \phi, P\phi\rangle_{L^2(\pi)}. \tag{1.8}$$

In (1.8), E denotes expectation with respect to the stationary distribution π. The heart of the idea is to express

$$\phi(x) = \tfrac{1}{2}\{(\phi(x) + \phi(y)) - (\phi(y) + \phi(w)) + \cdots + (\phi(z) + \phi(x))\}$$

and use the Cauchy–Schwarz inequality as before. This shows how paths with an odd number of edges enter the argument. To continue, write

$$\begin{aligned} E(\phi^2) &= \sum_x \phi(x)^2\pi(x) = \sum_x \frac{\pi(x)}{4}\left\{\sum_{e\in\sigma_x}\sqrt{\frac{Q(e)}{Q(e)}}\,(-1)^{l(e)}(\phi(e^+) + \phi(e^-))\right\}^2 \\ &\leq \sum_x \frac{\pi(x)}{4}|\sigma_x|_Q \sum_{e\in\sigma_x} (\phi(e^+) + \phi(e^-))^2 Q(e) \\ &= \frac{1}{4}\sum_e (\phi(e^+) + \phi(e^-))^2 Q(e) \sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x) \\ &\leq \frac{\iota}{4}\sum_e (\phi(e^+) + \phi(e^-))^2 Q(e) = \frac{\iota}{2}\left(E(\phi^2) + \langle\phi, P\phi\rangle_{L^2(\pi)}\right). \end{aligned}$$

After the second equality, the sum is over directed edges $e = (e^-, e^+)$ and $l(e)$ is the distance of e^- from x in σ_x. Dividing through by $E(\phi^2)$ gives a lower bound on any eigenvalue of P. □

For a random walk on a graph (1.6) the result specializes to

COROLLARY 2. *Let (X, E) be a connected graph which is not bipartite. Then, for any choice Σ of paths of odd length,*

$$\beta_{\min} \geq -1 + \frac{2}{d_* \sigma_* b_*}$$

with d_* the maximum degree, σ_* the maximum number of edges in any $\sigma \in \Sigma$ and

$$b_* = \max_e \#\{\sigma \in \Sigma : e \in \sigma\}.$$

REMARK 1.1. The upper bound on β_1 in Proposition 1 and the lower bound on $\beta_{\min}$ in Proposition 2 give a bound on the spectral gap $1 - \beta_*$.

REMARK 1.2. Another approach to bounding the smallest eigenvalue uses Proposition 1 on the second eigenvalue of P^2. Sometimes an auxiliary argument gives all eigenvalues positive. Then β_1 is all that is needed.

REMARK 1.3. The argument can be varied. For example, for each $(x, y) \in X \times X$, let σ_{xy} be a path from x to y with an odd number of edges. Let $\tilde{\iota} = \max_e \sum_{\sigma_{xy} \ni e} |\sigma_{xy}|_Q \pi(x)\pi(y)$. Then $\beta_{\min} \geq -1 + 1/\tilde{\iota}$. Since the paths σ_{xy} can also be used in Proposition 1, $\beta_* \leq 1 - 1/\tilde{\iota}$.

D. *Bounds on variation distance.* The variation distance between probabilities μ, π on a finite set X is defined as

$$\|\mu - \pi\|_{\mathrm{Var}} = \max_{A \subset X} |\mu(A) - \pi(A)| = \tfrac{1}{2} \sum_{x \in X} |\mu(x) - \pi(x)|.$$

This probabilist's version is $\frac{1}{2}$ the usual operator norm on measures as the dual of bounded continuous functions. The following arguments give bounds on the variation distance to stationarity for reversible Markov chains in terms of the second largest eigenvalue β_1 and $\beta_* = \max(\beta_1, |\beta_{m-1}|)$. As is well known, in continuous time (say exponential waiting time with rate 1), no parity problems arise and bounds on β_1 are all that is necessary. The transition kernel for continuous time is denoted by

$$P_t(x, y) = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} P^n(x, y), \qquad x, y \in X.$$

PROPOSITION 3. *Let $\pi(x), P(x, y)$ be a reversible Markov chain on a finite set X. Assume P is irreducible with eigenvalues $1 = \beta_0 > \beta_1 \geq \cdots \geq$*

$\beta_{m-1} \geq -1$. *Then for all* $x \in X$, $n \in \mathbb{N}$ *and* $t \in (0, \infty)$,

$$(1.9)\quad 4\|P^n(x,\cdot) - \pi\|_{\mathrm{Var}}^2 \leq \frac{1-\pi(x)}{\pi(x)}\beta_*^{2n} \quad \textit{with } \beta_* = \max(\beta_1, |\beta_{m-1}|),$$

$$(1.10)\qquad 4\|P_t(x,\cdot) - \pi\|_{\mathrm{Var}}^2 \leq \frac{1-\pi(x)}{\pi(x)}\exp[-2(1-\beta_1)t].$$

PROOF.

$$\begin{aligned} 4\|P^n(x,\cdot) - \pi(\cdot)\|^2 &= \left(\sum_y |P^n(x,y) - \pi(y)|\right)^2 \\ &= \left(\sum_y \sqrt{\frac{\pi(y)}{\pi(y)}}\,|P^n(x,y) - \pi(y)|\right)^2 \\ &\leq \sum_y \frac{1}{\pi(y)}\left((P^n(x,y))^2 - 2P^n(x,y)\pi(y) + \pi(y)^2\right) \\ &= \sum_y \frac{1}{\pi(y)}(P^n(x,y))^2 - 1 = \frac{1}{\pi(x)}P^{2n}(x,x) - 1. \end{aligned}$$

Above, the inequality is Cauchy–Schwarz and the identity

$$1/(\pi(x))P^{2n}(x,x) = \sum_y (P^n(x,y))^2/\pi(y)$$

follows from reversibility.

Let D be a diagonal matrix with xth diagonal entry $\sqrt{\pi(x)}$. The matrix DPD^{-1} has (x,y) entry $\sqrt{(\pi(x)/\pi(y))}\,P(x,y)$ and so is symmetric. It can thus be orthogonally diagonalized as $\Gamma B \Gamma^t$, with $\Gamma\Gamma^t = I$ and B a diagonal matrix consisting of the eigenvalues of P. All of the entries of B are real in $(-1,1]$ and one, say B_{zz}, equals 1. Thus the transition matrix P^{2n} can be written as

$$P^{2n} = D^{-1}\Gamma B^{2n}\Gamma^t D.$$

Multiplying out, the (x,y) entry of the matrix is

$$P^{2n}(x,y) = \sqrt{\frac{\pi(y)}{\pi(x)}}\sum_w \Gamma_{xw}B_{ww}^{2n}\Gamma_{yw}.$$

The rows of $\Gamma^t D$ are left eigenvectors of P. Since $\Lambda_{zz} = 1$, the zth row of Γ^t has entries $(\cdots \sqrt{\pi(x)} \cdots)$. It follows that the x, x entry of P^{2n} is

$$\pi(x) + \sum_{w \neq z} B_{ww}^{2n}\Gamma_{xw}^2.$$

Bounding B_{ww}^{2n} by β_*^{2n} and using the orthogonality of Γ, the inequality (1.9) of the lemma follows. The continuous time version can be proved by a virtually identical argument. □

REMARK. There are two places in the preceding argument for error to enter. The first place is our use of the Cauchy–Schwarz inequality in (1.12). Of greater concern is our estimate for $P^{2n}(x,x)$. In general, there is no reason that $\bar{\delta}_x$ will be in the eigenspace of β_1 and unless it is, the estimate for $P^{2n}(x,x)$ can be poor. For example, consider the case when, for each $n \in \mathbb{N}$, $P^n(x,x)$ is independent of $x \in X$. This arises for random walk on a group. Then π is also independent of $x \in X$ and so

$$\begin{aligned} P^{2n}(x,x) - 1 &= \frac{1}{|X|}\sum_{y\in X} P^{2n}(y,y) - 1 = \sum_{y \in X} P^{2n}(y,y)\pi(y) - 1 = \mathrm{tr}(P^{2n}) - 1 \\ &= \sum_{i=1}^{m-1} \beta_i^{2n}. \end{aligned}$$

This is certainly better than the estimate $(m-1)\beta_*^{2n}$ from (1.10) which essentially bounds β_i by β_*. In Section 2 we will discuss, for comparison purposes, some examples for which the entire spectrum is known.

E. *Related bounds*. There is a large literature offering bounds on the spectral gap in terms of the entries $P(x,y)$ and some aspect of the geometry of an associated graph. One promising development, the use of Cheeger-like inequalities, is reviewed in Section 3 and will not be discussed further here.

There are a variety of coefficients of ergodicity which bound $\beta_* = \max_{1 \le i \le m-1} |\beta_i|$. These include Dobrushin's bound

$$\beta_* \le \max_{x,y} \|P_x(\cdot) - P_y(\cdot)\|_{\mathrm{Var}},$$

where $P_x(\cdot)$ is the probability distribution given by the xth row of $P(x,y)$. A refinement is the Deutsch–Zenger bound

$$\beta_* \le \max_{x,y} \|P_x(\cdot) - P_y(\cdot)\| - R(x,y),$$

where

$$\begin{aligned} 2R(x,y) = {} & |P(x,x) - P(y,x)| + |P(x,y) - P(y,y)| \\ & - P(x,x) - P(y,y) + P(x,y) - P(y,x). \end{aligned}$$

Seneta (1981) and Rothblum and Tan (1985) contain extensive reviews of this subject. Horn and Johnson (1985) review the closely related subject of Geršgorin disks. These results are sharp in that there are examples where equality holds. The bounds can be far off. For example, consider simple random walk on a p point circle (Example 2.1). There are many rows of the transition matrix which are singular as measures, so both bounds above give $\beta_* \le 1$. To make use of ergodic coefficients, high powers of $P(x,y)$ must be considered, and approximating the entries of such powers seems like a formidable task.

Landau and Odlyzko (1981) offer a bound for random walk on a connected graph as in (1.6). They show

$$\beta_* \le 1 - \left[|X| d_*(\gamma_* + 1)\right]^{-1} < 1 - \frac{1}{|X|^3},$$

where $|X|$ is the number of vertices in the graph, d_* is the maximum degree and γ_* is the diameter. The first inequality gives the correct order for simple random walk on the circle and they show it is sharp for dumbbell-shaped graphs. For random walk on the d-cube (Example 2.2) it gives a bound of the form $1 - (\text{const.}/d^2 2^d)$ which is quite far from the right answer $(1 - (\text{const.}/d))$. One real advantage of this bound is that it is easy to compute compared to the geometric quantities involved in Propositions 1 and 2.

Very recent work by Milena Mihail and by Jim Fill allows the techniques used in the present paper to be applied to nonreversible chains as well. Mihail (1989) works directly with distance to stationarity, avoiding the use of eigenvalues in a novel way. Fill (1991) translates Mihail's ideas into probabilistic language and relates them to the techniques introduced here. One of the results is the following: If P is an aperiodic irreducible Markov chain on the finite set X, let $M(P) = P\tilde{P}$, with $\tilde{P}(x, y) = P(y, x)\pi(y)/\pi(x)$. This $M(P)$ is reversible with nonnegative eigenvalues and the same stationary distribution π. If $\beta^*(M)$ is the second largest eigenvalue of $M(P)$, Fill shows $4\|P^n(x, \cdot\,) - \pi\|^2 \le (\beta^*(M))^n/\pi(x)$. He offers a variety of examples where $\beta^*(M)$ can be approximated by the geometric techniques of the present paper.

Finally, we must mention that there are a variety of other techniques available for bounding rates of convergence of Markov chains to their stationary distributions. Aldous (1983) and Diaconis (1988) review and illustrate techniques such as coupling and stationary times. Diaconis and Fill (1990) develop a duality theory which gives useful bounds for examples. The question of how these ideas relate to the present paper seems tantalizing.

2. Examples. This section presents some simple examples where bounds are easy to obtain and compare with the exact answer. All of the examples involve random walks on graphs.

EXAMPLE 2.1 (The circle $\mathbb{Z}_p$). Let p be an odd number and consider the integers mod p as p points around a circle. For x and y in $\mathbb{Z}_p$, choose γ_{xy} as the shorter of the two paths from x to y. Here $\pi(x) = 1/p$, $d_* = 2$, $|E| = p$ and $\gamma_* = (p - 1)/2$. By symmetry, any edge has the same number of paths crossing over it. Take the edge from 0 to 1. A point at distance i to the left of 0 is connected to $(p - 1)/2 - i$ points by paths crossing from 0 to 1, $0 \le i \le (p - 3)/2$. Thus

$$b = \sum_{0 \le i \le (p-3)/2} \left(\frac{p-1}{2} - i \right) = \frac{p^2 - 1}{8}.$$

Corollary 1 gives

$$\beta_1 \le 1 - \frac{8p}{(p-1)^2(p+1)}.$$

The eigenvalues of $\mathbb{Z}_p$ are $\cos(2\pi j/p)$, $0 \le j \le p - 1$; see, e.g., Chapter 3C of Diaconis (1988). For p large, $\beta_1 = 1 - 2\pi^2/p^2 + O(1/p^4)$ compared to

$1 - 8/p^2$. In other words, the Poincaré technique gives the right order of magnitude but the constant term is off by a factor of about 2.

To get a lower bound on $\beta_{\min}$, choose σ_x as a clockwise path, going once around, starting and ending at x. The quantities in Corollary 2 are easily seen to be $d_* = 2$, $\sigma_* = p$ and $b_* = p$. The bound becomes $\beta_{\min} \geq -1 + 1/p^2$. This is of the right order for p large.

EXAMPLE 2.2 (The cube $\mathbb{Z}_2^d$). The classical Ehrenfest urn can also be described as nearest neighbor random walk on a d-dimensional cube with vertices the 2^d binary d-tuples. For background on this well-studied model, see Kac (1947), Letac and Takacs (1979) or Diaconis, Graham and Morrison (1989) and the references cited therein. Here $\pi(x) = 1/2^d$, $d_* = d$ and $|E| = d2^{d-1}$. For x and y in $\mathbb{Z}_2^d$, choose γ_{xy} by changing the coordinates where x differs from y to their opposite mod 2, working left to right, one coordinate at a time. Clearly $\gamma_* = d$ and for this choice of paths, $b = 2^{d-1}$. To see this, consider an edge (w, z). These differ in only one coordinate, say the jth. A path γ_{xy} crossing over this edge can begin at any x that coincides with w in coordinates after the $(j-1)$st (2^{j-1} choices) and ends in any y that coincides with z in coordinates up to the jth (2^{d-j} choices). Thus there are 2^{d-1} paths γ_{xy} crossing an edge. Corollary 1 gives

$$\beta_1 \leq 1 - \frac{2}{d^2}.$$

The eigenvalues of $\mathbb{Z}_2^d$ are $1 - 2j/d$ with multiplicity $\binom{d}{j}$, $0 \leq j \leq d$; see, for example, Chapter 3C of Diaconis (1988). Thus the bound here is off by a factor of d. This example appears in Jerrum and Sinclair (1988) in slightly different language.

Note that in this example, the graph is bipartite (after an even number of steps the walk started at 0 is at an even position). A frequently used variation eliminates parity problems by allowing the walk to hold in place with probability $1/(d+1)$ and choose a nearest neighbor with uniform probability. To use the bound in Proposition 2, take σ_x as a self-loop from x to x. The quantity ι of (1.7) equals $(d+1)$ and the bound becomes $\beta_{\min} \geq -1 + 2/(d+1)$. In this example, $\beta_{\min} = -1 + 2/(d+1)$.

REMARK 2.1. The paths chosen above give the best possible value of b, namely 2^{d-1}. To see this, note that there are $d2^d$ oriented edges on the cube. Any choice of paths has $2^d\binom{d}{i}$ ordered pairs of vertices at distance i and so

$$2^d \sum_{i=0}^{d} \binom{d}{i} = d2^{d-1} \text{ edges}.$$

Now the pigeonhole principle implies that some edge must be covered by 2^{d-1} paths.

REMARK 2.2. In investigating potential improvements of the bounds, we considered using random paths Γ_{xy} chosen from among all geodesic paths from x to y. For the cube, if the distance from x to y is j there are $j!$ paths and Γ_{xy} is chosen uniformly from these and independently for every x and y. The argument below shows that random paths do essentially the same as the paths chosen above.

The argument works in exactly the same way for any distance transitive graph $G = (V, E)$. Recall that this means that if x, y and x', y' are the same distance apart, then there is an automorphism taking x to x' and y to y'. Biggs (1974) gives background material, and Saxl (1981) gives lists of candidates. Such graphs are connected.

PROPOSITION 4. *Let G be a distance transitive graph. Then the second largest eigenvalue of the random walk on G is bounded above by*

$$\beta_1 \le 1 - \frac{1}{D} \le 1 - \frac{1}{(\gamma_*)^2},$$

where D is the expected squared distance of a random point in G from a fixed point and γ_* is the diameter of G.

PROOF. For any choice of paths γ_{xy}, the Cauchy–Schwarz inequality shows

$$\operatorname{Var}(\phi) \le \frac{1}{2|V|^2} \sum_e \phi(e)^2 \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|.$$

If γ_{xy} are randomly chosen geodesic paths, the expectation of the inner sum does not depend on e. Averaging over edges,

$$\begin{aligned} E\left\{ \sum_{\gamma_{xy} \ni e} |\gamma_{xy}| \right\} &= \frac{1}{2|E|} E\left\{ \sum_{e, x, y: e \in \gamma_{xy}} |\gamma_{xy}| \right\} = \frac{1}{2|E|} E\left\{ \sum_{x, y} |\gamma_{xy}|^2 \right\} \\ &= \frac{1}{2|E|} \sum_j j^2 \#\{(x, y) \in V \times V: \operatorname{dist}(x, y) = j\} \\ &= \frac{|V|}{2|E|} \sum_j j^2 \#\{y \in V: \operatorname{dist}(x_0, y) = j\} \\ &= \frac{|V|^2}{2|E|} D. \end{aligned}$$

Thus $\operatorname{Var}(\phi) \le D/4|E|\Sigma_e \phi(e)^2 = D\mathscr{E}(\phi, \phi)$ because $Q(x, y) = 1/2|E|$ if $x, y \in V$. $\square$

For the cube $\mathbb{Z}_2^d$, $D = (d(d+1))/4$ by a simple binomial calculation. This gives $\beta_1 \le 1 - 4/(d(d+1))$ which is asymptotically better by a factor of 2 than the $1 - 2/d^2$ achieved by the bound using fixed paths.

Here is a problem where random paths do better than any currently available deterministic paths. The graph consists of the $\binom{n}{k}$ k element subsets of $\{1, 2, \ldots, n\}$. A metric on these k sets is $d(x, y) = k - |x \cap y|$. A graph is formed by connecting sets at distance 1. Random walk on this graph is also known as the Bernoulli–Laplace diffusion model and is the original chain analyzed by Markov. This is a distance transitive graph with

$$D = \frac{1}{\binom{n}{k}} \sum_{j=0}^{k} j^2 \binom{k}{j}\binom{n-k}{j} = k^2\left(\frac{k}{n}\right)^2 + k\frac{k}{n}\left(1 - \frac{k}{n}\right)\left(1 - \frac{k-1}{n-1}\right).$$

The classical case has $n = 2k$. With k large $D = k^2/4 + O(k)$, so Proposition 3 gives

$$\beta_1 \leq 1 - \frac{4}{k^2} + O\left(\frac{1}{k^3}\right).$$

Diaconis and Shahshahani (1987) determine all the eigenvalues for this chain. In particular, if $n = 2k$, $\beta_1 = 1 - 2/k$. The best deterministic paths we know give $\beta_1 \leq 1 - C/k^3$ for a constant C.

Proposition 4 carries over as stated to graphs with automorphism groups acting transitively on the set of oriented edges. Aldous (1987) gives a similar result for graphs with automorphism groups acting transitively on vertices. In unpublished work, Fill has shown that Proposition 4 holds for distance regular graphs.

REMARK 2.3. The best Poincaré upper bound on β_1 for nearest neighbor random walk on $\mathbb{Z}_2^d$ is of the wrong order of magnitude. For this example, there is a further idea that gives the correct answer. After any number of steps, random walk on the cube is uniform over level sets with a constant number of ones. Thus the rate of convergence is the same as for the "distance chain" which records the distance from zero. Of course, this distance chain is simply the original Ehrenfest chain. What we will now show is that the Poincaré inequalities applied to the distance chain give the correct bounds for the second eigenvalue.

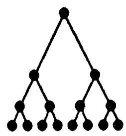
The distance chain has state space $\{0, 1, 2, \ldots, d\}$. The transition probability is $p(j, j-1) = j/d$, $p(j, j+1) = 1 - j/d$. The stationary probability is $\pi(j) = \binom{d}{j}/2^d$. Here, there is a unique path from x to y. One verifies easily that the maximum for both K and κ occurs for the "middle edge" $d/2 - 1$ to $d/2$ if d is even and $(d-1)/2$ to $(d+1)/2$ if d is odd. From here, it is a straightforward if tedious exercise to bound both K and κ. The results are

$$\kappa = \frac{1}{4} d \log d(1 + o(1)), \qquad K = \frac{d}{2}.$$

This is instructive in providing an example where the two versions of the Poincaré inequality differ. The version involving K being better—it gives the right rate with the right constant. Jim Fill has shown us examples where κ does better. The matter needs further investigation.

The main point is that using symmetry helps dramatically here; using the Poincaré inequality directly on the cube gives $\beta_1 \le 1 - 4/d(d+1)$, while using it on the orbit chain gives $1 - 2/d$.

EXAMPLE 2.3 (Trees). Consider the full binary tree of depth d.



For $d \ge 1$, such a tree has $2^{d+1} - 1$ vertices, $2^{d+1} - 2$ edges and the maximum degree is 3. Consider the Markov chain arising from nearest neighbor random walk on this tree. Each pair of points is connected by a unique path. The longest path is of length $2d$ and an edge connected to the root vertex is covered by

$$b = (2^d - 1)2^d$$

paths. The bound of Corollary 1 is therefore

$$\beta_1 \le 1 - \frac{1}{9d2^{d-1}}.$$

It can be shown that $\beta_1 = 1 - (1 + o(1))/2^{d+2}$.

REMARK 2.4. For a lower bound, take

$$\phi(x) = \begin{cases} -1 & \text{on the left subtree,} \\ 0 & \text{on the central vertex,} \\ 1 & \text{on the right subtree.} \end{cases}$$

This has

$$\mathscr{E}(\phi, \phi) = \frac{1}{2(2^d - 1)},$$

$$\operatorname{Var}(\phi) = 1 - \frac{1}{2^{d+1} - 2}.$$

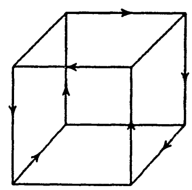
Now (1.2) gives

$$1 - \frac{1}{2^{d+1}} \sim 1 - \frac{1}{2(2^d - 1)\left(1 - \dfrac{1}{2^{d+1} - 2}\right)} \le \beta_1.$$

This bound will be compared with other approaches in Example 3.3.

REMARK 2.5. Similar bounds hold for less symmetric trees. The techniques involved are reasonably robust. Bounds for trees provide crude bounds for any connected graph by using a spanning tree. Unfortunately, trees have "bottlenecks" which lead to extremely weak bounds.

As an example, consider the cube $\mathbb{Z}_2^d$. This has a spanning path, e.g.,

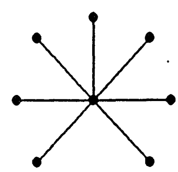


For d large, the bound of Corollary 1 based on paths gives

$$\beta_1 \leq 1 - \frac{4}{d^2 2^d}.$$

This is exponentially worse than the bounds derived in Example 2.2.

A second tree example which we find instructive is a "star" with a central vertex and n outside vertices:



$n = 7$

Random walk on this graph has eigenvalues 1, 0 and -1, with 0 having multiplicity $n - 1$. To cure periodicity, consider the Markov chain that holds with probability θ at every point. This has eigenvalues 1, θ and $2\theta - 1$, with θ having multiplicity $n - 1$.

The symmetry group of this graph operates transitively on the edges. The stationary distribution puts mass $1/2$ at the central vertex and mass $1/2n$ at each outside vertex. The quantity κ in the Poincaré bound of Proposition 1 is

$$\kappa = \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_Q \pi(x)\pi(y) = \frac{3n - 2}{2n} \frac{1}{1 - \theta}.$$

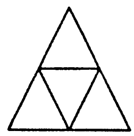
The bound from Proposition 1 becomes

$$\beta_1 \leq 1 - \frac{(1 - \theta)2n}{3n - 2}.$$

This is uninformative if θ is small (e.g., $\theta = 1/n$).

This example shows that the bound of Corollary 1 can be far from $1 - (1/\kappa)$. Corollary 1 gives $\beta_1 \leq 1 - 2|E|/d_*^2 \gamma_* b$ which becomes $\beta_1 \leq 1 - 1/n^2$ in the present case.

EXAMPLE 2.4 (Equilateral subdivision). The next example arose in an application. Ulf Grenander needed to put a grid of points on the surface of the usual sphere S_2 in three dimensions. He began with an icosahedron. This has faces which are equilateral triangles. Consider one face. If the midpoints of the face are connected, four equilateral triangles result. Connecting their midpoints, and continuing recursively, gives a sequence of triangular subdivisions. Grenander suggested carrying out such a subdivision of each face of the icosahedron, and then projecting the vertices of the graph obtained onto the surface of a circumscribing sphere.

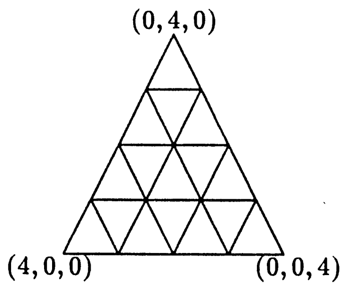


This grid is used as part of a smoothing algorithm. To analyze its asymptotic behavior, the second eigenvalue is needed. In the present example, the second eigenvalue of a single face is approximated. As explained at the end of the example, the same argument gives the same rate (up to constants) for the icosahedron.

Let G_n be the graph with vertex and edge sets

$$V_n = \{(x_1, x_2, x_3): 0 \le x_i \le n,\ x_1 + x_2 + x_3 = n,\ x_i \in \mathbb{N}\},$$
$$E_n = \{\{(x_1, x_2, x_3), (x_1', x_2', x_3')\}: |x_i - x_i'| \le 1\}. \tag{2.1}$$

Thus G_4 appears as



When n is a power of 2, G_n is the result of successive equilateral subdivision. It is easy to see that

$$|V_n| = \frac{(n+1)(n+2)}{2},$$
$$|E_n| = \frac{3n(n+1)}{2}.$$

PROPOSITION 5. *If β_1 is the second largest eigenvalue of G_n defined in* (2.1), *then*

$$1 - \frac{9}{n^2} \le \beta_1 \le 1 - \frac{1}{12n(n+2)}.$$

PROOF. Upper bound. For $x, y \in V_n$, let γ_{xy} be the line from x to y, if it exists, or the unique shortest path from x to y with one 60° counterclockwise turn:

Fix an edge e. Paths that cross e have at most one turn. There are at most $n|V_n|$ such paths with the turn at or before e and at most $n|V_n|$ such paths with the turn after e. Thus at most $n(n+1)(n+2)$ paths cross any edge. The maximum path length is n; the maximum degree is 6. The upper bound now follows from Corollary 1 of Section 2.

Lower bound. A lower bound can be derived by bounding $\mathscr{E}(\phi,\phi)/\mathrm{Var}(\phi)$ for any specific ϕ. The graph G_n has vertex set $\{(x_1, x_2, x_3)\colon x_1 + x_2 + x_3 = n\}$. Take $\phi(x) = x_1 - x_2$ so $\sum_x \phi(x) = 0$. Now

$$\mathscr{E}(\phi,\phi) = \frac{1}{2}\sum_e \phi^2(e)Q(e) = \frac{1}{4|E_n|}\left\{4\frac{|E_n|}{3} + \frac{2|E_n|}{3}\right\} = \frac{1}{2}$$

because $Q(e) = 1/2|E_n|$ and $\phi^2(e)$ takes value 4 on $\frac{1}{3}$ of the edges and 1 on $\frac{2}{3}$ of the edges.

For the variance, write

$$\begin{aligned}\mathrm{Var}(\phi) &= \frac{1}{2|E_n|}\sum_x (x_1 - x_2)^2 \deg(x) \geq \frac{1}{|E_n|}\sum_x (x_1 - x_2)^2 \\ &= \frac{1}{12|E_n|}(n^4 + O(n^3)) \\ &= \frac{n^2}{18} + O(n).\end{aligned}$$

Thus, from this choice of ϕ, $\mathscr{E}(\phi,\phi)/\mathrm{Var}(\phi) < 9/n^2 + O(1/n^3)$. The lower bound part of the proposition follows. □

REMARK. The argument above is based on the ideas of Richard Stong (private communication). He has carried out the calculations more carefully and shown that

$$1 - \frac{6}{n^2} + O\left(\frac{1}{n^3}\right) \leq \beta_1 \leq 1 - \frac{8}{7n^2} + O\left(\frac{1}{n^3}\right).$$

Using similar arguments for the triangulation of the full icosohedron with each face a G_n, Stong has shown that random walk on this graph has second largest eigenvalue satisfying

$$1 - \frac{C_1}{n^2} \leq \beta_1 \leq 1 - \frac{C_2}{n^2} + O\left(\frac{1}{n^3}\right) \quad \text{for explicit constants } C_1, C_2.$$

3. Cheeger's inequality.

A. *Introduction.* Let X be a finite set and $P(x, y)$ an ergodic Markov chain with stationary distribution π. Assume P is reversible and define a probability Q by

$$Q(x, y) = \pi(x)P(x, y) = \pi(y)P(y, x).$$

Inequalities on the second largest eigenvalue of P have been derived in terms of the geometric quantity

$$h = \min_{\pi(S) \leq 1/2} \frac{Q(S \times S^C)}{\pi(S)}, \tag{3.1}$$

where S^C denotes the compliment of S. Heuristically, $Q(S \times S^C)/\pi(S)$ is a measure of the relative flow out of S when the chain is in stationarity. If this is large for all S, the Markov chain should converge to π rapidly since there are no bottlenecks. This is made precise in the following result.

PROPOSITION 6 (Cheeger's inequality). *Let β_1 be the second largest eigenvalue of a reversible, ergodic Markov chain. Then*

$$1 - 2h \leq \beta_1 \leq 1 - h^2$$

with h defined by (3.1).

A short proof, along with references to work of Cheeger, Alon, Alon and Milman, Dodziuck and others is given in Section 3C. Early applications of Cheeger's inequality began with a symmetric graph where group theory could be used to bound β_1. This gave a bound on the expansion coefficient h and allowed construction of "expanders." Lubotzky (1989) gives a highly readable survey.

The point of view taken here is that h can sometimes be bounded directly, thus giving bounds on β_1.

EXAMPLE 3.1. Let p be an odd number and consider the graph $\mathbb{Z}_p$ introduced in Example 2.1. This has $P(x, y) = 1/2$ if $|x - y| = 1$ and $\pi(x) = 1/p$. If S is an interval $[a, b]$, $Q(S \times S^C) = 1/p$, $\pi(S) = |S|/p$. Elementary considerations show that h is achieved by taking S as an interval of size $(p - 1)/2$, so

$$h = \frac{2}{p - 1}.$$

Here $\beta_1 = \cos(2\pi/p) = 1 - 2\pi^2/p^2 + O(1/p^4)$ compared with the Cheeger bounds $1 - 4/(p - 1) \leq \beta_1 \leq 1 - 2/(p - 1)^2$. Thus the upper bound gives the right order but the constant is off by a factor of π^2. The bound is slightly worse than the Poincaré inequality of Section 2.

EXAMPLE 3.2. Let $\mathbb{Z}_2^d$ be the graph of the cube considered in Example 2.2. Thus $P(x, y) = 1/d$ if x and y differ in precisely one bit and $\pi(x) = 1/2^d$. An inductive argument shows that h is achieved by taking S to be the face $\{x\colon x_1 = 0\}$. This gives

$$h = \frac{1}{d}.$$

Here $\beta_1 = 1 - 2/d$, compared with the Cheeger bounds $1 - 2/d \le \beta_1 \le 1 - 1/2d^2$. Thus the lower bound is sharp and the upper bound is of the same order as the Poincaré bound but with a slightly worse constant. A different proof for the value of h appears in Section 3B.

EXAMPLE 3.3. Consider the full binary tree of depth d. An elementary argument shows that h is achieved by taking S to be all vertices in the left-hand subtree (excluding the root). This gives

$$h = \frac{1}{2^{d+1} - 3}.$$

Here the second largest eigenvalue satisfies $\beta_1 = 1 - 1/2^{d-2}(1 + o(1))$. Cheeger's inequality gives

$$1 - \frac{2}{2^{d+1} - 3} \le \beta_1 \le 1 - \frac{1}{2(2^{d+1} - 3)^2}.$$

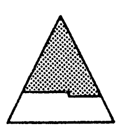
The Poincaré inequality of Example 2.3 gives

$$1 - \frac{1}{2^{d+1}} + O\left(\frac{1}{2^{2d}}\right) \le \beta_1 \le 1 - \frac{1}{9d2^{d-1}}.$$

For large d, the Poincaré upper bound is much smaller.

For a star with n outer vertices and holding probability θ, $h = 1 - \theta$. The associated Cheeger bound is $2\theta - 1 \le \beta_1 \le 1 - (1 - \theta)^2/2$. Here, $\beta_1 = \theta$ and, for θ small (e.g., $1/n$), both sides of the bound are uninformative.

EXAMPLE 3.4. Let G_n be the triangulated equilateral triangle introduced in Example 2.4. By a clever combinatorial argument Richard Stong has determined the set S where h is achieved. He shows S may be taken to be a convex set containing roughly half the points and meeting two sides and one vertex.



He proves

$$h(G_n) = \min\left\{\frac{2}{3m - \delta}, \frac{m + 1}{\left\lceil \frac{3n(n+1)}{4} \right\rceil}\right\}$$

with

$$m = \left\lfloor \sqrt{\frac{n(n+1)}{2} + \frac{1}{36}} + \frac{1}{6} \right\rfloor,$$

$$\delta = \begin{cases} 1 & \text{if } (m+1)(3m+2) - \dfrac{3n(n+1)}{2} = 2, 6 \text{ or } 12, \\ 0 & \text{otherwise.} \end{cases}$$

This gives the Cheeger constant as $h = 2\sqrt{2}/3n + O(n^{-2})$ and the bound

$$1 - \frac{4\sqrt{2}}{3n} + O\left(\frac{1}{n^2}\right) \le \beta_1 \le 1 - \frac{4}{9n^2} + O\left(\frac{1}{n^3}\right).$$

Stong has shown that the Poincaré technique gives $1 - 8/7n^2 + O(1/n^3)$ as an upper bound for β_1.

B. *Jerrum's and Sinclair's canonical paths.* Computing h seems difficult in general. In a sequence of papers, Jerrum and Sinclair have shown how a simple geometric idea gives a bound on h. Let X be a finite set and $P(x, y)$ be the transition probability for a reversible, irreducible Markov chain with stationary distribution $\pi(x)$ and reversing measure $Q(x, y) = \pi(x)P(x, y)$. For each ordered pair (x, y), let γ_{xy} be a path connecting x to y. The following geometric quantity arises:

$$\eta = \max_e Q(e)^{-1} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y). \tag{3.2}$$

Here the max is over all oriented edges $e = (x, y)$ and the sum is over all paths containing e. Proposition 7 is a version of Jerrum's and Sinclair's result.

PROPOSITION 7. *For a reversible, irreducible Markov chain P, the second largest eigenvalue satisfies*

$$\beta_1 \le 1 - \frac{1}{8\eta^2}$$

with η defined by (3.2).

PROOF. Associate with the path γ_{xy} the weight $\pi(x)\pi(y)$. Let $S \subset X$ be any set with $\pi(S) \le \frac{1}{2}$ and let W denote the aggregated weight of all paths which cross the cut from S to S^C. Clearly $W = \pi(S)\pi(S^C) \ge \frac{1}{2}\pi(S)$. Summing over cut edges $e \in \partial S$ gives the complementary bound

$$W \le \sum_{e \in \partial S} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) \le \eta \sum_{e \in \partial S} Q(e).$$

Combining the two bounds on W gives $h \ge 1/2\eta$ and Cheeger's inequality completes the proof. □

REMARK 3.1. For random walk on a connected graph $G = (V, E)$ the bound becomes $\beta_1 \le 1 - \frac{1}{2}(1/4\eta|E|)^2$. If d_* denotes the maximum degree in G, β_1 is bounded above by

$$1 - \frac{1}{2}\left(\frac{|E|}{d_*^2 b}\right)^2 \tag{3.3}$$

with $b = \max_e \#\{\gamma_{xy}: e \in \gamma_{xy}\}$ as in the Poincaré inequality of Corollary 1. The bound there was $\beta_1 \le 1 - 2|E|/d_*^2 0\gamma_* b$. As was shown in Section 2, usually the Poincaré inequality is smaller. For graphs, this happens if and only if $\gamma_*|E| \le 4d_*^2 b$. We have been unable to decide if this inequality, perhaps with the right side multiplied by 2, always holds.

REMARK 3.2. For highly symmetric graphs it is possible to use random paths to prove the following result.

PROPOSITION 8. *Let G be a distance transitive graph. Then the second largest eigenvalue of random walk on G is bounded above by*

$$\beta_1 \le 1 - \frac{1}{8d^2} \le 1 - \frac{1}{8\gamma_*^2},$$

where d is the expected distance of a random point on G from a fixed point and γ_* is the diameter of G.

PROOF. From the proof of Proposition 7 specialized to this case, for any choice of paths γ_{xy} and any set $S \subset V$ with $\pi(S) \le \frac{1}{2}$,

$$\pi(S)\pi(S^C) \le \frac{2|E|}{|V|^2} \sum_{e \in \partial S} \sum_{\gamma_{xy} \ni e} 1,$$

where ∂S is the set of $e = (x, y)$ with $x \in S$, $y \in S^C$. Let γ_{xy} be chosen as random geodesics from x to y. The expected value of the inner sum on the right is independent of the oriented edge e. It can be written as

$$E\left(\sum_{\gamma_{xy} \ni e} 1\right) = \frac{1}{2|E|} E\left(\sum_{\substack{e,\gamma \\ e \in \gamma_{xy}}} 1\right) = \frac{1}{2|E|} \sum_{\gamma} |\gamma| = \frac{1}{2|E|} \sum_j j|\{x, y\}: d(x, y) = j\}|$$

$$= \frac{|V|}{2|E|} \sum_j j|\{y: d(x_0, y) = j\}| = \frac{d|V|^2}{2|E|}.$$

Thus $\pi(S)\pi(S^C) \le Q(S \times S^C)d$, so $h \ge 1/(2d)$ and the result follows from Cheeger's inequality. □

In Proposition 3, the Poincaré technique was used to prove $\beta_1 \le 1 - 1/D \le 1 - 1/\gamma_*^2$. The diameter bound from the Poincaré inequality is better by the factor 8. There is no clear comparison for the bounds involving expected values since $d^2 \le D$.

The argument above showed $h \geq 1/2\gamma_*$. Aldous [(1987), page 39] showed $h \geq 1/2\gamma_*$ for Cayley graphs of groups. It is not clear how much symmetry is required for such a bound. Mohar (1989a, b) contains several results for general graphs which shed light on these inequalities.

We now briefly run through the four examples, using the canonical paths described in Section 2.

EXAMPLE 3.5 (The circle $\mathbb{Z}_p$). The bound (3.3) becomes $1 - 2p^2/(p^2 - 1)^2$. For p large, this is the best that can be done from Cheeger's inequality $(h = 2/(p-1))$, of the right order of magnitude $(\beta_1 = 1 - 2\pi^2/p^2 + O(1/p^4))$ and slightly worse than the $1 - 8/p^2$ from Poincaré's inequality.

EXAMPLE 3.6 (The cube $\mathbb{Z}_2^d$). The bound (3.3) becomes $1 - 1/2d^2$. This is the best that can be done using Cheeger's inequality $(h = 1/d)$, off by a factor of $d(\beta_1 = 1 - 2/d)$ and slightly worse than the $1 - 2/d^2$ from Poincaré's inequality.

EXAMPLE 3.7 (Binary trees). For a binary tree of depth d as in Example 3.3, the bound (3.3) becomes $1 - 2/81d2^{2d}$. This is asymptotically the same order of magnitude as the Cheeger bound, but far worse than the bound from Poincaré's inequality $(\beta_1 \leq 1 - 1/9d2^{d-1})$.

EXAMPLE 3.8 (Triangulated equilateral triangle). Here all bounds give the same answer: $\beta_1 \leq 1 - C/n^2$. Again the Poincaré inequality gives a slightly better constant than the Cheeger inequality.

C. *A short proof of Cheeger's inequality.* In this section we give a brief explanation of Cheeger's inequality for reversible Markov chains on a finite state space. Our notation will be the same as it was in Section 1. Thus, what we want to do is derive the estimate

$$(3.4)\quad \frac{h^2}{2} \leq \lambda_1 \leq 2h \quad \text{where} \quad h \equiv \inf\left\{\frac{Q(S \times S^c)}{\pi(S)} : 0 < \pi(S) \leq \frac{1}{2}\right\}.$$

The upper bound in (3.4) is very nearly a triviality. Namely, given $S \subseteq X$ with $0 < \pi(S) \leq \frac{1}{2}$, set

$$\psi_S(x) = \begin{cases} \pi(S^c) & \text{for } x \in S, \\ -\pi(S) & \text{for } x \in S^c, \end{cases}$$

note that ψ_S has π-mean-value 0 and use (1.2) to conclude that

$$\lambda_1 \leq \frac{\mathscr{E}(\psi_S, \psi_S)}{\operatorname{Var}(\psi_S)} = \frac{Q(S \times S^c)}{\pi(S)\pi(S^c)} \leq 2\frac{Q(S \times S^c)}{\pi(S)}.$$

The lower bound in (3.4) is more challenging. Indeed, it rests on two observations which are interesting in their own right. To explain the first of these, for any $\psi \in L^2(\pi)$ use ψ_+ to denote $\psi \vee 0$, the positive part of ψ, and set $S(\psi) = \{x \in X: \psi(x) > 0\}$. Assuming that $S(\psi) \neq \varnothing$, the first observation is that, for any $\psi \in L^2(\pi)$ and $\lambda \in [0, \infty)$,

$$\lambda \|\psi_+\|^2_{L^2(\pi)} \geq \mathscr{E}(\psi_+, \psi_+) \quad \text{if } L\psi \leq \lambda \psi \text{ on } S(\psi). \tag{3.5}$$

To see (3.5), simply note that

$$\lambda \|\psi_+\|^2_{L^2(\pi)} \geq (\psi_+, L\psi)_{L^2(\pi)} = \mathscr{E}(\psi_+, \psi) \geq \mathscr{E}(\psi_+, \psi_+),$$

where we have used polarization to see that

$$(\phi, L\psi)_{L^2(\pi)} = \mathscr{E}(\phi, \psi) \equiv \tfrac{1}{2} \sum_{x, y \in X} (\phi(y) - \phi(x))(\psi(y) - \psi(x)) Q(x, y)$$

$$\text{for } \phi, \psi \in L^2(\pi)$$

and have used

$$(\psi_+(y) - \psi_+(x))(\psi(y) - \psi(x)) \geq (\psi_+(y) - \psi_+(x))^2$$

to get the last inequality. [Simple as (3.5) may be, it makes essential use of the structure of L as reflected in the properties of Dirichlet forms.]

The second observation underlying the lower bound in (3.4) is that, for any $\psi \in L^2(\pi)$ with $S(\psi) \neq \varnothing$,

$$\mathscr{E}(\psi_+, \psi_+) \geq \frac{h(\psi)^2 \|\psi_+\|^2_{L^2(\pi)}}{2} \quad \text{where}$$

$$h(\psi) \equiv \inf\left\{ \frac{Q(S \times S^c)}{\pi(S)} : \varnothing \neq S \subseteq S(\psi) \right\}. \tag{3.6}$$

In proving (3.6), we may and will assume that $\psi \geq 0$ everywhere. Next, by an application of the Cauchy–Schwarz inequality, we write

$$\begin{aligned}
\sum_{x, y} |\psi^2(x) - \psi^2(y)| Q(x, y) &\leq \sqrt{2}\, \mathscr{E}(\psi, \psi)^{1/2} \left\{ \sum_{x, y} (\psi(x) + \psi(y))^2 Q(x, y) \right\}^{1/2} \\
&\leq 2\mathscr{E}(\psi, \psi)^{1/2} \left\{ \sum_{x, y} \left(\psi(x)^2 + \psi(y)^2\right) Q(x, y) \right\}^{1/2} \\
&= 2^{3/2} \mathscr{E}(\psi, \psi)^{1/2} \|\psi\|_{L^2(\pi)}.
\end{aligned}$$

At the same time, the left side of the above inequality can be written as

$$\begin{aligned}
2 \sum_{\psi(y) > \psi(x)} (\psi^2(y) - \psi^2(x)) Q(x, y) &= 4 \sum_{\psi(y) > \psi(x)} \left(\int_{\psi(x)}^{\psi(y)} t\, dt \right) Q(x, y) \\
&= 4 \int_0^\infty t \left(\sum_{\psi(x) \leq t < \psi(y)} Q(x, y) \right) dt,
\end{aligned}$$

which, because

$$\sum_{\psi(x) \leq t < \psi(Y)} Q(x, y) = Q(S \times S^c) \quad \text{with } S \equiv \{x : \psi(x) > t\} \subseteq S(\psi),$$

shows that

$$\int_0^\infty t\Bigg(\sum_{\psi(x)\le t<\psi(y)} Q(x,y)\Bigg)\,dt \ge h(\psi)\int_0^\infty t\pi(\{x: \psi(x) > t\})\,dt$$

$$= \frac{h(\psi)\|\psi\|^2_{L^2(\pi)}}{2}.$$

By combining (3.5) with (3.6), we arrive at

$$\lambda \ge \frac{h(\psi)^2}{2} \quad \text{if } L\psi \le \lambda\psi \text{ on } S(\psi), \tag{3.7}$$

for any $\lambda \in [0,\infty)$ and any $\psi \in L^2(\pi)$ with $S(\psi) \ne \varnothing$. To get the lower bound in (3.4) from here, take $\lambda = \lambda_1$ and ψ to be a normalized eigenfunction for λ_1. Because ψ must have π-mean-value 0, we can always arrange that $0 < \pi(S(\psi)) \le \frac{1}{2}$ and therefore that $h(\psi) \ge h$. Hence, the desired lower bound comes directly from (3.7) with this choice of λ and ψ. □

HISTORICAL REMARK. Cheeger's inequality was originally proved as a lower bound for the eigenvalues of the Laplacian on a compact Riemannian manifold. A host of mathematicians have refined and applied these ideas. Fiedler (1973), Alon (1986), Alon and Milman (1985) and Dodziuck (1984) developed geometric inequalities for Markov chains using a variety of closely related geometric quantities. The argument we have given above is a modification of an argument in Sinclair and Jerrum (1989).

There are further refinements possible for the upper bound. For example, F. R. K. Chung (1989) and P. Doyle (1989) (personal communications) have shown

$$\beta_1 \le \sqrt{1-h^2}.$$

A more careful history and extension of these ideas to Markov processes is given by Lawler and Sokal (1988).

4. Approximating the permanent. This final section treats a complex example of interest in theoretical computer science. Let A be an $n \times n$ matrix with 0–1 entries a_{ij}. The permanent of A is defined just like the determinant but without sign,

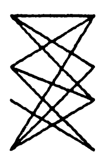
$$\text{Per}(A) = \sum_\sigma \prod_{i=1}^n a_{i\sigma}(i).$$

The permanent counts the number of permutations σ consistent with the restrictions imposed by A. The best available algorithm for computing Per(A) takes order $n2^n$ steps. Valiant (1979) has proved that computing permanents is $\sharp p$ complete and so equivalent to a host of other currently intractable problems. It is unlikely a faster algorithm will become available soon.

Broder (1986) introduced a stochastic algorithm for approximating Per(A) for matrices which are symmetric and dense in the sense that each row and each column contains at least $n/2$ nonzero entries. Jerrum and Sinclair (1989) analyzed Broder's algorithm. The central part of their analysis proves that an associated Markov chain converges sufficiently rapidly to give an approximation of $\operatorname{Per}(A)$ using a number of steps to within a factor of $1 \pm \varepsilon$, which is bounded by a polynomial in n and $1/\varepsilon$. This is an important result in theoretical computer science as the first example of a realistic, provably hard problem with a provably polynomial approximation.

Jerrum and Sinclair introduced the path arguments described in Section 3 to solve this problem. We show how the Jerrum–Sinclair construction coupled with the Poincaré inequality gives an improved rate of convergence. The original bound on the second largest eigenvalue $1 - C/n^{12}$ can be reduced to $1 - C/n^7$.

Jim Fill showed that the Poincaré inequality would give improved results. To describe his result, we work with an equivalent formulation in terms of matchings. Let $G = (V_1, V_2, E)$ be a bipartite graph with $|V_1| = |V_2| = n$ and $E \subseteq V_1 \times V_2$. A matching in G is a set of edges of G, no pair of which shares an endpoint. A perfect matching contains n edges. If the vertices in one set are boys and the second set girls and if an edge indicates approval, then a perfect matching "marries" all of the boys and girls in such a way that each person approves of his or her partner.



Given a bipartite graph, let $A_{ij} = 1$ if $(i, j) \in E$ and 0 otherwise. Clearly $\operatorname{Per}(A)$ counts the number of perfect matchings. Let M_n be the set of perfect matchings and let M_{n-1} be the set of matchings containing $n - 1$ edges. Broder's algorithm constructs a Markov chain on $X = M_n \cup M_{n-1}$.

If the process is at x, the next step is determined by choosing an edge (u, v) in the original bipartite graph uniformly at random. Then:

(4.1)

(a) If $x \in M_n$ and $(u, v) \in x$, delete (u, v) from x.
(b) If $x \in M_{n-1}$ and u and v are unmatched in x, add (u, v) to x.
(c) If $x \in M_{n-1}$, $(u, w) \in x$, and v is unmatched in x, add (u, v) and delete (u, w).
(d) If $x \in M_{n-1}$, $(w, v) \in x$ and u is unmatched in x, add (u, v) and delete (w, v).
(e) In all other cases stay at x.

Broder (1986) showed that this is a connected, symmetric Markov chain and that it converges to the uniform distribution on X. This allows one to choose points in M_n uniformly to good approximation and Broder showed how to convert this into a good estimate of the size of M_n.

To show that the algorithm outlined above is efficient, the rate of convergence of the chain described above must be bounded.

PROPOSITION 9. *Let a bipartite graph with each vertex of degree at least $n/2$ be given on two sets of n vertices. For the Markov chain described in* (4.1), *the second eigenvalue is bounded above by*

$$\beta_1 \leq 1 - \frac{1}{6n^7}.$$

PROOF. The argument uses canonical paths constructed by Jerrum and Sinclair (1988). We refer to their paper for details. The maximum degree of the graph associated to X is $d_* \leq n^2$. The minimum degree is bounded below by $n - 1$. It follows that $|E| \geq (n-1)|X|$.

Jerrum and Sinclair show that $b \leq 3n^4|X|$. To complete the analysis the longest path length γ_* must be bounded. From the Jerrum–Sinclair construction, the worst case that can arise is connecting two almost matchings. These are connected to well defined closest-matchings (at most length 2) and then these matchings are connected by an unwinding algorithm. This takes at most $2n$ steps, so $\gamma_* \leq 2(n+1)$.

The chain here has substantial holding probability, so Proposition 1 must be used directly (rather than Corollary 1). It yields

$$\beta_1 \leq 1 - \frac{1}{6n^7}. \qquad \square$$

REMARK. This example brings out the really new aspect of Jerrum and Sinclair's ideas. In the application, they are trying to estimate $|X|$ which in principle appears in the upper bound. They bound b by constructing a 1–1 map from the paths covering an edge into the set of vertices crossed with some extra information. This gives $b \leq 3n^4|X|$ (the $3n^4$ being the "extra information").

Acknowledgments. We thank David Aldous, Laszlo Babai, Fan Chung, Jim Fill, Anil Gangolli, Ulf Grenander, Alistair Sinclair, Laurent Salloff, Laurie Smith and Richard Stong for help with this paper.

REFERENCES

ALDOUS, D. (1983). Random walk on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII. Lecture Notes in Math.* **986** 243–297. Springer, New York.

ALDOUS, D. (1987). On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probab. Engrg. Inform. Sci.* **1** 33–46.

ALON, N. (1986). Eigenvalues and expanders. *Combinatorica* **6** 83–96.

ALON, N. and MILMAN, V. D. (1985). Isoperimetric inequalities for graphs and superconcentrators. *J. Combin. Theory Ser. B* **38** 73–88.

BANDLE, C. (1980). *Isoperimetric Inequalities and Applications*. Pitman, Boston.

BIGGS, W. (1974). *Algebraic Graph Theory*. Cambridge Univ. Press.

BRODER, A. Z. (1986). How hard is it to marry at random? On the approximation of the permanent. *Proc. 18th ACM Symp. Theoretical Computation* 50–58. [Erratum: *Proc. 20th ACM Symp. Theoretical Computation* (1988) 551.]

DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, Calif.

DIACONIS, P. and FILL, J. A. (1990). Strong stationary times via a new form of duality. *Ann. Probab.* **18** 1483–1522.

DIACONIS, P., GRAHAM, R. and MORRISON, J. (1989). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures and Algorithms* **1**. To appear.

DIACONIS, P. and SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J. Math. Anal.* **18** 208–218.

DODZIUCK (1984). Difference equations, isoperimetric inequality, and transience of certain random walks. *Trans. Amer. Math. Soc.* **284** 787–794.

FIEDLER, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J.* **23** 298–305.

FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 64–87.

HOLLEY, R. and STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 553–569.

HORN, R. and JOHNSON, C. (1985). *Matrix Analysis*. Cambridge Univ. Press.

INGRASSIA, S. (1990). Ph.D. dissertation, Dipartimento di Matematica, Citta Univ., Catania, Italy.

JERRUM, M. and SINCLAIR, A. (1988). Fast uniform generation of regular graphs. Technical Report, Dept. Computer Science, Univ. Edinburgh.

JERRUM, M. and SINCLAIR, A. (1989). Approximating the permanent. *SIAM J. Comput.* **18** 1149–1178.

KAC, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54** 369–391.

LANDAU, H. and ODLYZKO A. (1981). Bounds for eigenvalues of certain stochastic matrices. *Linear Algebra Appl.* **38** 5–15.

LAWLER, G. F. and SOKAL, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309** 557–580.

LETAC, G. and TAKACS, L. (1979). Random walk on an m-dimensional cube. *J. Reine Angew. Math.* **310** 187–195.

LUBOTZKY, A. (1989). *Discrete Groups, Expanding Graphs, and Invariant Measures*. SIAM, Philadelphia. To appear.

MIHAIL, M. (1989). Combinatorial aspects of expanders. Ph.D. dissertation, Dept. Computer Science, Harvard Univ.

MOHAR, B. (1989a). Eigenvalues, diameter, and mean distance in graphs. Technical Report, Dept. Mathematics, Univ. E. K. Ljubljanos, Jadranska, Yugoslavia.

MOHAR, B. (1989b). The Laplacian spectrum of graphs. Technical Report, Dept. Mathematics, Univ. E. K. Ljubljanos, Jadranska, Yugoslavia.

ROTHBLUM, U. and TAN, C. (1985). Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices. *Linear Algebra Appl.* **66** 45–86.

SAXL, J. (1981). On multiplicity-free permutation representations. *Finite Geometrics and Designs. London Math. Soc. Lecture Note Ser.* **48** 337–353. Cambridge Univ. Press.

SENETA, E. (1981). *Non-negative Matrices and Markov Chains*, 2nd ed. Springer, New York.

SINCLAIR, A. (1990). Improved bounds for mixing rates of Markov chains on combinatorial structures. Technical Report, Dept. Computer Science, Univ. Edinburgh.

SINCLAIR, A. and JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82** 93–133.

STANTON, D. (1984). Orthogonal polynomials and Chevalley groups. In *Special Functions: Group Theoretical Aspects and Applications* (R. A. Askey, T. H. Koorwinder and W. Schempp, eds.). Reidel, Dordrecht.

VALIANT, L. G. (1979). The complexity of computing the permanent. *Theoret. Comput. Sci.* **8** 189–201.

DEPARTMENT OF MATHEMATICS
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139