Elect. Comm. in Probab. **10** (2005), 282–296

ELECTRONIC COMMUNICATIONS in PROBABILITY

BOUNDING FASTEST MIXING

SÉBASTIEN ROCH[1]
Department of Statistics, University of California, Berkeley, Berkeley, CA 94720–3860
email: sroch@stat.berkeley.edu

Submitted 22 September 2004, *accepted in final form* 8 November 2005

AMS 2000 Subject classification: 60J10, 90C22, 65C05
Keywords: Rapidly mixing Markov chains, fastest mixing, semidefinite programming, canonical paths, conductance.

Abstract
In a recent work, Boyd, Diaconis and Xiao introduced a semidefinite programming approach for computing the fastest mixing Markov chain on a graph of allowed transitions, given a target stationary distribution. In this paper, we show that standard mixing time analysis techniques—variational characterizations, conductance, canonical paths—can be used to give simple, nontrivial lower and upper bounds on the fastest mixing time. To test the applicability of this idea, we consider several detailed examples including the Glauber dynamics of the Ising model.

1 Introduction

Sampling from a complex collection of objects is a basic procedure in physics, statistics and computer science. A widely used technique, known as Markov chain Monte Carlo (MCMC), consists in designing a Markov chain on the set to be sampled such that the law of the chain converges to the desired distribution. The chain is run long enough for a sample to be picked from a good approximation to the stationary distribution—a process known as mixing. The time one has to wait in order for this approximation to be satisfactory can be measured by various quantities [AF04], including the so-called relaxation time. In practice, it is crucial that this parameter be small. See e.g. [J03] for a survey of theoretical results on MCMC.

One way to picture a Markov chain (MC) on a combinatorial structure is to think of the states as nodes and of the transitions as edges. For a chain to be implementable, the neighborhood structure surrounding each node must be relatively simple. Under this constraint, one has to choose a set of allowed transitions that is most likely to produce fast convergence. This is usually done in a heuristic manner.

Once a graph of transitions has been chosen, there is still room for improvement. Indeed, one has some freedom in assigning transition probabilities to each edge under the requirement,

[1]RESEARCH SUPPORTED BY NSERC, FQRNT, CIPRES (NSF ITR GRANT NSF EF 03-31494), AND A LOÈVE FELLOWSHIP.

however, that the stationary distribution be of the right form. It turns out that choosing appropriately those probabilities can lead to a sizable decrease in the relaxation time.

In a seminal work, Boyd et al. [BDX04] recently showed that minimizing the relaxation time of an MC on a graph of transitions with a given stationary distribution can be formulated as a semidefinite program (SDP), a well-known generalization of linear programming to matrices. See e.g. [BV03]. This formulation enables the numerical computation of the *fastest* mixing chain on a graph. Boyd et al. [BDX04] solved numerically several examples.

A further benefit of this approach is that it provides a tight lower bound on the optimal relaxation time through the dual of the SDP. In a follow-up paper, Boyd et al. [BD$^+$04b] used this bound to exhibit an analytic expression for the fastest chain—and prove its optimality—when the graph is made of a simple path under uniform distribution.

However, only rather small graphs can be studied thoroughly in the SDP formulation because numerical solvers run in time polynomial in the size of the graph; in practice, chains have prohibitively large state spaces. As for the dual, it is potentially useful from a theoretical point of view even for complex chains, but the lack of an intuitive interpretation makes it difficult to apply.

In this paper, we propose a different approach arising from the following observation. One can obtain lower and upper bounds on the relaxation time of *completely specified* chains by way of well-known techniques such as path coupling, conductance, canonical paths etc. [J03]. Formally, those bounds are *parameterized by transition probabilities*. This prompts the questions: can one optimize those bounds as functions of the transition probabilities, and how close to optimum can one get by doing so?

1.1 Our results

We show through general results and examples that, for well-structured problems, the above scheme can be implemented, and that it is capable of providing nontrivial, sharp bounds.

On the lower bound side, we use a standard extremal characterization to derive a general lower bound which has a simple geometrical interpretation. It consists in embedding the nodes of the graph into an Euclidean space so as to stretch the nodes as much as possible under constraints on the distance separating nodes connected by an edge. We show through convex optimization arguments that it is actually tight. The simple interpretation makes it easier to apply than previous bounds. Our result is similar to a bound obtained recently by Sun et al. [SB$^+$04] in a different context. We apply our bound to several examples obtaining close-to-optimal lower bounds.

On the upper bound side, it seems much harder to derive useful, general results. A trivial bound can be obtained by considering any chain on the graph, e.g. a canonical Metropolis-Hastings chain, and computing an upper bound on its relaxation time. But as was shown by Boyd et al. [BDX04], there can be a large (unbounded) gap between standard and optimal chains. Instead, we show through examples that one can obtain almost tight bounds by studying closely standard canonical paths arguments and minimizing the bound over transition probabilities. Put differently, our technique consists in identifying bottleneck edges and increasing the flow on them. The fact that this scheme can work on nontrivial Markov chains is not obvious *a priori*, and this constitutes our main result in the upper bound case. Moreover, this technique is constructive and it allows to design a chain which might be close to the fastest one. Our scheme is likely to work only on well-structured problems but, even in that case, there is no other non-numerical approach known.

Our main example is the Glauber dynamics of the Ising model, a problem which we believe is beyond the reach of the numerical SDP approach. In the case of the tree, by a judicious choice of rates at which nodes are updated, we improve the relaxation time by an "optimal" factor.

1.2 Organization of the paper

We begin in Section 2 with a description of the setting and approach of [BDX04]. We introduce our main techniques in Sections 3 and 4. Section 5 is devoted to optimal rates of the Glauber dynamics of the Ising model.

2 Preliminaries

2.1 Setting

We are given an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a probability distribution π defined on the nodes of $\mathcal{G}$. We seek to sample from π and do so by running a reversible Markov chain $(X_t)_{t\geq 0}$ on the state space $\mathcal{V}$ with stationary distribution π, i.e. if $P = (P(i,j))_{i,j\in\mathcal{V}}$ denotes the transition matrix of $(X_t)_{t\geq 0}$, we must have $\pi(i)P(i,j) = \pi(j)P(j,i)$, $\forall i,j \in \mathcal{V}$. We also require that the only transitions allowed are those given by edges of $\mathcal{G}$, i.e. $P(i,j) = 0$, $\forall (i,j) \notin \mathcal{E}$. For convenience, we assume that all self-loops are present (see below).

The time to reach stationarity is governed by the second largest eigenvalue of P. More precisely, let $n = |\mathcal{V}|$ and $1 = \lambda_1(P) > \lambda_2(P) \geq \cdots \geq \lambda_n(P) \geq -1$ be the eigenvalues of P. We measure the speed at which stationarity is reached by the *relaxation time* $\tau_2(P) = \frac{1}{1-\lambda_2(P)}$. See [AF04] for a thorough discussion of other related quantities. The smaller $\lambda_2(P)$—and therefore $\tau_2(P)$—is, the faster (X_t) approaches π. Given this observation, it is natural to define the *fastest mixing chain* on $(\mathcal{G}, \pi)$ as the solution to the optimization problem

$$\begin{array}{lll} \min\limits_{P\geq 0} & \lambda_2(P) & \\ \text{s.t.} & P(i,j) = 0, & \forall (i,j) \notin \mathcal{E} \\ & \sum_{j\in\mathcal{V}} P(i,j) = 1, & \forall i \in \mathcal{V} \\ & \pi(i)P(i,j) = \pi(j)P(j,i), & \forall (i,j) \in \mathcal{E}. \end{array} \qquad (1)$$

In the remainder of this paper, we save the notation $P^\star$ for a solution of (1)—which might not be unique—and let $\lambda_2^\star = \lambda_2(P^\star)$, and $\tau_2^\star = \tau_2(P^\star)$. Note that our definition of fastest mixing differs slightly from that in [BDX04]. Here, we take the usual approach of ignoring the *smallest* eigenvalue by considering the possibility of adding a constant probability to each self-loop afterwards in order to bound the smallest eigenvalue away from -1.

2.2 Fastest mixing via SDP

A crucial observation in [BDX04] is that (1) is actually a semidefinite program (SDP). See e.g. [BV03] for background on convex and semidefinite programming. This formulation makes possible the numerical computation of optimal transition matrices. However, since the running time of SDP algorithms is polynomial in the size of the state space, this allows only to study rather small graphs—where sampling could possibly be performed using simpler techniques. One idea put forward by Boyd et al. [BDX04] is to solve the SDP on small instances of large combinatorial problems and try and guess the structure of the optimal matrix from the results.

This is the approach used in [BD$^+$04b] to identify the optimal chain on the path. This type of exact result might be hard to obtain in most cases.

From a theoretical point of view, an interesting consequence of the SDP formulation is the existence of a dual which can be used to give lower bounds on the optimal relaxation time. Let $\|Y\|_*$ be the sum of the singular values of Y. Then, in the case of the uniform stationary distribution, the dual (of the more general version taking into account the smallest eigenvalue) has the form [BDX04]

$$\begin{array}{ll} \max\limits_{z,Y} & \sum\limits_{i=1}^{n} z(i) \\ \text{s.t.} & z(i)+z(j) \leq 2Y(i,j), \quad \forall (i,j) \in \mathcal{E} \\ & \sum_{j\in\mathcal{V}} Y(i,j) = 0, \qquad\quad \forall i \in \mathcal{V} \\ & Y = Y^T, \quad \|Y\|_* \leq 1. \end{array} \tag{2}$$

Any feasible solution of (2) provides a lower bound on the best relaxation time achievable on $(\mathcal{G},\pi)$. Moreover, strong duality holds. In [BD$^+$04b], this is used to prove optimality of a conjectured fastest chain when the graph is a path. Note that giving an intuitive interpretation of this optimization problem is not straightforward. This is a potential obstacle to the construction of good feasible solutions.

3 Lower bounds

In this section, we discuss general lower bounds on fastest mixing that can be derived from common techniques for completely specified chains. We apply our bounds to several examples.

3.1 Variational characterization

The standard lower bound for completely specified chains is based on a variational characterization of the second eigenvalue of the transition matrix. See e.g. [AF04]. To reveal the geometric flavor of our result, we will consider a more general bound. Let $\psi_1,\ldots,\psi_n : \mathcal{V} \to \mathbb{R}$ be functions with 0 expectation under π, i.e. $\sum_{i\in\mathcal{V}} \pi(i)\psi_l(i) = 0$ for all l (where, as before, n is the number of nodes). For all $i \in \mathcal{V}$, think of $\Psi(i) = (\psi_1(i),\ldots,\psi_n(i))$ as a vector associated to node i. Therefore, $\Psi(1),\ldots,\Psi(n)$ is an embedding of the graph into $\mathbb{R}^n$. For each l separately, we have the inequality

$$(1-\lambda_2(P))\sum_{k\in\mathcal{V}} \pi(k)\psi_l^2(k) \leq \sum_{(i,j)\in\mathcal{E}} (\psi_l(i)-\psi_l(j))^2 Q(i,j),$$

where $Q(i,j) = \pi(i)P(i,j)$. Summing over l we get the bound

$$1-\lambda_2(P) \leq \frac{\sum_{(i,j)\in\mathcal{E}} \|\Psi(i)-\Psi(j)\|^2 Q(i,j)}{\sum_{k\in\mathcal{V}} \pi(k)\|\Psi(k)\|^2},$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. To turn the r.h.s. into a bound on $1-\lambda_2^\star$, we maximize over Q. But note that, for $\psi_1,\ldots,\psi_n$ fixed, the r.h.s. is linear in Q so this can be expressed as the *linear* program

$$\begin{array}{rcll} 1-\lambda_2^\star & \leq & \max\limits_{Q\geq 0} & \sum\limits_{(i,j)\in\mathcal{E}} \frac{\|\Psi(i)-\Psi(j)\|^2}{\sum_{k\in\mathcal{V}}\pi(k)\|\Psi(k)\|^2} Q(i,j) \\ & & \text{s.t.} & Q(i,j)=0, \qquad \forall (i,j)\notin\mathcal{E} \\ & & & \sum_{j\in\mathcal{V}} Q(i,j)=\pi(i), \quad \forall i\in\mathcal{V} \\ & & & Q(i,j)=Q(j,i), \qquad \forall (i,j)\in\mathcal{E}. \end{array} \tag{3}$$

The dual of this linear program is[2]

$$\begin{array}{rcll} 1-\lambda_2^\star & \leq & \min\limits_{\mathbf{z}} & \sum\limits_{i=1}^{n} \pi(i)z(i) \\ & & \text{s.t.} & z(i)+z(j)\geq \frac{\|\Psi(i)-\Psi(j)\|^2}{\sum_{k\in\mathcal{V}}\pi(k)\|\Psi(k)\|^2}, \quad \forall (i,j)\in\mathcal{E}. \end{array} \tag{4}$$

where variable $z(i)\in\mathbb{R}$ is dual to the *equality* constraint $\sum_{j\in\mathcal{V}} Q(i,j)=\pi(i)$. Note the similarity with (2). Note also that we can now minimize over $\psi_1,\ldots,\psi_n$ as well to get the best bound possible. Make the change of variables $w(i)=z(i)\sum_{k\in\mathcal{V}}\pi(k)\|\Psi(k)\|^2$ for all $i\in\mathcal{E}$, assume w.l.o.g. that $\sum_{i\in\mathcal{V}}\pi(i)w(i)=1$ and take the multiplicative inverse of the objective function (one can always renormalize the Ψ's by $\sum_{i\in\mathcal{V}}\pi(i)w(i)$; in particular, one can check that if $\pi(k)>0$ for all k and the $\Psi(k)$'s are not all equal to $\mathbf{0}$, then the inequalities in (4) ensure $\sum_{i\in\mathcal{V}}\pi(i)w(i)>0$). This finally leads to the lower bound implicit in:

Proposition 1 *The optimal relaxation time on $(\mathcal{G},\pi)$ admits the following characterization*

$$\begin{array}{rcll} \tau_2^\star & = & \max\limits_{\mathbf{w},\Psi(1),\ldots,\Psi(n)} & \sum\limits_{k\in\mathcal{V}} \pi(k)\|\Psi(k)\|^2 \\ & & \text{s.t.} & \|\Psi(i)-\Psi(j)\|^2 \leq w(i)+w(j), \quad \forall (i,j)\in\mathcal{E} \\ & & & \sum_{k=1}^n \pi(k)\Psi(k)=\mathbf{0} \\ & & & \sum_{i=1}^n \pi(i)w(i)=1. \end{array} \tag{5}$$

In particular, any feasible solution to the optimization problem above yields a lower bound on $\tau_2^\star$.

Informally, to obtain a lower bound on $\tau_2^\star$ we seek to embed the graph into $\mathbb{R}^n$ so as to "spread" the nodes as much as possible under constraints over the distances separating nodes connected by edges. The w's give some slack in choosing which edges are bound by stronger or weaker constraints. See the examples. This bound is similar to that obtained recently by [SB$^+$04] in a continuous-time context. There, however, the r.h.s. in the inter-node distance constraint is a *fixed* weight d_{ij} (instead of $w(i)+w(j)$), giving rise to a quite different problem.

Proof (of tightness): This follows from convex optimization duality. We refer the reader to [BV03] for background on convex programming. We go back to formulation (4). Note that we can assume w.l.o.g. that $\sum_{k\in\mathcal{V}}\pi(k)\|\Psi(k)\|^2=1$. Make the change of variables $h(i)=\|\Psi(i)\|^2-z(i)$ for all $i\in\mathcal{V}$, change the objective to $1-\sum_{i=1}^n\pi(i)z(i)=\sum_{i=1}^n\pi(i)h(i)$, and set $Y(i,j)=\Psi(i)^T\Psi(j)$ for all $i,j\in\mathcal{V}$ (where $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$). Then, using the Gram matrix representation for symmetric positive semidefinite matrices (an $n\times n$

[2] To obtain this particular form, one needs to consider only those $Q(i,j)$'s such that $(i,j)\in\mathcal{E}$ and then only one of $Q(i,j)$ and $Q(j,i)$.

matrix M is symmetric positive semidefinite if and only if there is a set $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of vectors in $\mathbb{R}^n$ such that $M_{ij} = \mathbf{x}_j^T \mathbf{x}_i$; see e.g. [HJ85]), we get the equivalent formulation

$$\begin{array}{ll} \max\limits_{\mathbf{h}, Y=Y^T \succeq 0} & \sum\limits_{i=1}^{n} \pi(i)h(i) \\ \text{s.t.} & h(i) + h(j) \leq 2Y(i,j), \quad \forall (i,j) \in \mathcal{E} \\ & \sum_{i,j \in \mathcal{V}} \pi(i)\pi(j)Y(i,j) = 0,\ \sum_{k=1}^{n} \pi(k)Y(k,k) = 1, \end{array} \tag{6}$$

where $A \succeq 0$ indicates that A is positive semidefinite. We now check that the dual of this convex optimization problem is equivalent to minimizing the second largest eigenvalue over reversible transition matrices on $(\mathcal{G}, \pi)$. The Lagrangian (for the minimization version) is

$$\begin{aligned} \mathcal{L}(\mathbf{h}, Y; Q, Z, \lambda, \gamma) &= -\boldsymbol{\pi}^T \mathbf{h} - \operatorname{tr}(YZ) + \sum_{(i,j) \in \mathcal{E}} Q(i,j)[h(i) + h(j) - 2Y(i,j)] \\ &\quad + \gamma \operatorname{tr}(Y \boldsymbol{\pi}\boldsymbol{\pi}^T) + \lambda \left[\operatorname{tr}(Y\Pi) - 1\right], \end{aligned}$$

where $\boldsymbol{\pi}$ is the distribution π seen as a vector, Π is the $n \times n$ diagonal matrix with diagonal $\boldsymbol{\pi}$, Z is a symmetric matrix, γ, λ and $Q(i,j)$ are real variables, and $\operatorname{tr}(A)$ denotes the trace of A. Extending Q to all pairs (i,j) by fixing $Q(i,j) = Q(j,i)$ if $(j,i) \in \mathcal{E}$ and $Q(i,j) = 0$ if $(i,j), (j,i) \notin \mathcal{E}$, we get to

$$\mathcal{L}(\mathbf{h}, Y; Q, Z, \lambda, \gamma) = \mathbf{h}^T(Q\mathbb{1} - \boldsymbol{\pi}) + \operatorname{tr}(Y[-Z - Q + \lambda\Pi + \gamma\boldsymbol{\pi}\boldsymbol{\pi}^T]) - \lambda,$$

where $\mathbb{1}$ is the all-one vector. To obtain the dual, we minimize the Lagrangian over $\mathbf{h}, Y$, and get

$$\begin{array}{ll} \max\limits_{Q \geq 0, Z=Z^T \succeq 0, \lambda, \gamma} & -\lambda \\ \text{s.t.} & Q\mathbb{1} - \boldsymbol{\pi} = 0, \\ & -Z - Q + \lambda\Pi + \gamma\boldsymbol{\pi}\boldsymbol{\pi}^T = 0, \end{array} \tag{7}$$

where Q is implicitly of the form defined above. Make the change of variables $Q(i,j) = \pi(i)P(i,j)$ for all i, j. Then the first constraint above ensures that P is a stochastic matrix satisfying the same conditions as in (1). To reveal the content of the second constraint, we remove the auxiliary variable Z and divide entry (i,j) of all matrices by $\sqrt{\pi(i)\pi(j)}$. We then have the constaint

$$\Pi^{1/2} P \Pi^{-1/2} - \gamma \mathbf{q}\mathbf{q}^T \preceq \lambda I, \tag{8}$$

where $\mathbf{q} = (\sqrt{\pi(1)}, \ldots, \sqrt{\pi(n)})$ and I is the identity matrix. A standard result in eigenvalue optimization [BV03] indicates that minimizing λ over the constraint (8) is equivalent to minimizing the largest eigenvalue of $\Pi^{1/2} P \Pi^{-1/2} - \gamma \mathbf{q}\mathbf{q}^T$. It is well-known [AF04] that $\Pi^{1/2} P \Pi^{-1/2}$ is symmetric, has the same eigenvalues as P and admits $\mathbf{q}$ as eigenvector with eigenvalue 1. Therefore, the largest eigenvalue of $\Pi^{1/2} P \Pi^{-1/2} - \gamma \mathbf{q}\mathbf{q}^T$ is $\max\{\lambda_2(P), 1-\gamma\}$. We can always choose γ so that this is $\lambda_2(P)$, and therefore (7) is equivalent to (1) as claimed. Because all the constraints in (1) are linear, the linear version of Slater's condition [BV03] implies that strong duality follows from feasibility. ■

Contrary to the standard setting, the multidimensionality of the embedding seems necessary in the fastest mixing context. In particular, plugging the eigenvector corresponding to the second largest eigenvalue of the optimal matrix as ψ_1 (with all other coordinates 0) into (5) does not necessarily give a tight bound because there is no guarantee that the optimal w's will allow enough room for a 1-dimensional embedding to spread sufficiently.

Remark 1 *The above bound is actually very similar to that in the case of completely specified chains which can be reformulated as*

$$\begin{array}{rcl} \tau_2(P) & = & \max_{\Psi(1),\ldots,\Psi(n)} \sum_{k\in\mathcal{V}} \pi(k)\|\Psi(k)\|^2 \\ & & \text{s.t.} \quad \sum_{(i,j)\in\mathcal{E}} \|\Psi(i)-\Psi(j)\|^2 Q(i,j) = 1 \\ & & \qquad \sum_{k=1}^{n} \pi(k)\Psi(k) = \mathbf{0}. \end{array} \tag{9}$$

Here the "slack" takes the form of a fixed weighted average over inter-node distances. The multidimensionality turns out not to be necessary in this case.

Remark 2 *The same scheme can be applied to the log-Sobolev constant. In that case, one maximizes the entropy instead of the variance. See also [BDX04].*

Remark 3 *If one were to maximize the smallest eigenvalue instead, the bound would be similar but with the term $\|\Psi(i)-\Psi(j)\|^2$ in the inter-node distance constraint replaced by $\|\Psi(i)+\Psi(j)\|^2$, giving rise to a different geometrical problem. The formulation (2) is equivalent to a combination of the two problems (smallest and second largest eigenvalues).*

3.2 Conductance

As an illustration of Proposition 1 we give a simple adaptation of the conductance bound to the context of fastest mixing.

Proposition 2 *Let Υ be the weighted vertex expansion of $(\mathcal{G},\pi)$*

$$\Upsilon \equiv \min\left\{ \frac{\pi(\delta S)}{\pi(S)\wedge\pi(S^c)} \; : \; S\subseteq\mathcal{V} \right\},$$

where $a\wedge b = \min\{a,b\}$ and δS is the set of nodes $i\in S^c$ such that there is a $j\in S$ with $(i,j)\in\mathcal{E}$. We have the following bound

$$\tau_2^\star \geq \frac{1}{2\Upsilon}.$$

This bound is easily derived from the usual conductance bound and is often used to obtain lower bounds on completely specified chains. Here we give a direct proof.

Proof: We use the bound in Proposition 1. Let $S\subseteq\mathcal{V}$ be any subset and assume w.l.o.g. that $\pi(\delta S^c)\leq\pi(\delta S)$. A simple embedding of $\mathcal{G}$ in $\mathbb{R}^n$ is to map each node to one of only 2 points $\mathbf{x}_0,\mathbf{x}_1$ in which case we can assume $\mathbf{x}_0=(x_0,0,\ldots,0)$ and $\mathbf{x}_1=(x_1,0,\ldots,0)$. Say subset S (resp. S^c) is mapped to $\mathbf{x}_0$ (resp. $\mathbf{x}_1$) and consider the choice $x_0=\frac{1-\pi(S)}{\sqrt{\pi(\delta S^c)}}$ and $x_1=-\frac{\pi(S)}{\sqrt{\pi(\delta S^c)}}$. Then, the constraint $\pi(S)\mathbf{x}_0+\pi(S^c)\mathbf{x}_1=\mathbf{0}$ is satisfied. Also, since the distance between the nodes in S (resp. S^c) is 0, we can set the w's of nodes *not on the boundary* of S (resp. S^c) to 0. We assign the value $w_0=(\pi(\delta S^c))^{-1}$ (resp. $w_1=0$) to the points in δS^c (resp. δS). This choice satisfies $\pi(\delta S^c)w_0+\pi(\delta S)w_1=1$ and $\|\mathbf{x}_0-\mathbf{x}_1\|^2=(\pi(\delta S^c))^{-1}=w_0+w_1$. Finally, we have $\pi(S)\|\mathbf{x}_0\|^2+\pi(S^c)\|\mathbf{x}_1\|^2=\frac{\pi(S)\pi(S^c)}{\pi(\delta S^c)}$ and the result follows from Proposition 1 and the fact that S is arbitrary. ■

3.3 Examples

3.3.1 $K_n - K_n$

This is the graph made of two n-node complete graphs joined by an edge. We denote the nodes on one side of the linking edge by $1, \ldots, n$ and those on the other side by $1', \ldots, n'$. The linking edge is $(1, 1')$. The stationary distribution is uniform. The vertex expansion bound gives $\Upsilon = \frac{1/(2n)}{n/(2n)} = \frac{1}{n}$ and $\tau_2^\star \geq n/2$. To get something sharper, we appeal to our more general bound. The bottleneck in this graph is intrinsically one-dimensional, so we take all coordinates except the first one to be 0, i.e. we consider only ψ_1. By symmetry, it is natural to map the nodes to $\psi_1(1) = -\psi_1(1') = x_0$ and $\psi_1(i) = -\psi_1(i') = x_1$, for $i \neq 1$, with $0 \leq x_0 \leq x_1$. The main insight here is that we should make the distance between 1 and $1'$ as large as possible because that pushes away from 0 all the other points at the same time (because of the local constraints). So we take $w(i) = w(i') = 0$, for all $i \neq 1$, and $w(1) = w(1') = n$, which gives $x_0 = \frac{\sqrt{2}}{2}\sqrt{n}$ and $x_1 = \frac{\sqrt{2}+2}{2}\sqrt{n}$. Summing the squares leads to a lower bound asymptotic to $(\frac{3}{2}+\sqrt{2})n \geq 2.914n$. In Section 4, we give an almost matching upper bound. See also [BD+04a] for a similar upper bound.

3.3.2 n-cycle and d-dimensional torus

In constrast to our preceding example, the n-cycle gives rise naturally to a multidimensional embedding. We let the stationary distribution be uniform. By symmetry we choose all w's equal. So all pairs of consecutive nodes have to be embedded to points at distance (at most) $\sqrt{2}$. Our goal of maximizing the sum of the squared norms—and the natural symmetry— leads to spreading the points evenly on a circle centered around the origin (in any 2-dimensional subspace of $\mathbb{R}^n$). That is, we take all coordinates except the first two to be 0 and, numbering the nodes from 1 to n in order of traversal, we let $(\psi_1(i), \psi_2(i)) = (R\cos(2\pi i/n), R\sin(2\pi i/n))$, $i = 1, \ldots, n$, for a value of R which remains to be determined. The distance between consecutive points has to be $\sqrt{2}$ so a little geometry suggests $R = \frac{\sqrt{2}}{2\sin(\pi/n)} \geq \frac{\sqrt{2}n}{2\pi}$. Thus the lower bound is $\tau_2^\star \geq \frac{n^2}{2\pi^2}$, matching the relaxation time of the symmetric walk. See e.g. [AF04].

One can generalize this result to the m^d-point grid on a d-dimensional torus by considering a $2d$-dimensional embedding. For $1 \leq i_1, \ldots, i_d \leq m$, node $(i_1, \ldots, i_d)$ is mapped to

$$(R\cos(2\pi i_1/m), R\sin(2\pi i_1/m), \ldots, R\cos(2\pi i_d/m), R\sin(2\pi i_d/m)),$$

with R as above. Thus, $\tau_2^\star \geq \frac{dm^2}{2\pi^2}$, again matching the relaxation time of the symmetric walk. See [AF04].

3.3.3 Geometric random graphs

In their analysis of random walks on geometric random graphs, Boyd et al. [BG+04] consider, in a key step, a variant of the d-dimensional grid of the previous example. Let k be a fixed integer smaller than m. Again, our graph is made of the m^d points of the d-dimensional torus $\mathbb{Z}_m^d$ (integers modulo m) with uniform stationary distribution. Two nodes $(i_1, \ldots, i_d)$ and $(j_1, \ldots, j_d)$ are connected by an edge if $i_l - j_l$ modulo m is less or equal to k for all $1 \leq l \leq d$ (the points are at most k cells apart in every dimension). Because of the "diagonal" edges, it seems natural to collapse all nodes on a single m-cycle. More precisely, we map $(i_1, \ldots, i_d)$ to $(R\cos(2\pi i_1/m), R\sin(2\pi i_1/m))$. We take uniform w's. Because some edges connect nodes k

steps apart, the radius (which is constrained by the fact that points connected by an edge are at most $\sqrt{2}$ apart) is now $R = \frac{\sqrt{2}}{2\sin(k\pi/m)} \geq \frac{\sqrt{2}m}{2k\pi}$ (assume that k divides n for convenience). Thus $\tau_2^\star \geq \frac{m^2}{2k^2\pi^2} = \Theta((n/D_d)^{2/d})$, where D_d is the degree of each node and n is the number of nodes. Asymptotically, this bound matches the lower bound in [BG[+]04] (the constant is off by a factor of d). There, exact expressions for the eigenvalues of tensor products of circulant matrices and the analysis of a linear program lead to a lower bound on fastest mixing on this graph. Our geometric method is simpler.

Remark 4 *In the previous two examples, plugging the same embeddings into the completely specifed setting (9) gives tight lower bounds on the symmetric walks. More generally, the lower bound in Proposition 1 applies to any completely specified chain—as do all lower bounds on fastest mixing—and it could prove useful as an alternative to the standard variational characterization when the precise details of the transition matrix appear too cumbersome.*

4 Upper bounds

It seems difficult to give general upper bounds on fastest mixing. An obvious technique is to pick an arbitrary chain and compute an upper bound on its relaxation time. For example, one might use the canonical (max-degree like) chain defined by the transition probabilities $P_d(i,j) = \pi(j)/\pi_*$ if $(i,j) \in \mathcal{E}$ (and 0 otherwise) with $\pi_* = \max\{\sum_{j:(i,j)\in\mathcal{E}} \pi(j) \,:\, i \in \mathcal{V}\}$. It is easy to check using the standard Cheeger inequality that

$$\tau_2^\star \leq \left(\frac{\pi_*}{\pi_0}\right)^2 \frac{2}{\Upsilon^2}.$$

where $\pi_0 = \min_{i\in\mathcal{V}} \pi(i)$. A different chain would have provided a different—and possibly better—bound. Anyhow, this Cheeger-type bound is very unlikely to lead to useful results, and moreover it provides no insight on the optimal chain.

Instead, the goal of this section is to illustrate the computation of a nontrivial upper bound through a canonical paths argument [JS89, DS91, S92]. The underlying idea is similar to that used for the lower bound above. That is, we think of a standard upper bound for completely specified chains as parameterized by transition probabilities and attempt to minimize the bound over those probabilities. It turns out that because of its straightforward dependence on the transition matrix, the canonical paths bound appears to be the most manageable. In this section and the next one, we show by way of examples that it can actually lead to sharp results.

4.1 Canonical paths: $K_n - K_n$ example continued

We illustrate our proposed technique on the $K_n - K_n$ graph with uniform distribution. This chain is analyzed in details in [BD[+]04a], where using sophisticated group-theoretic-based symmetry analysis, all eigenvalues are computed. Here, we give a different, more elementary treatment which might be applicable more generally. We proceed as follows: we write down the canonical paths upper bound as a function of P; we then choose P among π-reversible chains so as to minimize the bound. Given a set Γ of paths γ_{xy} in $\mathcal{G}$ for all pairs of nodes x, y, the canonical paths upper bound is

$$\tau_2(P) \leq \bar{\rho}(P, \Gamma),$$

with

$$\bar{\rho}(P,\Gamma) \quad = \quad \max_{e} \frac{\sum_{\gamma_{xy}\ni e} \pi(x)\pi(y)|\gamma_{xy}|}{Q(e)}, \tag{10}$$

where $|\gamma_{xy}|$ is the number of edges in γ_{xy}. Notice that the choice of paths depends—crucially—only on the graph and is therefore valid for any transition matrix consistent with $(\mathcal{G},\pi)$. Let $W(e)$ be the numerator in (10). On $K_n - K_n$, the natural choice of paths is to let γ_{xy} be the shortest path (in terms of number of edges) between x and y. Then

$$W(i,j) = \frac{1}{(2n)^2}, \quad \forall i,j \neq 1, \qquad W(i,1) = \frac{1}{(2n)^2}(1+2+3(n-1)) = \frac{3n}{(2n)^2}, \quad \forall i,$$
$$W(1,1') = \frac{1}{(2n)^2}(1+4(n-1)+3(n-1)^2) = \frac{3n^2-2n}{(2n)^2}.$$

Similar values hold for the other complete subgraph. The largest contribution to the maximum above clearly comes from $W(1,1')$. In order to decrease the ratio in $\bar{\rho}(P,\Gamma)$, we need to choose a large value for $Q(1,1')$. But as we increase $Q(1,1')$, the $Q(1,i)$'s and $Q(1',i')$'s have to be lowered accordingly. We do so until congestion is the same on edges $(1,1')$, $(1,i)$'s and $(1',i')$'s. That is, we require

$$\frac{W(1,1')}{Q(1,1')} = \frac{W(1,i)}{Q(1,i)}, \quad \forall i \neq 1 \qquad Q(1,1') + \sum_{i=2}^{n} Q(1,i) = \frac{1}{2n},$$

and similarly for the other side. The solution is

$$Q(1,i) = Q(1',i') = \frac{1}{2n(2n-5/3)}, \quad \forall i \neq 1, \qquad Q(1,1') = \frac{n-2/3}{2n(2n-5/3)}.$$

We extend this to all edges by

$$Q(i,j) = Q(i',j') = \frac{1}{n-1}\left[1 - \frac{1}{2n(2n-5/3)}\right], \quad \forall i,j \neq 1.$$

The upper bound becomes

$$\tau_2^\star \leq 3n(1-5/(6n)).$$

Recall that our lower bound was $\tau_2^\star \geq 2.9n$. Note that the standard chain would have consisted in choosing a neighbor uniformly at random at each step. The same calculation gives an upper bound of $\Omega(n^2)$ in that case.

Remark 5 *In summary, our upper bound technique consists in two steps: identifying transitions contributing to slow mixing by computing the congestion ratio in (10); then increasing as much as possible the probability of transition on those bottleneck edges. Instead, one might try to use the same idea with conductance (or other upper bounds). But in that case, the fact that all cuts—instead of edges—have to be accounted for simultaneously makes the task more difficult.*

5 Optimal rates for Glauber dynamics

In this section, we show that the framework discussed so far can be applied to large, well-structured combinatorial problems where the numerical SDP method has little chance of being helpful.

5.1 Glauber dynamics

Let $G = (V, E)$ be a finite graph[3]. A configuration on G is a map $\sigma : V \to C$, where C is a finite set. Typically, σ is a spin or a color. We consider the following stationary distribution on C^V

$$\pi(\sigma) \quad = \quad \frac{1}{Z} \prod_{(v,w)\in E} \alpha_{vw}(\sigma(v), \sigma(w)),$$

where Z is a normalization constant and (v, w) is an undirected edge with endpoints v, w. Let $\mathcal{S} \subseteq C^V$ be the subset of C^V on which π is nonzero. We wish to sample from π by running a reversible MC on $\mathcal{S}$, but allow only transitions that change the state of one node at a time, i.e. the transition graph is $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathcal{S}$ and $(\sigma, \sigma') \in \mathcal{E}$ if and only if $\sigma(v) = \sigma'(v)$ for all but at most one node $v \in V$. Let σ_v^a be the configuration

$$\sigma_v^a(w) = \begin{cases} \sigma(w), & \text{if } w \neq v, \\ a, & \text{if } w = v. \end{cases}$$

One such "local" MC is the so-called Glauber dynamics which, at each step, picks a node v of G uniformly at random and updates the value $\sigma(v)$ according to the transition probability distribution

$$K(\sigma, \sigma_v^a) \quad = \quad \frac{\displaystyle\prod_{w:(v,w)\in E} \alpha_{vw}(a, \sigma(w))}{\displaystyle\sum_{a'\in C} \prod_{w:(v,w)\in E} \alpha_{vw}(a', \sigma(w))}.$$

One can check that the resulting (discrete-time) chain is π-reversible. We actually consider a generalization of the Glauber dynamics by allowing the update rates[4] to vary. More precisely, at each step, we pick a node v of G with probability $\rho(v)$ for some distribution $\rho : V \to [0, 1]$, and we update $\sigma(v)$ according to K as above. The standard chain corresponds to uniform ρ.

Predictably the question we ask is: can we compute the rates ρ minimizing the relaxation time? Or at least can we get reasonable lower and upper bounds on fastest mixing in this restricted setting? We do so by following the methodology put forward in the previous sections.

We first give an elementary bound on the best achievable improvement. This result is essentially due to [BDX04].

Proposition 3 *Let $P^\star$ be the fastest chain on $(\mathcal{G}, \pi)$ (not necessarily of the Glauber dynamics type). Also, let $\rho^\star$ (resp. U) be the optimal (resp. uniform) rates for the Glauber dynamics. Denote by P_ρ the Glauber dynamics with rates ρ and let $\overline{K} = \max_{\sigma,v,a} K^{-1}(\sigma, \sigma_v^a)$. Then,*

$$\tau_2(P_U) \quad \leq \quad \overline{K}\, |V| \tau_2(P^\star), \qquad \textit{and} \qquad \tau_2(P_U) \leq |V| \tau_2(P_{\rho^\star}).$$

[3]We now have two graphs. As before, calligraphic letters are used to denote the transition graph (see below).
[4]In what follows, the term *rate* actually refers to a *discrete-time* analog of a rate.

Proof: By the variational characterization of $\lambda_2(P^\star)$ and the fact that $P^\star(\sigma,\sigma') \le 1$,

$$\begin{aligned}
1-\lambda_2(P^\star) &= \inf_g \frac{\displaystyle\sum_\sigma \sum_{v\in V}\sum_{a\in C}(g(\sigma)-g(\sigma_v^a))^2\pi(\sigma)P^\star(\sigma,\sigma_v^a)}{\displaystyle\sum_{\sigma,\sigma'}(g(\sigma)-g(\sigma'))^2\pi(\sigma)\pi(\sigma')}\\
&\le \overline{K}\,|V|\,\inf_g \frac{\displaystyle\sum_\sigma \sum_{v\in V}\sum_{a\in C}(g(\sigma)-g(\sigma_v^a))^2\pi(\sigma)U(v)K(\sigma,\sigma_v^a)}{\displaystyle\sum_{\sigma,\sigma'}(g(\sigma)-g(\sigma'))^2\pi(\sigma)\pi(\sigma')}\\
&= \overline{K}\,|V|\,(1-\lambda_2(P_U)).
\end{aligned}$$

A similar argument gives the second inequality. ■

Say $\overline{K}$ is $O(1)$. In that case, the best improvement over P_U one can hope for is a factor of $O(|V|)$.

We now use a canonical paths argument similar to that in Section 4 to obtain a general upper bound on fastest mixing for Glauber dynamics.

Proposition 4 *Let Γ be a set of paths $\gamma_{\sigma,\sigma'}$ in $\mathcal{G}$ for each pair σ,σ' in $\mathcal{S}$. Assume we have a bound B_v (depending only on v) on the ratio appearing in the canonical paths bound (10) for edges of the form (σ,σ_v^a) in the uniform rates case. Then,*

$$\tau_2(P_U) \le \max_v B_v, \qquad \textit{and} \qquad \tau_2(P_{\tilde\rho}) \le |V|^{-1}\sum_v B_v,$$

with the choice of rates $\tilde\rho(v) = B_v/\sum_u B_u$.

Proof: The first inequality is the canonical paths bound. For the second one, note that the ratio in (10) is multiplied by $(|V|B_v/\sum_u B_u)^{-1}$ when replacing uniform rates with $\tilde\rho(v)$. We then apply the canonical paths bound to $P_{\tilde\rho}$ using the bound B_v and the previous observation. Note that $\tilde\rho$ is the choice of rates that makes all bounds on the ratio in (10) equal. ■

The point of Proposition 4 is that optimal improvement can be attained if most B_v's are small compared to $\max_v B_v$. We give such an example in the next subsection.

5.2 Special case: the Ising model

We apply the previous result to the case of the Ising model on a finite graph. Here $C = \{-1,+1\}$, $\mathcal{S} = C^V$, and $\alpha_{vw}(\sigma(v),\sigma(w)) = \exp(\beta\sigma(v)\sigma(w))$, where $\beta > 0$ is some constant.

As shown in [BK+04], the relaxation time of the Glauber dynamics on a graph depends on its cut-width. Let $I : V \to \{1,\ldots,|V|\}$ be some ordering of the nodes (not necessarily optimal), and let $\xi^I(v)$ be the number of edges from $\{w : I(w) \le I(v)\}$ to $\{w : I(w) > I(v)\}$. The *cut-width* of a graph G is $\xi(G) = \min_I \max_v \xi^I(v)$. Let Δ be the maximum degree of G. Then it follows from [BK+04] that a bound as required in Proposition 4 is

$$B_v \equiv |V|^2 \exp\left((4\xi^I(v) + 2\Delta)\beta\right),$$

with in particular $\max_v B_v = |V|^2 \exp((4\xi(G) + 2\Delta)\beta)$ if I is an optimal ordering.

One can try and compute $\sum_v B_v/|V|$ in special cases. A rather uninteresting graph is the $s \times s$ grid. There, a natural ordering is to start from a corner, move horizontally as far as one can, then go to the next line and start over. In this ordering, the width of most nodes, including the maximum-width node, is approximately s and therefore using non-uniform rates has essentially no effect.

Here is a more interesting example. Let $T_r^{(b)} = (V_r, E_r)$ be the complete rooted b-ary tree with r levels (the root is at level 0 and the leafs, at level r). Let n_r be the number of vertices in $T_r^{(b)}$. In this case, the maximum degree is $\Delta = b + 1$.

Proposition 5 *For β large enough, an appropriate choice of rates leads to the estimate*

$$\tau_2(P_{\rho^\star}) = O\left(n_r\, e^{4(b-1)\beta r}\right),$$

as r tends to $+\infty$. In constrast, the best known upper bound on the uniform Glauber dynamics [BK+04] is

$$\tau_2(P_U) = O\left(n_r^2\, e^{4(b-1)\beta r}\right).$$

Proof: A good ordering of nodes of $T_r^{(b)}$, say I, is given by a depth-first search (DFS) traversal of the tree starting from the root. Note that the width of a node v is the number of unvisited neighbors of previously visited vertices when the DFS search reaches v. Therefore, the width of the root is b. Then, say vertex v is on level $1 \le l < r$ and is the q-th child of its parent w (in the DFS traversal order). Then $\xi^I(v) = \xi^I(w) + b - q$ because (1) v has b children, (2) q children of w have now been visited, and (3) all descendants of the first $q-1$ children of w have been visited—so these add nothing to the width. As for nodes on level r, we have $\xi^I(v) = \xi^I(w) - q$ if v is the q-th child of w. Thus, the contribution to $\sum_v B_v$ of the l-th level, $1 \le l < r$, is

$$B^{(l)} = B^{(l-1)}\left(e^{4(b-1)\beta} + \cdots + e^{4(0)\beta}\right) = B^{(l-1)}\left(\frac{e^{4b\beta}-1}{e^{4\beta}-1}\right) \equiv B^{(l-1)}\zeta(b,\beta),$$

and for $l = r$ we have

$$B^{(r)} = B^{(r-1)} e^{-4b\beta}\zeta(b,\beta).$$

Summing over all levels, we get

$$\begin{aligned}\frac{\sum_v B_v}{|V_r|} &= n_r^2\,\frac{e^{(4b+2\Delta)\beta}}{n_r}\left(1 + \zeta(b,\beta) + \cdots + \zeta(b,\beta)^{r-1} + e^{-4b\beta}\zeta(b,\beta)^r\right)\\ &= n_r\, e^{(6b+1)\beta}\left(\frac{\zeta(b,\beta)^r - 1}{\zeta(b,\beta)-1} + e^{-4b\beta}\zeta(b,\beta)^r\right).\end{aligned}$$

In the low-temperature regime, i.e. for β large (we actually assume $e^{4\beta} \gg 1$), this is

$$\frac{\sum_v B_v}{|V_r|} = O\Big(n_r\, \exp\Big\{4(b-1)r\beta + (6b+1)\beta\Big\}\Big),$$

whereas

$$\max_v B_v = n_r^2\, \exp\Big((4\xi(T_r^{(b)}) + 2\Delta)\beta\Big) \le n_r^2\, \exp\Big\{4(b-1)r\beta + (6b+1)\beta\Big\},$$

because $\xi(T_r^{(b)}) \leq (b-1)r + 1$ [BK$^+$04] (in the DFS ordering, the worst-case vertex is the parent of the first leaf encountered). Therefore, we get an optimal improvement of $O(n_r)$ over the usual Glauber dynamics. ■

For completeness sake, we note that it is also possible to obtain a lower bound on the relaxation time of the fastest mixing chain, as follows. We assume $b = 3$ for convenience. Let $\epsilon = (1 + e^{2\beta})^{-1}$. The parameter ϵ has the following interpretation. In [EK$^+$00], it is shown that the Ising model on a tree can be simulated by picking the spin of the root uniformly at random and then from top to bottom assigning to each vertex the spin of its parent with probability $1 - \epsilon$ and the opposite spin with probability ϵ.

Proposition 6 *Assume $b = 3$. Then*

$$\tau_2^\star = \Omega\left(n_r^{-\ln(2\epsilon+8\epsilon^2)-1}\right),$$

as r tends to $+\infty$. In constrast, the best known lower bound in the uniform case [BK$^+$04] is

$$\tau_2(P_U) = \Omega\left(n_r^{-\ln(2\epsilon+8\epsilon^2)}\right).$$

Proof: Berger et al. [BK$^+$04] use recursive majority to define a cut in the space of configurations and apply the conductance bound. The recursive majority $m(\sigma)$ of a configuration σ is computed as follows: set $M(v) = \sigma(v)$ for all v on level r; starting from level $r-1$ and up, compute M on each node by taking the majority of the values of M at the children of that node; output the value of M at the root. Let S be the set of configurations σ with $m(\sigma) = +1$. It follows from [BK$^+$04] that, under π, the probability that a configuration is such that its recursive majority is flipped by changing the value at a fixed leaf is at most $(2\epsilon + 8\epsilon^2)^{r-1}$. The union bound and the $\{-1, +1\}$ symmetry imply that $\pi(\delta S^c) \leq \frac{3^r}{2}(2\epsilon + 8\epsilon^2)^{r-1}$ and $\pi(S) = \frac{1}{2}$. By Proposition 2, we deduce $\lambda_2^\star \geq 1 - 2(3)^r(2\epsilon + 8\epsilon^2)^{r-1}$. On the other hand, the usual conductance bound applied to the uniform case gives that $\lambda_2(P_U) \geq 1 - 2\Phi_S$, with

$$\Phi_S = \pi(S)^{-1} \sum_{\sigma\in S, \tau\in S^c} \pi(\sigma) P_U(\sigma,\tau) \leq 2(3)^{-r} \sum_{\substack{\sigma\in S,\tau\in S^c\\(\sigma,\tau)\in\mathcal{E}}} \pi(\sigma) \leq (2\epsilon + 8\epsilon^2)^{r-1},$$

where we have used that $P_U(\sigma, \tau) \leq 3^{-r}$ for neighbors σ, τ [BK$^+$04]. Finally note that $3^r = O(n_r)$. ■

Obtaining tighter bounds would require a sharper analysis in the standard setting.

Remark 6 *We are not claiming that this choice of rates leads to the fastest sampling algorithm for this model. Indeed, in the case of the Ising model on a tree, a very simple propagation algorithm (described before Proposition 6) is much faster [EK$^+$00]. Rather, our point is to establish that fastest mixing analysis is feasible on nontrivial, large-scale chains. It remains to be seen whether fastest mixing ideas will find useful applications in sampling.*

Acknowledgments

This work was motivated by a talk of Persi Diaconis. We thank David Aldous, Elchanan Mossel, and Santosh Vempala for comments. Part of this work was done while visiting CSAIL at MIT.

References

[AF04] Aldous, D. and Fill, J., Reversible Markov Chains and Random Walks on Graphs, monograph in preparation, 2004.

[BK+04] Berger, N., Kenyon, C., Mossel, E., and Peres, Y., Glauber dynamics on trees and hyperbolic graphs, Probab. Theory Related Fields **131** (2005), no. 3, 311–340. MR2123248 (2005k:82066)

[BD+04a] Boyd, S., Diaconis, P., Parrilo, P., and Xiao, L., Symmetry analysis of reversible Markov chains, Internet Mathematics, **2** (2005), no. 1, 31-71.

[BD+04b] Boyd, S., Diaconis, P., Sun, J., and Xiao, L., Fastest mixing Markov chain on a path. To appear in The American Mathematical Monthly, 2005.

[BDX04] Boyd, S., Diaconis, P., Sun, J., and Xiao, L., Fastest mixing Markov chain on a graph, SIAM Rev. **46** (2004), no. 4, 667–689 (electronic). MR2124681 (2005j:90058)

[BG+04] Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D., Mixing Times for Random Walks on Geometric Random Graphs. To appear in SIAM ANALCO 2005.

[BV03] Boyd, S. and Vandenberghe, L., *Convex optimization*, Cambridge Univ. Press, Cambridge, 2004. MR2061575 (2005d:90002)

[DS91] Diaconis, P. and Stroock, D., Geometric bounds for eigenvalues of Markov chains, Ann. Appl. Probab. **1** (1991), no. 1, 36–61. MR1097463 (92h:60103)

[EK+00] Evans, W., Kenyon, C., Peres, Y., and Schulman, L. J., Broadcasting on trees and the Ising model, Ann. Appl. Probab. **10** (2000), no. 2, 410–433. MR1768240 (2001g:60243)

[HJ85] Horn, R. A. and Johnson, C. R., *Matrix analysis*, Cambridge Univ. Press, Cambridge, 1985. MR0832183 (87e:15001)

[J03] Jerrum, M., *Counting, sampling and integrating: algorithms and complexity*, Birkhäuser, Basel, 2003. MR1960003 (2004a:68044)

[JS89] Jerrum, M. and Sinclair, A., Approximating the permanent, SIAM J. Comput. **18** (1989), no. 6, 1149–1178. MR1025467 (91a:05075)

[S92] Sinclair, A., Improved bounds for mixing rates of Markov chains and multicommodity flow, Combin. Probab. Comput. **1** (1992), no. 4, 351–370. MR1211324 (94f:60087)

[SB+04] Sun, J., Boyd, S., Xiao, L., and Diaconis, P., The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. To appear in SIAM Review, 2005.