Fastest mixing Markov chain on a path *

Stephen Boyd† Persi Diaconis‡ Jun Sun† Lin Xiao†

October 2003

Simulation using Markov chain Monte Carlo is a mainstay of scientific computing; see, *e.g.*, [4, 5] for pointers to the literature. Thus the analysis and design of fast mixing Markov chains, with given stationary distribution, has become a research area. In [2], we show how to *numerically* find the fastest mixing Markov chain (*i.e.*, the one with smallest second-largest eigenvalue modulus) on a given underlying graph using tools of convex optimization, in particular, *semidefinite programming* [8, 3]. The present note presents a simple, self contained example where the optimal Markov chain can be identified analytically.

We consider a random walk on a path with $n \geq 2$ nodes, labeled $0, 1, \ldots, n-1$. There are $n-1$ edges connecting pairs of adjacent nodes; in addition, we allow a loop at each node, as shown in figure 1. Let P_{ij} denote the transition probability from node i to node j. We consider symmetric transition probabilities, *i.e.*, those that satisfy $P_{ij} = P_{ji}$. Since P is symmetric, the uniform distribution is stationary. The requirement that transitions can only occur along an edge or loop of the path is equivalent to $P_{ij} = 0$ for $|i - j| > 1$, *i.e.*, P is tridiagonal. Thus, P is a symmetric, stochastic, tridiagonal matrix.

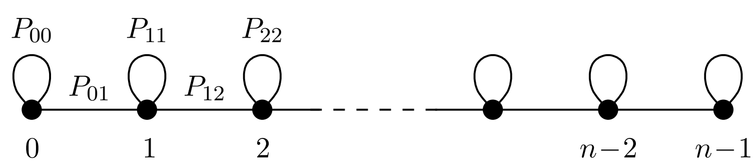

Figure 1: A path with loops at every node, with transition probabilities labeled.

The eigenvalues of P are real (since it is symmetric), and no more than one in modulus (since it is stochastic). We denote them in nonincreasing order:

$$1 = \lambda_0(P) \geq \lambda_1(P) \geq \cdots \geq \lambda_{n-1}(P) \geq -1.$$

The asymptotic rate of convergence of the Markov chain to the stationary distribution, *i.e.*, its mixing rate, depends on the second-largest eigenvalue modulus of P, which we denote $\mu(P)$:

$$\mu(P) = \max_{i=1,\ldots,n-1} |\lambda_i(P)| = \max \{\lambda_1(P),\ -\lambda_{n-1}(P)\}.$$

*Submitted to *The American Mathematical Monthly*, October 2003. Authors listed in alphabetical order.

†Information Systems Laboratory, Department of Electrical Engineering, Stanford University, CA 94305-9510. E-mail addresses: `boyd@stanford.edu`, `sunjun@stanford.edu`, `lxiao@stanford.edu`.

‡Department of Statistics and Department of Mathematics, Stanford University, CA 94305-4065.

The smaller $\mu(P)$ is, the faster the Markov chain converges to its stationary distribution. For more rigorous statements and background, see, *e.g.*, [6, 4, 1, 2] and references therein. The second-largest eigenvalue modulus can also be expressed as the spectral norm of P restricted to the subspace $\mathbf{1}^\perp$, *i.e.*, the subspace of all vectors whose components have zero sum:

$$\mu(P) = \|(I - (1/n)\mathbf{1}\mathbf{1}^T)P(I - (1/n)\mathbf{1}\mathbf{1}^T\|_2 = \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2.$$

(Here $\mathbf{1}$ denotes the vector all of whose components are one, and $\|\cdot\|_2$ denotes the spectral norm, which is the maximum modulus eigenvalue here since the matrices are symmetric.)

The question we address is: What choice of P minimizes $\mu(P)$ among all symmetric stochastic tridiagonal matrices? In other words, what is the fastest mixing (symmetric) Markov chain on a path? We will show that the transition matrix

$$P^\star = \begin{bmatrix} 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/2 & 0 & 1/2 \\ & & & 1/2 & 1/2 \end{bmatrix}$$

achieves the smallest possible value of $\mu(P)$, $\cos(\pi/n)$, among all symmetric stochastic tridiagonal matrices. Thus, to obtain the fastest possible mixing Markov chain on a path, we assign a probability of $1/2$ of moving left, a probability $1/2$ of moving right, and a probability $1/2$ of staying at each end node. (For the nodes not at either end, the probability of staying at the node is zero.) This optimal Markov chain is shown in figure 2. For $n = 2$, we have $\mu(P^\star) = \cos(\pi/2) = 0$, and this is clearly the optimal solution. For $n \geq 3$, while $P^\star$ is the transition matrix one would guess yields fastest mixing, we are not aware of a simpler proof of its optimality than the one we give below.

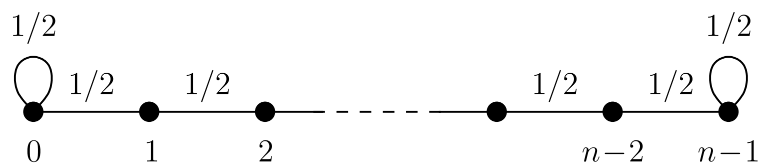

Figure 2: Fastest mixing Markov chain on a path.

Lemma. *Let $P \in \mathbf{R}^{n\times n}$ be a symmetric stochastic matrix, and suppose $Y \in \mathbf{R}^{n\times n}$ and $z \in \mathbf{R}^n$ satisfy*

$$Y\mathbf{1} = 0, \qquad Y = Y^T, \qquad \|Y\|_* \leq 1, \tag{1}$$

$$(z_i + z_j)/2 \leq Y_{ij} \quad \text{if} \quad P_{ij} \neq 0, \tag{2}$$

where $\|Y\|_* = \sum_{i=0}^{n-1} |\lambda_i(Y)|$. ($\|\cdot\|_*$ is dual of the spectral norm.) Then we have $\mu(P) \geq \mathbf{1}^T z$.

Proof. For any P, Y and z that satisfy the assumptions in the lemma, we have

$$\begin{array}{rcl}
\mu(P) & = & \|P-(1/n)\mathbf{1}\mathbf{1}^T\|_2 \\
& \geq & \|Y\|_*\|P-(1/n)\mathbf{1}\mathbf{1}^T\|_2 \\
& \geq & \mathbf{tr}\, Y\left(P-(1/n)\mathbf{1}\mathbf{1}^T\right) \\
& = & \mathbf{tr}\, YP = \sum_{i,j} Y_{ij}P_{ij} \\
& \geq & \sum_{i,j}(1/2)(z_i+z_j)P_{ij} \\
& = & (1/2)(z^TP\mathbf{1}+\mathbf{1}^TPz) \\
& = & \mathbf{1}^Tz.
\end{array}$$

Here $\mathbf{tr}$ denotes the trace. The second inequality follows from the Hölder inequality for matrix inner product and norms, $\mathbf{tr}\, A^TB \leq \|A\|_*\|B\|_2$. For the last inequality, we use assumption (2). □

Theorem. *The matrix* $P^\star$ *attains the smallest value of* μ, $\cos(\pi/n)$, *among all symmetric stochastic tridiagonal matrices.*

Proof. The result is clear for $n=2$. We assume now that $n>2$. The eigenvalues and associated orthonormal eigenvectors of $P^\star$ are

$$\begin{array}{lll}
\lambda_0 = 1, & v_0 = (1/\sqrt{n})\mathbf{1} & \\
\lambda_j = \cos\left(\dfrac{j\pi}{n}\right), & v_j(k) = \sqrt{\dfrac{2}{n}}\cos\left(\dfrac{(2k+1)j\pi}{2n}\right), & \begin{array}{l} j=1,\ldots,n-1 \\ k=0,\ldots,n-1. \end{array}
\end{array}$$

(See, *e.g.*,[7, §16.3].) Therefore we have

$$\mu(P^\star) = \lambda_1 = -\lambda_{n-1} = \cos(\pi/n).$$

We show that this is the smallest μ possible by constructing a pair Y and z that satisfy the assumptions (1) and (2) in the lemma, for all symmetric tridiagonal stochastic matrices P, and $\mathbf{1}^Tz = \cos(\pi/n)$.

We will take $Y = v_1v_1^T$. The entries of Y are $Y_{ij} = v_1(i)v_1(j)$. Since $P_{ij}=0$ for $|i-j|>1$, we only need to list the diagonal and superdiagonal entries:

$$\begin{array}{rcl}
Y_{ii} & = & v_1(i)v_1(i) = \dfrac{2}{n}\cos\left(\dfrac{(2i+1)\pi}{2n}\right)\cos\left(\dfrac{(2i+1)\pi}{2n}\right) \\
& = & \dfrac{1}{n}\left[1+\cos\left(\dfrac{\pi(2i+1)}{n}\right)\right], \qquad i=0,\ldots,n-1 \\
Y_{i,i+1} & = & v_1(i)v_1(i+1) = \dfrac{2}{n}\cos\left(\dfrac{(2i+1)\pi}{2n}\right)\cos\left(\dfrac{(2i+3)\pi}{2n}\right) \\
& = & \dfrac{1}{n}\left[\cos\left(\dfrac{\pi}{n}\right)+\cos\left(\dfrac{\pi(2i+2)}{n}\right)\right], \qquad i=0,\ldots,n-2.
\end{array}$$

The matrix Y has rank one. Its only nonzero eigenvalue is $\|v_1\|^2 = 1$, so its dual norm is $\|Y\|_* = 1$. Since $v_0 = (1/\sqrt{n})\mathbf{1}$ and $v_1 \perp v_0$, we have $Y\mathbf{1} = 0$. Thus, the assumptions (1) hold.

We take z to be

$$z_i = \frac{1}{n}\left[\cos\left(\frac{\pi}{n}\right) + \cos\left(\frac{\pi(2i+1)}{n}\right) \Big/ \cos\left(\frac{\pi}{n}\right)\right], \qquad i = 0, \ldots, n-1.$$

It is easy to verify that $\mathbf{1}^T z = \cos(\pi/n)$.

It remains to check that Y and z satisfy (2) for all symmetric tridiagonal matrices P. Let's first check the superdiagonal entries. For $i = 0, \ldots, n-2$, we have

$$\begin{aligned}
\frac{z_i + z_{i+1}}{2} &= \frac{1}{n}\left[\cos\left(\frac{\pi}{n}\right) + \frac{1}{2}\left(\cos\left(\frac{\pi(2i+1)}{n}\right) + \cos\left(\frac{\pi(2i+3)}{n}\right)\right) \Big/ \cos\left(\frac{\pi}{n}\right)\right] \\
&= \frac{1}{n}\left[\cos\left(\frac{\pi}{n}\right) + \cos\left(\frac{\pi(2i+2)}{n}\right)\right] \;=\; Y_{i,i+1}.
\end{aligned}$$

Therefore equality always holds for the superdiagonal (and subdiagonal) entries. For the diagonal entries, we need to check $(z_i + z_i)/2 = z_i \leq Y_{ii}$, *i.e.*,

$$\cos\left(\frac{\pi}{n}\right) + \cos\left(\frac{\pi(2i+1)}{n}\right) \Big/ \cos\left(\frac{\pi}{n}\right) \leq 1 + \cos\left(\frac{\pi(2i+1)}{n}\right), \qquad i = 0, \ldots, n-1,$$

which is equivalent to

$$\left[1 - \cos\left(\frac{\pi}{n}\right)\right]\left[1 - \cos\left(\frac{\pi(2i+1)}{n}\right) \Big/ \cos\left(\frac{\pi}{n}\right)\right] \geq 0, \qquad i = 0, \ldots, n-1.$$

But this is certainly true because

$$\cos\left(\frac{\pi(2i+1)}{n}\right) \leq \cos\left(\frac{\pi}{n}\right), \qquad i = 0, \ldots, n-1.$$

This completes the proof. □

Our proof is based on *duality theory* for semdefinite programming, applied to the fastest mixing Markov chain problem, as formulated in [2, 3]. The lemma below is exactly the weak duality result for this problem, and the conditions in the Lemma are that Y and z are feasible dual variables.

References

[1] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. `stat-www.berkeley.edu/users/aldous/RWG/book.html`, 2003. Forthcoming book.

[2] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing Markov chain on a graph. To appear in *SIAM Review*, problems and techniques section, 2004. Available at `www.stanford.edu/~boyd/fmmc.html`.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. Available at `www.stanford.edu/~boyd/cvxbook.html`.

[4] P. Brémaud. *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues*. Texts in Applied Mathematics. Springer-Verlag, Berlin-Heidelberg, 1999.

[5] G. Cobb and Y. Chen. An application of Markov chain Monte Carlo to community ecology. *The American Mathematical Monthly*, 110(4):265–288, 2003.

[6] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[7] W. Feller. *An Introduction to Probability and Its Applications*, volume I. Wiely, New York, 3rd edition, 1968.

[8] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.