Fastest Mixing Markov Chain on a Path

Stephen Boyd, Persi Diaconis, Jun Sun, and Lin Xiao

We consider the problem of assigning transition probabilities to the edges of a path in such a way that the resulting Markov chain or random walk mixes as rapidly as possible. In this note we prove that fastest mixing is obtained when each edge has a transition probability of 1/2. Although this result is intuitive (it was conjectured in [6]) and can be found numerically using convex optimization methods [2], [3], we give a self-contained proof here.

Consider a path with n (≥ 2) nodes, labeled $1, 2, \ldots, n$, with $n-1$ edges connecting pairs of adjacent nodes, and with a loop at each node, as shown in Figure 1. We consider a Markov chain (or random walk) on this path, with transition probability from node i to node j denoted P_{ij}. The requirement that transitions can occur only along an edge or loop of the path is equivalent to $P_{ij} = 0$ when $|i - j| > 1$ (i.e., P is a tridiagonal matrix). Since the P_{ij} are transition probabilities, we have $P_{ij} \geq 0$ and $\sum_j P_{ij} = 1$ (i.e., P is a stochastic matrix). This can be expressed as $P\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the vector with all components one.

We consider symmetric transition probabilities, meaning those that satisfy $P_{ij} = P_{ji}$. Thus, P is a symmetric, (doubly) stochastic, tridiagonal matrix. Since $P\mathbf{1} = \mathbf{1}$, we have $(\mathbf{1}/n)^T P = \mathbf{1}^T/n$, which means that the uniform distribution, given by $\mathbf{1}^T/n$, is stationary.

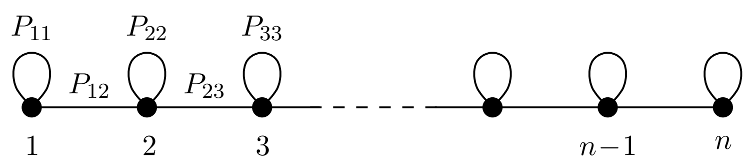

Figure 1: A path with loops at each node, with transition probabilities labeled.

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

$$1 = \lambda_1(P) \geq \lambda_2(P) \geq \cdots \geq \lambda_n(P) \geq -1.$$

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 (SLEM) of P, which we denote $\mu(P)$:

$$\mu(P) = \max_{i=2,\ldots,n} |\lambda_i(P)| = \max \{\lambda_2(P),\ -\lambda_n(P)\}.$$

The smaller $\mu(P)$ is, the faster the Markov chain converges to its stationary distribution. For example, we have the following bound:

$$\|\pi(t) - \mathbf{1}^T/n\|_{\mathrm{TV}} \leq (1/2)\sqrt{n}\mu^t,$$

where $\pi(t) = \pi(0)P^t$ is the probability distribution at time t and $\|\cdot\|_{\rm TV}$ denotes the total variation distance. (The total variation distance between two probability distributions π and $\hat{\pi}$ is the maximum of $|\,\mathbf{prob}_\pi(S) - \mathbf{prob}_{\hat{\pi}}(S)|$ over all subsets S of $\{1, 2, \ldots, n\}$.) For more background, see [5], [4], [1], or [2], and the references therein.

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 show that the transition matrix

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

achieves the smallest possible value of $\mu(P)$ (namely, $\cos(\pi/n)$) among all symmetric stochastic tridiagonal matrices. Thus, to obtain the fastest 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 of the two end nodes. (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.

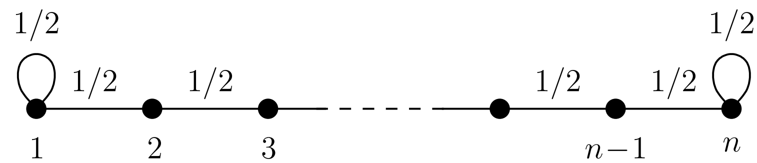

Figure 2: Fastest mixing Markov chain on a path.

For $n = 2$, we have $\mu(P^\star) = \cos(\pi/2) = 0$, which is clearly the optimal solution: in one step the distribution is exactly uniform for any initial distribution $\pi(0)$. When $n \geq 3$, $P^\star$ is the transition matrix one would guess yields fastest mixing; indeed, this was conjectured in [6]. But we are not aware of a simpler proof of its optimality than the one we give here.

Before proceeding, we describe another context in which the same mathematical problem arises. We imagine that there is a processor at each node of our path and that each link represents a direct network connection between the adjacent processors. Processor i has a job queue or load $q_i(t)$ (which we approximate as a positive real number) at time t. The goal is, at each step, to shift jobs across the links in such a way as to balance the load. In other words, we would like to have $q_i(t) \to \overline{q}$ as $t \to \infty$, where $\overline{q} = (1/n)\sum_i q_i(0)$ is the average of the initial queues. We ignore the reduction in the queues due to processing (or, equivalently, assume that the load balancing is done before the processing begins). We use the following simple scheme to balance the load: at each step, we compute the load imbalance $q_{i+1}(t) - q_i(t)$ across each link. We then transfer a fraction θ_i in $[0, 1]$ of the load imbalance from the more loaded to the less loaded processor. We must have $\theta_i + \theta_{i+1} \leq 1$ to ensure that we are not asked to transfer more than the load on a processor to its neighbors. It can be shown that, if the θ_i is positive and satisfy $\theta_i + \theta_{i+1} \leq 1$, then this iterative scheme achieves asymptotic balanced loads (i.e., $q_i(t) \to \overline{q}$ as $t \to \infty$). The problem is to find the fractions θ_i that result in the fastest possible load balancing.

It turns out that this optimal iterative load balancing problem is identical to the problem of finding the fastest mixing Markov chain on a path, with $P_{i,i+1} = \theta_i$. In particular, the evolution of the loads at the processors is given by $q(t) = P^t q(0)$. The speed of convergence of $q(t)$ to $\overline{q}\mathbf{1}$ is given by the second-largest eigenvalue modulus $\mu(P)$. By the basic result in this paper, the fastest possible load balancing is accomplished by shifting one-half of the load imbalance on each edge from the more loaded to the less loaded processor. More discussion of this load balancing problem can be found in [6].

We now proceed to prove the basic result.

Lemma 1. *If P is an $n \times n$ symmetric stochastic matrix, then*

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

where $\|\cdot\|_2$ denotes the spectral norm (maximum singular value).

Proof. To see this, we note that $\mathbf{1}$ is the eigenvector of P associated with the eigenvalue $\lambda_1 = 1$. Therefore the eigenvalues of $P - \mathbf{1}\mathbf{1}^T/n$ are $0, \lambda_2, \ldots, \lambda_n$. Since $P - \mathbf{1}\mathbf{1}^T/n$ is symmetric, its spectral norm is equal to the maximum magnitude of its eigenvalues (i.e., the norm is $\max\{\lambda_2, -\lambda_n\}$, which is $\mu(P)$). □

Lemma 2. *If P is an $n \times n$ symmetric stochastic matrix and if y and z in $\mathbf{R}^n$ satisfy*

$$\mathbf{1}^T y = 0, \qquad \|y\|_2 = 1, \tag{2}$$

$$(z_i + z_j)/2 \le y_i y_j \quad (P_{ij} \ne 0), \tag{3}$$

then $\mu(P) \ge \mathbf{1}^T z$.

Proof. For any P, y, and z as specified, we have

$$\begin{aligned} \mu(P) &= \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2 \\ &\ge y^T\left(P - (1/n)\mathbf{1}\mathbf{1}^T\right)y = y^T P y = \sum_{i,j} P_{ij} y_i y_j \\ &\ge \sum_{i,j} (1/2)(z_i + z_j) P_{ij} = (1/2)(z^T P \mathbf{1} + \mathbf{1}^T P z) = \mathbf{1}^T z. \end{aligned}$$

The first inequality follows from the assumption $\|y\|_2 = 1$ and Lemma 1. The second inequality is a consequence of assumption (3) and the fact that $P_{ij} \ge 0$. □

Theorem. *The value $\cos(\pi/n)$ of μ for the matrix $P^\star$ in (1) is the smallest among all symmetric stochastic tridiagonal matrices.*

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

$$\lambda_1 = 1, \qquad v_0 = (1/\sqrt{n})\mathbf{1},$$
$$\lambda_j = \cos\left(\frac{(j-1)\pi}{n}\right), \qquad v_j(k) = \sqrt{\frac{2}{n}}\cos\left(\frac{(2k-1)(j-1)\pi}{2n}\right)$$

for $j = 2, \ldots, n$ and $k = 1, \ldots, n$ (see, for example, [7, sec. 16.3]). Therefore we have

$$\mu(P^\star) = \lambda_2 = -\lambda_n = \cos(\pi/n).$$

We show that this is the smallest μ possible by constructing a pair of vectors y and z, with $\mathbf{1}^T z = \cos(\pi/n)$, that satisfy the assumptions (2) and (3) in Lemma 2 for any symmetric tridiagonal stochastic matrix P.

We take $y = v_2$, so the assumptions (2) in the second lemma clearly hold. We take z to be

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

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

It remains to check that y and z satisfy (3) for any symmetric tridiagonal matrix P. We first check the superdiagonal entries. For $i = 1, \ldots, n-1$ we have

$$\begin{aligned}
\frac{z_i + z_{i+1}}{2} &= \frac{1}{n}\left[\cos\left(\frac{\pi}{n}\right) + \frac{1}{2}\left(\cos\left(\frac{(2i-1)\pi}{n}\right) + \cos\left(\frac{(2i+1)\pi}{n}\right)\right) \Big/ \cos\left(\frac{\pi}{n}\right)\right] \\
&= \frac{1}{n}\left[\cos\left(\frac{\pi}{n}\right) + \cos\left(\frac{2i\pi}{n}\right)\right] \\
&= \frac{2}{n}\cos\left(\frac{(2i-1)\pi}{2n}\right)\cos\left(\frac{(2i+1)\pi}{2n}\right) \quad = \quad y_i y_{i+1}.
\end{aligned}$$

Therefore equality always holds for the superdiagonal (and subdiagonal) entries. For the diagonal entries we need to check that $(z_i + z_i)/2 = z_i \le y_i^2$—i.e.,

$$\cos\left(\frac{\pi}{n}\right) + \cos\left(\frac{(2i-1)\pi}{n}\right) \Big/ \cos\left(\frac{\pi}{n}\right) \le 2\cos^2\left(\frac{(2i-1)\pi}{2n}\right) = 1 + \cos\left(\frac{(2i-1)\pi}{n}\right)$$

—for $i = 1, \ldots, n$. This is equivalent to

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

which is certainly true because

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

This completes the proof. □

References

[1] D. Aldous and J. Fill, *Reversible Markov Chains and Random Walks on Graphs* (to appear); available at `stat-www.berkeley.edu/users/aldous/RWG/book.html`.

[2] S. Boyd, P. Diaconis, and L. Xiao, Fastest mixing Markov chain on a graph, *SIAM Review* **46** (2004) 667-689.

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

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

[5] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains, *Ann. Appl. Prob.* **1** (1991) 36–61.

[6] R. Diekmann, S. Muthukrishnan, and M. V. Nayakkankuppam, Engineering diffusive load balancing algorithms using experiments, in *Lecture Notes in Computer Science*, no. 1253, Springer-Verlag, Berlin, 1997, pp. 111-122.

[7] W. Feller. *An Introduction to Probability and Its Applications*, vol. 1, 3rd ed., Wiley, New York, 1968.

Stephen Boyd, Jun Sun
Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9510.
(Emails: `boyd@stanford.edu`, `sunjun@stanford.edu`, `lxiao@stanford.edu`)

Lin Xiao
Center for the Mathematics of Information, Mail Code 136-93, California Institute of Technology, Pasadena, CA 91125-9300. (Email: `lxiao@caltech.edu`)

Persi Diaconis
Department of Statistics, Stanford University, Stanford, CA 94305-4065.