

A Summary: *Fastest Mixing Markov Chain on a Graph*

Prepared for: Dr.Kartik Sivaramakrishnan

Prepared by: Susan Osborne

1 Introduction.

This paper summarizes the ideas presented in *Fastest Mixing Markov Chain on a Graph* written by Stephen Boyd, Persi Diaconis, and Lin Xiao. The author of this paper makes no claim of originality. The intent is to review some of the ideas set forth in the aforementioned article and to expound upon its basic content. This is accomplished through describing details of the intricacies of the problem and by illustrating an example for which data is given in the article.

2 Fastest Mixing Markov Chain on a Graph.

2.1 Markov Chain on a Graph.

Let $G = (V, E)$ denote a graph with vertex set $V = \{1, 2, \dots, n\}$ and edge set E , where edge $(i, j) \in E$ whenever there exists an edge that joins vertex v_i to vertex v_j . A graph G is an undirected graph whenever edge $(j, i) \in E$ if and only if $(i, j) \in E$. A vertex v_i has a self-loop whenever the edge $(i, i) \in E$. A *Markov Chain* is a stochastic process $\{X(n), n = 0, 1, 2, \dots\}$ such that if $X(n) = i$, the process is said to be in state i and there is a fixed probability P_{ij} that the process will be next in state j . A markov chain that takes on only a finite or countable number of states is called a discrete-time markov chain. Consider a connected undirected graph on n vertices with self-loops at each vertex. Then define a discrete-time markov chain on the vertices of the graph by letting $X(t) \in V$ for $t = 0, 1, \dots$ be the actual state of the markov chain at time t . An edge of the graph is assigned a transition probability with which $X(t)$ can move to a neighboring vertex or possibly stay in the current position due to the fact that each vertex has a self-loop. Note that the transition probabilities are nonnegative and the sum of the probabilities assigned to the edges connected to any particular vertex must be equal to 1. The markov chain is represented by a transition probability matrix P , where P is an $n \times n$ matrix defined by

$$P_{i,j} = \text{Prob}(X(t+1) = j | X(t) = i), \quad i, j = 1, \dots, n, \quad (1)$$

where $P_{i,j} = 0$ if $(i, j) \notin E$ (i.e. the probability of moving to a nonadjacent vertex is 0). Additionally, the matrix must be doubly stochastic (ie. the rows and columns both sum to 1) and symmetric. These conditions imply that the matrix satisfies $P \geq 0$ (i.e. $P_{ij} \geq 0$ for all $i, j = 1, \dots, n$), $Pe = e$, where

e is a vector of all ones, since P is doubly stochastic and $P = P^T$ due to the symmetry of P . Define the probability distribution of the state at time t as $\pi(t) \in R^n$ such that $\pi_i(t) = \text{Prob}(X(t) = i)$. The state distribution satisfies the recursive equation $\pi(t+1) = P^T \pi(t)$, which implies that the distribution at time t is given by $\pi(t+1) = \hat{P}^T \pi(0)$, where \hat{P}^T is the t -step transition probability matrix which is obtained by raising the matrix P to the t power. A markov chain is *irreducible* if all states can be accessed from any given state and *aperiodic* if starting in state i the process can return to state i in one step. The markov chain defined as above is both irreducible and aperiodic. Since the above process satisfies these two properties, then as t increases, the probability distribution of the state time $\pi(t)$ converges to the uniform distribution $(\frac{1}{n})e$.

2.2 The Second Largest Eigenvalue Modulus.

The eigenvalue structure of the matrix P determines the rate of convergence of $\pi(t)$ to the uniform distribution $(\frac{1}{n})e$. Since P is a stochastic symmetric matrix its spectral radius $\rho(P) = 1$ and all of its eigenvalues are real. According to Perron-Frobenius theory, the eigenvalues must also be less than or equal to 1 in magnitude. Thus, the eigenvalues of P can be arranged in a nondecreasing order as follows: $1 = \lambda_1(P) \geq \lambda_2(P) \geq \dots \geq \lambda_n(P) \geq -1$. Then rate of convergence is given by the second largest eigenvalue modulus (SLEM), which is defined by

$$\mu(P) = \max \{ \lambda_2(P), -\lambda_n(P) \}. \quad (2)$$

Thus, $\mu(P) < 1$ for the irreducible aperiodic stochastic matrix P . Some typical measures for fast mixing of a markov chain are the *mixing rate* $\log(\frac{1}{\mu})$, the *mixing time* $\frac{1}{\log(\frac{1}{\mu})}$, and the *spectral gap* $1 - \mu$. The focus here is on the SLEM, the smaller the SLEM the faster the markov chain reaches its equilibrium distribution.

2.3 Fastest Mixing Markov Chain Problem.

Consider the problem of determining the edge transition probabilities that provide the fastest mixing markov chain. This can be achieved by minimizing the SLEM as follows:

$$\begin{aligned} \min \quad & \mu(P) \\ \text{s.t.} \quad & P \geq 0 \\ & P^T e = e \\ & P = P^T \\ & P_{i,j} = 0, (i,j) \notin E. \end{aligned} \quad (3)$$

Note the optimization variable is P and the graph is the data. The optimal SLEM is $\mu^* = \inf \{ \mu(P) | P \geq 0, P^T e = e, P = P^T, P_{i,j} = 0, (i,j) \notin E \}$. The existence of one optimal transition matrix is guaranteed by the continuity on μ and the fact the set of all transition matrices is compact (see [1]).

3 Heuristic Methods.

The Maximum-Degree Chain and the Metropolis-Hastings Chain are two heuristic methods that attempt to obtain the transition probabilities for the fastest

mixing markov chain. Although neither is guaranteed to produce the transition probabilities that give the fastest mixing markov chain, these simple heuristics tend to provide probabilities for a fast mixing markov chain. Both methods are illustrated on the simple graph given in figure 1 (shown without self-loops) in the sections below.

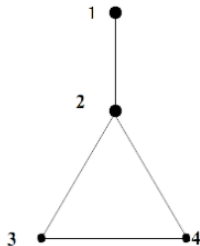


Figure 1: Simple graph

3.1 The Maximum-Degree Chain

Vertices v and u are adjacent vertices if the edge uv is in the edge set E of the graph G . The degree of vertex v is the number of vertices adjacent with vertex v , not counting self loops. Let the degree of vertex i be d_i and let d_{max} be the maximum degree of the graph. Now assign to each edge of the graph, excluding self-loops, the probability $\frac{1}{d_{max}}$. For the graph in figure 1, notice that node 2 has 3 adjacent vertices and no other node has a degree greater than 3. Thus, $d_{max} = 3$ and the probability for each edge that is not a self loop is equal to $\frac{1}{3}$. The probability of each self-loop is then assigned to ensure that the sum of the probabilities at particular vertex is equal to 1. Shown in Figure 2 are two graphs, the first showing the assignment of the probabilities without the self-loops and second including the self-loops and their probabilities. Once the

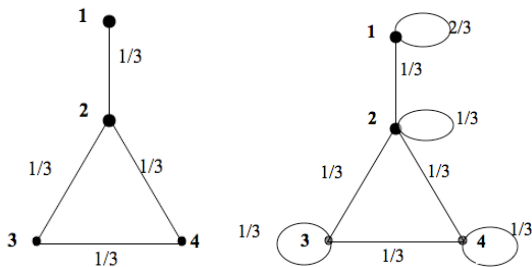


Figure 2: Application of Maximum-Degree Chain Heuristic

probabilities have been assigned to the graph, they can be transformed into a matrix called the maximum-degree transition probability matrix, $P_{(ij)}^{md}$. This matrix is given by

$$P_{ij}^{md} = \begin{cases} \frac{1}{d_{max}} & (i, j) \in E, i \neq j \\ 1 - \frac{d_i}{d_{max}} & i = j \\ 0 & (i, j) \notin E \end{cases} \quad (4)$$

For the graph in figure 1,

$$P^{md}_{(ij)} = \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \end{pmatrix} \quad (5)$$

Since the second largest eigenvalue of P^{md} is $\frac{2}{3}$, the SLEM for the matrix is $\mu(P^{md}) = \frac{2}{3}$.

3.2 The Metropolis-Hastings Chain

A random walk on a graph can be visualized by assuming an individual is located at some vertex v_i on the graph at some point in time. Suppose that the individual must move to another vertex, but has no preference as to which to move to next. Thus, any vertex that is adjacent to v_i is equally likely to be chosen. For example, consider the simple graph in Figure 1. If the person is at vertex 4, then there are two adjacent vertices and thus the probability is $\frac{1}{2}$ that the individual will move to vertex 2. Likewise, the probability that the individual moves to vertex 3 from vertex 4 is also $\frac{1}{2}$. Clearly, the transition probabilities of a simple random walk on a graph is given by

$$P^{rw}_{ij} = \begin{cases} \frac{1}{d_i} & (i, j) \in E, i \neq j \\ 0 & (i, j) \notin E \end{cases} \quad (6)$$

The symmetry of the probability matrix is necessary to ensure that the eigenvalues are real. Note that the random walk probability matrix is not symmetric. However, the transition probabilities can be modified by applying the Metropolis-Hastings algorithm to the random walk probability matrix (see [4] for details on the algorithm). If the equilibrium distribution is the uniform distribution, then the resulting transition probability matrix is symmetric. This matrix is given in a simplified form by

$$P^{mh}_{ij} = \begin{cases} \min \left\{ \frac{1}{d_i}, \frac{1}{d_j} \right\} & (i, j) \in E, i \neq j \\ \sum_{(i,k) \in E} \max \left\{ 0, \frac{1}{d_i} - \frac{1}{d_k} \right\} & i = j \\ 0 & (i, j) \notin E \end{cases} \quad (7)$$

Figure 3 shows the results of applying this function to the graph in Figure 1.

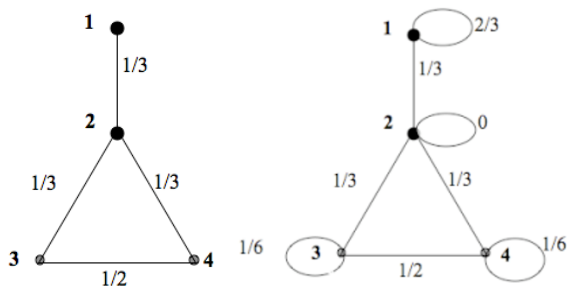


Figure 3: Application of Metropolis-Hastings Chain Heuristic

The resulting transition probability matrix is

$$P_{(ij)}^{mh} = \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/6 & 1/2 \\ 0 & 1/3 & 1/2 & 1/6 \end{pmatrix} \quad (8)$$

The eigenvalues of this matrix are $1, \frac{2}{3},$ and $-\frac{1}{3}$. Thus, the SLEM for the matrix $\mu(P^{mh})$ is $\frac{2}{3}$. It is appropriate to note that in this example, both heuristics gave the same SLEM, yet this is not always the case (see [1] for an example for which the results differ).

4 Convexity and Convex Optimization.

4.1 Convexity of the SLEM.

The convexity of μ is established through the use of the spectral norm of P restricted to the subspace $e^\perp = \{u \in R^n | e^T u = 0\}$. This requires the determination of the orthogonal projection on e^\perp . Normalizing the vector e gives $\frac{1}{\sqrt{n}} \|e\|$. Then, the elementary orthogonal projection for e^\perp is given by the matrix $I - (\frac{1}{n})ee^T$ (see [3] for more on orthogonal projectors). Now, let $\mu(P) = \|(I - (\frac{1}{n})ee^T)P(I - (\frac{1}{n})ee^T)\|_2 = \|P(I - (\frac{1}{n})ee^T)\|_2$. This is the spectral norm of an affine function of P , and thus a convex function.

4.2 Convex Problem Formulation

Consider the constraint set for the Fastest Mixing Markov Chain problem along with the objective of minimizing the norm $\mu(P)$. Formulate the problem as follows:

$$\begin{aligned} \min \quad & \mu(P) = \|P(I - (\frac{1}{n})ee^T)\|_2 \\ \text{s.t.} \quad & P \geq 0 \\ & P^T e = e \\ & P = P^T \\ & P_{i,j} = 0, (i,j) \notin E. \end{aligned} \quad (9)$$

Since this formulation has a convex objective function with all linear constraints, the problem of minimizing the spectral norm over the set of symmetric stochastic matrices is a convex optimization problem.

4.3 Semidefinite Programming Formulation.

Introducing a scalar s as an upper bound on the norm, the convex optimization problem (9) can be rewritten as a semidefinite programming (SDP) problem. The SDP is as follows:

$$\begin{aligned} \min \quad & s \\ \text{s.t.} \quad & -sI \preceq P(I - (\frac{1}{n})ee^T) \preceq sI \\ & P \geq 0 \\ & P^T e = e \\ & P = P^T \\ & P_{i,j} = 0, (i,j) \notin E. \end{aligned} \quad (10)$$

The SDP (10) can be rewritten and a linear matrix inequality (LMI). Let $\text{Diag}(\text{svec}(P))$ denote the diagonal matrix consisting of the $\frac{n(n+1)}{2}$ upper triangle positions of the symmetric matrix P . Then $P - (\frac{1}{n})ee^T + sI$, $sI - P + (\frac{1}{n})ee^T$, and $\text{Diag}(\text{svec}(P))$ are positive semidefinite matrices. A matrix with block diagonal form in which each block is positive semidefinite is also positive semidefinite. Thus, rewrite the SDP (10) as follows:

$$\begin{aligned} & \min s \\ & \text{s.t.} \\ & \begin{pmatrix} P - (\frac{1}{n})ee^T + sI & 0 & 0 \\ 0 & sI - P + (\frac{1}{n})ee^T & 0 \\ 0 & 0 & \text{Diag}(\text{svec}(P)) \end{pmatrix} \succeq 0, \end{aligned} \quad (11)$$

For graphs with up to approximately one thousand edges, this SDP can be solved to ϵ -optimality by a standard SDP solver. Larger problems can be solved by tailoring a solver to take advantage of the sparsity of the matrix that is particular to this application. The simple graph in Figure 1 was solved exactly by an SDP solver to obtain the transition probability matrix

$$P_{(ij)} = \begin{pmatrix} 6/11 & 5/11 & 0 & 0 \\ 5/11 & 0 & 3/11 & 3/11 \\ 0 & 3/11 & 4/11 & 4/11 \\ 0 & 3/11 & 4/11 & 4/11 \end{pmatrix} \quad (12)$$

The SLEM of this matrix is $\mu(P) = \frac{7}{11}$, which is smaller than the SLEM for the heuristic methods for this graph. It should be noted that the matrix (12) is not unique.

5 Conclusion

Several heuristic methods are available to determine a fast mixing for a discrete-time markov chain defined on a undirected connected graph. The Fastest Mixing Markov Chain Problem can be formulated as convex optimization problem, more specifically as a SDP. The typical measures of fast mixing are based on the SLEM. Using the SLEM as the measure, the SDP formulation of the problem out performs both heuristics methods.

References

- [1] S. BOYD, P DIACONIS, AND L. XIAO, *Fastest Mixing Markov Chain on a Graph*, SIAM Review, 46(2004), pp. 667-689.
- [2] G. CHARTRAND AND L. LESNIAK, *Graphs and Digraphs, 3rd. ed.*, Chapman and Hall, Suffolk, Great Britain 1996.
- [3] C. MEYER, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2004.
- [4] S. ROSS, *Introduction to Probability Models, 8th ed.*, Academic Press, San Diego CA, 2003.
- [5] K. SIVARAMAKRISHNAN, *Class Notes*, North Carolina State Univeristy, Raleigh NC, 2007.