A NEW EIGENVALUE BOUND FOR REVERSIBLE MARKOV CHAINS WITH APPLICATIONS TO THE TEMPERATURE-ASYMPTOTICS OF SIMULATED ANNEALING †

Madhav P. Desai and Vasant B. Rao

Coordinated Science Laboratory
and Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1101 W. Springfield Avenue,
Urbana, Illinois 61801.

ABSTRACT

This paper presents a novel upper bound for the second largest eigenvalue of a finite *reversible* time-homogeneous Markov chain as a function of three parameters, namely, the smallest transition probability, the *underlying structure* of the chain, and the *skewness* of the equilibrium distribution. Simulated Annealing (SA), is an example of a probabilistic algorithm that is widely used for solving combinatorial optimization problems, wherein the transition probabilities are controlled by a certain *temperature* parameter $T>0$. Using the results of this paper, we can bound the time constant of convergence of SA to equilibrium at any fixed temperature $T>0$, and also study the *temperature asymptotics*, namely, the growth of this bound as $T\rightarrow 0$. The eigenvalue bound of this paper is also compared with the bound derived by Jerrum and Sinclair in [4].

1. INTRODUCTION

Let $\Omega = \{1,2, \ldots, N\}$ be a discrete *state space*, and consider a *time-homogeneous* Markov chain ($X(k)$) on Ω with an $N\times N$ probability *transition matrix* $P = [p_{ij}]$ such that for any $i, j \in \Omega$, and time $k \geq 0$, $p_{ij} = Prob\ (X(k+1)=j \mid X(k)=i)$. Let $\nu(k) = [\nu_i(k)]$ be the $1\times N$ *distribution vector* describing the chain at time k such that $\nu_i(k) = Prob(X(k)=i)$; it follows that $\nu(k+1) = \nu(k)\,P$. Suppose the Markov chain converges to an *equilibrium vector* π, i.e.,

$$\lim_{k\rightarrow\infty} \nu(k) = \pi = \pi P. \tag{1.1}$$

In this paper, we are primarily interested in the speed of convergence of $\nu(k)$ to π. Let $1=\lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_N|$ denote the eigenvalues of P arranged in descending order of magnitude. It is then well known [1] that the error at time k can be bounded by

$$\|\nu(k) - \pi\| \leq A_N\ |\lambda_2|^k \tag{1.2}$$

where A_N is a constant independent of time k. If we define

$$\tau = -(\log|\lambda_2|)^{-1} \tag{1.3}$$

as the *time constant of convergence*, then (1.2) can be written as $\|\nu(k) - \pi\| \leq A_N\, e^{-k/\tau}$. It follows that if $|\lambda_2| \leq 1-1/q$ for some $q \gg 1$, then $\tau \leq q$. Furthermore, given any $0<\delta<1$ we will have $\|\nu(k) - \pi\| \leq \delta$ whenever $k > [\log(A_N) + \log(1/\delta)]\,\tau$. Therefore, the rate at which the Markov chain achieves equilibrium is determined by the time constant τ and hence by the eigenvalue of second largest magnitude λ_2.

The main result of this paper is the derivation of an upper bound on the eigenvalue of second largest magnitude of a *reversible* Markov chain. In his remarkable paper [3], Alon established the relationship between the second smallest eigenvalue $\mu_2(Q)$ of the *Laplacian matrix* Q of a graph G, and a certain *expansion parameter* $c(G)$ of the graph. A direct application of his ideas to Markov chains leads to a useful bound only for the case of *symmetric* Markov chains as shown in [2]. A symmetric Markov chain, however, can only have the *uniform* equilibrium vector, namely, $\pi_i = 1/N$ for all $i \in \Omega$. In this paper we seek a useful bound for *reversible* Markov chains which, in general, could have *non-uniform* equilibrium distributions.

The bound derived in this paper is of the form $|\lambda_2| \leq 1-1/q$, where q is related to the minimum non-zero off-diagonal entry in P, the *skewness* of its equilibrium vector, and $\mu_2(Q)$. Recently, Jerrum and Sinclair [4] have derived an alternate bound of the form $|\lambda_2| \leq 1-\phi^2/2$, where ϕ is a certain *conductance parameter* associated with the reversible Markov chain which is an extension of the expansion idea for edge-weighted graphs. We compare the two bounds and exhibit a class of Markov chains for which our bound, treated as a function of skewness alone, is asymptotically tighter than the Jerrum and Sinclair bound. Reversible Markov chains are of interest because they can be used to model stochastic algorithms for combinatorial optimization such as Simulated Annealing (SA) [6]. As an application of our results, we will consider using SA at a fixed temperature to solve some specific combinatorial optimization problems and derive bounds on the time constant of convergence of such chains.

† This work was supported in part by a grant from the Semiconductor Research Corporation under contract # SRC 86-12-109, and in part by a grant from the United States Air Force Office of Scientific Research under contract # AFOSR 88-0181.

2. PRELIMINARIES AND DEFINITIONS

We study a time-homogeneous Markov chain ($X(k)$) on a finite state space $\Omega=\{1,2, \ldots, N\}$ with transition matrix $P = [p_{ij}]$. In this paper we are using the standard graph-theoretic terminology from [5].

Definition 2.1 : The *underlying directed graph* of P is a directed graph $G_d(V, E_d)$ with vertex set $V=\Omega$, and an arc (i,j) directed from vertex i to vertex j if and only if $p_{ij} \neq 0$. The matrix P is *irreducible* if G_d is strongly connected. For an irreducible matrix, let r denote the greatest common divisor of the lengths of all the directed cycles in its underlying directed graph. If $r=1$ the matrix is said to be *primitive*. A primitive matrix P also has the property that there exists an integer $m>0$ such that P^m has all strictly positive entries. Some authors refer to Markov chains with irreducible P as ergodic chains, and to chains with primitive P as regular chains.

From the Perron-Frobenius theory of nonnegative matrices [1], it can be shown that $\lambda=1$ is the largest eigenvalue of an irreducible P. Moreover, 1 is a simple eigenvalue. Let π be the *left* eigenvector corresponding to the eigenvalue 1 of P, i.e., $\pi = \pi P$, satisfying $\pi \mathbf{1} = 1$, where $\mathbf{1}$ is a column vector with each entry = 1. We refer to π as the *equilibrium vector* of the Markov chain as motivated by the following result.

Theorem 2.2 (Perron-Frobenius [1]) : Consider a Markov chain with irreducible P, distribution vector $\nu(k)$ at time k, and equilibrium vector π. Then,

(1) $\pi_i > 0$ for all $i \in \Omega$ and all *right* eigenvectors corresponding to any other eigenvalue $\lambda<1$ of P must be orthogonal to π. Furthermore, $P\mathbf{1} = \mathbf{1}$ and all *left* eigenvector corresponding to any eigenvalue $\lambda<1$ of P must be orthogonal to $\mathbf{1}$.

(2) If P is primitive, then the distribution vector $\nu(k)$ converges to π as $k\rightarrow\infty$ for any starting distribution vector $\nu(0)$.

Definition 2.3 : If P is *structurally-symmetric*, i.e., $p_{ij}>0$ if and only if $p_{ji}>0$, then its *underlying undirected graph* is a simple undirected graph $G(V, E)$ obtained from the underlying directed graph $G_d(V, E_d)$ by deleting all self-loops and replacing all directed 2-cycles by simple edges. Thus, arcs (i,j) and (j,i) in G_d are replaced by a single edge $\{i,j\}$ in G.

Definition 2.4 : For an irreducible P with equilibrium vector π, the *skewness* s_π is defined to be

$$s_\pi = \max_{i,j\in\Omega} \frac{\pi_i}{\pi_j} \tag{2.1}$$

Clearly, s_π for an irreducible P is well defined from part (1) of Theorem 2.2. The main result of this paper deals with reversible Markov chains, which we now define as follows.

Definition 2.5 : A Markov chain with an irreducible transition matrix P and equilibrium vector π is said to be *reversible*, if

$$p_{ij}\,\pi_i = p_{ji}\,\pi_j \quad \text{for all } i,j\in\Omega \tag{2.2}$$

Proposition 2.6 : Consider a reversible Markov chain with transition matrix P and equilibrium vector π. Define $d_i = \sqrt{\pi_i}$ for each $i\in\Omega$, and $D = diag[\,d_1,d_2,\cdots,d_N]$. Then,

(i) P is structurally symmetric, while $D^2 P$ and $D\,P\,D^{-1}$ are both symmetric.

(iii) Consequently, P is diagonalizable and has real eigenvalues.

In general, for any $K\times K$ matrix M with real eigenvalues, let $\lambda_1(M)\geq\lambda_2(M)\geq\cdots\geq\lambda_K(M)$ denote the eigenvalues of M arranged in *descending order*. Using this notation, Theorem 2.2, and Proposition 2.6, it is clear that for a reversible Markov chain

$$1=\lambda_1(P) > \lambda_2(P) \geq \lambda_3(P) \geq \cdots \geq \lambda_N(P). \tag{2.3}$$

Definition 2.7 : Given a simple undirected graph $G(V,E)$ on N vertices. Let $deg(i)$ denote the *degree* of vertex $i\in V$. Then the *Laplacian matrix* $Q(G)$ is an $N\times N$ matrix with *off-diagonal* entries defined as $q_{ij}=-1$ if $\{i,j\}\in E$ or 0 otherwise, and *diagonal* entries as $q_{ii} = deg(i)$ for all $i\in V$. Clearly, the Laplacian matrix $Q(G)$ is a symmetric matrix.

Proposition 2.8 : If $G(V,E)$ is a connected simple graph on N vertices with a Laplacian matrix Q, then

(1) $Q\mathbf{1}=\underline{0}$, and $rank(Q)=N-1$. Moreover, there exists a $N-1\times N$ matrix B of full rank such that $Q = B^T B$.

(2) The quadratic form $x^T Qx = \sum_{\{i,j\}\in E} (x_i - x_j)^2$.

In general, for any $K\times K$ matrix M with real eigenvalues let $\mu_1(M)\leq\mu_2(M)\leq\cdots\leq\mu_K(M)$ denote the eigenvalues of M arranged in *ascending order*. From Proposition 2.8, we have Q is *positive semi-definite*, $\mu_1(Q)=0$, and $\mu_2(Q)>0$.

Lemma 2.9 (Min-max principle [13]) : If A and B are any two symmetric $K\times K$ matrices such that $A-B$ is positive semi-definite, then $\mu_i(B)\leq\mu_i(A)$ for each $i=1,2,\ldots,K$.

Lemma 2.10 : Let B be any $N-1\times N$ matrix of full rank. Then $\mu_i(BB^T)=\mu_{i+1}(B^TB)$ for each $i=1,2,\ldots,N-1$.

Theorem 2.11 : Let Q be any $N\times N$ symmetric and positive semi-definite matrix with $rank(Q)=N-1$, Σ be a $N\times N$ diagonal matrix with entries $\sigma_i>0$, and let $\sigma_{min}>0$ denote the smallest diagonal entry in Σ. Then, $\mu_2(\Sigma Q\Sigma)\geq\sigma_{min}^2\mu_2(Q)$.

Proof : From Proposition 2.8, $Q = B^TB$, where B is an $N-1\times N$ matrix of full rank. Define $C = B\Sigma$. Clearly, $\Sigma Q\Sigma = C^TC$. Therefore, by Lemma 2.10,

$$\mu_2(\Sigma Q\Sigma) = \mu_2(C^TC) = \mu_1(CC^T) \tag{2.4}$$

But $CC^T = B\Sigma^2B^T$. Also, for any vector $x\in\mathbf{R}^{N-1}$

$$x^T(CC^T - \sigma_{min}^2 BB^T)\,x = \sum_{i=1}^{N-1}(\sigma_{ii}^2-\sigma_{min}^2)y_i^2 \geq 0 \tag{2.5}$$

where we have defined $y = B^Tx$. Therefore the matrix $CC^T-\sigma_{min}^2BB^T$ is positive semi-definite by definition; hence, by Lemma 2.9, we conclude that

$$\mu_1(CC^T)\geq\sigma_{min}^2\mu_1(BB^T) \tag{2.6}$$

Applying Lemma 2.10 once again, we get

$$\mu_1(BB^T)=\mu_2(B^TB)=\mu_2(Q) \tag{2.7}$$

Combining (2.4), (2.6), and (2.7) proves this theorem. □

3. A NEW EIGENVALUE BOUND

A reversible Markov chain has a structurally-symmetric irreducible transition matrix P with an underlying undirected graph G which is both connected and simple. From Proposition 2.6, all eigenvalues of P are real. Moreover, from Perron-Frobenius theory [1], P has the second largest eigenvalue $\lambda_2<1$. The main result of this paper is to obtain a tighter upper bound for λ_2 of P.

Theorem 3.1 : Let $\Omega=\{1,2,\cdots,N\}$, and consider a reversible Markov chain on the state space Ω with transition matrix P, and equilibrium vector π. If $\lambda<1$ is any eigenvalue of P, then

$$\lambda \leq 1 - \alpha\,\mu_2(Q)\,/\,s_\pi \tag{3.1}$$

where s_π = the skewness of the equilibrium vector π of P, $\mu_2(Q)$ = the second smallest eigenvalue of the Laplacian matrix Q of the underlying undirected graph $G(V,E)$ of P, and $\alpha=\min\{p_{ij}:\{i,j\}\in E\}$ is the smallest non-zero off-diagonal entry in P.

Proof : Let $d_i=\sqrt{\pi_i}$ for each $i\in\Omega$, and define the $N\times N$ diagonal matrix $D = diag\,[\,d_1,d_2,\cdots,d_N\,]$. Since P is irreducible, $\pi_i>0$ for each $i\in\Omega$ from part (1) of Theorem 2.2. Therefore $d_i>0$, D is invertible, and $D^{-1} = diag\,[\,d_1^{-1},\cdots,d_N^{-1}\,]$.

Let $\lambda<1$ be any eigenvalue of P and let $x\in\mathbf{R}^N$ be the corresponding right eigenvector, i.e., $Px=\lambda x$. Therefore,

$$1-\lambda = \frac{x^T(D^2-W)x}{x^TD^2x} \tag{3.2}$$

where we have defined the matrix $W=D^2P$, whose entries satisfy $w_{ij}=d_i^2\,p_{ij}=\pi_i\,p_{ij}$. The reversibility condition of (2.2) implies that W is symmetric. Also, $W\mathbf{1}=D^2P\mathbf{1}=D^2\mathbf{1}$, by part (1) of Theorem 2.2. Now, consider the quadratic form in the numerator of (3.2) which can be written, after some algebraic manipulation, as

$$x^T(D^2-W)x = \sum_{\{i,j\}\in E} w_{ij}\,(x_i-x_j)^2 \geq \beta \sum_{\{i,j\}\in E}(x_i-x_j)^2 \tag{3.3}$$

where

$$\beta=\min\{w_{ij}:\{i,j\}\in E\} \geq \alpha\,\pi_{min}, \tag{3.4}$$

denotes the smallest non-zero off-diagonal entry in W, $\pi_{\min}$ is the smallest entry in π, and α is defined in the statement of the theorem. Applying Proposition 2.8 part (2) to the right hand side of (3.3) and using (3.4) we get

$$x^T(D^2-W)x \geq \alpha\,\pi_{\min}\,x^TQx \tag{3.5}$$

where Q is the Laplacian matrix associated with the underlying graph G. Combining (3.2) and (3.5) results in

$$1-\lambda \geq \alpha\,\pi_{\min}\,x^TQx\,/\,x^TD^2x. \tag{3.6}$$

It must be noted that x is a right eigenvector of P with eigenvalue $\lambda<1$, while $\pi=\mathbf{1}^TD^2$ is a left eigenvector of P with eigenvalue 1. Part (1) of Theorem 2.2 immediately shows that x and π must be orthogonal, i.e., $\mathbf{1}^TD^2x=0$. So consider the following constrained optimization problem :

Minimize $z^TQ\,z$ over all $z\in\mathbf{R}^N$

such that $\mathbf{1}^TD^2z=0$ and $z^TD^2z=1$.

Setting $y=Dz$ or $z=D^{-1}y$, the problem becomes equivalent to

Minimize $y^TD^{-1}QD^{-1}y$ over all $y\in\mathbf{R}^N$

such that $\mathbf{1}^TDy=0$ and $y^Ty=1$.

Recall from Section 2, that Q is a symmetric positive semi-definite matrix with eigenvalues $0=\mu_1(Q)<\mu_2(Q)$ and $Q\mathbf{1}=\underline{0}$. Moreover, $D^{-1}QD^{-1}$ is a symmetric positive semi-definite matrix with nullvector $D\mathbf{1}$, i.e., $D^{-1}QD^{-1}\,D\mathbf{1}=\underline{0}$. Therefore, the above optimization problem is to minimize the quadratic form $y^TD^{-1}QD^{-1}y$ over all *normalized* vectors $y\in\mathbf{R}^N$ that are orthogonal to $D\mathbf{1}$, the eigenvector corresponding to the smallest eigenvalue 0 of the matrix $D^{-1}QD^{-1}$. The required minimum value of the quadratic form is clearly $\mu_2(D^{-1}QD^{-1})$ from quadratic programming theory [14]. Hence, we get

$$x^TQx\,/\,x^TD^2x \geq \mu_2(D^{-1}QD^{-1}) \tag{3.7}$$

Applying Theorem 2.11 to the right hand side of (3.7), with $\Sigma=D^{-1}$, gives

$$\mu_2(D^{-1}QD^{-1}) \geq \mu_2(Q)\,/\,\pi_{\max} \tag{3.8}$$

where $\pi_{\max}$ is the largest entry in π.

Finally, combining (3.6), (3.7), and (3.8), we get

$$1-\lambda \geq \alpha\,\pi_{\min}\,\mu_2(Q)\,/\,\pi_{\max} = \alpha\,\mu_2(Q)\,/\,s_\pi \tag{3.9}$$

thus proving the theorem. □

For some graphs G, the second smallest eigenvalue $\mu_2(Q(G))$ is easy to compute analytically. Two examples are given below.

Cycle graphs: If G is a simple-cycle on N vertices, then the eigenvalues of its Laplacian matrix Q can be shown to be [10]

$$\mu_i(Q) = 2(1 - \cos(2\pi(i-1)/N)) \text{ for all } 1 \le i \le N \tag{3.10}$$

Consequently, the second smallest eigenvalue $\mu_2(Q) = 2(1 - \cos(2\pi/N))$ which approaches 0 as $N \to \infty$.

Hypercube graphs: If G is an n-dimensional hypercube having $N = 2^n$ vertices, then its Laplacian matrix Q has $n+1$ distinct eigenvalues [11] given by

$$\xi_m = 2m \text{ for all } 0 \le m \le n \tag{3.11}$$

where eigenvalue $2m$ has an algebraic multiplicity of $\binom{n}{m}$. Consequently, the second smallest eigenvalue $\mu_2(Q) = 2$ which is independent of N, the size of the matrix.

For graphs G in which $\mu_2(Q(G))$ is not easy to compute, one can use a lower bound derived by Alon [3], which depends on the *expansion* properties of a graph.

4. APPLICATIONS OF THE EIGENVALUE BOUND

As an application of the results of Section 3, we consider the Simulated Annealing (SA) algorithm. This algorithm was first proposed as a probabilistic algorithm for solving difficult combinatorial optimization problems [6]. It has been used with some success in problems such as VLSI layout optimization, the design of FIR filters with finite precision, and image restoration.

We describe the SA algorithm briefly. Let $\Omega = \{1, \cdots, N\}$ be a set of states with a cost function $C: \Omega \to \mathbf{R}$. The SA algorithm attempts to find a state with *globally minimum* cost. Let $x(k)$ denote the state of the algorithm at time k. With each state $i \in \Omega$, we associate a set of neighboring states $N_i \subset \Omega$, which satisfy the following assumptions:

(A1) The neighboring sets are symmetric; that is, $j \in N_i$ if and only if $i \in N_j$.

(A2) Given any two states i and j in Ω, there exists a finite sequence of states $i_0, i_1, \cdots, i_m$ such that $i_0 = i$, $i_m = j$, and $i_{l+1} \in N_{i_l}$, for each $l = 0, 1, \cdots, m-1$. This condition is often referred to as the *reachability* requirement.

(A3) A simplifying assumtion that $|N_i| = \rho$ for each $i \in \Omega$, i.e., all neighbor sets are of the same size.

Let $T > 0$ be a fixed parameter that is analogous to the *temperature* in the physical annealing process. We define

$$\epsilon = e^{-1/T} \tag{4.1}$$

to simplify notation. Note that $T = (\log \epsilon^{-1})^{-1}$. So, if $0 < T < \infty$ then $0 < \epsilon < 1$. Also, as $T \to 0$ we have $\epsilon \to 0$. The SA algorithm thus simulates a time-homogeneous Markov chain on state space Ω with transition matrix $P = [\, p_{ij} \,]$ with off-diagonal entries ($i \neq j$) given by

$$p_{ij} = \begin{cases} \rho^{-1} \epsilon^{[C(j)-C(i)]^+} & \text{if } j \in N_i \\ 0 & \text{if } j \in \Omega - N_i \end{cases} \tag{4.2}$$

where $[z]^+$ denotes the positive part of a real number z. The diagonal entries of P are given by

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij} \tag{4.3}$$

The equilibrium vector $\pi(\epsilon) = [\pi_i(\epsilon)]$ satisfies the *Boltzmann* distribution with $\pi_i(\epsilon) = \gamma \epsilon^{C(i)}$ where γ is a normalizing constant. It is easy to check that P is primitive and and the Markov process is reversible for $0 < \epsilon < 1$.

It must be emphasized that we have assumed a *fixed temperature* $T > 0$ for all time k of the SA algorithm. This is often referred to as Fixed-Temperature-Simulated-Annealing, (FTSA) as opposed to a situation wherein the temperature is allowed to vary with time k according to a prespecified *cooling schedule* (see [7,8,9] for details) which results in a *time-inhomogeneous* Markov chain. In this paper, however we focus only on the FTSA algorithm.

Temperature Asymptotics : For a chosen $0 < \epsilon < 1$, let $\nu_\epsilon(k)$ denote the distribution vector of the FTSA chain at time $k \ge 0$ as defined in Section 1. From Theorem 2.2 part (2) we have

$$\lim_{k \to \infty} \| \nu_\epsilon(k) - \pi(\epsilon) \| = 0 \tag{4.4}$$

In this section we are primarily interested in the rate of convergence of (4.4) as a function of $\epsilon \to 0$, in which case it can be shown that $\| \pi(\epsilon) - \pi^* \| \to 0$, where π^* denotes the *optimal distribution* vector whose entries are uniformly distributed over the states with globally minimum cost. We refer to this as the *temperature asymptotics* of FTSA. From the discussion in Section 1, it is clear that for a particular $\epsilon > 0$, the rate of convergence of (4.4) is governed by the time-constant of convergence τ, defined by (1.3). Using (1.3) and (3.1) we now derive a bound for τ and study the behavior of this bound as $\epsilon \to 0$.

From Theorem 3.1 we can obtain an upper bound on the eigenvalue of the transition P with second largest *algebraic* value. However, to obtain a meaningful bound on τ, the time-constant of convergence, we need an upper bound on the eigenvalue of P of second largest *magnitude*. To avoid this problem, we consider a new Markov chain corresponding to the matrix $\hat{P} = \frac{1}{2}(I + P)$. Clearly, $\hat{P}$ has non-negative eigenvalues and $\hat{P}$ has the same equilibrium vector as P. Furthermore, the off-diagonal entries of $\hat{P}$ are half the corresponding entries of P; hence, $\hat{P}$ is also reversible by (2.2) and has the same underlying undirected graph as P. We will therefore work with the new $\hat{P}$ instead of P.

Let us now relate the parameters used in the bound of Theorem 3.1 to the parameters of the optimization problem being solved by an FTSA Markov chain. Define

$$\Delta = \max_{i,j \in \Omega} | C(i) - C(j) | \tag{4.5}$$

as the maximum cost difference between any two states. Let $G(V,E)$ be the underlying undirected graph of $\hat{P}$ (or P) with Laplacian matrix Q and define

$$\delta = \max_{\{i,j\} \in E} | C(i) - C(j) | \tag{4.6}$$

as the maximum difference in costs between any two *neighboring* states in the Markov chain. Then, the skewness of the chain is given by

$$s_\pi = \epsilon^{-\Delta} \tag{4.7}$$

The smallest non-zero off-diagonal entry of $\hat{P}$ can be computed to be

$$\alpha = \epsilon^{\delta} / 2\rho \tag{4.8}$$

where ρ is the number of neighboring states for each state as given by assumption (A3). Using (3.1), (4.7), and (4.8) we get

$$0 \le \lambda_2(\hat{P}) \le 1 - \mu_2(Q)\, \epsilon^{(\Delta+\delta)} / 2\rho \tag{4.9}$$

from which the time-constant for convergence for sufficiently small ϵ using (1.3) can be bounded by

$$\tau \le \frac{2\rho}{\mu_2(Q)\, \epsilon^{(\Delta+\delta)}} \le \frac{2\rho}{\mu_2(Q)}\, s_\pi^{1+\delta/\Delta} \tag{4.10}$$

For a fixed optimization problem (i.e., fixed N, ρ, Q, etc.), (4.10) shows that the time-constant for convergence of the FTSA Markov chain to its equilibrium vector with skewness s_π is $\tau = O(s_\pi^2)$, since $\delta < \Delta$ by definition. In practice, usually $\delta \ll \Delta$ which yields $\tau = O(s_\pi)$. Furthermore, the bound given by Theorem 3.1 may not be tight suggesting an even slower growth of τ as a function of the skewness s_π.

We now compare our eigenvalue bound of (3.1) with that of Jerrum and Sinclair derived in [4]. To this end we need the following definitions.

Definition 4.1 : [4] Given a reversible Markov chain on state space Ω with transition matrix P and equilibrium vector π. The *conductance parameter* is defined as

$$\phi(P) = \min \frac{\sum_{i \in S, j \in V-S} p_{ij} \pi_i}{\sum_{i \in S} \pi_i} \tag{4.11}$$

where the above minimization is performed over all subsets S of states with $0 < \sum_{i \in S} \pi_i \le 1/2$.

Theorem 4.2 : [4] For a reversible Markov chain with transition matrix P satisfying $p_{ii} \ge 1/2$ for all $i \in \Omega$, and conductance parameter ϕ, we

have

$$1 - 2\phi \leq \lambda_2(P) \leq 1 - \phi^2/2 \tag{4.12}$$

We will consider two examples. The first one will demonstrate that our eigenvalue bound of (3.1) is fairly tight as a function of of the skewness s_π. The second example will show that the parameters required to compute our bound are much easier to estimate than the conductance parameter ϕ required by Jerrum and Sinclair to compute their bound given by (4.12).

Example 4.3 : Consider a simple cycle on $N = 4n$ vertices as the underlying graph of a FTSA Markov chain with a cost function defined as follows :

$$C(i) = \begin{cases} i & \text{if } 1 \leq i \leq n \\ 2n+1-i & \text{if } n+1 \leq i \leq 2n \\ i-2n & \text{if } 2n+1 \leq i \leq 3n \\ 4n+1-i & \text{if } 3n+1 \leq i \leq 4n \end{cases} \tag{4.13}$$

Using these costs, $\rho = 2$, and some $\epsilon > 0$, define the transition matrix P using (4.2) and (4.3) and set $\hat{P} = \frac{1}{2}(I+P)$. For transition matrix $\hat{P}$ it can be shown that $\Delta = n-1$, $\delta = 1$ $\alpha = \epsilon/4$, skewness $s = \epsilon^{-(n-1)}$ of the corresponding equilibrium vector $\pi(\epsilon)$. Since the underlying graph of $\hat{P}$ is a cycle graph, using (3.10) and (4.11) gives

$$\mu_2(Q) = 2(1 - \cos(\frac{\pi}{2n})) \quad , \quad \phi = \frac{\epsilon^{n-1} - \epsilon^n}{4(1-\epsilon^n)} \tag{4.14}$$

On this example, using our eigenvalue bound from (3.1) we get

$$1 - \lambda_2(\hat{P}) \geq \frac{\epsilon^n}{2}(1 - \cos(\frac{\pi}{2n})) \tag{4.15}$$

while the Jerrum and Sinclair bound from (4.12) gives

$$1 - \lambda_2(\hat{P}) \geq \epsilon^{2n-2}/32 \tag{4.16}$$

for sufficiently small ϵ and large n. If we look at the time constant τ as a function of the skewness s, then combining our upper-bound of (3.1) with the lower-bound of (4.12) results in

$$2s \leq \tau \leq \frac{2}{1 - \cos(\pi/2n)} s^{(1+1/(n-1))} \tag{4.17}$$

whereas using both the bounds from (4.12) yields

$$2s \leq \tau \leq 32 s^2 \tag{4.18}$$

This clearly shows that, the upper-bound in (4.17) is fairly tight for large n. Furthermore, for large skewness s, the upper-bound in (4.17) is considerably better than that of (4.18).

The purpose of Example 4.3 was merely to illustrate an example of a reversible Markov chain for which the eigenvalue bound (hence, a bound on the rate of convergence) is fairly tight. The corresponding optimization problem, however, is very easy, since, by construction, the states 1, $2n$, $2n+1$, and $4n$ have the globally minimum cost of 1. The following example illustrates a difficult and more realistic optimization problem for which one can still use our eigenvalue bound of (3.1) to obtain a meaningful bound on the time-constant of convergence of the corresponding FTSA Markov chain.

Example 4.4 : Let $\{a_1 \leq a_2 \leq \cdots \leq a_n\}$ be a set of n positive integers in ascending order and define $K = \frac{1}{2}(a_1 + a_2 + \cdots + a_n)$. Let Ω denote the state space of all binary vectors of length n and consider a state $u = (u_1, u_2, \ldots, u_n)$ where $u_i \in \{0,1\}$. Define the cost of the state as $C(u) = |K - \sum_{1 \leq i \leq n} a_i u_i|$. Define the neighbors of a state u as all states differing from u in exactly one bit and consider an FTSA algorithm to find the state of minimum cost. This is the optimization version of the SET_PARTITION problem that is known to be NP-Complete [12]. Clearly, $N = 2^n$, $\rho = n$, $\delta = a_n$, $\Delta = K$, skewness $s = \epsilon^K$, and the underlying graph is the n-dimensional hypercube for which $\mu_2(Q) = 2$. Using (4.10), we immediately get an upper-bound for the time constant of convergence as

$$\tau \leq n\, s^{1+a_n/K} \tag{4.19}$$

For example, if the given integers are $\{3,5,6,11,15\}$, we have $n = 5$, $a_n = 15$, and $K = 20$. From (4.19), we see that the time-constant for an FTSA algorithm to reach an equilibrium vector of skewness $s = 10^4$ is bounded above by $\tau \leq 5 \times 10^7$ iterations.

Estimating the conductance parameter for the chain in Example 4.4 is not straight forward; hence, the Jerrum and Sinclair eigenvalue bound of (4.12) is not directly useful in obtaining a meaningful bound for the time-constant in this case. However, with considerable ingenuity, Jerrum and Sinclair have been successful in obtaining good lower bounds for the conductance of certain classes of reversible chains [4]. Indeed, for these chains, the conductance is much larger than $O(1/s_\pi)$; hence, our upper-bound by (3.1) is not tight in this case. Our bound, on the other hand is very simple to compute in general, as demonstrated by Example 4.4, and is also tight on certain chains as considered in Example 4.3.

5. CONCLUSIONS

In this paper we have derived a new upper bound on the second largest eigenvalue of a reversible Markov chain. The bound is a simple function of the skewness of the equilibrium vector of the chain and we give examples of reversible chains where the upper bound is fairly tight. The upper bound on the eigenvalue enables us to study the time constant of convergence of the Markov chain to its equilibrium vector. In particular, we can bound the time constant of convergence of a fixed temperature simulated annealing (FTSA) algorithm solving a particular instance of an optimization problem. Moreover, we can study the growth of this bound as the temperature approaches zero or skewness becomes arbitrarily large.

REFERENCES

[1] E. Senata, *Nonnegative Matrices and Markov Chains,* Springer Verlag, 1980.

[2] D. Aldous, "On the Markov Chain Simulation Method for Uniform Combinatorial Distributions and Simulated Annealing," *Probability in Engineering and Informational Sciences,* Vol. 1, 1987.

[3] N. Alon, "Eigenvalues and Expanders," *Combinatorica,* Vol. 6, No. 2, pp. 83-96, 1986.

[4] A. Sinclair, and M. Jerrum, "Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains," *Information and Computing,* Vol. 82, No.1, pp 93-133, July 1989.

[5] J. Bondy, and U. S. R. Murthy, *Graph Theory with Applications,* North Holland, New York, 1979.

[6] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by Simulated Annealing," *Science,* Vol. 220, No. 4598, pp. 671-680, May 1983.

[7] B. Hajek, "A Tutorial Survey of Theory and Applications of Simulated Annealing," *Proc. 24th Conf. on Decision and Control,* Ft. Lauderdale, Florida, pp 755-760, December 1985.

[8] D. P. Connors and P. R. Kumar, "Simulated Annealing and Balance of Recurrence Order in Time-Inhomogeneous Markov Chains," *Proc. 26th Conf. on Decision and Control,* Los Angeles, California, pp. 2261-2263, December 1987.

[9] D. Mitra, F. Romeo, and A. L. Sangiovanni-Vincentelli, "Convergence and Finite-Time Behavior of Simulated Annealing," *Proc. 24th Conf. on Decision and Control,* Ft. Lauderdale, Florida, pp 761-767, December 1985.

[10] N. Biggs, *Algebraic Graph Theory,* Cambridge University Press, Cambridge 1974.

[11] N. Alon, and V. D. Milman, " λ_1, Isoperimetric Inequalities for Graphs and Superconcentrators," *J. Combinatorial Theory Ser. B,* Vol. 38, pp. 73-88, 1985.

[12] M. R. Garey, and D. S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness,* Freeman 1979.

[13] T. Kato, *Perturbation Theory for Linear Operators,* Springer-Verlag, Berlin, 1980.

[14] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming,* Addison-Wesley, Reading Mass., 1973.