The Annals of Applied Probability
1994, Vol. 4, No. 2, 347–389

ON THE RATE OF CONVERGENCE OF THE METROPOLIS ALGORITHM AND GIBBS SAMPLER BY GEOMETRIC BOUNDS[1]

BY SALVATORE INGRASSIA

Università di Catania

In this paper we obtain bounds on the spectral gap of the transition probability matrix of Markov chains associated with the Metropolis algorithm and with the Gibbs sampler. In both cases we prove that, for small values of T, the spectral gap is equal to $1 - \lambda_2$, where λ_2 is the second largest eigenvalue of P. In the case of the Metropolis algorithm we give also two examples in which the spectral gap is equal to $1 - \lambda_{\min}$, where $\lambda_{\min}$ is the smallest eigenvalue of P. Furthermore we prove that random updating dynamics on sites based on the Metropolis algorithm and on the Gibbs sampler have the same rate of convergence at low temperatures. The obtained bounds are discussed and compared with those obtained with a different approach.

1. Introduction. The Metropolis algorithm and the Gibbs sampler are Monte Carlo methods for the generation of samples from a finite set Ω (with $|\Omega| = N$) with a given probability distribution π which charges every point of Ω [Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953), Hastings (1970) and Geman and Geman (1984)]. They have many applications in pattern analysis and synthesis, image restoration and the implementation of simulated annealing. Both algorithms use an aperiodic and irreducible Markov chain P reversible with respect to the distribution π [i.e, $\pi(x)P(x, y) = \pi(y)P(y, x)$]. These hypotheses imply that π is the unique stationary distribution for P and that its spectrum $\{\lambda_i\}_{i=1,\dots,N}$ is real with $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_N > -1$; for simplicity we shall write $\lambda_{\min}$ rather than λ_N. Let $\rho(P) = \max\{\lambda_2, |\lambda_{\min}|\}$ be the second largest eigenvalue in absolute value of P; it is well known that $\rho(P)$ gives bounds on the rate of convergence of P toward its stationary distribution. Often the eigenvalue $\lambda_{\min}$ is not considered and only λ_2 is studied. In fact a slower Markov chain with transition probability matrix $\frac{1}{2}(I + P)$ rather than P is considered, where I is the $N \times N$ identity matrix; that is, one considers the random walk in which there is introduced extra holding probability of $\frac{1}{2}$ in each state. Obviously all the eigenvalues of this matrix are nonnegative and hence $\rho[\frac{1}{2}(I + P)] = (1 + \lambda_2)/2$. However, the main problem remains, because even if we know $\rho[\frac{1}{2}(I + P)]$, it would be interesting to know more information

Received August 1992; revised April 1993.

[1]Research partially supported by a grant of the Italian National Research Council.

AMS 1991 *subject classifications.* Primary 60J10; secondary 60J15, 15A42.

Key words and phrases. Gibbs sampler, Markov chains, Metropolis algorithm, rate of convergence.

about $\rho(P)$. Hanlon (1992) gives some Metropolis chains which can be explicitly diagonalized and for which $\rho(P)$ is equal to λ_2 in some cases and equal to $\lambda_{\min}$ in other cases.

In this paper we develop upper bounds on λ_2 and lower bounds on $\lambda_{\min}$ and hence on the spectral gap $1 - \rho(P)$ for the transition probability matrices associated with the two algorithms. These bounds depend on some geometric quantities of the underlying graph $G(P)$ of the Markov chain with transition matrix P. Related ideas were used by Chiang and Chow (1988) and Desai (1992). We show that in both the algorithms $\rho(P)$ is equal to λ_2 for small values of T. Furthermore, we prove that the random updating dynamics on sites based on the Metropolis algorithm and on the Gibbs sampler have the same rate of convergence at low temperatures.

The paper is organized as follows. In Section 2 we introduce notation and definitions that we shall use throughout this paper and, moreover, we recall some recent results of Diaconis and Stroock (1991) which constitute the basis of our main results. In Section 3 we give some preliminary results; in Sections 4 and 5, bounds on λ_2 and $\lambda_{\min}$ and, hence, on $\rho(P)$ are proved, respectively, for the Metropolis algorithm and for the Gibbs sampler. The bounds for λ_2 relative to the Metropolis algorithm are not completely new; in fact, we find similar results in Holley and Stroock [(1988), Theorem 2.1]. The other bounds, as far as we know, are new. In Section 6 we discuss the results: We compare our bounds with those of Desai (1992) and give some examples of applications both in cases in which λ_2 and $\lambda_{\min}$ are known and other cases of interest. Finally, we discuss the rates of convergence of the Metropolis algorithm and of the Gibbs sampler in random updating dynamics.

2. Geometric bounds for eigenvalues of reversible chains. In this section we introduce some basic notation and definitions and recall some recent results proved by Diaconis and Stroock (1991) in a quite general context: P is the transition probability matrix of an aperiodic and irreducible Markov chain on a state space Ω, and it is reversible with respect to a probability distribution π. Recently many authors [Sinclair and Jerrum (1989), Diaconis and Stroock (1991) and Sinclair (1991)], following a functional approach, have proved bounds on the eigenvalues λ_2 and $\lambda_{\min}$ of P depending on certain geometric quantities of the graph underlying the transition matrix P. We show that Diaconis and Stroock's (1991) results in our case lead to the tightest bounds.

Let P be an aperiodic, irreducible transition probability matrix which is reversible with respect to its stationary distribution π. Throughout this paper we assume these hypotheses on P and we shall not repeat them in the propositions and theorems.

We define the *underlying graph* $G(P) = [\Omega, E]$ *of the Markov chain with transition probability matrix* P as the graph with set of vertices Ω and set of edges E given by the pairs $\{x, y\} \in \Omega \times \Omega$ such that $P(x, y) > 0$ and $x \neq y$. The hypothesis of irreducibility of P implies that the graph $G(P)$ is connected.

We state that $x, y \in \Omega$ are *adjacent* or *neighbours* if $P(x, y) > 0$ (i.e., if $\{x, y\} \in E$). Furthermore, let $N(x)$ be the set of the vertices adjacent to x

$$N(x) = \{y \in \Omega: \{x, y\} \in E\},$$

and let $d(x) = |N(x)|$ be the degree of the vertex x.

Given any two vertices $x, y \in \Omega$ we denote by γ_{xy} a path from x to y: $\gamma_{xy} = (x = x_0, x_1, \dots, x_K = y)$ such that $P(x_{k-1}, x_k) > 0$, for each $k = 1, 2, \dots, K$; furthermore, we say that $z \in \gamma_{xy}$ if z is an element of the sequence $(x_k)_{k=0,\dots,K}$. Finally, we denote by $\Gamma = \{\gamma_{xy}\colon x, y \in \Omega\}$ a set of paths which join pairs of vertices of Ω (in such a way that Γ contains exactly one path for each pair of vertices).

Let Q be the matrix defined, for each $x, y \in \Omega$, by $Q(x, y) = \pi(x)P(x, y)$; the detailed balance condition implies that the matrix Q is symmetric and that $Q(x, y) = 0$ if and only if $P(x, y) = 0$. The *Q-length* of a path $\gamma_{xy} \in \Gamma$ is defined by

$$|\gamma_{xy}|_Q = \sum_{e \in \gamma_{xy}} (Q(e))^{-1},$$

where the sum is extended over the set of the edges which constitute the path γ_{xy}. Often it is better to use unit weights [see Diaconis and Stroock (1991), Proposition 1′].

NOTE 2.1. In the following we shall consider also directed edges; in this case we shall write $e = (e^-, e^+)$ when the edge e is directed from the vertex e^- to the vertex e^+. The set of directed edges is denoted $\vec{E}$.

Let a fixed set of paths Γ on $G(P)$ be given. The first quantity of interest is the parameter κ, which is related to the set Γ as follows:

$$\kappa = \max_{e \in \vec{E}} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_Q \pi(x)\pi(y), \tag{1}$$

where the maximum is over directed edges in the graph and the sum is over all the paths which travel a certain edge e. We note immediately that κ increases both with the number of paths which cross the same edge e (the sum is over all the paths which travel the edge e) and with the length of each of them (κ is a sum of lengths of paths).

Analogously let $\Sigma = \{\sigma_x\}$ be a collection of cycles with an odd number of edges (Σ contains only one cycle for each vertex), and let us define the Q-length of a cycle $\sigma_x \in \Sigma$ by

$$|\sigma_x|_Q = \sum_{e \in \sigma_x} (Q(e))^{-1},$$

and hence the parameter ι,

$$\iota = \max_{e \in \vec{E}} \sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x), \tag{2}$$

where the maximum is over directed edges in the graph and the sum is over all the cycles which travel a certain edge e.

By an application of the Cauchy–Schwarz inequality, Diaconis and Stroock (1991) have proved the following bounds on λ_2 and $\lambda_{\min}$ in terms of the parameters κ and ι.

PROPOSITION 2.1. *The eigenvalues λ_2 and $\lambda_{\min}$ of P satisfy*

$$\lambda_2 \le 1 - \frac{1}{\kappa}, \tag{3}$$

$$\lambda_{\min} \ge -1 + \frac{2}{\iota}, \tag{4}$$

with κ and ι defined, respectively, in (1) *and* (2).

The bound (3) is a discrete version of the Poincaré inequality. The next proposition gives a lower bound for λ_2. Let W be a proper subset of Ω and define

$$Q(W \times W^c) = \sum_{x \in W,\, y \in W^c} Q(x, y) = \sum_{x \in W,\, y \in W^c} \pi(x) P(x, y).$$

Then [e.g., Diaconis and Stroock (1991)].

PROPOSITION 2.2. *Let W be a proper subset of Ω. Then we have*

$$\lambda_2 \ge 1 - \frac{Q(W \times W^c)}{\pi(W)\pi(W^c)}. \tag{5}$$

In the next section, with regard to the problem under examination, we shall give a choice of the sets Γ and Σ, and we shall prove some preliminary results.

3. Preliminary results: A choice of paths and cycles. Let H be a real-valued function defined on Ω and, without loss of generality, assume that H is a nonconstant function with $\min_{x \in \Omega} H(x) = 0$. Let x and y be two states of the set Ω. We shall say that x and y are *equivalent* if there exists a path of constant energy from x to y, that is, if there exists a path $\gamma_{xy} = (x_k)_{k=0,\dots,K}$, with $x_0 = x$ and $x_K = y$, such that $H(x_k) = H(x_{k-1})$, for each $k = 1, \dots, K$. Let us denote by Γ_{xy} the set of all paths from x to y. We shall say that x is a *point of local minimum for the function H* either if the set $\{y \in \Omega | H(y) < H(x)\}$ is empty (in such a case x is a point of absolute minimum) or if, for each $y \in \Omega$ with $H(y) < H(x)$ and for each path $\gamma_{xy} \in \Gamma_{xy}$, there exists $z \in \gamma_{xy}$ such that $H(z) > H(x)$.

For each $x, y \in \Omega$ let $\gamma_{xy} = (x_k)_{k=0,\dots,K}$ be any path from x to y and let $\{H(x_k)\}_{k=0,\dots,K}$ be the corresponding set of values assumed by the function in the points of γ_{xy}. For each $\gamma_{xy} \in \Gamma_{xy}$, let us define

$$\text{elev}(\gamma_{xy}) \equiv \max_k \{H(x_k)\}$$

and then

$$H_{xy} \equiv \min_{\gamma_{xy} \in \Gamma_{xy}} \operatorname{elev}(\gamma_{xy}).$$

Hence H_{xy} is the lowest possible elevation along any path from x to y. Afterwards we set

$$m \equiv \max_{x,\, y \in \Omega} \{H_{xy} - H(x) - H(y)\}. \tag{6}$$

The parameter m is the least total elevation gain of the function H in the sense described preceding Theorem 2.1 in Holley and Stroock (1988).

We shall consider the following set Γ: For each $x, y \in \Omega$, the set contains a path having an elevation $\operatorname{elev}(\gamma_{xy})$ such that

$$\operatorname{elev}(\gamma_{xy}) - H(x) - H(y) \leq m. \tag{7}$$

We shall refer to these paths as *admissible paths*.

NOTE 3.1. Holley and Stroock (1988) consider a particular set of admissible paths, the set of paths having the lowest possible elevation. Actually this choice is too restrictive; in fact, for the proof of Theorem 2.1 in Holley and Stroock (1988), condition (7) is sufficient.

The following proposition gives a characterization of the function H in terms of the parameter m now defined.

PROPOSITION 3.1. *Let Ω be a finite set and let H be a real-valued function defined on Ω. Then the parameter m defined in* (6) *is nonnegative. Furthermore, $m = 0$ if and only if the function H has only one point of local minimum* (*up to equivalence*).

PROOF. The first statement of the proposition is obvious. It is sufficient to choose either x or y as a point of absolute minimum. Regarding the second assertion, at first let us suppose that

$$\max_{x,\, y \in \Omega} \{H_{xy} - H(x) - H(y)\} = 0,$$

and we shall prove that the function H has only one point of local minimum (up to equivalence). We prove it by contradiction. Accordingly, suppose that, in addition to a point x_0 of absolute minimum, there exists a distinct point y_0 of local minimum that is not equivalent to x_0. We must consider the following two cases: (1) $0 = H(x_0) = H(y_0)$; (2) $0 = H(x_0) < H(y_0)$.

In the first case, the hypothesis $m = 0$ implies that $H_{x_0 y_0} = 0$ and hence that there exists a path of constant energy from x_0 to y_0, contradicting the hypothesis that the points are not equivalent. In the second case, the hypothesis $m = 0$ implies that $H_{x_0 y_0} = H(y_0)$; but this is a contradiction because, from the definition of a point of local minimum given at the beginning of this section, along any path $\gamma_{x_0 y_0} \in \Gamma_{x_0 y_0}$ there must exist a point z such that $H(z) > H(y_0)$, and hence $H_{x_0 y_0} > H(y_0)$.

Conversely, let us suppose that the function H has only one point of local minimum (up to equivalence) and let us prove that $H_{xy} - H(x) - H(y) \leq 0$, for all $x, y \in \Omega$. We need only to show that, for each $x_0, x, y \in \Omega$ with $0 = H(x_0) \leq H(x) \leq H(y)$, the hypotheses imply that $H_{xy} = H(y)$. Obviously $H_{xy} = H(y)$ if y is a point of absolute maximum for the function H. In general we suppose, again for contradiction, that for any $\gamma_{xy} \in \Gamma_{xy}$ there exists $z \in \gamma_{xy}$ such that $H_{xy} = H(z) > H(y)$ and prove that this hypothesis implies that the function H has at least two nonequivalent points of local minimum. We can prove it by the construction of two finite and decreasing sequences which terminate in two nonequivalent points of local minima, say, y_J and x_I.

In fact, if y is not a point of local minimum for H, then there exists a point $y_1 \in N(y) \setminus \{y\}$ such that $H(y_1) < H(y)$; if y_1 is not a point of local minimum for H, then there exists a point $y_2 \in N(y_1) \setminus \{y, y_1\}$ such that $H(y_2) < H(y_1)$ and so on. In such a way we obtain a finite sequence $\{y_j\}_{j=0,\dots,J}$, where $y_0 = y$ and y_J is a point of local minimum. With the same argument for the point x, we can obtain another point of local minimum from the sequence $\{x_i\}_{i=0,\dots,I}$. Obviously these two points cannot coincide or be equivalent. This completes the proof. □

As regards the choice of the set Σ of cycles with an odd number of edges, first let us consider the set Λ_P defined by

$$\Lambda_P = \{x \in \Omega | P(x, x) > 0\}. \tag{8}$$

We denote by Σ_x the set of all cycles from x to x having an odd number of edges, and we denote by $r(\sigma_x, \Lambda_P)$ the number of edges which must be travelled along the (general) cycle σ_x, starting at x, to reach a vertex $y \in \sigma_x \cap \Lambda_P$. Then set

$$r_x = \min_{\sigma_x \in \Sigma_x} r(\sigma_x, \Lambda_P), \tag{9}$$

with $r_x = 0$ if $x \in \Lambda_P$.

We take for $\Sigma = \{\sigma_x\colon x \in \Omega\}$ the set of cycles σ_x having the minimum number of edges indispensable to join the vertex x to a vertex of Λ_P.

In other words, $\sigma_x \in \Sigma$ if and only if either (1) $x \in \Lambda_P$ and $\sigma_x = x \to x$ or (2) $x \notin \Lambda_P$ and $\sigma_x = \{x_k\}_{k=0,\dots,K}$ is an odd cycle with the minimum number of edges K such that

$$x_0 \equiv x_K, \dots, x_{(K-3)/2} \equiv x_{(K+3)/2} \notin \Lambda_P \quad \text{and} \quad x_{(K-1)/2} \equiv x_{(K+1)/2} \in \Lambda_P.$$

NOTE 3.2. Unlike the set of paths Γ, the set of cycles Σ depends on the transition probability matrix P.

Finally, let us set

$$r^* = \max_{x \in \Omega} r_x. \tag{10}$$

In particular, each cycle $\sigma_x \in \Sigma$ has a number of edges not greater than $2r^* + 1$. We shall give some characterizations of the set Λ_P in the next section.

NOTE 3.3. Obviously, more than one set of admissible paths Γ and more than one set of shortest cycles Σ can exist. In the following we shall always consider one of the possible choices.

4. Results relating to the Metropolis algorithm. Let H be a real-valued function defined on Ω and, without loss of generality, assume that H is a nonconstant function with $\min_{x \in \Omega} H(x) = 0$. From now on π is the Gibbs distribution depending on the function H and on a parameter $T > 0$:

$$\pi(x) \equiv \frac{e^{-H(x)/T}}{Z_T} \quad \forall\, x \in \Omega, \tag{11}$$

where $Z_T = \sum_{x \in \Omega} e^{-H(x)/T}$ is the *partition function*. The set Ω is the set of *states* or *configurations* of a given physical system, the function H is called *energy* [hence $H(x)$ is the energy of the state $x \in \Omega$] and the positive parameter T is the *temperature*.

Let $G = [\Omega, E]$ be a connected graph; for simplicity, we suppose that G contains neither self-loops nor multiple edges. Let us denote by $d^* = \max_{x \in \Omega} d(x)$ the maximum degree of the graph. Let $R = (R(x, y))_{x, y \in \Omega}$ be the symmetric transition probability matrix on the set Ω defined for each $x, y \in \Omega$ by

$$R(x, y) \equiv \begin{cases} \dfrac{1}{d^*}, & \text{if } y \in N(x) \text{ and } y \neq x, \\ \dfrac{d^* - d(x)}{d^*}, & \text{if } y = x, \\ 0, & \text{otherwise}. \end{cases} \tag{12}$$

The transition probability matrix R is irreducible and reversible with respect to the uniform distribution on Ω.

The Metropolis algorithm is based on a Markov chain with transition probability matrix $P = (P(x, y))_{x, y \in \Omega}$ defined as follows:

$$P(x, y) = \begin{cases} R(x, y), & \text{if } \pi(y) \geq \pi(x) \text{ and } y \neq x, \\ \dfrac{\pi(y)}{\pi(x)} R(x, y), & \text{if } \pi(y) < \pi(x), \\ 1 - \sum_{z \neq x} P(x, z), & \text{if } y = x, \end{cases} \tag{13}$$

where π is the Gibbs distribution defined in (11) and the transition matrix R has been defined in (12).

The transition probability P is irreducible and reversible with respect to the Gibbs measure. In fact, for $\pi(x) \leq \pi(y)$ and $y \neq x$, we have

$$\pi(x)P(x,y) = \pi(x)R(x,y) = \pi(x)R(y,x) = \pi(y)P(y,x).$$

Furthermore, the transition probability matrix P is aperiodic [e.g., see Frigessi, Hwang, Sheu and Di Stefano (1993), Proposition 3].

Note that $G(P) = G = [\Omega, E]$ is the underlying graph of the Markov chain with transition probability matrix P. In this section we obtain a bound on the spectral gap of the transition matrix P in terms of some geometric quantities of the underlying graph $G(P)$ of the Markov chain and of the function H and prove that $\rho(P) = \lambda_2$ for small values of T. Two other quantities of interest are

$$b_\Gamma \equiv \max_e \#\{\gamma \in \Gamma | e \in \gamma\}, \tag{14}$$

$$\gamma_\Gamma \equiv \text{maximum number of edges in any path of } \Gamma. \tag{15}$$

4.1. *Bounds on* λ_2. In the next two theorems we get results similar to those of Theorem 2.1 in Holley and Stroock (1988). The difference consists in the fact that we are able to specify the constant which appears in front of $e^{-m/T}$.

THEOREM 4.1. *Let P be the transition matrix of the Markov chain* (13). *Then its second largest eigenvalue λ_2 satisfies the relation*

$$\lambda_2 \leq 1 - \frac{Z_T}{b_\Gamma \gamma_\Gamma d^*} e^{-m/T}, \tag{16}$$

where $Z_T = \sum_{x \in \Omega} e^{-H(x)/T}$.

PROOF. We choose a set of admissible paths Γ on the graph $G(P)$, that is, a set of paths satisfying (7). Let $e \in \vec{E}$; by the definition of R it follows that

$$Q(e) \equiv \pi(e^-)P(e) = \begin{cases} \dfrac{\pi(e^-)}{d^*}, & \text{if } \pi(e^-) \leq \pi(e^+), \\ \dfrac{\pi(e^+)}{d^*}, & \text{if } \pi(e^-) > \pi(e^+), \end{cases}$$

where $P(e) = P(e^-, e^+)$ and $Q(e) = Q(e^-, e^+)$. The Q-length of the path γ_{xy} from x to y is given by

$$|\gamma_{xy}|_Q \equiv \sum_{e \in \gamma_{xy}} (Q(e))^{-1} = \sum_{e \in \gamma_{xy}} [\pi(e^-)P(e)]^{-1},$$

and, if we set $\pi(e_*) = \min\{\pi(e^-), \pi(e^+)\}$, it follows that

$$Q(e) \equiv \pi(e^-)P(e) = \frac{\pi(e_*)}{d^*}.$$

Hence

$$\begin{aligned}|\gamma_{xy}|_Q \pi(x)\pi(y) &= \sum_{e \in \gamma_{xy}} [\pi(e^-)P(e)]^{-1} \pi(x)\pi(y) \\ &= \frac{1}{Z_T} \sum_{e \in \gamma_{xy}} d^* \exp\left[\frac{H(e_*) - H(x) - H(y)}{T}\right],\end{aligned}$$

where $H(e_*) = \max\{H(e^-), H(e^+)\}$. Thus, by the definition (1) of κ, it follows that

$$\begin{aligned}\kappa &\equiv \max_{e'} \sum_{\gamma_{xy} \ni e'} |\gamma_{xy}|_Q \pi(x)\pi(y) \\ &= \frac{1}{Z_T} \max_{e'} \sum_{\gamma_{xy} \ni e'} \sum_{e \in \gamma_{xy}} d^* \exp\left[\frac{H(e_*) - H(x) - H(y)}{T}\right].\end{aligned}$$

Since $H(e_*) - H(x) - H(y) \le \text{elev}(\gamma_{xy}) - H(x) - H(y) \le m$, for each $e \in \gamma_{xy}$, by our choice of Γ, from the definitions of m, b_Γ and γ_Γ—respectively, (6), (14) and (15)—we obtain

$$\kappa \le \frac{b_\Gamma \gamma_\Gamma d^*}{Z_T} e^{m/T}.$$

The theorem follows from Proposition 2.1. □

NOTE 4.1. If $H \equiv 0$, then P here is a random walk, and the bound (16) reduces to

$$\lambda_2 \le 1 - \frac{N}{b_\Gamma \gamma_\Gamma d^*},$$

which is exactly the bound of Corollary 1 in Diaconis and Stroock (1991), if G is assumed regular, since then $2|E|$ equals Nd^*. Note also that, in the special case $H \equiv 0$, *any* choice of paths is permissible.

NOTE 4.2. The fundamental difference between the bounds obtained in this paper and those of Diaconis and Stroock (1991) obtained in the case of simple random walks is exhibited by comparing Theorem 4.1 with Corollary 1 in Diaconis and Stroock (1991). In fact they give a bound on λ_2 for *any* choice of the set Γ and hence the problem is to find an optimal choice of this set (the optimal set may not be unique, as we remark in Note 3.3). By contrast the bounds (16) and all the bounds obtained in this paper are obtained corresponding to the set Γ of admissible paths discussed in the previous section. Analogous comments apply to the set of cycles Σ.

Other geometric bounds on λ_2 can be obtained with other approaches. With the notation of Section 3, let

$$\eta = \max_{e \in \vec{E}} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y), \tag{17}$$

$$\bar{\kappa} = \max_{e \in \vec{E}} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}| \pi(x)\pi(y), \tag{18}$$

where $|\gamma_{xy}|$ is the number of edges of the path γ_{xy} and the maximum is over directed edges. Sinclair and Jerrum (1989) and Sinclair (1991), respectively, have proved the following bounds on λ_2:

$$\lambda_2 \le 1 - \frac{1}{8\eta^2}, \tag{19}$$

$$\lambda_2 \le 1 - \frac{1}{\bar{\kappa}}, \tag{20}$$

based, respectively, on the Cheeger inequality and the Poincaré inequality. Hence we have the following theorem.

THEOREM 4.2. *Let P be the transition matrix of the Markov chain* (13). *Then its second largest eigenvalues λ_2 satisfies the relation*

$$\lambda_2 \le 1 - \frac{Z_T^2}{8(b_\Gamma d^*)^2} e^{-2m/T}. \tag{21}$$

PROOF. The proof is the same of that of Theorem 4.1. From (17) we have

$$\begin{aligned}
\eta &= \max_{e \in \vec{E}} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) \\
&= \max_{e \in \vec{E}} \frac{d^*}{\pi(e^*)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) \\
&\le \frac{b_\Gamma d^*}{Z_T} e^{m/T}.
\end{aligned}$$

The theorem follows from (19). □

Hence if

$$e^{-m/T} < \frac{8 b_\Gamma d^*}{\gamma_\Gamma},$$

then the bound of Theorem 4.1 is tighter than the bound obtained in Theorem 4.2. On the contrary, with easy calculations it can be proved that the bound (20) gives the same result as Theorem 4.1.

Now we achieve a lower bound for λ_2. In this theorem some inequalities are quite crude, but here we are mainly interested in the rate of convergence of λ_2.

THEOREM 4.3. *Let P be the transition matrix of the Markov chain* (13). *Then its second largest eigenvalue satisfies*

$$\lambda_2 \geq 1 - \frac{Z_T}{d^*}\left(\frac{N}{2}\right)^2 e^{-m/T}, \tag{22}$$

where N is the cardinality of the set Ω and m has been defined in (6).

PROOF. The first part of the proof follows as in Holley and Stroock [(1988), Lemma 2.3]. Let x_0 and y_0 be such that $H(x_0) \leq H(y_0)$ and $H_{x_0 y_0} - H(x_0) - H(y_0) = m$. Let W be the following subset of Ω:

$$W = \{z \in \Omega | H_{x_0 z} < H_{x_0 y_0}\}.$$

The result is that $x_0 \in W$ and $y_0 \in W^c$. Let $(x, y) \in \vec{E} \cap (W \times W^c)$, where $\vec{E}$ is the set of directed edges. Then, by the choice of the set W we have $H(y) \geq H_{x_0 y_0}$ and $H(y) > H(x)$. In fact, the path $\gamma_{x_0 y}$ can be chosen so that $\gamma_{x_0 y} = (\gamma_{x_0 x}, y)$ and hence, from the definition of W, $H_{x_0 x} < H_{x_0 y_0} \leq H_{x_0 y}$. Therefore $H(y) > H(x)$ and $H(y) \geq H_{x_0 y_0}$, and hence

$$Q(x, y) = \pi(x) P(x, y) = \frac{\pi(y)}{d^*}.$$

Then, by Proposition 2.2,

$$\begin{aligned}
\lambda_2 &\geq 1 - \frac{Q(W \times W^c)}{\pi(W)\pi(W^c)} \\
&= 1 - \frac{\sum_{(x, y) \in \vec{E} \cap (W \times W^c)} \pi(y)}{d^* \pi(W)\pi(W^c)} \\
&= 1 - \frac{1}{Z_T} \frac{\sum_{(x, y) \in \vec{E} \cap (W \times W^c)} e^{-H(y)/T}}{d^* \pi(W)\pi(W^c)}.
\end{aligned}$$

Since

$$H(y) \geq H_{x_0 y_0},$$
$$\exp[-H(x_0)/T] \leq \sum_{x \in W} \exp[-H(x)/T]$$

and

$$\exp[-H(y_0)/T] \leq \sum_{y \in W^c} \exp[-H(y)/T],$$

it follows that

$$\lambda_2 \geq 1 - Z_T \frac{1}{d^*} \sum_{(x,y) \in \vec{E} \cap (W \times W^c)} \exp\left\{ \frac{-\left(H_{x_0 y_0} - H(x_0) - H(y_0)\right)}{T} \right\}.$$

Consequently, we have

$$\lambda_2 \geq 1 - Z_T \frac{1}{d^*} \sum_{(x,y) \in \vec{E} \cap (W \times W^c)} e^{-m/T}.$$

Finally, since the set $\vec{E} \cap (W \times W^c)$ contains at most $(N/2)^2$ edges, we have

$$\lambda_2 \geq 1 - \frac{Z_T}{d^*} \left(\frac{N}{2} \right)^2 e^{-m/T}$$

and hence the theorem. □

4.2. *Bounds on* $\lambda_{\min}$. Now we turn to the bound on $\lambda_{\min}$. First we must provide a characterization of the set Λ_P.

PROPOSITION 4.1. *Let P be the transition matrix of the Markov chain* (13). *Then the set Λ_P does not depend on the temperature T.*

PROOF. The proposition follows immediately by considering that $P(x, x) = 1 - \sum_{y \neq x} P(x, y)$. □

PROPOSITION 4.2. *Let P be the transition matrix of the Markov chain* (13) *and $x \in \Omega$. Then $x \notin \Lambda_P$ if and only if $d(x) = d^*$ and $H(y) \leq H(x)$, for each $y \in N(x)$.*

PROOF. The statement follows from the definition of P since $x \notin \Lambda_P$ if and only if $P(x, x) = 0$, that is, $\sum_{y \neq x} P(x, y) = 1$. By definition of P, this holds if and only if $\pi(y) \geq \pi(x)$ for all $y \in N(x)$ and $d(x) = d^*$. □

The next two corollaries follows from Proposition 4.1.

COROLLARY 4.1. *Let P be the transition matrix of the Markov chain* (13), *and let r^* be the parameter defined in* (10). *Then $r^* \geq 1$.*

PROOF. Obvious, as $\Lambda_P \subset \Omega$. □

COROLLARY 4.2. *Let P be the transition matrix of the Markov chain* (13). *If x and y are two distinct vertices of Ω and $\gamma_{xy} = (x_k)_{k=0,\ldots,K}$ is a path from x to y composed entirely of vertices not belonging to Λ_P, then γ_{xy} is a path of constant energy.*

PROOF. It follows immediately from Corollary 4.1. □

Now let us introduce two other quantities, namely,

$$\delta = \min_{x \in \Omega} \min_{y \in N(x):\, H(y) \neq H(x)} |H(y) - H(x)|, \tag{23}$$

$$b_\Sigma = \max_{e \in \vec{E}:\, e \text{ is not a self-loop}} \#\{\sigma \in \Sigma | e \in \sigma\}. \tag{24}$$

We have assumed $H \not\equiv 0$ so that $0 < \delta < \infty$. The parameter δ is the minimum gap of the function in any transition; b_Σ, with respect to the set of the shortest cycles Σ, has the same role as b_Γ with respect to the set of paths Γ. Furthermore, let us denote with e_Σ an edge which is crossed by the most number of $\sigma \in \Sigma$, that is,

$$e_\Sigma = \arg\max_{e \in \vec{E}} \left\{ \sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x) \right\}. \tag{25}$$

THEOREM 4.4. *Let P be the transition matrix of a Markov chain associated with the Metropolis algorithm given in* (13). *Then the least eigenvalue $\lambda_{\min}$ of P satisfies*

$$\lambda_{\min} \geq -1 + \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T} - 1}\right)^{-1}, \tag{26}$$

where δ has been defined in (23); *A and B are two constants given by $A = 2b_\Sigma r^* + b_\Sigma + 1$ and $B = b_\Sigma + 1$, with r^* and b_Σ given, respectively, in* (10) *and* (24).

PROOF. Let us choose a set of shortest cycles Σ with an odd number of edges as defined in Section 3:

(i) If $x \in \Lambda_P$, then

$$\sigma_x = x \to x \equiv \gamma_{xx}.$$

(ii) Otherwise, let $\gamma_{xy} = (x_k)_{k=0,\dots,K}$ (with $x_0 = x$ and $x_K = y$) be a shortest path from x to y such that $x_k \notin \Lambda_P$, for $k = 0, \dots, K-1$, and $y \in \Lambda_P$.

For simplicity throughout this proof, we shall use the same notation γ to denote subpaths of the cycle σ_x, but these paths should not be confused with the earlier ones.

The proof follows the scheme of Theorem 4.1: First we shall obtain an upper bound on the parameter ι given in (2),

$$\iota = \max_{e \in \vec{E}} \sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x),$$

and then we shall apply Proposition 2.1.

In Section 3 we introduced the quantity r_x as the smallest number of edges which must be travelled along the cycle σ_x starting at x to reach the nearest vertex $y \in \sigma_x \cap \Lambda_P$, and then $r^* = \max_{x \in \Omega} r_x$. Hence for each $e \in \vec{E}$ we can write

$$\sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x) = \sum_{\sigma_x \ni e} \left(1_{\{r_x \geq 1\}} |\sigma_x|_Q \pi(x) + 1_{\{r_x = 0\}} |\sigma_x|_Q \pi(x) \right), \tag{27}$$

where the real function $1_{\{\cdot\}}$ is the indicator function of the set $\{\cdot\}$. In other words we split the sum $\sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x)$ of the cycles which share the edge e into two parts. The first one corresponds to the cycles $\sigma_x \in \Sigma$ with $x \notin \Lambda_P$; the second one corresponds to the self-loop $e = \{x, x\}$.

First let us consider the case $x \notin \Lambda_P$. Any cycle σ_x with $r_x \geq 1$ can be described as

$$\sigma_x = (\gamma_{xy}, \gamma_{yy}, \gamma_{yx}),$$

where $y \in \Lambda_P$ and γ_{yx} is simply γ_{xy} reversed. To keep the proof simple (the other case is harder), we consider here the case $d(y) = d^*$. The subpath γ_{xy} can be further split as

$$\gamma_{xy} = (\gamma_{xw}, \gamma_{wy}),$$

where γ_{xw} is the subset of γ_{xy} with vertices not belonging to Λ_P having a constant energy [and equal to $H(x)$] and $r_x - 1$ edges; γ_{wy} is the edge (w, y). Then we can write

$$\sigma_x = (\gamma_{xw}, \gamma_{wy}, \gamma_{yy}, \gamma_{yw}, \gamma_{wx}), \tag{28}$$

and hence

$$\begin{aligned} |\sigma_x|_Q \pi(x) = {} & |\gamma_{xw}|_Q \pi(x) + |\gamma_{wy}|_Q \pi(x) + |\gamma_{yy}|_Q \pi(x) \\ & + |\gamma_{yw}|_Q \pi(x) + |\gamma_{wx}|_Q \pi(x). \end{aligned} \tag{29}$$

We have to estimate each part of $|\sigma_x|_Q$ in (29), considering that $|\gamma_{xw}|_Q = |\gamma_{wx}|_Q$ and $|\gamma_{wy}| = |\gamma_{yw}|_Q$.

To begin, as γ_{xy} is a path of constant energy equal to $H(x)$ with $r_x - 1$ edges, we have

$$\begin{aligned} |\gamma_{xw}|_Q \pi(x) &= \sum_{e \in \gamma_{xw}} (Q(e))^{-1} \pi(x) = \sum_{e \in \gamma_{xw}} \frac{d^*}{\pi(x)} \pi(x) \\ &= (r_x - 1) d^* \leq (r^* - 1) d^*. \end{aligned} \tag{30}$$

Furthermore, from Corollary 4.1, it follows that $H(x) = H(w) \geq H(y)$ and hence

$$|\gamma_{wy}|_Q \pi(x) = [\pi(w) P(w, y)]^{-1} \pi(x) = \frac{d^*}{\pi(w)} \pi(x) = d^*. \tag{31}$$

The bound on $|\gamma_{yy}|_Q \pi(x)$ is more complicated:

$$\begin{aligned} |\gamma_{yy}|_Q \pi(x) &= [\pi(y) P(y, y)]^{-1} \pi(x) \\ &= \left[\pi(y) \left(1 - \sum_{z \in N(y) \setminus \{y\}} P(y, z) \right) \right]^{-1} \pi(x) \\ &= \left[1 - \frac{1}{d^*} \sum_{z \in N(y) \setminus \{y\}} \exp\left(\frac{-(H(z) - H(y))^+}{T} \right) \right]^{-1} \frac{\pi(x)}{\pi(y)}. \end{aligned}$$

As $y \in \Lambda_P$ and $d(y) = d^*$, Corollary 4.1 implies that there exists at least one $z_0 \in N(y) \setminus \{y\}$ such that $H(z_0) > H(y)$; then we have

$$
\begin{aligned}
|\gamma_{yy}|_Q \pi(x) &\le \left[1 - \frac{1}{d^*}\left(d^* - 1 + \exp\left[\frac{-(H(z_0) - H(y))^+}{T}\right]\right)\right]^{-1} \frac{\pi(x)}{\pi(y)} \\
&= d^* \frac{\exp\left[-(H(x) - H(y))^+/T\right]}{1 - \exp\left[-(H(z_0) - H(y))^+/T\right]} \\
&\le d^* \frac{1}{1 - \exp\left[-(H(z_0) + H(y))^+/T\right]} \\
&\le d^* \frac{1}{1 - \exp(-\delta/T)} \\
&= d^* \left(1 + \frac{1}{\exp(\delta/T) - 1}\right),
\end{aligned} \tag{32}
$$

where δ is defined in (23).

Finally, from (29) by means of (30)–(32) we obtain

$$
1_{\{r_x \ge 1\}} |\sigma_x|_Q \pi(x) \le d^* \left(2r^* + 1 + \frac{1}{e^{\delta/T} - 1}\right) \tag{33}
$$

and then a bound on the first part of (27);

$$
\sum_{\sigma_x \ni e} \left(1_{\{r_x \ge 1\}} |\sigma_x|_Q \pi(x)\right) \le b_\Sigma d^* \left(2r^* + 1 + \frac{1}{e^{\delta/T} - 1}\right) \tag{34}
$$

where b_Σ is defined in (24).

Now let us turn to the case $x \in \Lambda_P$. Again, for simplicity, we consider $d(x) = d^*$ (the other case is harder). A bound on the second term of the right-hand side of (27) can be obtained with arguments similar to that of the bound (32). Thus we have (we shall omit details)

$$
\begin{aligned}
|\gamma_{xx}|_Q \pi(x) &= [\pi(x) P(x,x)]^{-1} \pi(x) \\
&= (P(x,x))^{-1} \\
&= \left[1 - \sum_{z \in N(x)\setminus\{x\}} P(x,z)\right]^{-1} \\
&= \left[1 - \frac{1}{d^*} \sum_{z \in N(x)\setminus\{x\}} \exp\left(\frac{-(H(z) - H(x))^+}{T}\right)\right]^{-1} \\
&\le d^* \frac{\exp(\delta/T)}{\exp(\delta/T) - 1} \\
&= d^* \left(1 + \frac{1}{\exp(\delta/T) - 1}\right).
\end{aligned} \tag{35}
$$

Hence, from (34) and (35), we have obtained a bound on (27):

$$\sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x) \le b_\Sigma d^* \left(2r^* + 1 + \frac{1}{e^{\delta/T} - 1}\right) + d^* \left(1 + \frac{1}{e^{\delta/T} - 1}\right) \tag{36}$$

and, with some calculations, we obtain the following upper bound on ι:

$$\begin{aligned} \iota &\equiv \max_e \sum_{\sigma_x \ni e} |\sigma_x|_Q \pi(x) \\ &\le d^* \left(2b_\Sigma r^* + b_\Sigma + 1 + \frac{b_\Sigma + 1}{e^{\delta/T} - 1}\right). \end{aligned} \tag{37}$$

The theorem follows from Proposition 2.1, setting $A = 2b_\Sigma r^* + b_\Sigma + 1$ and $B = b_\Sigma + 1$. □

NOTE 4.3. If we can find a set of shortest cycles Σ such that e_Σ is not a self-loop, then we shall obtain a tighter bound on ι than (37), as $1_{\{r_x = 0\}} = 0$ in (27). In fact in this case we have

$$\iota \le d^* \left(2b_\Sigma r^* + b_\Sigma + \frac{b_\Sigma}{e^{\delta/T} - 1}\right).$$

In this case set $A = 2r^* b_\Sigma + b_\Sigma$ and $B = b_\Sigma$ to obtain the result of Theorem 4.4.

4.3. *Bounds on $\rho(P)$.* The results of the previous two subsections give a bound on $\rho(P)$. Now we prove that the bound on $\rho(P)$ uses (16) for small values of T and the (26) for large values of T. Finally, we shall prove that $\rho(P) = \lambda_2$ for small values of T.

THEOREM 4.5. *Let P be the transition matrix of the Markov chain associated with the Metropolis algorithm* (13), *and let ξ be the number of absolute minima of the function H.*

Claim 1. *If $m > 0$ and $T \le T_1$, then we have*

$$\rho(P) \le 1 - \frac{\xi}{b_\Gamma \gamma_\Gamma d^*} e^{-m/T}, \tag{38}$$

where

$$T_1 = \min\left[m\left(\log\frac{N(A+B)}{2b_\Gamma\gamma_\Gamma}\right)^{-1}, \frac{\delta}{\log 2}\right]. \tag{39}$$

Claim 2. *If $m > 0$ and $T \ge T_2$, then we have*

$$\rho(P) \le 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T} - 1}\right)^{-1}, \tag{40}$$

where

$$T_2 = \max\left[m\left(\log\frac{\xi(A+B)}{2b_\Gamma\gamma_\Gamma}\right)^{-1}, \frac{\delta}{\log 2}\right]. \tag{41}$$

Claim 3. *If* $m = 0$, *then*

$$\rho(P) \le 1 - \min\left[\frac{\xi}{b_\Gamma\gamma_\Gamma d^*}, \frac{2}{Ad^*}\right], \tag{42}$$

for each $T > 0$.

PROOF OF CLAIM 1. According to Theorem 4.1 we have

$$\lambda_2 \le 1 - \frac{Z_T}{b_\Gamma\gamma_\Gamma d^*}e^{-m/T},$$

and according to Theorem 4.4 we have

$$\lambda_{\min} \ge -1 + \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T}-1}\right)^{-1}.$$

The proof of the claim is completed by showing that

$$1 - \frac{Z_T}{b_\Gamma\gamma_\Gamma d^*}e^{-m/T} \ge 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T}-1}\right)^{-1}$$

for $T \le T_1$, and observing that $\xi \le Z_T$. Indeed, from (39) we have

$$\begin{aligned}
1 - \frac{Z_T}{b_\Gamma\gamma_\Gamma d^*}e^{-m/T} &\ge 1 - \frac{N}{b_\Gamma\gamma_\Gamma d^*}e^{-m/T} \\
&\ge 1 - \frac{N}{b_\Gamma\gamma_\Gamma d^*}e^{-m/T_1} \\
&\ge 1 - \frac{N}{b_\Gamma\gamma_\Gamma d^*}\frac{2b_\Gamma\gamma_\Gamma}{N(A+B)} \\
&= 1 - \frac{2}{d^*}(A+B)^{-1} \\
&= 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/(\delta/\log 2)}-1}\right)^{-1} \\
&\ge 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T_1}-1}\right)^{-1} \\
&\ge 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T}-1}\right)^{-1}.
\end{aligned}$$

□

PROOF OF CLAIM 2. For $T \geq T_2$ we have

$$\begin{aligned}
1 - \frac{Z_T}{b_\Gamma \gamma_\Gamma d^*} e^{-m/T} &\leq 1 - \frac{\xi}{b_\Gamma \gamma_\Gamma d^*} e^{-m/T} \\
&\leq 1 - \frac{\xi}{b_\Gamma \gamma_\Gamma d^*} e^{-m/T_2} \\
&\leq 1 - \frac{\xi}{b_\Gamma \gamma_\Gamma d^*} \frac{2 b_\Gamma \gamma_\Gamma}{\xi(A+B)} \\
&= 1 - \frac{2}{d^*}(A+B)^{-1} \\
&= 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/(\delta/\log 2)} - 1}\right)^{-1} \\
&\leq 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T_2} - 1}\right)^{-1} \\
&\leq 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T} - 1}\right)^{-1}. \qquad \square
\end{aligned}$$

PROOF OF CLAIM 3. This claim follows from the same estimates as for Claim 1 with the observation $B/(e^{\delta/T} - 1) \geq 0$. □

Now we shall prove that $\rho(P) = \lambda_2$ for small values of T.

THEOREM 4.6. *Let P be the transition matrix associated with the Metropolis algorithm* (13), *and let r^*, d^*, A, and B be the quantities defined above. Then $\rho(P) = \lambda_2$, for $T < T_*$, where*

$$T_* = \min\left[m\left(\log\frac{N^3(A+B)}{8}\right)^{-1}, \frac{\delta}{\log 2}\right].$$

PROOF. According to Theorem 4.3 we have

$$\lambda_2 \geq 1 - \frac{Z_T}{d^*}\left(\frac{N}{2}\right)^2 e^{-m/T},$$

and according to Theorem 4.4 we have

$$\lambda_{\min} \geq -1 + \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T} - 1}\right)^{-1}.$$

The proof of the claim is completed by showing that

$$1 - \frac{Z_T}{d^*}\left(\frac{N}{2}\right)^2 e^{-m/T} \geq 1 - \frac{2}{d^*}\left(A + \frac{B}{e^{\delta/T} - 1}\right)^{-1}$$

for $T \le T_*$. The proof is essentially the same as that of Claim 1 of Theorem 4.5. □

As we observed before, the lower bound (22) on λ_2 is quite rough. Improving this bound leads to a higher value of T_*.

5. Results relating to the Gibbs sampler. In this section we wish to obtain two main results. First we prove that for the Gibbs sampler $\rho(P) = \lambda_2$ for small values of T; second, we prove that the random updating dynamics on sites based on the Gibbs sampler and on the Metropolis algorithm have the same rate of convergence for small values of T.

Let S be a lattice of sites and let $\mathscr{G}$ be a *neighbourhood system of sites*, that is, a collection $\mathscr{G} = \{\mathscr{G}_s,\ s \in S\}$ with $\mathscr{G}_s \subset S$, $s \notin \mathscr{G}_s$ and $s \in \mathscr{G}_t$ if and only if $t \in \mathscr{G}_s$. Let $X = \{X_s,\ s \in S\}$ be a random Markov field with respect to $\mathscr{G}$ and assume that all X_s have the same state space $V = \{0, 1, \dots, c-1\}$. Finally, denote by $\Omega = V^S$ the configuration space, that is,

$$\Omega = \{(x_s)_{s \in S},\ x_s \in V\},$$

with $N_s(x)$ the set of the configurations $x' \in \Omega$ which differ from x at most in the coordinate s:

$$N_s(x) = \{x' \in \Omega:\ x'_j = x_j, \text{ for } j \ne s\} \tag{43}$$

and

$$N(x) = \bigcup_{s \in S} N_s(x). \tag{44}$$

Obviously $y \in N_s(x)$ if and only if $x \in N_s(y)$. For simplicity we shall suppose that all the sets $N_s(x)$, for each $s \in S$ and $x \in \Omega$, have the same cardinality c, that is, $|N_s(x)| = c$, and hence $d^* = |N(x)| = c|S|$.

Let H be a given real-valued function defined on Ω, and let π be the Gibbs measure on Ω with respect to the function H and a parameter $T > 0$. The function H is the energy and it usually can be written as a sum of potentials depending on the neighbourhood system of sites $\mathscr{G}$.

The Gibbs sampler algorithm [Geman and Geman (1984), Geman (1991)] is a Monte Carlo method for the generation of samples from Ω with probability distribution π. Given a site $s \in S$, first let us define the transition matrix

$$RG_s(x, y) \equiv \begin{cases} \dfrac{e^{-H(y)/T}}{\sum_{z \in N_s(x)} e^{-H(z)/T}}, & \text{if } y \in N_s(x), \\ 0, & \text{otherwise,} \end{cases}$$

and introduce the real-valued function 1_s,

$$1_s(x, y) = \begin{cases} 1, & \text{if } y \in N_s(x), \\ 0, & \text{otherwise.} \end{cases}$$

The Gibbs sampler relative to a random updating of sites is based on a Markov chain with state space Ω and transition matrix $RG = (RG(x,y))_{x,y\in\Omega}$ defined as follows:

$$
\begin{aligned}
RG(x,y) &\equiv \frac{1}{|S|}\sum_{s\in S} RG_s(x,y)\\
&= \frac{1}{|S|}\sum_{s\in S}\frac{e^{-H(y)/T}}{\sum_{z\in N_s(x)} e^{-H(z)/T}}1_s(x,y).
\end{aligned}
\tag{45}
$$

In this notation, the letter R is used as mnemonics for *random updating of sites*.

It can immediately be shown that each RG_s (and hence RG) is reversible with respect to the Gibbs distribution π given by (11). In fact if $y \in N_s(x)$, then, by (43), the two sets $N_s(x)$ and $N_s(y)$ are equal and hence we get

$$
\begin{aligned}
\pi(x)RG_s(x,y) &= \frac{e^{-H(x)/T}}{Z_T}\frac{e^{-H(y)/T}}{\sum_{z\in N_s(x)} e^{-H(z)/T}}\\
&= \frac{e^{-H(y)/T}}{Z_T}\frac{e^{-H(x)/T}}{\sum_{z\in N_s(y)} e^{-H(z)/T}} = \pi(y)RG_s(y,x).
\end{aligned}
$$

Then also the transition matrix RG defined by (45) is reversible with respect to the Gibbs distribution π. Moreover the transition probability matrix RG is irreducible and aperiodic [e.g., see Proposition 3 in Frigessi, Hwang, Sheu and Di Stefano (1993)].

Now let $G(RG) = [\Omega, E]$ be the graph underlying the transition matrix RG. In this section we obtain bounds on the eigenvalues of RG as we did previously. Now λ_2 and $\lambda_{\min}$ refer to RG. The results will be discussed in subsection 6.3. The methodology is the same but we have to consider a suitable set of admissible paths Γ; the quantities m, b_Γ and γ_Γ have been introduced before.

5.1. *A suitable set of admissible paths.* In order to obtain an upper bound on the second largest eigenvalue λ_2 of RG, we must select a suitable set of admissible paths Γ. Let $x, y \in \Omega$ and $e \in \gamma_{xy}$, with the notation introduced above that $e = (e^-, e^+)$ is the directed edge from the vertex e^- to the vertex e^+. Thus e^- and e^+ differ only in the value of a site, say, s. In Theorem 4.1, for each $e \in \gamma_{xy}$, we have

$$
H(e_*) - H(x) - H(y) \le m, \tag{46}
$$

where $H(e_*) = \max\{H(e^-), H(e^+)\}$.

Now we require a stronger condition than (46). For each $x, y \in \Omega$ we require a path such that, for each $e \in \gamma_{xy}$,

$$\{H(e^-) - H(\tilde{e})\} + \{H(e^+) - H(x) - H(y)\} \leq m \tag{47}$$

holds, where $\tilde{e}$ is any neighbour of e^- at site s, that is, $\tilde{e} \in N_s(e^-)$, s being the site in which e^- and e^+ differ. This choice is always possible by means of a suitable sequence of updating of the sites along the path. Now we prove the existence of a path satisfying (47) in the most critical case, that is, the case in which the path from x to y is such that $\text{elev}(\gamma_{xy}) - H(x) - H(y) = H_{xy} - H(x) - H(y) = m$ and either $H(e^-) = H_{xy}$ or $H(e^+) = H_{xy}$. The other cases may be treated in an analogous way.

In this case (47) becomes

$$\{H(e^-) - H(\tilde{e})\} + \{H(e^+) - H(x) - H(y)\} \leq H_{xy} - H(x) - H(y),$$

for each $e \in \gamma_{xy}$.

In fact, with the notation of Section 3, let $z \in \Omega$ such that $H(z) = H_{xy}$, and let v' and v'' be the vertices immediately preceding and following z along the admissible path γ_{xy}. The vertices v' and z differ only in the value of a site, say, s'; similarly the vertices v'' and z differ only in a certain site, say, s''. Then (47) is guaranteed if we consider a path such that

$$H(v') - H(t') \leq 0 \quad \text{for each } t' \in N_{s'}(v'), \tag{48}$$

$$H(v'') - H(t'') \leq 0 \quad \text{for each } t'' \in N_{s''}(v''). \tag{49}$$

In other words, the path must cross the vertex $v' \in N_{s'}(z)$ such that $H(v') - H(t') \leq 0$, for each $t' \in N_{s'}(z)$, immediately before the vertex z such that $H(z) = H_{xy}$. The vertex v'' must be crossed immediately after the vertex z. The elevation of the paths does not change. This choice can be made clearer by Figure 1. The two paths (x, w', z, w'', y) and (x, w', v', z, v'', y) have the

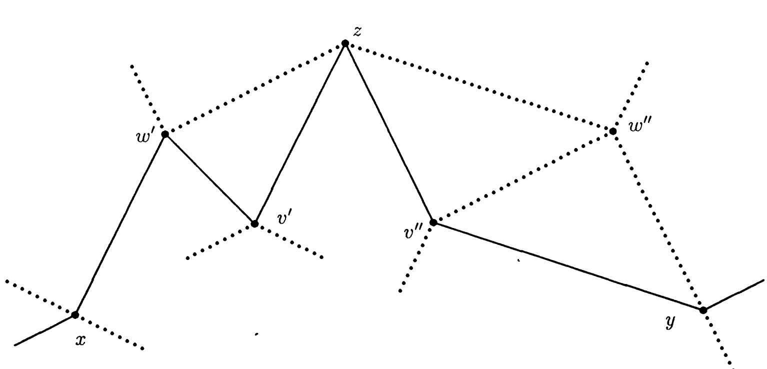

FIG. 1. *Example of paths from x to y (in this case each vertex has the same degree $d^* = 4$): The path from x to y to be considered is (x, w', v', z, v'', y). All other paths (in which at least one edge is drawn in dotted line) are to be rejected.*

same elevation, but the latter is the one to be considered. In fact, for example, if $e^- = w'$ and $e^+ = z$, when $\tilde{e} = v'$, then

$$H(w') - H(v') > 0$$

and (47) is *not* satisfied; similarly, if $e^- = z$ and $e^+ = w''$, when $\tilde{e} = v''$, then

$$H(w'') - H(v'') > 0$$

and again (47) is *not* satisfied. The set of paths Γ introduced above shall be used in the proof of Theorem 5.1.

5.2. *Bounds on eigenvalues.* In this section we get bounds on the eigenvalues λ_2 and $\lambda_{\min}$ of RG.

First, to obtain an upper bound on λ_2, we shall consider a set of admissible paths Γ such that the condition (47) holds.

THEOREM 5.1. *Let RG be the transition probability matrix of the local updating dynamic based on the Gibbs sampler* (45). *Then its second largest eigenvalue satisfies the relation*

$$\lambda_2 \le 1 - \frac{Z_T}{b_\Gamma \gamma_\Gamma c|S|} e^{-m/T}. \tag{50}$$

PROOF. Let Γ be the set of paths which we have introduced above and $e \in E$, and set $RG(e) = RG(e^-, e^+)$. We have

$$\begin{aligned} |\gamma_{xy}|_Q \pi(x)\pi(y) &\equiv \sum_{e \in \gamma_{xy}} Q^{-1}(e)\pi(x)\pi(y) \\ &= \sum_{e \in \gamma_{xy}} [\pi(e^-)RG(e)]^{-1} \pi(x)\pi(y) \\ &= \sum_{e \in \gamma_{xy}} \left(\frac{1}{|S|} \sum_{s \in S} \pi(e^-) RG_s(e) 1_s(e) \right)^{-1} \pi(x)\pi(y) \\ &= \sum_{e \in \gamma_{xy}} \left(\frac{1}{|S|} \sum_{s \in S} \frac{e^{-H(e^+)/T}}{\sum_{\tilde{e} \in N_s(e^-)} e^{-H(\tilde{e})/T}} \frac{e^{-H(e^-)/T}}{Z_T} 1_s(e) \right)^{-1} \\ &\qquad \times \pi(x)\pi(y). \end{aligned}$$

For each $e \in \gamma_{xy}$ the sum in $(\sum_{s \in S} \cdots)^{-1}$ contains only one term different from zero, and so

$$\begin{aligned} |\gamma_{xy}|_Q \pi(x)\pi(y) = \frac{|S|}{Z_T} \sum_{e \in \gamma_{xy}} \sum_{s \in S} \sum_{\tilde{e} \in N_s(e^-)} &\exp\left(\frac{\{H(e^-) - H(\tilde{e})\}}{T} \right) \\ &\times \exp\left(\frac{\{H(e^+) - H(x) - H(y)\}}{T} \right) 1_s(e). \end{aligned}$$

Furthermore, our choice of paths assures that

$$\exp\left(\frac{\{H(e^-) - H(\tilde{e})\}}{T}\right)\exp\left(\frac{\{H(e^+) - H(x) - H(y)\}}{T}\right)$$
$$\leq \exp\left(\frac{m}{T}\right)$$

holds for each $e \in \gamma_{xy}$. Then

$$\sum_{s \in S} \sum_{\tilde{e} \in N_s(e^-)} \exp\left(\frac{\{H(e^-) - H(\tilde{e})\}}{T}\right)\exp\left(\frac{\{H(e^+) - H(x) - H(y)\}}{T}\right)1_s(e)$$
$$\leq c\exp\left(\frac{m}{T}\right),$$

and thus

$$|\gamma_{xy}|_Q \pi(x)\pi(y) \leq \frac{|S|}{Z_T} \sum_{e \in \gamma_{xy}} ce^{m/T}.$$

Hence, from the definition of γ_Γ it follows that

$$|\gamma_{xy}|_Q \pi(x)\pi(y) \leq \frac{|S|}{Z_T} c\gamma_\Gamma e^{m/T}.$$

Thus we obtain

$$\kappa \equiv \max_{e' \in \vec{E}} \sum_{\gamma_{xy} \ni e'} |\gamma_{xy}|_Q \pi(x)\pi(y)$$
$$\leq \frac{b_\Gamma \gamma_\Gamma c|S|}{Z_T} e^{m/T}.$$

Proposition 2.1 completes the proof:

$$\lambda_2 \leq 1 - \frac{Z_T}{b_\Gamma \gamma_\Gamma c|S|} e^{-m/T}. \qquad \square$$

A lower bound for λ_2 is obtained in the next theorem. In this theorem some minorizations are quite crude, but we wish to fix the rate of convergence of λ_2.

THEOREM 5.2. *Let RG be the transition probability matrix of a local updating dynamic based on the Gibbs sampler* (45). *Then its second largest eigenvalue satisfies the relation*

$$\lambda_2 \geq 1 - \frac{Z_T}{|S|}\frac{N^2}{4} e^{-m/T}. \tag{51}$$

PROOF. Let W be the proper subset of Ω defined in Theorem 4.3:

$$W = \{z \in \Omega | H_{x_0 z} < H_{x_0 y_0}\},$$

where x_0 and y_0 are such that $H(x_0) \le H(y_0)$ and $H_{x_0 y_0} - H(x_0) - H(y_0) = m$. Then

$$\frac{Q(W\times W^c)}{\pi(W)\pi(W^c)} = \left(\sum_{(x,y)\in\vec{E}\cap(W\times W^c)} \pi(x)RG(x,y)\right)\left(\sum_{x\in W}\pi(x)\sum_{y\in W^c}\pi(y)\right)^{-1}$$
$$= \left(\sum_{(x,y)\in\vec{E}\cap(W\times W^c)} \frac{1}{|S|}\sum_{s\in S}\pi(x)RG_s(x,y)1_s(x,y)\right)$$
$$\times\left(\sum_{x\in W}\pi(x)\sum_{y\in W^c}\pi(y)\right)^{-1}.$$

Now in the sum over $(x, y) \in \vec{E} \cap (W \times W^c)$ we may suppose that x and y differ at only one site, say, s [otherwise $RG_{s'}(x, y) = 0$, for all s']; hence we have

$$\frac{Q(W\times W^c)}{\pi(W)\pi(W^c)} = \frac{1}{|S|}\left(\sum_{(x,y)\in\vec{E}\cap(W\times W^c)} \frac{e^{-H(y)/T}}{\sum_{z\in N_s(x)} e^{-H(z)/T}}\frac{e^{-H(x)/T}}{Z_T}\right)$$
$$\times\left(\sum_{x\in W}\frac{e^{-H(x)/T}}{Z_T}\sum_{y\in W^c}\frac{e^{-H(y)/T}}{Z_T}\right)^{-1},$$

and, as $x_0 \in W$ and $y_0 \in W^c$, it follows that

$$\frac{Q(W\times W^c)}{\pi(W)\pi(W^c)} \le \frac{Z_T}{|S|}\sum_{(x,y)\in\vec{E}\cap(W\times W^c)} \frac{\exp[-H(x)/T]\exp[-H(y)/T]}{\sum_{z\in N_s(x)}\exp[-H(z)/T]}$$
$$\times\left(\exp[-H(x_0)/T]\exp[-H(y_0)/T]\right)^{-1}$$
$$= \frac{Z_T}{|S|}\sum_{(x,y)\in\vec{E}\cap(W\times W^c)} \frac{\exp[-H(y)/T]}{\sum_{z\in N_s(x)}\exp[-\{H(z)-H(x)\}/T]}$$
$$\times\exp\left(\frac{\{H(x_0)+H(y_0)\}}{T}\right)$$
$$= \frac{Z_T}{|S|}\sum_{(x,y)\in\vec{E}\cap(W\times W^c)} \frac{\exp[-H(y)/T]}{1+\sum_{z\in N_s(x)\setminus\{x\}}\exp[-\{H(z)-H(x)\}/T]}$$
$$\times\exp\left(\frac{\{H(x_0)+H(y_0)\}}{T}\right).$$

In the proof of Theorem 4.3 we showed that if $(x, y) \in \vec{E} \cap (W \times W^c)$, then $H_{x_0 y_0} \le H(y)$. Hence

$$\frac{Q(W \times W^c)}{\pi(W)\pi(W^c)} \le \frac{Z_T}{|S|} \sum_{(x,y) \in \vec{E} \cap (W \times W^c)} \exp\left(\frac{-\{H_{x_0 y_0} - H(x_0) - H(y_0)\}}{T} \right)$$
$$= \frac{Z_T}{|S|} \sum_{(x,y) \in \vec{E} \cap (W \times W^c)} \exp\left(-\frac{m}{T} \right).$$

Since the set $\vec{E} \cap (W \times W^c)$ has less than or equal to $(N/2)^2$ edges, we have

$$\frac{Q(W \times W^c)}{\pi(W)\pi(W^c)} \le \frac{Z_T}{|S|} \frac{N^2}{4} e^{-m/T},$$

which, using Proposition 2.2, completes the proof. □

Now let us turn to the smallest eigenvalue. We begin with a proposition which is a companion to Corollary 4.1.

PROPOSITION 5.1. *Let RG be the transition probability matrix of a local updating dynamic based on the Gibbs sampler* (45). *Then* $r^* = 0$.

PROOF. The definition of RG shows that $RG(x, x) > 0$, for each $x \in \Omega$. Hence $x \in \Lambda_{RG}$, for each $x \in \Omega$, and then $r^* = 0$. □

In order to prove a bound on the smallest eigenvalue of RG, we define another quantity, the maximum jump Δ of the function H in any transition:

$$\Delta = \max_{x \in \Omega} \max_{y \in N(x)} |H(y) - H(x)|. \tag{52}$$

THEOREM 5.3. *Let RG be the transition probability matrix of a local updating dynamic based on the Gibbs sampler* (45). *Then its smallest eigenvalue satisfies the relation*

$$\lambda_{\min} \ge -1 + \frac{2}{1 + (c-1)e^{\Delta/T}},$$

where Δ *is as defined in* (52).

PROOF. Proposition 5.1 states that self-loops are allowed in each site s, that is, $x \in \Lambda_{RG}$ for each $x \in \Omega$. In this case the set of shortest cycles Σ is constituted by self-loops. Then we have

$$|\sigma_x|_Q \pi(x) = \left(\frac{1}{|S|} \sum_{s \in S} \frac{\exp[-H(x)/T]}{\sum_{z \in N_s(x)} \exp[-H(z)/T]} \right)^{-1}$$
$$= \left(\frac{1}{|S|} \sum_{s \in S} \left(1 + \sum_{z \in N_s(x) \setminus \{x\}} \exp\left[\frac{\{H(x) - H(z)\}}{T} \right] \right)^{-1} \right)^{-1}.$$

As the harmonic mean is always less than or equal to the arithmetic mean, we have

$$\begin{aligned}|\sigma_x|_Q\pi(x) &\le \frac{1}{|S|}\sum_{s\in S}\left(1+\sum_{z\in N_s(x)\setminus\{x\}}\exp\left[\frac{\{H(x)-H(z)\}}{T}\right]\right)\\ &= 1+\frac{1}{|S|}\sum_{s\in S}\sum_{z\in N_s(x)\setminus\{x\}}\exp\left(\frac{\{H(x)-H(z)\}}{T}\right)\\ &\le 1+(c-1)\exp\left(\frac{\Delta}{T}\right).\end{aligned}$$

Finally, as each cycle contains only one edge (i.e., the self-loop), we get

$$\begin{aligned}\iota &= \max_{e\in\vec{E}}\sum_{\sigma_x\ni e}|\sigma_x|_Q\pi(x)\\ &\le 1+(c-1)e^{\Delta/T},\end{aligned}$$

and Proposition 2.1 completes the proof. □

5.3. *Bounds on the spectral gap.* We show how results of the previous subsection give a useful bound on $\rho(RG)$ for small values of T when $m > \Delta$.

THEOREM 5.4. *Let RG be the transition matrix of the Markov chain associated with the Gibbs sampler algorithm* (45), *and let ξ be the number of absolute minima of the function H. Suppose $m > \Delta$. Then if $T \le T_1$, we have*

$$\rho(RG) \le 1 - \frac{\xi}{c|S|b_\Gamma\gamma_\Gamma}e^{-m/T}, \tag{53}$$

where

$$T_1 = \begin{cases}(m-\Delta)\left(\log\dfrac{N}{2|S|b_\Gamma\gamma_\Gamma}\right)^{-1}, & \textit{if } N > 2|S|b_\Gamma\gamma_\Gamma,\\ \infty, & \textit{otherwise}.\end{cases} \tag{54}$$

PROOF. According to Theorem 5.1 we have

$$\lambda_2 \le 1 - \frac{Z_T}{c|S|b_\Gamma\gamma_\Gamma}e^{-m/T},$$

and according to Theorm 5.3 we have

$$\lambda_{\min} \ge -1 + \frac{2}{1+(c-1)e^{\Delta/T}}.$$

The proof of the claim is completed by showing that

$$1 - \frac{2}{1+(c-1)e^{\Delta/T}} \le 1 - \frac{Z_T}{c|S|b_\Gamma\gamma_\Gamma}e^{-m/T},$$

for $T \le T_1$. For each $T > 0$, as $\xi < Z_T < N$ and $e^{\Delta/T} > 1$, we have

$$1 - \frac{N}{c|S|b_\Gamma\gamma_\Gamma}e^{-m/T} < 1 - \frac{Z_T}{c|S|b_\Gamma\gamma_\Gamma}e^{-m/T},$$

$$1 - \frac{2}{1 + (c-1)e^{\Delta/T}} < 1 - \frac{2}{ce^{\Delta/T}}.$$

Straightforward calculations give, for $T \le T_1$,

$$1 - \frac{2}{c}e^{-\Delta/T} \le 1 - \frac{N}{c|S|b_\Gamma\gamma_\Gamma}e^{-m/T}.$$

This completes the proof. □

Now we prove that $\rho(RG) = \lambda_2$ for small values of T.

THEOREM 5.5. *Let RG be the transition matrix of the Markov chain associated with a Gibbs sampler algorithm* (45). *Suppose $m > \Delta$. Then if $T \le T_*$, we have*

$$\rho(RG) = \lambda_2, \tag{55}$$

where

$$T_* = (m - \Delta)\left(\log\frac{cN^3}{8|S|}\right)^{-1}.$$

PROOF. The proof goes along the same lines as Theorem 5.4, substituting the use of Theorem 5.2 for that of Theorem 5.1. Note that, as $N = c^{|S|}$, $cN^3 > 8|S|$ for $c \ge 2$. □

NOTE 5.1. Using other techniques,

$$\rho(RG) = \lambda_2,$$

for each $T > 0$, has been proved by Frigessi, Hwang, Sheu and Di Stefano (1993).

6. Discussion. In this section we discuss usefulness and drawbacks of our bounds and compare them with the geometric bounds on λ_2 and $\lambda_{\min}$ for reversible Markov chains recently obtained by Desai (1992) following an algebraic approach. First we summarize these results and then we give a comparison with the bounds previously obtained both in some cases in which we know the exact value of λ_2 and $\lambda_{\min}$ [Hanlon (1992)] and in another case of interest. Finally, we compare the rate of convergence of random updating dynamics based on the Metropolis algorithm and the Gibbs sampler. This problem has been recently approached by some authors. Frigessi, Hwang, Sheu and Di Stefano (1993) have proved that for Ising models the Metropolis algorithm is the best at low temperatures and the worst at high temperatures. Chiang and Chow (1991) have proved that the Metropolis algorithm and the Gibbs sampler are asymptotically equivalent in annealing for lattices.

6.1. *The geometric bounds of Desai* (1992). In his Ph.D. dissertation, Desai (1992) obtained geometric bounds on λ_2 and $\lambda_{\min}$ for reversible Markov chains, following an algebraic approach. These bounds, in conjunction with some results of Fiedler (1973), give easy-to-compute upper bounds on λ_2 for any reversible Markov chain with transition matrix P and a lower bound on $\lambda_{\min}$ for a reversible Markov chain with transition matrix P having a k-regular underlying graph (i.e., a graph in which each vertex has the same degree k).

Let $G = [\Omega, E]$ be a graph; denote by $A = (A(x, y)_{x, y \in \Omega})$ the *adjacency matrix* of the graph G and by $D = (D(x, x)_{x \in \Omega})$ the diagonal matrix indexed by Ω, where $D(x, x) = d(x)$. The matrix $L = D - A$ is named the *Laplacian matrix* of G [Mohar (1991), Fiedler (1989)]; another name, the *admittance matrix*, comes from the theory of electrical networks. The Laplacian matrix of a graph and its eigenvalues can be used in several areas of mathematical research and have an interpretation in various physical and chemical theories. For example, Maas (1987) considers the spectrum of the Laplacian matrix to describe the behaviour of a liquid flowing through a system of connected pipes. The second smallest eigenvalue of L is named the *algebraic connectivity* of the graph G, and it is related to several problems in graph theory and its applications.

In this section, P is an aperiodic, irreducible and reversible transition probability matrix with respect to a probability measure π, and we denote by $G(P)$ the underlying graph of P, by L the Laplacian matrix of G and by $\mu_2(L)$ the second smallest eigenvalue of L. Finally, let us set

$$\pi_{\max} = \max_{x} \pi(x), \tag{56}$$

$$Q_{\min} = \min_{x \neq y} \{\pi(x) P(x, y) : P(x, y) > 0\}. \tag{57}$$

Let λ_2 be the second largest eigenvalue of P. In Theorem 4.8 of Desai (1992) the following upper bound on λ_2 is obtained:

$$\lambda_2 \leq 1 - \frac{Q_{\min}}{\pi_{\max}} \mu_2(L). \tag{58}$$

Now we turn to the smallest eigenvalue $\lambda_{\min}$. Let us define the matrix $|L| = (|L(x, y)|)_{x, y \in \Omega}$, and denote by $\mu_1(|L|)$ the smallest eigenvalue of the matrix $|L|$, with $\theta = \min_{x \in \Omega} P(x, x)$ the smallest diagonal entry in P. Theorem 4.13 of Desai (1992) gives the following lower bound on $\lambda_{\min}$:

$$\lambda_{\min} \geq -1 + \frac{Q_{\min}}{\pi_{\max}} \mu_1(|L|) + 2\theta. \tag{59}$$

If the underlying graph of P is k-regular, that is, the degree of each vertex is k, then

$$\mu_1(|L|) = 2k - \mu_N(L), \tag{60}$$

where $\mu_N(L)$ is the largest eigenvalue of the Laplacian matrix L of G, and then

$$\lambda_{\min} \geq -1 + \frac{Q_{\min}}{\pi_{\max}}(2k - \mu_N(L)) + 2\theta. \tag{61}$$

In (58) and (61) the quantities $\mu_2(L)$ and $\mu_N(L)$ in general cannot be explicitly evaluated. However, bounds for $\mu_2(L)$ and $\mu_N(L)$ were obtained in a paper by Fiedler (1973) dependent on some classical parameters of graph theory. Let e_G be the *edge connectivity* of the graph G, that is, the minimal number of edges whose removal would result in losing connectivity of the graph G, and let d^* be the maximum degree of the graph G. Fiedler [(1973), Theorem 4.3] gives two lower bounds on $\mu_2(L)$ depending on these quantities:

$$\mu_2(L) \geq 2e_G[1 - \cos(\pi/N)], \tag{62}$$

$$\mu_2(L) \geq c_1 e_G - c_2 d^*, \tag{63}$$

where $c_1 = 2[\cos(\pi/N) - \cos(2\pi/N)]$ and $c_2 = 2\cos(\pi/N)[1 - \cos(\pi/N)]$ and the second bound is better if and only if $2e_G > d^*$. For k-regular graphs one can also apply Diaconis–Stroock bounds (or Cheeger) to obtain bounds on $\mu_2(L)$ and on $\mu_N(L)$.

An upper bound on $\mu_N(L)$ is given in Fiedler [(1973), Theorem 3.7],

$$\mu_N(L) \leq 2d^*. \tag{64}$$

Hence, from (58), (62) and (63), we obtain the following bounds on λ_2:

$$\lambda_2 \leq 1 - \frac{Q_{\min}}{\pi_{\max}} 2e_G\left[1 - \cos\left(\frac{\pi}{N}\right)\right], \tag{65}$$

$$\lambda_2 \leq 1 - \frac{Q_{\min}}{\pi_{\max}}(c_1 e_G - c_2 d^*) \tag{66}$$

(the second is better if and only if $2e_G > d^*$); if the underlying graph of P is k-regular, then, from (59), (60) and (64), we have the following bound on $\lambda_{\min}$:

$$\lambda_{\min} \geq -1 + 2\theta. \tag{67}$$

If $\theta = 0$, then the bound (67) does not give any information on $\lambda_{\min}$. We shall refer to (65), (66) and (67) as Desai–Fiedler bounds.

Finally, the algebraic bounds on λ_2 obtained in this subsection can be summarized as

$$\lambda_2 \leq 1 - C\frac{Q_{\min}}{\pi_{\max}}, \tag{68}$$

where

$$(69)\qquad C = \begin{cases} \mu_2(L), & \text{if this value is known,} \\ 2e_G\left[1 - \cos\left(\dfrac{\pi}{N}\right)\right], & \text{if the value of } \mu_2(L) \text{ is unknown} \\ & \quad \text{and } 2e_G \le d^*, \\ c_1 e_G - c_2 d^*, & \text{if the value of } \mu_2(L) \text{ is unknown} \\ & \quad \text{and } 2e_G > d^*. \end{cases}$$

6.2. *Metropolis chains: Examples and comparisons.* Let M be the oscillation of the function H, defined as

$$(70)\qquad M \equiv \max_{x,\, y \in \Omega} (H(x) - H(y));$$

in particular, $M = H_{\max} = \max_{x \in \Omega} H(x)$, as we supposed $H_{\min} = \min_{x \in \Omega} H(x) = 0$.

The algebraic bound on λ_2 given in the previous section can be specialized in the case of Metropolis chains as follows.

THEOREM 6.1. *Let P be the transition matrix of the Markov chain* (13). *Then its second largest eigenvalue λ_2 satisfies the relation*

$$(71)\qquad \lambda_2 \le 1 - C\frac{e^{-M/T}}{d^*},$$

where C and M have been introduced, respectively, in (69) *and* (70).

PROOF. The proof follows from the algebraic bound on λ_2 given in (68), evaluating suitably the quantities $Q_{\min}$ and $\pi_{\max}$. We have

$$\pi_{\max} = \max_{x \in \Omega} \pi(x) = \max_{x \in \Omega} \frac{e^{-H(x)/T}}{Z_T} = \frac{e^{-H_{\min}/T}}{Z_T} = \frac{1}{Z_T},$$

as $H_{\min} = 0$.

In the evaluation of the quantity $Q_{\min} = \min_{x \neq y}\{\pi(x)P(x, y) \colon P(x, y) > 0\}$, we have to consider two cases:

(1) $H(x) \ge H(y)$, that is, $\pi(x) \le \pi(y)$. Then

$$\pi(x)P(x, y) = \pi(x)R(x, y) = \frac{1}{d^*}\frac{e^{-H(x)/T}}{Z_T} = \frac{1}{d^*}\pi(x);$$

(2) $H(x) < H(y)$, that is, $\pi(x) > \pi(y)$. Then

$$\pi(x)P(x, y) = \pi(y)R(x, y) = \frac{1}{d^*}\frac{e^{-H(y)/T}}{Z_T} = \frac{1}{d^*}\pi(y).$$

Thus,

$$Q_{\min} = \min_{x \in \Omega}\left(\frac{1}{d^*}\pi(x)\right) = \frac{e^{-H_{\max}/T}}{d^* Z_T} = \frac{1}{d^*}\frac{e^{-M/T}}{Z_T}. \qquad \square$$

Now we can immediately compare this last bound with that obtained in Theorem 4.1.

PROPOSITION 6.1. *Let P be the transition matrix of the Markov chain* (13), *and let ξ be the number of absolute minima of the function H. Suppose $m > 0$.*

(i) *If $M > m$, then we have*

$$\lambda_2 \le 1 - \frac{Z_T}{b_\Gamma \gamma_\Gamma d^*} e^{-m/T} \le 1 - \frac{C}{d^*} e^{-M/T}, \tag{72}$$

for $T \le T_*$, where the constant C has been introduced in (69) *and T_* is given by*

$$T_* = \begin{cases} (M - m)\left(\ln \dfrac{C b_\Gamma \gamma_\Gamma}{\xi}\right)^{-1}, & \text{if } \xi < C b_\Gamma \gamma_\Gamma, \\ \infty, & \text{otherwise.} \end{cases}$$

(ii) *If $M = m$, then* (72) *holds for each $T > 0$, provided that $\xi \ge C b_\Gamma \gamma_\Gamma$* (*and bounds are in the other order if $N < C b_\Gamma \gamma_\Gamma$*).

PROOF. The proof follows with straightforward calculations from Theorems 4.1 and 6.1. □

In the rest of this subsection we consider some examples of applications of the obtained bounds. The first two cases come from Hanlon (1992), who considers a particular random walk on the symmetric group S_f in which the probability of moving from a permutation p to any permutation $p(i, j)$ depends on the change in the number of disjoint cycles between p and $p(i, j)$. Let $c(p)$ denote the number of cycles in the disjoint cycle decomposition of p. Then $c(p(i, j)) = c(p) \pm 1$. The probability of moving from p to $p(i, j)$ depends only on whether $c(p(i, j)) = c(p) + 1$ or $c(p(i, j)) = c(p) - 1$, and it is α times as likely to move to $p(i, j)$ if $c(p(i, j)) = c(p) - 1$ rather than $c(p(i, j)) = c(p) + 1$. For $\alpha > 1$, these chains are particular cases of Metropolis chains and in some cases they can be explicitly diagonalized: For these cases we wish to compare the bounds on λ_2 and $\lambda_{\min}$ with the exact values. In both the examples any set of shortest paths is also admissible. Further details are given in the Appendix.

In the case $\alpha = 1$ we have simple random walks.

EXAMPLE 6.1 (Random walk on S_3). Let $\Omega = \{x_1, x_2, \dots, x_6\}$. Let H be a real-valued function such that $H(x_1) = 2$, $H(x_2) = H(x_3) = H(x_4) = 1$, $H(x_5 = H(x_6) = 0$. For a fixed value of $\alpha > 1$, let $T > 0$ be such that

$$\frac{1}{T} = \log \alpha.$$

Let R be the transition probability matrix defined as

$$R = \frac{1}{3}\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

Then, from (13), we have

$$(73) \qquad P = \frac{1}{3\alpha}\begin{bmatrix} 0 & \alpha & \alpha & \alpha & 0 & 0 \\ 1 & \alpha - 1 & 0 & 0 & \alpha & \alpha \\ 1 & 0 & \alpha - 1 & 0 & \alpha & \alpha \\ 1 & 0 & 0 & \alpha - 1 & \alpha & \alpha \\ 0 & 1 & 1 & 1 & 3(\alpha - 1) & 0 \\ 0 & 1 & 1 & 1 & 0 & 3(\alpha - 1) \end{bmatrix}.$$

In this case, Hanlon (1992) has calculated that

$$(74) \qquad \lambda_2 = 1 - \frac{1}{\alpha},$$

$$(75) \qquad \lambda_{\min} = -\frac{1}{\alpha}.$$

Then $\rho(P)$ depends on α: In fact, $\lambda_2 > |\lambda_{\min}|$ for $\alpha > 2$; hence $\rho(P) = \lambda_2$ for $\alpha \geq 2$ and $\rho(P) = |\lambda_{\min}|$ for $1 < \alpha < 2$.

In Figure 2 we give the underlying graph of the transition matrix P. In this case we have immediately that $\gamma_\Gamma = 2$, $d^* = 3$, $\xi = 2$, $r^* = 1$ and $m = \delta = 1$. We see immediately that $x_2, \dots, x_6 \in \Lambda_P$ while $x_1 \notin \Lambda_P$, and hence

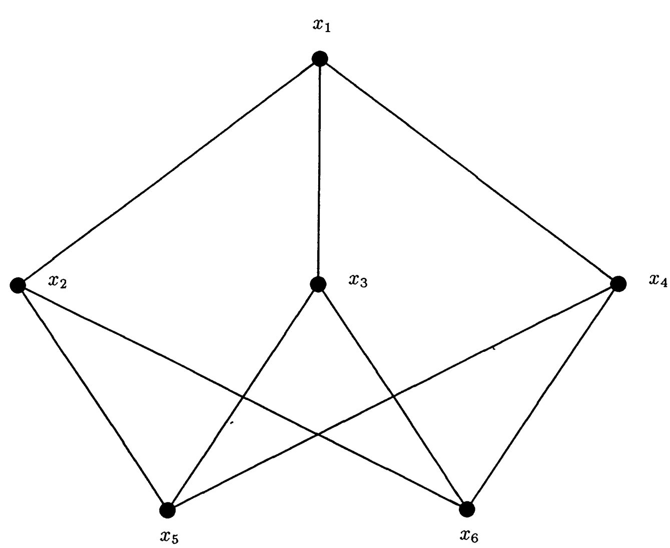

FIG. 2. *The underlying graph of the transition matrix P of Example* 1.

σ_{x_1} must include the self-loop of x_2 or x_3 or x_4; thus $b_\Sigma = 1$. The parameter b_Γ can be evaluated as in Diaconis and Stroock [(1991), Example 2.2]. The graph $G(P)$ has $Nd^* = 18$ oriented edges. Any choice of shortest paths is also a set of admissible paths, and it has $Nd^* = 18$ ordered pairs of vertices at distance 1 and $N(d^* - 1) = 12$ ordered pairs of vertices at distance 2; that is, $18 + 2 \cdot 12 = 42$ directed edges to be crossed. Thus some edge must be crossed by three paths, and hence $b_\Gamma \geq 3$. In the Appendix we give a set of paths Γ achieving $b_\Gamma = 3$.

Therefore the obtained bounds give

$$\lambda_2 \leq 1 - \frac{1}{9\alpha},$$

$$\lambda_{\min} \geq -1 + \frac{1}{3}\frac{\alpha - 1}{2\alpha - 1}.$$

The bound on λ_2 is not too imprecise and has the right order of magnitude as $\alpha \to \infty$; the bound on $\lambda_{\min}$ has the right order of magnitude as α goes to 1. The bound on $\rho(P)$ gives

$$\rho(P) = \max(\lambda_2, |\lambda_{\min}|) \leq \begin{cases} 1 - \dfrac{1}{9\alpha}, & \text{if } \alpha \geq \dfrac{5 + \sqrt{13}}{6}, \\ 1 - \dfrac{1}{3}\dfrac{\alpha - 1}{2\alpha - 1}, & \text{if } 1 < \alpha < \dfrac{5 + \sqrt{13}}{6}. \end{cases}$$

Note that $(5 + \sqrt{13})/6 \simeq 1.434$.

Now let us compute the algebraic bounds. From Hanlon (1992) we can obtain the values of $\mu_2(L)$ and $\mu_N(L)$ for the graph $G(P)$, which are $\mu_2(L) = 3$ and $\mu_N(L) = 6$ (see also the Appendix). As $d^* = 3$ and $M = H(x_1) = 2$, then Theorem 6.1 gives

$$\lambda_2 \leq 1 - \frac{1}{\alpha^2},$$

and its order of magnitude is off by a factor of $1/\alpha$ as $\alpha \to \infty$. As $\theta = 0$ in (67), the bound on $\lambda_{\min}$ is equal to -1 and so does not provide information on $\lambda_{\min}$. We can also compute the Desai–Fiedler bounds. The graph $G(P)$ has $e_G = d^* = 3$, $N = 6$ and $2e_G > d^*$; thus we consider the bound (66). As $H(x_1) = 2$, with straightforward calculations we get

$$\pi_{\max} = \frac{1}{Z_T},$$

$$Q_{\min} = \frac{1}{3\alpha^2}\frac{1}{Z_T},$$

$$c_1 = \sqrt{3} - 1,$$

$$c_2 = \sqrt{3}\left(1 - \frac{\sqrt{3}}{2}\right);$$

then

$$\lambda_2 \le 1 - \frac{1}{2\alpha^2}.$$

EXAMPLE 6.2 (Random walk on S_4). Let $\Omega = \{x_1, x_2, \ldots, x_{24}\}$. Let H be a real-valued function such that the following hold: $H(x_1) = 3$; $H(x_i) = 2$, for $i = 2, \ldots, 7$; $H(x_i) = 1$, for $i = 8, \ldots, 18$; $H(x_i) = 0$, for $i = 19, \ldots, 24$. As above, $1/T = \log \alpha$. For simplicity we do not display the matrix R; in the Appendix we give the graph underlying P. In this case the second largest eigenvalue and the smallest one depend on the value of α. For large values of α, the eigenvalue

$$\lambda' = \frac{9(\alpha - 1) + \sqrt{9\alpha^2 - 2\alpha + 9}}{12\alpha}$$

is arbitrarily close to 1; for values of α arbitrarily close to 1, the eigenvalue

$$\lambda'' = -\frac{1}{\alpha}$$

is close to -1. In the Appendix we show that $\lambda_2 = \lambda'$ for α sufficiently large, and $\lambda_{\min} = \lambda''$ for α sufficiently close to 1. For the graph underlying the transition matrix P we have $\gamma_\Gamma = 3$, $d^* = 6$, $\xi = 6$, $r^* = 1$ and $m = \delta = 1$. With arguments similar to those of Example 6.1, we obtain $b_\Sigma = 1$ and $b_\Gamma = 8$. Therefore the obtained bounds give

$$\lambda_2 \le 1 - \frac{1}{24\alpha},$$

$$\lambda_{\min} \ge -1 + \frac{1}{6}\frac{\alpha - 1}{2\alpha - 1}.$$

Hence

$$\rho(P) = \max(\lambda_2, |\lambda_{\min}|) \le \begin{cases} 1 - \dfrac{1}{24\alpha}, & \text{for } \alpha \ge \dfrac{3 + \sqrt{5}}{4}, \\ 1 - \dfrac{1}{6}\dfrac{\alpha - 1}{2\alpha - 1}, & \text{for } 1 < \alpha < \dfrac{3 + \sqrt{5}}{4}. \end{cases}$$

Note that $(3 + \sqrt{5})/4 \simeq 1.309$.

Also in this case the values of $\mu_2(L)$ and $\mu_N(L)$ are known and they are $\mu_2(L) = 4$ and $\mu_N(L) = 12$ (see the Appendix). Then Theorem 6.1 gives

$$\lambda_2 \le 1 - \frac{2}{3\alpha^3},$$

and its order of magnitude is off by a factor $1/\alpha^2$ as $\alpha \to \infty$. As $\theta = 0$, the bound on $\lambda_{\min}$ is equal to -1 and so does not provide information on $\lambda_{\min}$.

Now let us compute the Desai–Fiedler bound on λ_2. The graph $G(P)$ has $e_G = d^* = 6$, $N = 24$ and $2e_G > d^*$, and thus we consider the bound (66). As $H_{\max} = M = 3$ and

$$c_1 e_G - c_2 d^* = 12\sin^2(\pi/24),$$

we get

$$\lambda_2 \le 1 - \frac{2}{\alpha^3}\sin^2\left(\frac{\pi}{24}\right). \tag{76}$$

These bounds are worse than the bound derived from Theorem 4.1.

The main problem with our bounds concerns the estimate of the parameters in (38). In general, given the graph $G(P)$, only the quantity d^* can be readily evaluated, while the partition function Z_T can be bounded below by the number $\xi \ge 1$ of the points of absolute minima of the function H. On the contrary, the parameters b_Γ, γ_Γ and m are in general difficult to calculate except in some simple cases.

Nevertheless, calculable bounds on $\rho(P)$ can be obtained by choosing a set of paths Γ' with the smallest number of edges between each pair of states. In this case Theorem 4.1 can be reformulated as follows.

THEOREM 6.2. *Let P be the transition probability matrix of the Markov chain associated with the Metropolis algorithm defined in* (13). *Then its second largest eigenvalue λ_2 satisfies the relation*

$$\lambda_2 \le 1 - \frac{Z_T}{b'_\Gamma \gamma'_\Gamma d^*} e^{-M/T}, \tag{77}$$

where $M = \max_{x, y \in \Omega}(H(x) - H(y))$ is the oscillation of the function H, and b'_Γ and γ'_Γ are defined analogously as (14) *and* (15) *for the set Γ'.*

PROOF. The proof is the same as for Theorem 4.1. □

Furthermore, the results of Theorem 4.5 are changed accordingly. For example, Claim 1 becomes

$$\rho(P) \le 1 - \frac{\xi}{b'_\Gamma \gamma'_\Gamma d^*} e^{-M/T}, \tag{78}$$

for $T \le T'_1$, where

$$T'_1 = \min\left[M\left(\log\frac{N(A+B)}{2b'_\Gamma \gamma'_\Gamma}\right)^{-1}, \frac{\delta}{\log 2}\right]. \tag{79}$$

Now (78) does not give the best order of convergence to π, but it can be easily applied, at least in some cases of interest. Diaconis and Stroock (1991) have some examples in which the parameters b'_Γ and γ'_Γ are explicitly evaluated.

This fact shows us that in the choice of the set of paths, in general we have to decide between admissible paths and paths with the smallest number of edges, with the advantages and disadvantages given above. In another paper, we show that neither of the Poincaré bounds and the Desai–Fiedler bounds dominates the other in general [Ingrassia (1994)].

Finally, we compare the bounds of Theorems 6.1 and 6.2 in the case of the d-dimensional cube. Analogous bounds can be computed for the Gibbs sampler.

EXAMPLE 6.3. Let us consider a Metropolis chain on a d-dimensional cube. Given a set of paths Γ' with the smallest number of edges, in bound (77) we have $d^* = d$, $\gamma'_\Gamma = d$ and $b'_\Gamma = 2^{d-1}$. In many cases of interest ξ and M can be directly evaluated (with $1 \le \xi \le Z_T$); then bound (77) gives

$$\lambda_2 \le 1 - \frac{1}{d^2 2^{d-1}} e^{-M/T}. \tag{80}$$

In this case $\mu_2(L) = 2$ [see, e.g., Fiedler (1973)], and Theorem 6.1 gives

$$\lambda_2 \le 1 - \frac{2}{d} e^{-M/T}.$$

Thus it is much tighter than (80), while from the Desai–Fiedler bound (68), as $e_G = d$ and $N = 2^d$, we have

$$\lambda_2 \le 1 - 2d \sin^2(\pi/2^d) e^{-M/T}, \tag{81}$$

which is even worse than (80) for large d.

NOTE 6.1. For a Metropolis chain on a d-dimensional cube, we cannot compute the bound of Theorem 4.1 on λ_2 in general cases. In fact the set of paths Γ depends on the definition of the function H on the set Ω (i.e., on each vertex of the cube).

6.3. *A comparison between the rates of convergence of the Metropolis algorithm and the Gibbs sampler.* Recently Frigessi, Hwang, Sheu and Di Stefano (1993) investigated the speed of weak convergence of the Metropolis algorithm and the Gibbs sampler in terms of their second largest eigenvalues in absolute value and studied the stochastic Ising model in depth. In particular, it is proved that the Metropolis algorithm is the best at low temperature (and then, in this case, it is faster than the Gibbs sampler) and the worst at high temperature (see the original paper for details).

To begin, we reformulate the Metropolis algorithm as a random local updating dynamic on sites. As for the Gibbs sampler, first we consider the transition probability matrix RM_s relative to the site s. With the notation of Section 5, we have

$$RM_s(x,y) \equiv \begin{cases} \dfrac{\exp\left[-\{H(y) - H(x)\}^+/T\right]}{c-1}, & \text{if } y \in N_s(x) \setminus \{x\}, \\ 1 - \displaystyle\sum_{z \in N_s(x)\setminus\{x\}} RM_s(x,z), & \text{if } y = x, \\ 0, & \text{otherwise.} \end{cases}$$

Then the transition probability matrix RM relative to the random updating of sites, for each $x, y \in \Omega$, is given by

$$\begin{aligned}RM(x,y) &\equiv \frac{1}{|S|}\sum_{s\in S} RM_s(x,y)\\ &= \frac{1}{|S|}\sum_{s\in S}\frac{\exp\left[-\{H(y)-H(x)\}^+/T\right]}{c-1}1_s(x,y),\end{aligned} \tag{82}$$

where c is the cardinality of the set $N_s(x)$. As above, in this notation, the letter R means *random updating of sites*.

For each $s \in S$, the transition matrix RM_s is reversible with respect to π; thus also RM is reversible with respect to π. Now Theorem 4.1 becomes the following.

THEOREM 6.3. *Let RM be the transition probability matrix of a random updating dynamic based on the Metropolis algorithm. Then its second largest eigenvalue satisfies the relation*

$$1-\frac{Z_T}{(c-1)|S|}\left(\frac{N}{2}\right)^2 e^{-m/T} \le \lambda_2 \le 1-\frac{Z_T}{b_\Gamma\gamma_\Gamma(c-1)|S|}e^{-m/T}. \tag{83}$$

PROOF. The theorem follows from Theorems 4.1 and 4.3. □

Since at low temperature the dominant eigenvalue is λ_2, then we have proved that, for small values of T, the two algorithms have the same order of convergence $e^{-m/T}$ to the equilibrium distribution π.

APPENDIX

Let S_f be the symmetric group, let p denote any permutation of S_f, let $p(i,j)$ denote the permutation of p by exchanging i and j and let $c(p)$ denote the number of cycles in the disjoint cycle decomposition of p. Then $c(p(i,j)) = c(p) \pm 1$.

Let $H: S_f \to \mathbb{R}$ be a real-valued function defined as $H(p) = c(p) - 1$, for each $p \in S_f$, and let $R_f = (R_f(p,q))_{p,q\in S_f}$ be the transition probability matrix on S_f given by

$$R_f(p,q) \equiv \begin{cases}\binom{f}{2}^{-1}, & \text{if } pq^{-1} \text{ is a transposition},\\ 0, & \text{otherwise.}\end{cases} \tag{84}$$

For $\alpha > 1$, set $1/T = \log \alpha$; then from (11) we have

$$\pi(p) = \frac{\exp[-H(p)\log\alpha]}{Z_{T,\alpha}} = \frac{\alpha^{-H(p)}}{Z_{T,\alpha}},$$

where $Z_{T,\alpha} = \Sigma_{p \in S_f} \exp[-H(p)\log \alpha] = \Sigma_{p \in S_f} \alpha^{-H(p)}$. Then the transition probability matrix $P_{\alpha,f} = (P_{\alpha,f}(p,q))_{p,q \in S_f}$ of the Metropolis algorithm becomes:

$$(85)\qquad P_{\alpha,f}(p,q) \equiv \begin{cases} \binom{f}{2}^{-1} \frac{1}{\alpha}, & \text{if } q = p(i,j) \text{ and } c(q) = c(p) + 1, \\ \binom{f}{2}^{-1}, & \text{if } q = p(i,j) \text{ and } c(q) = c(p) - 1, \\ 1 - \sum_{r \neq p} P(p,r), & \text{if } q = p, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, $P_{1,f}$ gives a simple random walk on S_f. In Hanlon (1992) the transition probability matrices $P_{\alpha,3}$ and $P_{\alpha,4}$ are studied in depth.

NOTE A.1. For a fixed integer f, all transition probability matrices $\{P_{\alpha,f}\}_{\alpha \geq 1}$ have the same underlying graph $G(P_{\cdot,f})$. Let L_f denote the Laplacian matrix of $G(P_{\cdot,f})$ and $\{\mu_i(L_f)\}_{i=1,\dots,f!}$ denote the set of eigenvalues of L_f. As the graph $G(P_{\cdot,f})$ is k-regular, then the following relation holds between the eigenvalues of L_f and the eigenvalues $\{\lambda_i^{(1,f)}\}_{i=1,\dots,f!}$ of $P_{1,f}$:

$$(86)\qquad \mu_i(L_f) = \binom{f}{2}\left(1 - \lambda_i^{(1,f)}\right).$$

In the following we give further details on Examples 6.1 and 6.2.

TABLE 1
Eigenvalues of $P_{\alpha,3}$

Eigenvalue	**Multiplicity**
1	1
$1 - \frac{1}{\alpha}$	1
$\frac{1}{3}\left(1 - \frac{1}{\alpha}\right)$	3
$-\frac{1}{\alpha}$	1

TABLE 2
Eigenvalues of L_3

Eigenvalue	Multiplicity
0	1
3	4
6	1

CASE $f = 3$. The transition probability matrix $P_{\alpha,3}$ is given in (73). In Table 1 we list the eigenvalues of $P_{\alpha,3}$ [Hanlon (1992)]; hence $\lambda_2 = 1 - 1/\alpha$ and $\lambda_{\min} = -1/\alpha$.

Afterwards, from (86), we immediately obtain the eigenvalues of L_3 (Table 2); and hence $\mu_2(L_3) = 3$ and $\mu_6(L_3) = 6$.

The graph $G(P_{\cdot,3})$ is given in Figure 3, where we have the following:

$$(c = 3) \qquad x_1 = \text{id};$$
$$(c = 2) \qquad x_2 = (1,2), \qquad x_3 = (1,3), \qquad x_4 = (2,3);$$
$$(c = 1) \qquad x_5 = (1,2,3), \qquad x_6 = (1,3,2).$$

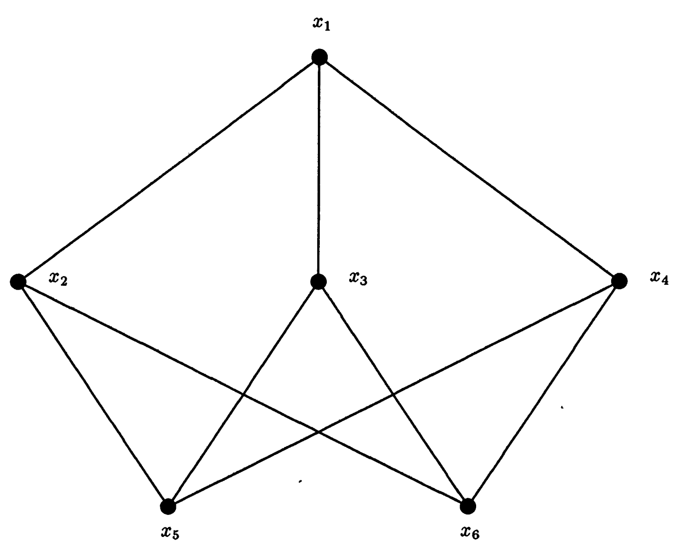

FIG. 3.

A set of admissible paths is the following (for simplicity, we shall write γ_{ij} rather than $\gamma_{x_i x_j}$):

$$\begin{array}{lllll}
\gamma_{12} = e_{12}; & \gamma_{13} = e_{13}; & \gamma_{14} = e_{14}; & \gamma_{15} = e_{12}, e_{25}; & \gamma_{16} = e_{14}, e_{46}; \\
\gamma_{21} = e_{21}; & \gamma_{23} = e_{21}, e_{13}; & \gamma_{24} = e_{25}, e_{54}; & \gamma_{25} = e_{25}; & \gamma_{26} = e_{26}; \\
\gamma_{31} = e_{31}; & \gamma_{32} = e_{31}, e_{12}; & \gamma_{34} = e_{36}, e_{64}; & \gamma_{35} = e_{35}; & \gamma_{36} = e_{36}; \\
\gamma_{41} = e_{41}; & \gamma_{42} = e_{45}, e_{52}; & \gamma_{43} = e_{46}, e_{63}; & \gamma_{45} = e_{45}; & \gamma_{46} = e_{46}; \\
\gamma_{51} = e_{53}, e_{31}; & \gamma_{52} = e_{52}; & \gamma_{53} = e_{53}; & \gamma_{54} = e_{54}; & \gamma_{56} = e_{53}, e_{36}; \\
\gamma_{61} = e_{64}, e_{41}; & \gamma_{62} = e_{62}; & \gamma_{63} = e_{63}; & \gamma_{64} = e_{64}; & \gamma_{65} = e_{63}, e_{35}.
\end{array}$$

By counting, one can verify that the most traveled directed edges $(e_{12}, e_{25}, e_{31}, e_{36}, e_{46}, e_{53}, e_{63}, e_{64})$ belong to three paths and thus $b_\Gamma = 3$. Note that the paths γ_{23} and γ_{32} do not have the lowest elevation; condition (7) holds—in fact, we have

$$\operatorname{elev}(\gamma_{xy}) - H(x) - H(y) = H_1 - H_2 - H_3 = 0 < m = 1.$$

CASE $f = 4$. In Figure 4 we give the graph $G(P_{\cdot,4})$, where we have the following:

$$\begin{array}{lllll}
(c = 4) & x_1 = \mathrm{id}; & & & \\
(c = 3) & x_2 = (1,2), & x_3 = (1,3), & x_4 = (1,4), & x_5 = (2,3), \\
 & x_6 = (2,4), & x_7 = (3,4); & & \\
(c = 2) & x_8 = (1,2,3), & x_9 = (1,2,4), & x_{10} = (1,3,4), & x_{11} = (2,3,4), \\
 & x_{12} = (1,3,2), & x_{13} = (1,4,2), & x_{14} = (1,4,3), & x_{15} = (2,4,3), \\
 & x_{16} = (1,2)(3,4), & x_{17} = (1,3)(2,4), & x_{18} = (1,4)(2,3); & \\
(c = 1) & x_{19} = (1,2,3,4), & x_{20} = (1,2,4,3), & x_{21} = (1,3,2,4), & x_{22} = (1,4,2,3), \\
 & x_{23} = (1,3,4,2), & x_{24} = (1,4,3,2). & &
\end{array}$$

The transition probability matrix $P_{\alpha,4}$ can be immediately written by (85). Unfortunately the list of eigenvalues of $P_{\alpha,4}$ given in Hanlon (1992) is incomplete since the sum of the multiplicities is 22 rather than 24. Moreover, some of these eigenvalues are given implicitly as roots of a certain equation depending on α, and these eigenvalues are incorrectly given, missing a factor 6α. In fact, in the case $\alpha = 1$ the roots of this equation are not in the interval $[-1, 1]$.

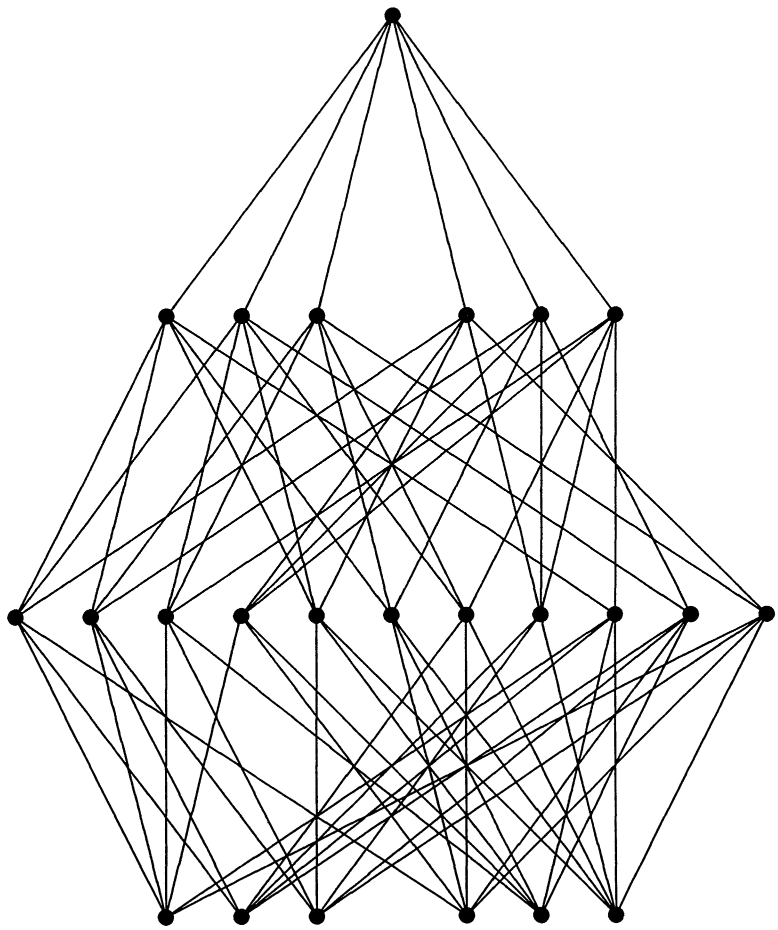

FIG. 4. *The graph $G(P_{\alpha,4})$: First line, x_1; second line (from left to right), $x_2, \dots, x_7$; third line, $x_8, \dots, x_{18}$; fourth line, $x_{19}, \dots, x_{24}$.*

We have computed the correct spectrum of $P_{\alpha,4}$, which is given in Table 3, where r_1, r_2 and r_3 are the roots of the equation

$$r^3 - 9(\alpha - 1)r^2 + 4(5\alpha^2 - 11\alpha + 5)r - (12\alpha^3 - 50\alpha^2 + 50\alpha - 12) = 0. \tag{87}$$

For large values of α, the eigenvalue

$$\lambda' = \frac{9(\alpha - 1) + \sqrt{9\alpha^2 - 2\alpha + 9}}{12\alpha} \tag{88}$$

TABLE 3
Eigenvalues of $P_{\alpha,4}$

Eigenvalue	Multiplicity
1	1
$\frac{1}{2} - \frac{1}{6\alpha}$	4
$-\frac{1}{\alpha}$	1
$\frac{1}{2} - \frac{1}{2\alpha}$	1
$\frac{1}{6} - \frac{1}{2\alpha}$	4
$\frac{1}{3} - \frac{1}{3\alpha}$	1
$\frac{9(\alpha-1) + \sqrt{9\alpha^2 - 2\alpha + 9}}{12\alpha}$	3
$\frac{9(\alpha-1) - \sqrt{9\alpha^2 - 2\alpha + 9}}{12\alpha}$	3
$\frac{r_1}{6\alpha}, \frac{r_2}{6\alpha}, \frac{r_3}{6\alpha}$	2 each

is arbitrarily close to 1; for values of α arbitrarily close to 1, the eigenvalue

$$\lambda'' = -\frac{1}{\alpha}$$

is close to -1. In fact, for $\alpha = 1$, from (87), we obtain the eigenvalues $-\frac{1}{3}$, 0 and $\frac{1}{3}$, while for large values of α, the eigenvalues given by (87) are smaller than the eigenvalue λ' given by (88).

Afterwards, from (86), we immediately obtain the eigenvalues of L_4 given in Table 4, and hence $\mu_2(L_4) = 4$ and $\mu_{24}(L_4) = 12$.

Acknowledgments. I am indebted to Paolo Baldi both for calling my attention to the problem considered here and for his help in the preparation of this paper. Thanks also are due to Arnoldo Frigessi, Riccardo Re, Julian

TABLE 4
Eigenvalues of L_4

Eigenvalue	Multiplicity
0	1
4	9
6	4
8	9
12	1

Stander and Ezio Venturino for their suggestions and remarks. Special thanks are due to one of the referees for his very careful revision and interesting suggestions (in particular for the idea of Theorem 4.5), which have contributed to improving the results and the clarity with respect to a previous version.

REFERENCES

CHIANG, T. S. and CHOW, Y. (1988). On eigenvalues and annealing rates. *Math. Oper. Res.* **13** 508–511.

CHIANG, T. S. and CHOW, Y. (1991). A comparison of simulated annealing of Gibbs sampler and Metropolis algorithms. Technical report, Institute of Mathematics, Academia Sinica, Taipei, Taiwan.

DESAI, M. P. (1992). An eigenvalue-based approach to the finite time behaviour of simulated annealing. Ph.D. dissertation, Coordinate Science Laboratory, Univ. Illinois, Urbana-Champaign.

DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, CA.

DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.

FIEDLER, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J.* **23** 298–305.

FIEDLER, M. (1989). Laplacian of graphs and algebraic connectivity. *Combinatorics Graph Theory* **25** 57–70.

FRIGESSI, A., HWANG, C. R., SHEU, S. J. and DI STEFANO, P. (1993). Convergence rates of the Gibbs sampler, the Metropolis algorithm, and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B* **55** 205–220.

GEMAN, D. (1991). Random fields and inverse problems in imaging. *Ecole d'Eté de Probabilités de Saint-Flour XVIII. Lecture Notes in Math.* **1427** 115–193. Springer. Berlin.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6** 721–741.

HANLON, P. (1992). A Markov chain on the symmetric group and Jack symmetric functions. *Discrete Math.* **99** 123–140.

HASTINGS, W. K. (1970). Monte Carlo sampling using Markov chains and their application. *Biometrika* **57** 77–109.

HOLLEY, R. and STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 553–569.

INGRASSIA, S. (1994). Geometric approaches to the estimation of the spectral gap of reversible Markov chains. *Combinatorics, Probab. Comput.* To appear.

MAAS, C. (1987). Transportation in graphs and the admittance spectrum. *Discrete Appl. Math.* **16** 31–49.

METROPOLIS, M., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

MOHAR, B. (1991). The Laplacian spectrum of graphs. In *Graph Theory, Combinatorics, and Applications* (Y. Alavi, ed.). Wiley, New York.

SINCLAIR, A. (1991). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probab. Comput.* **1** 351–370.

SINCLAIR, A. and JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82** 93–133.

ISTITUTO DI STATISTICA
FACOLTÀ DI
ECONOMIA E COMMERCIO
CORSO ITALIA, 55
25129 CATANIA
ITALY