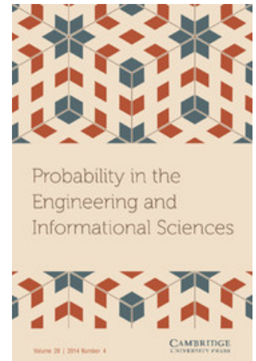

ON THE CONVERGENCE OF METROPOLIS-TYPE RELAXATION AND ANNEALING WITH CONSTRAINTS

Marc C. Robini, Yoram Bresler and Isabelle E. Magnin

Probability in the Engineering and Informational Sciences / Volume 16 / Issue 04 / October 2002, pp 427 - 452
DOI: 10.1017/S0269964802164035, Published online: 10 October 2002

Probability in the Engineering and Informational Sciences, **16**, 2002, 427–452. Printed in the U.S.A.

ON THE CONVERGENCE OF METROPOLIS-TYPE RELAXATION AND ANNEALING WITH CONSTRAINTS

Marc C. Robini
CREATIS
UMR CNRS 5515, INSA Lyon
Villeurbanne, France
E-mail: marc.robini@creatis.insa-lyon.fr

Yoram Bresler
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801
E-mail: ybresler@uiuc.edu

Isabelle E. Magnin
CREATIS
UMR CNRS 5515, INSA Lyon
Villeurbanne, France

We discuss the asymptotic behavior of time-inhomogeneous Metropolis chains for solving constrained sampling and optimization problems. In addition to the usual inverse temperature schedule $(\beta_n)_{n\in\mathbb{N}^*}$, the type of Markov processes under consideration is controlled by a divergent sequence $(\theta_n)_{n\in\mathbb{N}^*}$ of parameters acting as Lagrange multipliers. The associated transition probability matrices $(P_{\beta_n,\theta_n})_{n\in\mathbb{N}^*}$ are defined by $P_{\beta,\theta} = q(x,y)\exp(-\beta(W_\theta(y) - W_\theta(x))^+)$ for all pairs (x,y) of distinct elements of a finite set Ω, where q is an irreducible and reversible Markov kernel and the energy function W_θ is of the form $W_\theta = U + \theta V$ for some functions $U, V : \Omega \to \mathbb{R}$. Our approach, which is based on a comparison of the distribution of the chain at time n with the invariant measure of P_{β_n,θ_n}, requires the computation of an upper bound for the second largest eigenvalue in absolute value of P_{β_n,θ_n}. We extend the geometric bounds derived by Ingrassia and we give new sufficient conditions on the control sequences for the algorithm to simulate a Gibbs distribution with energy U on the constrained set $\widetilde{\Omega} = \{x \in \Omega : V(x) = \min_{z\in\Omega} V(z)\}$ and to minimize U over $\widetilde{\Omega}$.

 0269-9648/02 $12.50

1. INTRODUCTION

Let Ω be a general but finite state space and let $\tilde{\Omega}$ be a proper subset of Ω defined by

$$\tilde{\Omega} := \left\{ x \in \Omega : V(x) = \min_{z \in \Omega} V(z) \right\}, \tag{1.1}$$

where V is a nonconstant real-valued function on Ω. Let $U : \Omega \to \mathbb{R}$ be another nonconstant function. Our primary goal is to study the asymptotic behavior of a class of discrete-time, nonhomogeneous Markov chains controlled by a temperature variable, together with a parameter acting as a Lagrange multiplier in order to solve the following two problems.

Problem 1 (Sampling with Constraints): Let $\beta_0 \in \mathbb{R}_+^*$. Sample from the Gibbs distribution

$$\tilde{\pi}_{\beta_0}(x) = \tilde{Z}_{\beta_0}^{-1} \mathbb{1}_{\{x \in \tilde{\Omega}\}} \exp(-\beta_0 U(x)), \quad x \in \Omega, \tag{1.2}$$

where the constant $\tilde{Z}_{\beta_0}$, the partition function, is given by $\tilde{Z}_{\beta_0} = \sum_{z \in \tilde{\Omega}} \exp(-\beta_0 U(z))$.

Problem 2 (Global Optimization with Constraints): Minimize U over $\tilde{\Omega}$; that is, determine the set

$$\tilde{\Omega}_{\min} := \left\{ x \in \tilde{\Omega} : U(x) = \min_{z \in \tilde{\Omega}} U(z) \right\}.$$

The class of Markov processes under consideration is based on an extension of the dynamics introduced by Metropolis, Rosenbluth, Rosenbluth, and Teller [29]. Let μ be a probability distribution on Ω that charges every point and let q be an irreducible transition probability matrix on Ω called the communication kernel. We assume that q is μ-reversible in the sense that $\mu(x)q(x,y) = \mu(y)q(y,x)$ for all $x, y \in \Omega$. Next, for any two parameters $\beta, \theta \in \mathbb{R}_+^*$, define the transition probability matrix $P_{\beta,\theta}$ on Ω by

$$P_{\beta,\theta}(x,y) = \begin{cases} q(x,y)\exp(-\beta(W_\theta(y) - W_\theta(x))^+) & \text{if } y \neq x \\ 1 - \sum_{z \neq x} P_{\beta,\theta}(x,z) & \text{if } y = x, \end{cases} \tag{1.3}$$

where $w^+ := w \vee 0$ and $W_\theta : \Omega \to \mathbb{R}$ is a nonconstant function parameterized by θ. Under our assumptions, $P_{\beta,\theta}$ is primitive (i.e., irreducible and aperiodic) and its unique equilibrium probability measure is the "μ-weighted" Gibbs distribution $\pi_{\beta,\theta}$ at temperature β^{-1} defined by

$$\pi_{\beta,\theta}(x) = Z_{\beta,\theta}^{-1} \mu(x) \exp(-\beta W_\theta(x)), \quad x \in \Omega, \tag{1.4}$$

where $Z_{\beta,\theta} = \sum_{z\in\Omega} \mu(z)\exp(-\beta W_\theta(z))$ (the letter Z is used throughout to denote appropriate normalizing constants; the corresponding expressions will be omitted in the sequel provided there is no ambiguity). Finally, let us put

$$W_\theta(x) = U(x) + \theta V(x) \quad \text{for all } x \in \Omega \tag{1.5}$$

and consider two strictly positive, nondecreasing sequences $(\beta_n)_{n\in\mathbb{N}^*}$ and $(\theta_n)_{n\in\mathbb{N}^*}$ such that $\lim_{n\to+\infty}\theta_n = +\infty$. We will study the family of nonhomogeneous Markov chains $(X_n)_{n\in\mathbb{N}}$ with the initial law of X_0 given by ν_0 and transitions $\mathrm{P}(X_n = y|X_{n-1} = x) = P_{\beta_n,\theta_n}(x,y)$, $x, y \in \Omega$. Such chains are Metropolis-type algorithms with state space Ω, energy function (1.5) controlled by $(\theta_n)_{n\in\mathbb{N}^*}$, communication kernel q, and inverse temperature schedule $(\beta_n)_{n\in\mathbb{N}^*}$. We shall use the notation $\mathbf{M}(\Omega, U, V, q, (\beta_n), (\theta_n))$ for short.

Substituting (1.5) into (1.4), we find that, for any $\beta_0 \in \mathbb{R}_+^*$ and for all $x \in \Omega$,

$$\lim_{\theta\to+\infty} \pi_{\beta_0,\theta}(x) = \pi_{\beta_0,\infty}(x) := Z^{-1}_{\beta_0,\infty}\mathbb{1}_{\{x\in\tilde{\Omega}\}}\mu(x)\exp(-\beta_0 U(x)), \tag{1.6}$$

which reduces to $\tilde{\pi}_{\beta_0}$ (1.2) when q is symmetric. It can also be checked that $\pi_{\beta,\theta}$ tends to a distribution which gives strictly positive mass to every configuration $x \in \tilde{\Omega}_{\min}$ as $\beta, \theta \to +\infty$. More precisely,

$$\lim_{\beta,\theta\to+\infty} \pi_{\beta,\theta}(x) = \pi_\infty(x) := Z^{-1}_\infty \mathbb{1}_{\{x\in\tilde{\Omega}_{\min}\}}\mu(x). \tag{1.7}$$

From these observations, by analogy with the unconstrained case (i.e., $V \equiv$ constant), $\mathbf{M}(\Omega, U, V, q, (\beta_n), (\theta_n))$ will be referred to as stochastic relaxation if $\beta_n \equiv$ constant and as simulated annealing if $\lim_{n\to+\infty}\beta_n = +\infty$. The key idea is that, for sufficiently slowly increasing control sequences, the law ν_n of X_n should be close to π_{β_n,θ_n} and we can expect that

$$\lim_{n\to+\infty} \sup_{\nu_0\in\mathcal{M}(\Omega)} \|\nu_n - \Pi\|_{\mathrm{Var}} = 0, \tag{1.8}$$

where $\mathcal{M}(\Omega)$ denotes the set of all probability measures on Ω, $\Pi = \pi_{\beta_0,\infty}$ (1.6) or π_∞ (1.7) depending on whether relaxation or annealing is considered, and the variation distance between probabilities ν and π on a finite set Ω is defined by

$$\|\nu - \pi\|_{\mathrm{Var}} := \max_{E\subset\Omega}|\nu(E) - \pi(E)| = \frac{1}{2}\sum_{x\in\Omega}|\nu(x) - \pi(x)|.$$

Discrete-time Metropolis algorithms have been extensively studied during the last 15 years. We refer to Metropolis et al. [29], Hastings [19], Peskun [31], and Kirkpatrick, Gelatt, and Vecchi [28] for original accounts, to Hajek [18], Chiang and Chow [7], Holley and Stroock [20], and Catoni [4–6] for theoretical work on annealing, to Ingrassia [22], and François [12] for results associated with relaxation, and to Azencott [1], Tierney [36], and Gidas [17] for synthetic reviews. However, most of the available convergence results concern the unconstrained case. Actually, when dealing with deterministic constraints, some theoretical results relating to the Gibbs sampler on a finite product space can be found in Geman and Geman [15] and Winkler [37] (see also Geman [14] and Gidas [17]), whereas for Metropolis chains, the only convergence results we are aware of were recently established by Yao [38] (see Gidas [17] for the continuous-time case). Adopting Dobrushin's contraction technique, Yao [38] proved that (1.8) holds for both relaxation and annealing if the control sequences satisfy

$$\beta_n \theta_n = \zeta \ln(n + c_1) + c_2$$

for some constants $c_1 > 0$, $c_2 \geq 0$, and $0 < \zeta < (\ell \mathfrak{d}_V)^{-1}$, where

$$\mathfrak{d}_V := \max\{|V(y) - V(x)| : x, y \in \Omega,\, q(x,y) > 0\}$$

and ℓ is the smallest integer such that $\prod_{n=k+1}^{k+\ell} P_{\beta_n,\theta_n}$ is a positive matrix for k large enough. Still, this sufficient condition can be greatly improved. For any $x, y \in \Omega$, let us denote by Γ_{xy} the set of all simple paths (admissible for q) from x to y and let

$$V(x,y) := \min_{\gamma \in \Gamma_{xy}} \max_{z \in \gamma} V(z) \tag{1.9}$$

be the minimal communication level between x and y on the constraints landscape (Ω, V, q). Let $\tilde{\Omega}_{\text{loc}}$ be the set of proper local minima of (Ω, V, q); that is, $x \in \tilde{\Omega}_{\text{loc}}$ if it is not a global minimum (i.e., $x \notin \tilde{\Omega}$) and no state y with $V(y) < V(x)$ is such that $V(x,y) = V(x)$. We shall show that (1.8) is guaranteed by the less restrictive condition $\zeta < h_V^{-1}$, where the constant

$$\begin{aligned} h_V &:= \max_{x,y \in \Omega}\{V(x,y) - V(x) - V(y)\} + \min_{z \in \Omega} V(z) \\ &= \max_{x \in \tilde{\Omega} \cup \tilde{\Omega}_{\text{loc}},\, y \in \tilde{\Omega} : x \neq y} V(x,y) - \min_{z \in \Omega} V(z), \end{aligned} \tag{1.10}$$

the critical height of (Ω, V, q), can be described as the minimal constraint barrier separating any local or global minimum of (Ω, V, q) from another state in $\tilde{\Omega}$. This nonnegative quantity was first introduced by Holley and Stroock [20] and Chiang and Chow [7]; it reduces to the constant of Hajek [18] when the involved energy landscape has a unique global minimum. Still, our results differ from the work of these authors in that the latter provide asymptotic annealing convergence conditions for the unconstrained case; that is, as V can then be assumed to be zero, conditions involving only $(\beta_n)_{n \in \mathbb{N}^*}$ and $h_{W_0} = h_U$.

Clearly, $h_V < \ell \mathfrak{d}_V$ whenever $h_V > 0$, and the ratio $\ell \mathfrak{d}_V / h_V$ is substantially large in many situations of practical interest (we are not interested in the special case $h_V = 0$ which can be tackled with more efficient approaches). Indeed, as far as annealing is concerned, paralleling the constrained optimization problem with the unconstrained one shows that the improvement we offer here is similar to the improvement provided by Holley and Stroock [20] and Chiang and Chow [7] over standard convergence results based on ergodicity theorems (see, e.g., Geman and

Geman [16] and Mitra, Sangiovanni-Vincentelli, and Sangiovanni-Vincentelli [30]). Before proceeding with the outline of this article, note that some related work about annealing with time-dependent energy function can be found in Frigerio and Grillo [13] and Del Moral and Miclo [8]. Nevertheless, our contribution does not fit into this framework, as the associated assumptions would force the sequence $(\theta_n)_{n\in\mathbb{N}^*}$ to be bounded above.

Let $L^2(1/\pi)$, where π is a strictly positive distribution on Ω, be the real vector space $\mathbb{R}^{|\Omega|}$ endowed with the inner product $\langle\phi,\psi\rangle_{1/\pi} := \sum_{x\in\Omega}\phi(x)\psi(x)(\pi(x))^{-1}$, from which we derive the vector norm $\|\phi\|_{1/\pi} := \langle\phi,\phi\rangle_{1/\pi}^{1/2}$. Then, let us put $\pi_n := \pi_{\beta_n,\theta_n}$ as defined by (1.4) and (1.5) and let

$$\xi_n := \|\nu_n - \pi_n\|_{1/\pi_n}, \tag{1.11}$$

that is, ξ_n^2 is the chi-square contrast of ν_n with respect to π_n (see, e.g., Brémaud [3]). Since $\|\nu_n - \pi_n\|_{\mathrm{Var}} \le \frac{1}{2}\xi_n$ and $\lim_{n\to+\infty}\|\pi_n - \Pi\|_{\mathrm{Var}} = 0$, it turns out that $\lim_{n\to+\infty}\xi_n = 0$ is a sufficient condition for (1.8) to hold. Starting from this straightforward observation, the purpose of Section 2 is to compute an upper bound on ξ_n in the general case of a nonhomogeneous Markov chain with transitions $(P_n)_{n\in\mathbb{N}^*}$ having the property that, for all n, P_n is irreducible and reversible relative to some probability distribution π_n. Our conclusions are contained in Theorem 2.1, the application of which goes through the estimation of an upper bound on the second largest eigenvalue in absolute value of each P_n. In order to meet this requirement, Section 3 is dedicated to the extension of the eigenvalue bounds computed by Ingrassia [22]; it gives rise to new spectral estimates that may be of interest in their own right. Finally, in Section 4, we make use of these estimates together with Theorem 2.1 in order to prove the main result of this article, Theorem 4.1, which states the basic conditions on the control sequences for the quantity ξ_n associated with P_{β_n,θ_n} to be bounded above by a strictly positive power of n^{-1}. The convergence properties of Metropolis-type relaxation and annealing then follow directly and we derive expressions for the convergence rates which allow the measurement of the benefits resulting from our improvement in the upper bound on ζ.

2. GENERAL RESULTS

Let Ω be a finite state space. We consider the general context of a discrete-time, nonhomogeneous Markov chain $(X_n)_{n\in\mathbb{N}}$ on Ω with irreducible and reversible transition probability matrices $(P_n)_{n\in\mathbb{N}^*}$. In other words, each operator P_n defined as $[P_n\phi](x) = \sum_{y\in\Omega} P_n(x,y)\phi(y)$, $x\in\Omega$, $\phi:\Omega\to\mathbb{R}$, is a self-adjoint contraction on $L^2(\pi_n)$, the real vector space $\mathbb{R}^{|\Omega|}$ endowed with the inner product $\langle\phi,\psi\rangle_{\pi_n} := \sum_{x\in\Omega}\phi(x)\psi(x)\pi_n(x)$. Also, recall that these hypotheses imply that, for all n, the spectrum of P_n, say $\{\lambda_{n,i}\}_{i=1,\dots,|\Omega|}$, is real with $1 = \lambda_{n,1} > \lambda_{n,2} \ge \cdots \ge \lambda_{n,|\Omega|} \ge -1$ and $\lambda_{n,|\Omega|} > -1$ if and only if P_n is aperiodic. We shall denote the second largest eigenvalue in absolute value of P_n by $\rho(P_n) := \lambda_{n,2} \vee |\lambda_{n,|\Omega|}|$.

Let ν_n be the law of X_n and let π_n be the unique invariant distribution of P_n. Our goal here is to compute an upper bound on $\xi_n := \|\nu_n - \pi_n\|_{1/\pi_n}$ in order to provide a starting point for the characterization of the variation distance between ν_n and π_n in the specific case of Metropolis algorithms. As will become clear, $\rho(P_n)$ plays a central role toward this end.

PROPOSITION 2.1: *For all $n \in \mathbb{N}\backslash\{0,1\}$, we have $\xi_n \leq \rho(P_n)(a_n\xi_{n-1} + b_n)$, with $a_n := \|D_n^{-1/2}D_{n-1}^{1/2}\|_2$ and $b_n := \|\pi_{n-1} - \pi_n\|_{1/\pi_n}$, where $D_n := \operatorname{diag}(\pi_n)$ and $\|\cdot\|_2$ is the matrix norm induced by the Euclidian vector norm.*

The proof appeals to the following lemma. We denote by $\mathbf{1}$ the function which is identically equal to 1 on Ω (in vector notation, $\mathbf{1} = (1,\ldots,1)^T$).

LEMMA 2.1: *Let P be an irreducible, π-reversible transition probability matrix on Ω. Then, for each $\phi:\Omega \to \mathbb{R}$ such that $\langle \phi, \mathbf{1}\rangle_\pi = 0$, we have $\|P\phi\|_\pi \leq \rho(P)\|\phi\|_\pi$.*

PROOF: Define $S = D^{1/2}PD^{-1/2}$ by $D := \operatorname{diag}(\pi)$. Clearly, S and P have the same eigenvalues $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1$. Moreover, from the reversibility of P, the matrix S is symmetric and therefore real orthogonally diagonalizable. Let $\{v_i\}_{i=1,\ldots,|\Omega|}$ be an orthonormal set of right eigenvectors of S such that v_i corresponds to the eigenvalue with the same subscript. The vectors $w_i := D^{-1/2}v_i$ $(i = 1,\ldots,|\Omega|)$ form an orthonormal basis in $L^2(\pi)$ and they also form a set of right eigenvectors of P. Consequently, for any $\phi:\Omega \to \mathbb{R}$, we can write $\phi = \sum_i \langle\phi, w_i\rangle_\pi w_i$ and $P\phi = \sum_i \lambda_i\langle\phi, w_i\rangle_\pi w_i$ so that $\|\phi\|_\pi^2 = \sum_i \langle\phi, w_i\rangle_\pi^2$ and $\|P\phi\|_\pi^2 = \sum_i \lambda_i^2\langle\phi, w_i\rangle_\pi^2$. Since $w_1 = \mathbf{1}$, it follows that if $\langle\phi,\mathbf{1}\rangle_\pi = 0$, then

$$\frac{\|P\phi\|_\pi^2}{\|\phi\|_\pi^2} = \frac{\displaystyle\sum_{i=2,\ldots,|\Omega|} \lambda_i^2\langle\phi, w_i\rangle_\pi^2}{\displaystyle\sum_{i=2,\ldots,|\Omega|} \langle\phi, w_i\rangle_\pi^2} \leq \rho^2(P). \qquad \blacksquare$$

PROOF OF PROPOSITION 2.1: Since $\nu_n - \pi_n = P_n^T\nu_{n-1} - D_n\mathbf{1} = D_nP_nD_n^{-1}\nu_{n-1} - D_nP_n\mathbf{1}$, we have $D_n^{-1/2}(\nu_n - \pi_n) = D_n^{1/2}P_n(D_n^{-1}\nu_{n-1} - \mathbf{1})$ and, hence, $\xi_n = \|P_n(D_n^{-1}\nu_{n-1} - \mathbf{1})\|_{\pi_n}$. Observe that $\langle D_n^{-1}\nu_{n-1} - \mathbf{1}, \mathbf{1}\rangle_{\pi_n} = \langle \nu_{n-1}, \mathbf{1}\rangle - 1 = 0$. Therefore, applying Lemma 2.1, we obtain

$$\begin{aligned}\xi_n &\leq \rho(P_n)\|D_n^{-1}\nu_{n-1} - \mathbf{1}\|_{\pi_n} = \rho(P_n)\|\nu_{n-1} - \pi_n\|_{1/\pi_n}\\ &\leq \rho(P_n)(\|\nu_{n-1} - \pi_{n-1}\|_{1/\pi_n} + b_n),\end{aligned}$$

where

$$\begin{aligned}\|\nu_{n-1} - \pi_{n-1}\|_{1/\pi_n} &= \|D_n^{-1/2}D_{n-1}^{1/2}D_{n-1}^{-1/2}(\nu_{n-1} - \pi_{n-1})\|_2\\ &\leq \|D_n^{-1/2}D_{n-1}^{1/2}\|_2\|D_{n-1}^{-1/2}(\nu_{n-1} - \pi_{n-1})\|_2\\ &= a_n\xi_{n-1}. \qquad \blacksquare\end{aligned}$$

Proposition 2.1 is all that is needed to initiate the proof of the following result.

THEOREM 2.1: *Assume that there exists a constant* $N \in \mathbb{N}\backslash\{0,1\}$ *and two bounded functions* $f, g : (\alpha, +\infty) \to \mathbb{R}_+^*$, *with* $\alpha < N$, *such that the following hold:*

(i) For all positive integers $n \geq N$, $f(n) \leq -\ln(\rho(P_n)a_n)$ *and* $g(n) \geq \rho(P_n)b_n$.
(ii) f *and* g *are continuously differentiable and decreasing in* $(\alpha, +\infty)$.
(iii) There exists a real constant $c \in (-1, +\infty)$ *such that* $(g(x))^{-1}(g/f)'(x) \geq c$ *for all* $x \in [N+1, +\infty)$.

Then, for all positive integers $n > N$,

$$\xi_n < \Xi_{N-1} \exp(-F(n+1) + F(N)) + \frac{\chi}{1+c} \exp(f(n+1)) \frac{g(n+2)}{f(n+2)}, \qquad \textbf{(2.1)}$$

where Ξ_{N-1} *(see (2.2)) is a positive constant determined by the initial law* ν_0 *and the transitions* $(P_n)_{2 \leq n < N}$, F *is a primitive of* f *in* $(\alpha, +\infty)$, *and* $\chi := \sup\{g(x)/g(x+2) : x \in [N, +\infty)\}$.

PROOF: Taking that $\prod_{k=l}^{n} \rho(P_k)a_k = 1$ if $l > n$, Proposition 2.1 gives

$$\xi_n \leq \xi_1 \prod_{k=2}^{n} \rho(P_k)a_k + \sum_{m=2}^{n} \rho(P_m)b_m \prod_{k=m+1}^{n} \rho(P_k)a_k =: \Xi_n \qquad \textbf{(2.2)}$$

for all $n \geq 2$. Observe that, for all $n \geq N$,

$$\begin{aligned} \prod_{k=m}^{n} \rho(P_k)a_k &=: \left(\prod_{k=m}^{N-1} \rho(P_k)a_k\right)\left(\prod_{k=m\vee N}^{n} \rho(P_k)a_k\right) \\ &\leq \left(\prod_{k=m}^{N-1} \rho(P_k)a_k\right) \exp\left(-\sum_{k=m\vee N}^{n} f(k)\right), \end{aligned}$$

with

$$\sum_{k=m\vee N}^{n} f(k) \geq \int_{m\vee N}^{n+1} f(x)\, dx =: F(n+1) - F(m \vee N).$$

Hence, setting $S_n := \sum_{m=N}^{n} g(m) \exp(F(m+1))$, we obtain

$$\xi_n \leq \Xi_{N-1} \exp(-F(n+1) + F(N)) + \exp(-F(n+1)) S_n. \qquad \textbf{(2.3)}$$

We have

$$\begin{aligned} S_n &\leq \chi \sum_{m=N}^{n} g(m+2) \exp(F(m+1)) \\ &\leq \chi \mathcal{I}_n, \\ \mathcal{I}_n &:= \int_{N}^{n+1} g(x+1) \exp(F(x+1))\, dx. \end{aligned} \qquad \textbf{(2.4)}$$

Our task now is to derive an upper bound on $\mathcal{I}_n$. The substitution $y = F(x+1)$ yields

$$\mathcal{I}_n = \int_{F(N+1)}^{F(n+2)} (G \circ F^{-1})'(y) \exp(y)\, dy,$$

where G is a primitive of g in $[N+1, +\infty)$. Note that $(G \circ F^{-1})'(y) = (g \circ F^{-1})(y)/(f \circ F^{-1})(y) > 0$ for all $y \in [F(N+1), +\infty)$. Integrating by parts, we get

$$\mathcal{I}_n = \left[\frac{g(y)}{f(y)} \exp(F(y)) \right]_{N+1}^{n+2} - \mathcal{J}_n,$$

with

$$\mathcal{J}_n = \int_{F(N+1)}^{F(n+2)} (G \circ F^{-1})''(y) \exp(y)\, dy \geq \mathcal{I}_n \min_{y \in [F(N+1), F(n+2)]} \frac{(G \circ F^{-1})''(y)}{(G \circ F^{-1})'(y)}$$
$$= \mathcal{I}_n \min_{x \in [N+1, n+2]} \frac{1}{g(x)} \left(\frac{g}{f} \right)'(x) \geq \mathcal{I}_n c,$$

and it follows that

$$\mathcal{I}_n \leq (1+c)^{-1} \left[\frac{g(y)}{f(y)} \exp(F(y)) \right]_{N+1}^{n+2} < (1+c)^{-1} \exp(F(n+2)) \frac{g(n+2)}{f(n+2)}. \tag{2.5}$$

Finally, from (2.3)–(2.5), we obtain

$$\xi_n < \Xi_{N-1} \exp(-F(n+1) + F(N)) + \frac{\chi}{1+c} \exp(F(n+2) - F(n+1)) \frac{g(n+2)}{f(n+2)}$$

and the theorem follows from the fact that since f is decreasing, F is concave and, hence, $F(n+2) - F(n+1) \leq f(n+1)$. ■

A straightforward example of functions satisfying conditions (ii) and (iii) in Theorem 2.1 is given by $f(x) = K_1 x^{-\varepsilon_1}$ and $g(x) = K_2 x^{-\varepsilon_2}$, where K_2, ε_2, and $\varepsilon_1 < \varepsilon_2 \wedge 1$ are strictly positive constants and $K_1 > (\varepsilon_2 - \varepsilon_1)(N+1)^{\varepsilon_1 - 1}$. In this case, the interesting thing is that the upper bound in (2.1) has the limit 0 as $n \to +\infty$. Indeed, in Section 4, a similar choice will be made for the application of Theorem 2.1 to the convergence of the class of algorithms $\mathbf{M}(\Omega, U, V, q, (\beta_n), (\theta_n))$. However, to be able to do so, we must provide some upper bounds for a_n, b_n, and $\rho(P_n)$. This task turns out to be easy when considering a_n and b_n (see Lemma 4.1), but the lack of spectral estimates for generic Metropolis chains makes it much harder for the second largest eigenvalue in absolute value. The following section is intended to introduce some results that will allow us to overcome this difficulty.

3. GEOMETRIC BOUNDS FOR EIGENVALUES OF METROPOLIS CHAINS

Consider an irreducible, reversible transition probability matrix P on a finite set Ω and let $\rho(P)$ be the second largest eigenvalue in absolute value of P; that is, $\rho(P) = \lambda_2 \vee |\lambda_{|\Omega|}|$, where $\lambda_2 < 1$ and $\lambda_{|\Omega|} \geq -1$ respectively stand for the second largest and the smallest eigenvalues. Motivated by the well-known fact that $\rho(P)$ governs the rate of convergence of the time-homogeneous Markov chain with transition P, several authors (Sinclair and Jerrum [35], Diaconis and Stroock [10], Sinclair [34], and Desai and Rao [9]) have developed upper bounds for λ_2 and lower bounds for $\lambda_{|\Omega|}$ (see Ingrassia [21] for a synthetic review on the subject). Nevertheless, these spectral estimates depend on some geometric quantities associated with the transition graph of P, which are difficult to compute for generic updating dynamics like the Gibbs sampler and Metropolis-type algorithms. In such situations, the tightest eigenvalue bounds known to us were derived by Ingrassia [22] from the estimates of Diaconis and Stroock [10].

When treating Metropolis chains, Ingrassia [22] restricted his study to specific symmetric communication kernels. Here, we extend the results of this author to any irreducible and reversible communication. We first introduce some definitions and we recall the estimates of Diaconis and Stroock [10] in order to state the problem clearly. Then, an upper bound on λ_2 and a lower bound on $\lambda_{|\Omega|}$ are proved for the purpose of studying the constrained relaxation and annealing processes described in the Introduction. Although these results allow us to bound the mixing time of general Metropolis chains, it is worthwhile noting that tighter convergence time bounds can be obtained for specific problems involving particular instances of either the Metropolis dynamics or the closely related Glauber dynamics. For instance, polynomial time bounds were proved by Jerrum and Sinclair [25] for sampling nearly perfect matchings in bipartite graphs and matchings in weighted graphs (see also Jerrum and Sinclair [26]). Positive results of a similar flavor can be found in Jerrum [24] and Dyer et al. [11] for sampling vertex colorings of a low-degree graph, and in Jerrum and Sorkin [27] for the graph bisection problem. Still, in addition to our preoccupation, general bounds can be useful for understanding complex chains such as the ones that arise in image processing inverse problems (see, e.g., Geman and Geman [16] and Robini, Rastello, and Magnin [32]) or for other problems that have been shown to produce negative (i.e., super-polynomial or exponential time) convergence results—such problems include finding maximum matchings in arbitrary graphs (Sasaki and Hajek [33]), reaching a maximum clique in a random graph (Jerrum [23]), or the q-state Potts model and the independent set problem on rectangular subsets of an hypercube lattice (Borgs et al. [2]).

3.1. Notation and Preliminaries

We assume that the irreducible Markov kernel $P : \Omega \times \Omega \to [0,1]$ is reversible with respect to its invariant distribution π; that is,

$$Q(x,y) := \pi(x)P(x,y) = Q(y,x) \quad \text{for all } x, y \in \Omega.$$

The transition graph $G(P) = [\Omega, \vec{\mathcal{A}}]$ associated with P is the directed graph with set of vertices Ω and set of arcs $\vec{\mathcal{A}} = \{x \to y : x, y \in \Omega, P(x,y) > 0\}$; we shall use the notation $\vec{a} = (a_-, a_+)$ for an arc with initial vertex a_- and end vertex a_+.

Given any two vertices $x, y \in \Omega$, we denote by γ_{xy} a path from x to y on $G(P)$ and we define its Q-length by

$$|\gamma_{xy}|_Q = \sum_{\vec{a} \in \gamma_{xy}} (Q(\vec{a}))^{-1}. \tag{3.1}$$

Since P is irreducible, we can construct a collection $\Gamma = \{\gamma_{xy} : x, y \in \Omega, x \neq y\}$ of paths on $G(P)$ containing one simple path for each ordered pair of distinct vertices $(x, y) \in \Omega \times \Omega$. The first geometric quantity of interest, the "Poincaré coefficient," is

$$\kappa_\Gamma := \max_{\vec{a} \in \vec{\mathcal{A}}} \sum_{\gamma_{xy} \in \Gamma \,:\, \gamma_{xy} \ni \vec{a}} |\gamma_{xy}|_Q \, \pi(x) \pi(y). \tag{3.2}$$

Assuming that P is aperiodic, we can choose a set $\Sigma = \{\sigma_x : x \in \Omega\}$ of cycles on $G(P)$ containing one cycle with an odd number of arcs for each vertex. The second geometric quantity to be considered is related to Σ as follows:

$$\iota_\Sigma := \max_{\vec{a} \in \vec{\mathcal{A}}} \sum_{\sigma_x \in \Sigma \,:\, \sigma_x \ni \vec{a}} |\sigma_x|_Q \, \pi(x), \tag{3.3}$$

where the Q-length $|\sigma_x|_Q$ is defined by analogy with (3.1).

This notation allows us to state the spectral estimates computed by Diaconis and Stroock [10].

PROPOSITION 3.1: *Let P be an irreducible, π-reversible transition probability matrix on Ω. Then, the second largest eigenvalue λ_2 of P satisfies*

$$\lambda_2 \leq 1 - \kappa_\Gamma^{-1} \tag{3.4}$$

with κ_Γ defined in (3.2). Moreover, if P is aperiodic, then its smallest eigenvalue $\lambda_{|\Omega|}$ satisfies

$$\lambda_{|\Omega|} \geq -1 + 2\iota_\Sigma^{-1} \tag{3.5}$$

with ι_Σ defined in (3.3).

Obviously, for some particular case of interest, the quality of the eigenvalue bounds that can be derived from the above results depends on making a judicious selection of the sets Γ and Σ. In the following subsections, we focus on Metropolis chains with transition probability matrix P on Ω defined by

$$P(x,y) = q(x,y) \exp(-\beta(W(y) - W(x))^+) \quad \text{if } y \neq x, \tag{3.6}$$

where the inverse temperature $\beta \in \mathbb{R}_+^*$ is fixed, the energy function $W : \Omega \to \mathbb{R}$ is nonconstant, and the communication kernel $q : \Omega \times \Omega \to [0,1]$ is taken to be irreducible and reversible with respect to its invariant distribution μ. Therefore, P is primitive and its equilibrium probability measure is the Gibbs distribution $\pi(x) = Z_\beta^{-1}\mu(x)\exp(-\beta W(x))$, $x \in \Omega$.

To compute our spectral estimates, we shall adopt the approach of Ingrassia [22] based on Proposition 3.1. Let

$$\mathcal{S}(x) := \{y \in \Omega\backslash\{x\} : x \to y \in \vec{\mathcal{A}}\}$$

be the set of proper neighbors of x and let $d_\star := \max_{x\in\Omega}|\mathcal{S}(x)|$. Ingrassia's results are limited to symmetric communication kernels of the form

$$q(x,y) = \begin{cases} (d_\star)^{-1} & \text{if } y \in \mathcal{S}(x) \\ (d_\star)^{-1}(d_\star - |\mathcal{S}(x)|) & \text{if } y = x \\ 0 & \text{otherwise,} \end{cases}$$

with the additional assumption that $|\mathcal{S}(x)| = d_\star$ for all $x \in \{y \in \Omega : (\exists z \in \mathcal{S}(y))[W(z) > W(y)]\} =: \Upsilon$ when considering the smallest eigenvalue. Our contribution is to extend these results to arbitrary irreducible and reversible q. The same proof techniques are used and this generalization mainly involves dealing with the two following awkward elements: $q(x,y)$ is no longer constant for all pairs (x,y) such that $y \in \mathcal{S}(x)$, and $q(x,x)$ is not necessarily zero if $x \in \Upsilon$.

3.2. An Upper Bound on λ_2

Given any pair of distinct vertices $(x,y) \in \Omega \times \Omega$, the set of all simple paths from x to y on $G(P)$ will be denoted by Γ_{xy}. Let h be the critical height of the energy landscape (Ω, W, q) defined by analogy with (1.9) and (1.10). A path γ_{xy} is said to be W-admissible if

$$\max_{z\in\gamma_{xy}} W(z) - W(x) - W(y) + \min_{z\in\Omega} W(z) \leq h \tag{3.7}$$

(it is said to be strictly W-admissible if the inequality is strict). Obviously, there is at least one W-admissible simple path between every pair of distinct vertices $(x,y) \in \Omega \times \Omega$. Like Ingrassia [22], we consider a set $\Gamma = \{\gamma_{xy} : x, y \in \Omega, x \neq y\}$ of W-admissible simple paths on $G(P)$. Let b_Γ be the maximum number of paths $\gamma \in \Gamma$ that use the same arc and let ℓ_Γ be the length of the longest path in Γ; that is,

$$b_\Gamma := \max_{\vec{a}\in\vec{\mathcal{A}}} |\{\gamma \in \Gamma : \gamma \ni \vec{a}\}| \tag{3.8}$$

and

$$\ell_\Gamma := \max_{\gamma\in\Gamma} |\gamma|, \tag{3.9}$$

where $|\gamma|$ denotes the number of arcs in γ. The following result is obtained by means of (3.4) in Proposition 3.1.

THEOREM 3.1: *Let P be a transition probability matrix of the form (3.6) with irreducible and μ-reversible communication kernel. Then, the second largest eigenvalue λ_2 of P satisfies*

$$\lambda_2 \leq 1 - \frac{|\Omega_{\min}|\,q_{\min}}{\mu_\star^2 b_\Gamma \ell_\Gamma} \exp(-\beta h), \qquad \textbf{(3.10)}$$

where $\Omega_{\min} = \{x \in \Omega : W(x) = \min_{z\in\Omega} W(z)\}$ *is the set of global minima of W,* $q_{\min} = \min\{q(\vec{a}) : \vec{a} \in \vec{\mathcal{A}}, a_- \neq a_+\}$ *denotes the minimum communication probability over proper transitions, and* $\mu_\star = \max\{\mu(x)/\mu(y) : x, y \in \Omega\}$ *measures the fractional communication dissymmetry.*

PROOF: We have $Q(\vec{a}) = \pi(a_-)P(\vec{a}) = Z_\beta^{-1}\mu(a_-)q(\vec{a})\exp(-\beta(W(a_-)\vee W(a_+)))$ and hence

$$|\gamma_{xy}|_Q \pi(x)\pi(y) = Z_\beta^{-1}\mu(x)\mu(y) \sum_{\vec{a}\in\gamma_{xy}} (\mu(a_-)q(\vec{a}))^{-1} \times \exp(\beta(W(a_-)\vee W(a_+) - W(x) - W(y)))$$

for any $\gamma_{xy} \in \Gamma$. From our choice of Γ (3.7), since $W(a_-)\vee W(a_+) \leq \max\{W(z) : z \in \gamma_{xy}\}$, we have

$$|\gamma_{xy}|_Q \pi(x)\pi(y) \leq Z_\beta^{-1}\mu(x)\mu(y)\exp\Big(\beta\Big(h - \min_{z\in\Omega} W(z)\Big)\Big) \sum_{\vec{a}\in\gamma_{xy}} (\mu(a_-)q(\vec{a}))^{-1}$$

$$\leq \frac{\mu_\star^2 |\gamma_{xy}|}{\underline{Z}_\beta q_{\min}} \exp(\beta h),$$

$$\underline{Z}_\beta := \sum_{z\in\Omega} \exp\Big(-\beta\Big(W(z) - \min_{x\in\Omega} W(x)\Big)\Big).$$

Thus, by the definitions of b_Γ (3.8) and ℓ_Γ (3.9), and since $\underline{Z}_\beta \geq |\Omega_{\min}|$, the quantity κ_Γ (3.2) satisfies

$$\kappa_\Gamma \leq \frac{\mu_\star^2 b_\Gamma \ell_\Gamma}{|\Omega_{\min}|\,q_{\min}} \exp(\beta h)$$

and the theorem follows from Proposition 3.1. ■

Note 3.1: The upper bound (3.10) is similar to the one computed by Holley and Stroock [20], the difference being that we can express the constant appearing in front of $\exp(-\beta h)$. In addition to its intrinsic importance, this specification is needed to set the basis for the proof of our main results (Section 4) because of the dependence upon θ in the case of relaxation and annealing with constraints.

Note 3.2: In place of (3.4), other geometric estimates can provide a starting point for computing an upper bound on λ_2. For instance, based on some previous work (Sinclair and Jerrum [35]), Sinclair [34] proved the following bounds:

$$\lambda_2 \leq 1 - \frac{1}{8\eta_\Gamma^2}, \quad \eta_\Gamma := \max_{\vec{a}\in\vec{\mathcal{A}}} (Q(\vec{a}))^{-1} \sum_{\gamma_{xy}\in\Gamma : \gamma_{xy}\ni\vec{a}} \pi(x)\pi(y), \tag{3.11}$$

$$\lambda_2 \leq 1 - \frac{1}{\bar{\eta}_\Gamma}, \quad \bar{\eta}_\Gamma := \max_{\vec{a}\in\vec{\mathcal{A}}} (Q(\vec{a}))^{-1} \sum_{\gamma_{xy}\in\Gamma : \gamma_{xy}\ni\vec{a}} |\gamma_{xy}|\pi(x)\pi(y). \tag{3.12}$$

Adopting a similar reasoning as in the proof of Theorem 3.1, the bound (3.11) gives

$$\lambda_2 \leq 1 - \frac{1}{8}\left(\frac{|\Omega_{\min}|q_{\min}}{\mu_\star^2 b_\Gamma}\right)^2 \exp(-2\beta h), \tag{3.13}$$

whereas (3.12) leads to the same upper bound as (3.10). Alternatively, provided that W has a unique global minimum $\tilde{x}$ and $\pi(\tilde{x}) > (c^{1/3} + \frac{4}{9}c^{-1/3} - \frac{2}{3})^2$, $c := 19/27 + \sqrt{33}/9$, the approach proposed by François [12] leads to an estimate of the form

$$\lambda_2 \leq 1 - K\exp(-2\beta h) \tag{3.14}$$

for some constant $K > 0$ and β large enough. Finally, one can resort to the results of Desai and Rao [9] from which it is possible to obtain

$$\lambda_2 \leq 1 - K' \exp(-\beta\Delta_W) \tag{3.15}$$

for some constant $K' > 0$ and $\Delta_W := \max\{W(x) - W(y) : x, y \in \Omega\}$. Still, except for the trivial case $h = 0$ and the extreme case $h = \Delta_W$, the upper bound of Theorem 3.1 becomes sharper than (3.13)–(3.15) as β increases. Furthermore, this observation holds in the context of relaxation and annealing with constraints.

3.3. A Lower Bound on $\lambda_{|\Omega|}$

Let Λ_P be the set defined by $\Lambda_P = \{x \in \Omega : P(x,x) > 0\}$. For any vertex $x \in \Omega$, we denote by r_x the minimum number of arcs of $G(P)$ that are needed to join x to an element of Λ_P. In other words,

$$r_x := \mathbb{1}_{\{x\notin\Lambda_P\}} \min_{y\in\Lambda_P} \min_{\gamma\in\Gamma_{xy}} |\gamma|.$$

Following Ingrassia [22], we consider a set $\Sigma = \{\sigma_x : x \in \Omega\}$ of odd cycles on $G(P)$ reaching Λ_P through a minimum number of arcs; that is,

$$\sigma_x = \begin{cases} \gamma_{xx} & \text{if } x \in \Lambda_P \\ (\gamma_{xy}, \gamma_{yy}, \gamma_{yx}) & \text{otherwise,} \end{cases} \tag{3.16}$$

where $\gamma_{xx} := x \to x$, the vertex y is an element of Λ_P, the path γ_{xy} is such that $|\gamma_{xy}| = r_x$, and γ_{yx} stands for γ_{xy} reversed. Finally, let δ be the minimum nonzero jump of W over proper transitions,

$$\delta := \min_{x\in\Omega} \min_{y\in\mathcal{S}(x) : W(y)\neq W(x)} |W(x) - W(y)|,$$

and let b_Σ be the maximum number of cycles $\sigma \in \Sigma' := \Sigma \backslash \{\gamma_{xx} : x \in \Lambda_P\}$ that use the same arc,

$$b_\Sigma := \max_{\vec{a} \in \vec{\mathcal{A}}} |\{\sigma \in \Sigma' : \sigma \ni \vec{a}\}|.$$

THEOREM 3.2: *Let P be a transition probability matrix of the form (3.6) with irreducible and μ-reversible communication kernel, and let μ_* and $q_{\min}$ be as defined in Theorem 3.1. Then, the smallest eigenvalue $\lambda_{|\Omega|}$ of P satisfies*

$$\lambda_{|\Omega|} \geq -1 + 2q_{\min}\left[A\left(\frac{q_{\min}}{\bar{q}_{\min}} \vee \frac{1}{1-\exp(-\beta\delta)}\right) + B\right]^{-1}, \qquad \textbf{(3.17)}$$

where $A = b_\Sigma \mu_* + 1$, $B = 2b_\Sigma r_* \mu_*$ with $r_* = \max\{r_x : x \in \Omega\}$, and $\bar{q}_{\min}$ is the minimum nonzero self-loop communication probability if such exists; that is, $\bar{q}_{\min} = \min\{q(x,x) : x \in \Lambda_q\}$ if $\Lambda_q := \{x \in \Omega : q(x,x) > 0\} \neq \emptyset$, otherwise $\bar{q}_{\min} = 1$.

Let us start with a few lemmas to keep the proof simple.

LEMMA 3.1: *$x \notin \Lambda_P$ if and only if $q(x,x) = 0$ and $W(z) \leq W(x)$ for all $z \in \mathcal{S}(x)$.*

PROOF: The lemma follows directly from the fact that

$$x \notin \Lambda_P \Leftrightarrow \sum_{z \in \mathcal{S}(x)} q(x,z)\exp(-\beta(W(z) - W(x))^+) = 1. \qquad \blacksquare$$

LEMMA 3.2: *Assume that $\Omega \backslash \Lambda_P$ is nonempty. Then, any path $\gamma = (x_l)_{l=0,\ldots,L}$ on $G(P)$ such that $x_l \notin \Lambda_P$ for all $l \in \{0,\ldots,L\}$ is a path of constant energy.*

PROOF: If $W(x_l) > W(x_{l-1})$ for some $l \in \{1,\ldots,L\}$, then $x_{l-1} \in \Lambda_P$ by Lemma 3.1. Likewise, as $x \in \mathcal{S}(y) \Leftrightarrow y \in \mathcal{S}(x)$ from the irreducibility and the reversibility of q, $W(x_l) < W(x_{l-1})$ implies that $x_l \in \Lambda_P$. Hence, γ must be of constant energy. $\blacksquare$

LEMMA 3.3: *For any $x \in \Lambda_P$, we have $P(x,x) \geq \bar{q}_{\min} \wedge (q_{\min}(1 - \exp(-\beta\delta)))$.*

PROOF: On one hand, if $W(z) \leq W(x)$ for all $z \in \mathcal{S}(x)$, then $q(x,x) > 0$ by Lemma 3.1 and

$$P(x,x) = 1 - \sum_{z \in \mathcal{S}(x)} q(x,z) = q(x,x) \geq \bar{q}_{\min}.$$

On the other hand, if there exists $z_0 \in \mathcal{S}(x)$ such that $W(z_0) > W(x)$, then

$$\begin{aligned} P(x,x) &\geq \left(1 - \sum_{z \in \mathcal{S}(x) \backslash \{z_0\}} q(x,z)\right) - q(x,z_0)\exp(-\beta(W(z_0) - W(x))) \\ &\geq q(x,z_0) - q(x,z_0)\exp(-\beta\delta) \\ &\geq q_{\min}(1 - \exp(-\beta\delta)). \end{aligned} \qquad \blacksquare$$

PROOF OF THEOREM 3.2: The approach is similar to the proof of Theorem 3.1 in the sense that we shall compute an upper bound on the geometric quantity ι_Σ (3.3) in order to apply (3.5) in Proposition 3.1.

For any $\vec{a} \in \vec{\mathcal{A}}$, we can write

$$\sum_{\sigma_x \in \Sigma : \sigma_x \ni \vec{a}} |\sigma_x|_Q \pi(x) = \sum_{\sigma_x \in \Sigma : \sigma_x \ni \vec{a}} (\mathbb{1}_{\{x \notin \Lambda_P\}} + \mathbb{1}_{\{x \in \Lambda_P\}})|\sigma_x|_Q \pi(x) \tag{3.18}$$

so that the cases $x \notin \Lambda_P$ and $x \in \Lambda_P$ can be considered separately.

Let us first assume that $x \notin \Lambda_P$. Then, by our choice of Σ (see (3.16)), $\sigma_x = (\gamma_{xz}, \gamma_{zy}, \gamma_{yy}, \gamma_{yz}, \gamma_{zx})$, where $y \in \Lambda_P$, γ_{yy} is a self-loop from y to y, the path $\gamma_{zy} = z \to y$ is simply an arc with $z \notin \Lambda_P$, and γ_{xz} has all its vertices in $\Omega \backslash \Lambda_P$ and length $|\gamma_{xz}| = r_x - 1$. Note that from Lemmas 3.1 and 3.2, $W(z) \geq W(y)$ and γ_{xz} is a path of constant energy. Since Q is symmetric, we have

$$|\sigma_x|_Q \pi(x) = 2(|\gamma_{xz}|_Q \pi(x) + |\gamma_{zy}|_Q \pi(x)) + |\gamma_{yy}|_Q \pi(x). \tag{3.19}$$

We can provide an upper bound for each of the three terms that appear in the right-hand side of (3.19). First, as γ_{xz} is of constant energy,

$$\begin{aligned} |\gamma_{xz}|_Q \pi(x) &= \sum_{\vec{a} \in \gamma_{xz}} (\pi(a_-) P(\vec{a}))^{-1} \pi(x) \\ &= \sum_{\vec{a} \in \gamma_{xz}} (\mu(a_-) q(\vec{a}))^{-1} \mu(x) \leq (r_* - 1) \frac{\mu_*}{q_{\min}}. \end{aligned}$$

Second, since $W(x) = W(z) \geq W(y)$,

$$|\gamma_{zy}|_Q \pi(x) = \frac{\pi(x)}{\pi(z) P(z,y)} = \frac{\mu(x)}{\mu(z) q(z,y)} \leq \frac{\mu_*}{q_{\min}}.$$

Third, appealing to Lemma 3.3, we have

$$|\gamma_{yy}|_Q \pi(x) = \frac{\pi(x)}{\pi(y) P(y,y)} \leq \frac{\mu(x)}{\mu(y) P(y,y)} \leq \mu_* (\bar{q}_{\min} \wedge (q_{\min}(1 - \exp(-\beta\delta)))^{-1}.$$

Consequently,

$$|\sigma_x|_Q \pi(x) \leq \frac{2 r_* \mu_*}{q_{\min}} + \mu_* \left(\frac{1}{\bar{q}_{\min}} \vee \frac{1}{q_{\min}(1 - \exp(-\beta\delta))} \right)$$

and it follows that

$$\sum_{\sigma_x \in \Sigma : \sigma_x \ni \vec{a}} \mathbb{1}_{\{x \notin \Lambda_P\}} |\sigma_x|_Q \pi(x) \leq \frac{1}{q_{\min}} \left[\mu_* b_\Sigma \left(\frac{q_{\min}}{\bar{q}_{\min}} \vee \frac{1}{1 - \exp(-\beta\delta)} \right) + B \right]. \tag{3.20}$$

Now, let us consider the case $x \in \Lambda_P$. Then, the cycle $\sigma_x \in \Sigma$ is a self-loop from x to x and the corresponding treatment is easier. Making use of Lemma 3.3, we have

$$|\sigma_x|_Q \pi(x) = (P(x,x))^{-1} \leq (\bar{q}_{\min} \wedge (q_{\min}(1 - \exp(-\beta\delta)))^{-1}.$$

Hence, as $\sum_{\sigma_x \in \Sigma : \sigma_x \ni \vec{a}} \mathbb{1}_{\{x \in \Lambda_P\}} \leq 1$ for all $\vec{a} \in \vec{\mathcal{A}}$,

$$\sum_{\sigma_x \in \Sigma : \sigma_x \ni \vec{a}} \mathbb{1}_{\{x \in \Lambda_P\}} |\sigma_x|_Q \pi(x) \leq \frac{1}{q_{\min}} \left(\frac{q_{\min}}{\bar{q}_{\min}} \vee \frac{1}{1 - \exp(-\beta\delta)} \right). \tag{3.21}$$

Finally, from (3.18), (3.20), and (3.21) and by the definition of ι_Σ (3.3), we obtain

$$\iota_\Sigma \leq \frac{1}{q_{\min}} \left[A \left(\frac{q_{\min}}{\bar{q}_{\min}} \vee \frac{1}{1 - \exp(-\beta\delta)} \right) + B \right]$$

and the theorem follows from Proposition 3.1. ■

3.4. An Upper Bound on the Mixing Time

Starting from a given configuration $x \in \Omega$, the rate of convergence of a primitive Markov chain with transition probability matrix P and equilibrium distribution π can be measured via the "mixing time" $\mathcal{T}_x : \mathbb{R}_+^* \to \mathbb{N}$ defined by

$$\mathcal{T}_x(\varepsilon) = \min\{n \in \mathbb{N}^* : (\forall m \geq n)[\|P^m(x, \cdot) - \pi\|_{\mathrm{Var}} \leq \varepsilon]\}.$$

As noted by Sinclair [34], $\mathcal{T}_x(\varepsilon)$ can be bounded above in terms of $\rho(P)$ and $\pi(x)$:

$$\mathcal{T}_x(\varepsilon) \leq (1 - \rho(P))^{-1} \ln\left(\frac{1}{\varepsilon \pi(x)} \right). \tag{3.22}$$

Hence, Theorems 3.1 and 3.2 give an upper bound on the time to reach (quasi-) equilibrium from a given initial state. The following result is noteworthy, although we shall not use it in the sequel.

COROLLARY 3.1: *Let P be a transition probability matrix of the form (3.6) with irreducible and μ-reversible communication kernel and let μ_*, b_Γ, ℓ_Γ, $q_{\min}$, $\bar{q}_{\min}$, δ, A, and B be the quantities defined as in Theorems 3.1 and 3.2. If*

$$\beta \geq (\delta^{-1} \ln 2) \vee \left(h^{-1} \ln\left[\frac{|\Omega_{\min}|}{2\mu_*^2 b_\Gamma \ell_\Gamma} \left(A \left(\frac{q_{\min}}{\bar{q}_{\min}} \vee 2 \right) + B \right) \right] \right) =: \beta^*,$$

then

$$\mathcal{T}_x(\varepsilon) \leq \frac{\mu_*^2 \ln(1/\varepsilon\pi(x))}{|\Omega_{\min}| q_{\min}} b_\Gamma \ell_\Gamma \exp(\beta h). \tag{3.23}$$

PROOF: Let λ_u and λ_l be the upper bound on λ_2 and the lower bound on $\lambda_{|\Omega|}$ given in (3.10) and (3.17), respectively. If $\beta \geq \beta^*$, then

$$\lambda_u \geq 1 - 2q_{\min}\left[A\left(\frac{q_{\min}}{\bar{q}_{\min}} \vee 2\right) + B\right]^{-1}$$
$$= 1 - 2q_{\min}\left[A\left(\frac{q_{\min}}{\bar{q}_{\min}} \vee \frac{1}{1-\exp(-\delta\delta^{-1}\ln 2)}\right) + B\right]^{-1} \geq -\lambda_l.$$

Therefore, $\rho(P) \leq \lambda_u$ and the corollary follows from (3.22). ∎

Note that (3.23) holds for any $\beta \in \mathbb{R}_+^*$ if one considers the slower Markov chain with transition probability $\frac{1}{2}(I + P)$ rather than P, where I is the $|\Omega| \times |\Omega|$ identity matrix (i.e., if one introduces an additional self-loop probability of $\frac{1}{2}$ for each state). In both situations, the mixing time is governed by the critical height of the energy landscape together with the geometric quantities b_Γ and ℓ_Γ associated with some set of admissible paths. This confirms two basic intuitions about necessary conditions for rapid convergence to equilibrium: The chain should not contain any bottleneck and the diameter of its transition graph should be small.

4. CONVERGENCE TOWARD EQUILIBRIUM AND THE GROUND STATES

We now have all of the necessary ingredients to study the class of Metropolis algorithms $\mathbf{M}(\Omega, U, V, q, (\beta_n), (\theta_n))$ defined in the Introduction; that is, the family of nonhomogeneous Markov chains $(X_n)_{n\in\mathbb{N}}$ with transitions $(P_{\beta_n,\theta_n})_{n\in\mathbb{N}^*}$ defined by $P_{\beta,\theta}(x,y) = q(x,y)\exp(-\beta(W_\theta(y) - W_\theta(x))^+)$ for all pairs (x,y) of distinct elements of Ω, where the following assumptions hold:

A1. q is an irreducible, μ-reversible Markov kernel on Ω.

A2. $W_\theta = U + \theta V$ is nonconstant for all $\theta \in \mathbb{R}_+^*$.

For each n, the law of X_n is denoted by ν_n and, for simplicity, we shall put $P_n := P_{\beta_n,\theta_n}$ and write π_n for the equilibrium distribution of P_n.

In order to be able to apply Theorem 2.1 for establishing the convergence of relaxation and annealing, we have to compute upper bounds on the quantities $a_n = \|D_n^{-1/2} D_{n-1}^{1/2}\|_2$ and $b_n = \|\pi_{n-1} - \pi_n\|_{1/\pi_n}$ as well as on the second largest eigenvalue in absolute value $\rho(P_{\beta,\theta})$. These intermediate steps form the subject of Section 4.1. They will allow us to prove our main theorem, Theorem 4.1, from which we deduce the convergence results summarized in Corollaries 4.1 and 4.2.

4.1. Upper Bounds on a_n, b_n, and $\rho(P_{\beta,\theta})$

Let us first consider a_n and b_n. For each $n \in \mathbb{N}\backslash\{0,1\}$, we set

$$\sigma_n := (\beta_n\theta_n - \beta_{n-1}\theta_{n-1})(\Delta_U\theta_n^{-1} + \Delta_V), \tag{4.1}$$

where Δ_J stands for the oscillation of the function $J:\Omega\to\mathbb{R}$, defined as $\Delta_J := \max_{x,y\in\Omega}(J(x)-J(y))$.

LEMMA 4.1: *Assume that the control sequences $(\beta_n)_{n\in\mathbb{N}^*}$ and $(\theta_n)_{n\in\mathbb{N}^*}$ are strictly positive and monotonic increasing. Then, $a_n \le \exp(\frac{1}{2}\sigma_n)$ and $b_n \le \exp(\sigma_n)-1$ for all $n\in\mathbb{N}\backslash\{0,1\}$.*

PROOF: Since

$$a_n = \max\{\lambda^{1/2} : \lambda \in \mathrm{spectrum}(D_n^{-1}D_{n-1})\} = \max_{x\in\Omega}\left(\frac{\pi_{n-1}(x)}{\pi_n(x)}\right)^{1/2}$$

and

$$b_n = \left(\sum_{x\in\Omega}\left(\frac{\pi_{n-1}(x)}{\pi_n(x)}-1\right)^2\pi_n(x)\right)^{1/2} \le \max_{x\in\Omega}\left|\frac{\pi_{n-1}(x)}{\pi_n(x)}-1\right|,$$

it suffices to show that $\exp(-\sigma_n)\le \pi_{n-1}(x)/\pi_n(x)\le\exp(\sigma_n)$ for all $x\in\Omega$.

For each $n\ge 1$, let Ψ_n be the real-valued function on Ω characterized by

$$\Psi_n(x) = \beta_n\Big[\Big(U(x)-\min_{z\in\Omega}U(z)\Big)+\theta_n\Big(V(x)-\min_{z\in\Omega}V(z)\Big)\Big].$$

The equilibrium distribution π_n can then be expressed as $\pi_n(x) = Z_n^{-1}\mu(x)\times\exp(-\Psi_n(x))$ with $Z_n=\sum_{z\in\Omega}\mu(z)\exp(-\Psi_n(z))$ and it is easy to check that

$$\exp\Big(-\max_{z\in\Omega}(\Psi_n(z)-\Psi_{n-1}(z))\Big)\le\frac{Z_n}{Z_{n-1}}\le\frac{\pi_{n-1}(x)}{\pi_n(x)}\le\exp(\Psi_n(x)-\Psi_{n-1}(x))$$

for all $n\ge 2$. The lemma follows from the fact that

$$\Psi_n(x)-\Psi_{n-1}(x)\le(\beta_n-\beta_{n-1})\Delta_U+(\beta_n\theta_n-\beta_{n-1}\theta_{n-1})\Delta_V\le\sigma_n$$

for all $x\in\Omega$. ■

Note 4.1: It can be shown that for n sufficiently large, the inequalities given in Lemma 4.1 still hold if we replace Δ_U by $\max\{U(x)-U(y): x\in\Omega, y\in\widetilde{\Omega}\}$ in (4.1), where $\widetilde{\Omega}$ is defined by (1.1). Moreover, when $\beta_n=\beta_0\in\mathbb{R}_+^*$ for all $n\in\mathbb{N}^*$ (i.e., in the relaxation case), we obtain $a_n\le\exp(\frac{1}{2}\sigma_n')$ and $b_n\le\exp(\sigma_n')-1$, with $\sigma_n'=\beta_0(\theta_n-\theta_{n-1})\Delta_V$ for all $n\ge 2$. However, we will not appeal to these tighter bounds, as they do not lead to better convergence results.

Now, let us turn to the second largest eigenvalue in absolute value of $P_{\beta,\theta}$. Some of the quantities involved in Theorems 3.1 and 3.2 are functions of θ so that a little care is needed when applying these results. In this context, it is important to keep in mind that the set of arcs on $G(P_{\beta,\theta})$ that are not self-loops is entirely defined by the communication kernel q and is therefore independent of U, V, β, and θ. A few definitions are needed before stating our upper bound on $\rho(P_{\beta,\theta})$.

Let $\mathcal{H}_V$ be the collection of ordered pairs of distinct states $(x, y) \in \Omega \times \Omega$ such that

$$V(x, y) - V(x) - V(y) + \min_{z \in \Omega} V(z) = h_V,$$

where $V(x, y)$ is the minimal communication level between x and y on (Ω, V, q) and h_V is the critical height of (Ω, V, q) (see (1.9) and (1.10)). Because q is irreducible, we can construct a set $\Gamma_a = \{\gamma_{xy} : (x, y) \notin \mathcal{H}_V, x \neq y\}$ consisting of exactly one strictly V-admissible simple path on $G(P_{\beta,\theta})$ for each ordered pair of distinct elements $(x, y) \in (\Omega \times \Omega)\backslash \mathcal{H}_V$. A simple path γ_{xy} on $G(P_{\beta,\theta})$ is said to be V-critical if $(x, y) \in \mathcal{H}_V$ and $\max\{V(z) : z \in \gamma_{xy}\} = V(x, y)$. We denote by Γ_c a set $\{\gamma_{xy} : (x, y) \in \mathcal{H}_V\}$ composed of exactly one V-critical simple path for each pair $(x, y) \in \mathcal{H}_V$, and by G_c, we denote the collection of all such sets of paths. Finally, let

$$b_* := \max_{\Gamma_c \in G_c} b_{\Gamma_a \cup \Gamma_c} \quad \text{and} \quad \ell_* := \ell_{\Gamma_a} \vee \left(\max_{\Gamma_c \in G_c} \ell_{\Gamma_c} \right),$$

where b_Γ and ℓ_Γ are respectively defined by (3.8) and (3.9).

PROPOSITION 4.1: *For any constant $\beta_m \in \mathbb{R}_+^*$, there exists a constant $\theta_m \in \mathbb{R}_+^*$ such that, for any $\beta \geq \beta_m$ and for any $\theta \geq \theta_m$, $\rho(P_{\beta,\theta}) \leq 1 - \tau \exp(-\beta(2\Delta_U + \theta h_V))$, where $\tau := q_{\min}(\mu_*^2 b_* \ell_*)^{-1}$.*

The proof resorts to the following simple lemmas.

LEMMA 4.2: *The critical height h_{W_θ} of the energy landscape (Ω, W_θ, q) satisfies $-2\Delta_U + \theta h_V \leq h_{W_\theta} \leq 2\Delta_U + \theta h_V$.*

PROOF: Let us respectively denote by $J_{\min}$ and $J_{\max}$ the minimum and the maximum of a function J on Ω. It is straightforward to show that $U_{\min} + \theta V(x, y) \leq W_\theta(x, y) \leq U_{\max} + \theta V(x, y)$ for all pair of distinct states $(x, y) \in \Omega \times \Omega$ and the lemma follows by considering that $U_{\min} + \theta V_{\min} \leq \min_{z \in \Omega} W_\theta(z) \leq U_{\max} + \theta V_{\min}$. ■

LEMMA 4.3: *There exists a constant $\theta'_m \in \mathbb{R}_+^*$ such that, for any $\theta \geq \theta'_m$, the collection of sets $\{\Gamma = \Gamma_a \cup \Gamma_c : \Gamma_c \in G_c\}$ contains some set $\{\gamma_{xy}(\theta) : x, y \in \Omega,\ x \neq y\}$ exclusively made up of W_θ-admissible paths.*

PROOF: By appealing to Lemma 4.2, we can make the following two observations.

i. Any strictly V-admissible path becomes strictly W_θ-admissible as θ increases. Therefore, there exists $t_a \in \mathbb{R}_+^*$ such that any path in Γ_a is W_θ-admissible if $\theta \geq t_a$.
ii. Let γ_{xy} be a path joining two states x and y such that $(x, y) \in \mathcal{H}_V$. If γ_{xy} is not V-critical, then

$$\max_{z \in \gamma_{xy}} V(z) - V(x) - V(y) + \min_{z \in \Omega} V(z) > h_V$$

and, hence, γ_{xy} is not W_θ-admissible for large values of θ. Consequently, for θ sufficiently large, any W_θ-admissible path linking $(x,y) \in \mathcal{H}_V$ is V-critical. Because for every $\theta \in \mathbb{R}_+^*$, there exists a W_θ-admissible path between every pair of distinct points $(x,y) \in \Omega \times \Omega$, we deduce that for θ large enough, at least one of the V-critical paths linking any pair $(x,y) \in \mathcal{H}_V$ is W_θ-admissible. It follows that there exists $t_c \in \mathbb{R}_+^*$ such that for every $\theta \geq t_c$, at least one of the sets $\Gamma_c \in G_c$ is exclusively composed of W_θ-admissible paths.

Setting $\theta_m' = t_a \vee t_c$ completes the proof. ■

PROOF OF PROPOSITION 4.1: We denote by $\Gamma(\theta)$ a set of W_θ-admissible simple paths on (Ω, W_θ, q) and the dependence of λ_2, $\Omega_{\min}$ (Theorem 3.1), $\lambda_{|\Omega|}$, A, B, δ (Theorem 3.2) upon θ is indicated by the subscript θ.

Clearly, there exists $\vartheta_m \in \mathbb{R}_+^*$ such that, for all $\theta \geq \vartheta_m$,

$$\delta_\theta \geq \min_{x\in\Omega} \min_{y\in\mathcal{S}(x)\,:\,U(y)\neq U(x)} |U(x) - U(y)| =: \delta_U > 0.$$

Then, applying Theorems 3.1 and 3.2 with $|\Omega_{\min,\theta}| \leq |\Omega|$, we obtain that, for any $\theta \geq \vartheta_m$, a sufficient condition for the upper bound on $\lambda_{2,\theta}$ (3.10) to be greater than the absolute value of the lower bound on $\lambda_{|\Omega|,\theta}$ (3.17) is given by

$$h_{W_\theta} \geq \beta^{-1} \ln\left[\frac{|\Omega|}{2\mu_*^2 b_{\Gamma(\theta)} \ell_{\Gamma(\theta)}} \left(A_\theta \left(\frac{q_{\min}}{\bar{q}_{\min}} \vee \frac{1}{1 - \exp(-\beta\delta_U)} \right) + B_\theta \right) \right] =: \mathcal{L}(\beta, \theta).$$

Hence, by Theorem 3.1 and using the fact that $|\Omega_{\min,\theta}| \geq 1$, if $\theta \geq \vartheta_m$ and $h_{W_\theta} \geq \mathcal{L}(\beta,\theta)$, then

$$\begin{aligned} \rho(P_{\beta,\theta}) &\leq 1 - \tau_\theta \exp(-\beta h_{W_\theta}), \quad \tau_\theta := q_{\min}(\mu_*^2 b_{\Gamma(\theta)} \ell_{\Gamma(\theta)})^{-1}, \\ &\leq 1 - \tau_\theta \exp(-\beta(2\Delta_U + \theta h_V)) \quad \text{(by Lemma 4.2).} \end{aligned}$$

Consequently, given any $\beta_m \in \mathbb{R}_+^*$, the above inequality holds for all pairs (β,θ) such that $\beta \geq \beta_m$ and $\theta \in \{\vartheta \geq \vartheta_m : h_{W_\vartheta} \geq \mathcal{L}(\beta_m, \vartheta)\} =: \mathcal{D}_\theta$. Since $\sup\{\mathcal{L}(\beta_m,\vartheta) : \vartheta \in \mathbb{R}_+^*\} < +\infty$ and $h_{W_\vartheta} \geq -2\Delta_U + \vartheta h_V$ by Lemma 4.2, $\mathcal{D}_\theta \supset \{\vartheta \geq \vartheta_m'\}$ for some constant $\vartheta_m' \geq \vartheta_m$. The proposition follows by appealing to Lemma 4.3 and then setting $\theta_m = \theta_m' \vee \vartheta_m'$. ■

4.2. Main Results

In addition to Assumptions A1 and A2 on q and W_θ, we shall make the following assumptions on the control sequences $(\beta_n)_{n\in\mathbb{N}^*}$ and $(\theta_n)_{n\in\mathbb{N}^*}$:

A3. $(\beta_n)_{n\in\mathbb{N}^*}$ and $(\theta_n)_{n\in\mathbb{N}^*}$ are strictly positive and monotonic increasing.

A4. $\lim_{n\to+\infty} \theta_n = +\infty$.

A5. There exists real constants $\omega \in (-1, +\infty)$ and $\zeta \in (0, h_V^{-1})$ such that $\beta_n\theta_n - \beta_{n-1}\theta_{n-1} \leq \zeta(n+\omega)^{-1}$ for all $n \in \mathbb{N}\backslash\{0,1\}$.

A6. The sequence $(\theta_n / \ln(n+\omega))_{n\in\mathbb{N}\backslash\{0,1\}}$ decreases eventually.

By applying Theorem 2.1 together with Lemma 4.1 and Proposition 4.1, we obtain the following theorem, which shows that the quantity ξ_n (1.11) has the limit zero as $n \to +\infty$ whenever assumptions 3–6 are satisfied.

THEOREM 4.1: *Under assumptions 3–6, for any positive real constant*

$$C > \frac{\zeta \Delta_V}{\tau(1+\omega)^{\zeta h_V} \exp(-\beta_1 \theta_1 h_V)} =: C_0,$$

there exists a constant $M \in \mathbb{N}\backslash\{0,1\}$ such that $\xi_n \leq C(n+\omega)^{\zeta(2\Delta_U \theta_n^{-1} + h_V) - 1}$ for all $n \geq M$.

PROOF: In order to be able to apply Theorem 2.1, we must first provide bounds for the quantities $-\ln(\rho(P_n)a_n)$ and $\rho(P_n)b_n$. From Proposition 4.1, we have $\ln(\rho(P_n)) \leq -\tau \exp(-\beta_n(2\Delta_U + \theta_n h_V))$ for n sufficiently large. Therefore, by considering Lemma 4.1 and noting that assumption 5 implies $\beta_n\theta_n - \beta_1\theta_1 < \zeta \ln((n+\omega)/(1+\omega))$, there exists $n_1 \geq 2$ such that, for all $n \geq n_1$,

$$-\ln(\rho(P_n)a_n) > \tau_n (n+\omega)^{-\zeta(2\Delta_U \theta_n^{-1} + h_V)} - \upsilon_n (n+\omega)^{-1}$$

with $\tau_n := \tau \exp((\zeta \ln(1+\omega) - \beta_1 \theta_1)(2\Delta_U \theta_n^{-1} + h_V))$ and $\upsilon_n := \frac{1}{2}\zeta(\Delta_U \theta_n^{-1} + \Delta_V)$. Since $\lim_{n\to+\infty} \theta_n = +\infty$ and $\zeta h_V < 1$, it follows that for any real constant δ_1 satisfying

$$0 < \delta_1 < \lim_{n \to +\infty} \tau_n = \tau(1+\omega)^{\zeta h_V} \exp(-\beta_1 \theta_1 h_V),$$

there exists $n_2 \geq n_1$ such that

$$-\ln(\rho(P_n)a_n) \geq \delta_1 (n+\omega)^{-\zeta(2\Delta_U \theta_n^{-1} + h_V)} \quad \text{for all } n \geq n_2. \tag{4.2}$$

At the same time, using Lemma 4.1 and assumption 5, we have $\rho(P_n)b_n < b_n \leq \exp(2\upsilon_n(n+\omega)^{-1}) - 1 =: s_n$ for all $n \geq 2$. Clearly, under assumptions 3 and 4, the sequence $(s_n)_{n\in\mathbb{N}^*}$ is monotonic decreasing and $\lim_{n\to+\infty} s_n/(\zeta \Delta_V (n+\omega)^{-1}) = 1$. Consequently, for any real constant $\delta_2 > \zeta \Delta_V$, there exists $n_3 \geq 2$ such that

$$\rho(P_n)b_n \leq \delta_2 (n+\omega)^{-1} \quad \text{for all } n \geq n_3. \tag{4.3}$$

We now examine the conditions (i)–(iii) stated in Theorem 2.1. Let $\Theta : [1, +\infty) \to \mathbb{R}_+^*$ be any monotonic increasing function such that $\Theta(n) = \theta_n$ for all $n \geq 2$ and $\Theta(x)/\ln(x+\omega)$ decreases eventually. We define the functions f and g by

$$f(x) = \delta_1 (x+\omega)^{-\zeta(2\Delta_U/\Theta(x) + h_V)} \quad \text{and} \quad g(x) = \delta_2 (x+\omega)^{-1} \tag{4.4}$$

so that, according to (4.2) and (4.3), it suffices to choose $N \geq n_2 \vee n_3$ to obtain condition (i). Then, because the derivative of f is

$$f'(x) = -\zeta f(x) \left[h_V (x+\omega)^{-1} + 2\Delta_U \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{\ln(x+\omega)}{\Theta(x)} \right) \right],$$

there exists $\alpha_0 \geq 1$ such that f is decreasing in $[\alpha_0, +\infty)$ and, hence, condition (ii) holds for any $\alpha \geq \alpha_0$. In addition,

$$\frac{1}{g(x)}\left(\frac{g}{f}\right)'(x) = \delta_2^{-1}\left(\frac{g}{f}\right)(x)\left[\zeta h_V - 1 + 2\zeta\Delta_U(x+\omega)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\ln(x+\omega)}{\Theta(x)}\right)\right]$$
$$> \frac{\zeta h_V - 1}{\delta_1}(x+\omega)^{\zeta(2\Delta_U/\Theta(x)+h_V)-1} \quad \text{for } x \text{ large enough,}$$

and, thus, for any real constant $c \in (-1,0)$, there exists $n_4 \geq 2$ such that condition (iii) is satisfied for any $N \geq n_4$.

Finally, applying Theorem 2.1 with f and g defined by (4.4) and $N \geq n_2 \vee n_3 \vee \lfloor \alpha_0 + 1 \rfloor \vee n_4$, we obtain that for all $n > N$,

$$\xi_n < \Xi_{N-1}\exp(-F(n+1) + F(N))$$
$$+ \frac{(1+2(N+\omega)^{-1})\delta_2}{(1+c)\delta_1}\exp(f(n+1))(n+\omega+2)^{\zeta(2\Delta_U\theta_{n+2}^{-1}+h_V)-1},$$

where the first term in the right-hand side tends to zero exponentially fast as $n \to +\infty$. The theorem follows from the fact that for any real constant $\varepsilon > 0$, the constants δ_1, δ_2, c, and N can be chosen in such a way that

$$0 < \frac{(1+2(N+\omega)^{-1})\delta_2}{(1+c)\delta_1}\exp(f(n+1)) - C_0 \leq \varepsilon$$

for n sufficiently large. ■

The following two corollaries give simple conditions on the control sequences for $\mathbf{M}(\Omega, U, V, q, (\beta_n), (\theta_n))$ to converge in variation to $\pi_{\beta_0,\infty}$ (1.6) and π_∞ (1.7) together with the associated asymptotic convergence rates.

COROLLARY 4.1 (Relaxation): *Assume that* $\beta_n = \beta_0 \in \mathbb{R}_+^*$ *and*

$$\theta_n = \zeta\beta_0^{-1}\ln(n+c_1) + c_2 \quad \textit{for all } n \in \mathbb{N}^*, \tag{4.5}$$

where $\zeta \in (0, h_V^{-1})$, $c_1 \in \mathbb{R}_+^*$, *and* $c_2 \in \mathbb{R}_+$ *are given constants. Then, for any initial distribution* ν_0,

$$\|\nu_n - \pi_{\beta_0,\infty}\|_{\mathrm{Var}} = O(n^{(\zeta h_V - 1)\vee(-\zeta\delta_r)}) \quad \textit{as } n \to +\infty,$$

where

$$\delta_r := \min\{V(z) : z \in \Omega\backslash\tilde{\Omega}\} - \min\{V(z) : z \in \Omega\}. \tag{4.6}$$

PROOF: According to (4.5), $\beta_0(\theta_n - \theta_{n-1}) < \zeta(n-1+c_1)^{-1}$ for all $n \geq 2$ and the sequence $(\theta_n/\ln(n-1+c_1))_{n\geq 2}$ is strictly decreasing. Therefore, assumptions 3–6 are satisfied and Theorem 4.1 gives

$$\|\nu_n - \pi_n\|_{\mathrm{Var}} = O(n^{\zeta(2\Delta_U\theta_n^{-1}+h_V)-1}) = O(n^{\zeta h_V - 1}). \tag{4.7}$$

For each $n \geq 1$, let Φ_n be the real-valued function on Ω defined by

$$\Phi_n(x) = \exp(-\beta_0[U(x) + \theta_n \underline{V}(x)]), \quad \underline{V}(x) := V(x) - \min\{V(z) : z \in \Omega\}.$$

Clearly, $\Phi_n(x) = O(\exp(-\beta_0 \theta_n \underline{V}(x))) = O(n^{-\zeta \underline{V}(x)})$ for all $x \in \Omega \backslash \tilde{\Omega}$ and we can write $\pi_n(x) = Z_n^{-1} \mu(x) \Phi_n(x)$ for all $x \in \Omega$. Hence, since $Z_n > Z_{\beta_0,\infty} > 0$,

$$\pi_n(x) = O(n^{-\zeta \underline{V}(x)}) \quad \text{for all } x \in \Omega \backslash \tilde{\Omega}. \tag{4.8}$$

Now, if $x \in \tilde{\Omega}$, we have

$$\begin{aligned} |\pi_n(x) - \pi_{\beta_0,\infty}(x)| &= Z_{\beta_0,\infty}^{-1} \mu(x) \exp(-\beta_0 U(x)) \frac{Z_n - Z_{\beta_0,\infty}}{Z_n} \\ &= \pi_{\beta_0,\infty}(x) \pi_n(\Omega \backslash \tilde{\Omega}), \end{aligned}$$

whereas $|\pi_n(x) - \pi_{\beta_0,\infty}(x)| = \pi_n(x)$ for all $x \in \Omega \backslash \tilde{\Omega}$. Consequently,

$$\begin{aligned} \|\pi_n - \pi_{\beta_0,\infty}\|_{\text{Var}} &= \pi_n(\Omega \backslash \tilde{\Omega}) \\ &= O(n^{-\zeta \delta_r}) \quad \text{(by (4.8))}. \end{aligned} \tag{4.9}$$

The corollary follows directly from (4.7) and (4.9) by the triangle inequality. ■

Note 4.2: Corollary 4.1 shows that whenever $\zeta \leq (h_V + \delta_r)^{-1}$, the asymptotic convergence rate of constrained relaxation is governed by the difference δ_r (4.6) between the smallest constraint outside the feasible set $\tilde{\Omega}$ and the minimum constraint value. It appears that the fastest convergence rate is achieved for $\zeta = (h_V + \delta_r)^{-1}$ and, hence, rather surprisingly, setting ζ arbitrarily close to h_V^{-1} takes us away from best performance. Still, it can be checked that our improvement in the upper bound on ζ with respect to Yao's result [38] allows for faster convergence if $h_V + \delta_r < \ell \mathfrak{d}_V$, a situation likely to arise in practice, and $(\ell \mathfrak{d}_V)^{-1} < \zeta < h_V^{-1}(1 - \delta_r(\ell \mathfrak{d}_V)^{-1})$; the associated asymptotic convergence speed gain is $O(n^{\epsilon})$, where $\epsilon \in (0, \delta_r((h_V + \delta_r)^{-1} - (\ell \mathfrak{d}_V)^{-1})]$.

COROLLARY 4.2 (Annealing): *Assume that $(\beta_n)_{n \in \mathbb{N}^*}$ and $(\theta_n)_{n \in \mathbb{N}^*}$ are strictly positive, monotonic increasing sequences such that $\lim_{n \to +\infty} \beta_n = \lim_{n \to +\infty} \theta_n = +\infty$ and*

$$\beta_n \theta_n = \zeta \ln(n + c_1) + c_2 \quad \textit{for all } n \in \mathbb{N}^*, \tag{4.10}$$

where $\zeta \in (0, h_V^{-1})$, $c_1 \in \mathbb{R}_+^*$, and $c_2 \in \mathbb{R}_+$ are given constants. Then, for any initial distribution ν_0,

$$\|\nu_n - \pi_\infty\|_{\text{Var}} = O(\exp(-\beta_n \delta_a)) \quad \textit{as } n \to +\infty,$$

where

$$\delta_a := \min\{U(z) : z \in \tilde{\Omega} \backslash \tilde{\Omega}_{\min}\} - \min\{U(z) : z \in \tilde{\Omega}\}. \tag{4.11}$$

PROOF: For each $n \geq 1$, let Φ'_n be the real-valued function on Ω defined by

$$\Phi'_n(x) = \exp(-\beta_n[\widetilde{U}(x) + \theta_n \underline{V}(x)]), \quad \widetilde{U}(x) := U(x) - \min\{U(z) : z \in \widetilde{\Omega}\}.$$

Then, $\Phi'_n(x) = \exp(-\beta_n \widetilde{U}(x))$ for all $x \in \widetilde{\Omega} \backslash \widetilde{\Omega}_{\min}$ and we can write $\pi_n(x) = (Z'_n)^{-1}\mu(x)\Phi'_n(x)$ for all $x \in \Omega$. Since $Z'_n > Z_\infty > 0$, we have

$$\pi_n(\widetilde{\Omega} \backslash \widetilde{\Omega}_{\min}) = O(\exp(-\beta_n \delta_a)) \tag{4.12}$$

and we shall complete the proof by showing that

$$\|\nu_n - \pi_\infty\|_{\text{Var}} \sim \pi_n(\widetilde{\Omega} \backslash \widetilde{\Omega}_{\min}) \quad \text{as } n \to +\infty. \tag{4.13}$$

If $x \in \widetilde{\Omega}_{\min}$, we have $|\pi_n(x) - \pi_\infty(x)| = Z_\infty^{-1}\mu(x)(Z'_n - Z_\infty)/Z'_n = \pi_\infty(x)\pi_n(\Omega \backslash \widetilde{\Omega}_{\min})$, whereas $|\pi_n(x) - \pi_\infty(x)| = \pi_n(x)$ for all $x \in \Omega \backslash \widetilde{\Omega}_{\min}$. Therefore, $\|\pi_n - \pi_\infty\|_{\text{Var}} = \pi_n(\Omega \backslash \widetilde{\Omega}_{\min})$ and it follows that

$$|t_n - (1 + v_n)| \leq \frac{\|\nu_n - \pi_\infty\|_{\text{Var}}}{\pi_n(\widetilde{\Omega} \backslash \widetilde{\Omega}_{\min})} \leq t_n + 1 + v_n, \tag{4.14}$$

where

$$t_n := \frac{\|\nu_n - \pi_n\|_{\text{Var}}}{\pi_n(\widetilde{\Omega} \backslash \widetilde{\Omega}_{\min})} \quad \text{and} \quad v_n := \frac{\pi_n(\Omega \backslash \widetilde{\Omega})}{\pi_n(\widetilde{\Omega} \backslash \widetilde{\Omega}_{\min})}.$$

It is easy to check that assumptions 5 and 6 are satisfied with $\omega = -1 + c_1$ and thus, by Theorem 4.1, there exists some real constants $K, \epsilon > 0$ such that $\|\nu_n - \pi_n\|_{\text{Var}} \leq Kn^{-\epsilon}$ for n sufficiently large. In the same way, as (4.10) implies that

$$\Phi'_n(x) = O(n^{-\zeta(\widetilde{U}(x)\theta_n^{-1} + \underline{V}(x))}) \quad \text{for all } x \in \Omega \backslash \widetilde{\Omega},$$

there exists some real constants $K', \epsilon' > 0$ such that $\pi_n(\Omega \backslash \widetilde{\Omega}) \leq K'n^{-\epsilon'}$ for n sufficiently large. Then, since (4.10) and (4.12) give $\pi_n(\widetilde{\Omega} \backslash \widetilde{\Omega}_{\min}) = O(n^{-\zeta\delta_a\theta_n^{-1}})$, we deduce that $t_n, v_n \to 0$ as $n \to +\infty$ so that (4.14) leads to (4.13). ■

Note 4.3: What emerges from Corollary 4.2 is that the asymptotic convergence rate of constrained annealing is controlled by the critical constant δ_a (4.11) defined as the difference between the second smallest energy value in the feasible set $\widetilde{\Omega}$ and the constrained minimum. Clearly, the convergence rate is better for problems with larger δ_a and for a choice of ζ close to h_V^{-1} together with a slowly increasing control sequence $(\theta_n)_{n \in \mathbb{N}^*}$. It turns out that our improved upper bound on ζ leads to faster convergence provided that β_n is taken to be strictly increasing with respect to ζ for

n sufficiently large; the resulting convergence speed gain depends on the sequence $(\theta_n)_{n\in\mathbb{N}^*}$ and is up to $O(n^{\epsilon\theta_n^{-1}})$, $\epsilon = h_V^{-1} - (\ell\mathfrak{d}_V)^{-1}$.

References

1. Azencott, R. (1992). Sequential simulated annealing: Speed of convergence and acceleration techniques. In R. Azencott (ed.), *Simulated annealing: Parallelization techniques*. New York: Wiley, pp. 1–10.
2. Borgs, C., Chayes, J.T., Frieze, A., Kim, J.H., Tetali, P., Vigoda, E., & Vu, V.H. (1999). Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In *Proceedings of the 40th IEEE Symposium on the Foundations of Computer Science*, pp. 218–229.
3. Brémaud, P. (1999). *Markov chains: Gibbs fields, Monte Carlo simulations, and queues*. New York: Springer-Verlag.
4. Catoni, O. (1991). Applications of sharp large deviations estimates to optimal cooling schedules. *Annales de l'Institut H. Poincaré, Probabilités et Statistiques* 27(4): 463–518.
5. Catoni, O. (1991). Sharp large deviations estimates for simulated annealing algorithms. *Annales de l'Institut H. Poincaré, Probabilités et Statistiques* 27(3): 291–383.
6. Catoni, O. (1992). Rough large deviation estimates for simulated annealing: Application to exponential schedules. *The Annals of Probability* 20(3): 1109–1146.
7. Chiang, T.-S. & Chow, Y. (1988). On the convergence rate of annealing processes. *SIAM Journal on Control and Optimization* 26(6): 1455–1470.
8. Del Moral, P. & Miclo, L. (1999). On the convergence and applications of generalized simulated annealing. *SIAM Journal on Control and Optimization* 37(4): 1222–1250.
9. Desai, M.P. & Rao, V.B. (1993). On the convergence of reversible Markov chains. *SIAM Journal on Matrix Analysis and Applications* 14(4): 950–966.
10. Diaconis, P. & Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability* 1(1): 36–61.
11. Dyer, M., Goldberg, L.A., Greenhill, C., Jerrum, M., & Mitzenmacher, M. (2001). An extension of path coupling and its application to the Glauber dynamics for graph colorings. *SIAM Journal on Computing* 30(6): 1962–1975.
12. François, O. (2000). Geometric inequalities for the eigenvalues of concentrated Markov chains. *Journal of Applied Probability* 37: 15–28.
13. Frigerio, A. & Grillo, G. (1993). Simulated annealing with time-dependent energy function. *Mathematische Zeitschrift* 213: 97–116.
14. Geman, D. (1990). *Random fields and inverse problems in imaging*. Lecture Notes in Mathematics, Vol. 1427. Berlin: Springer-Verlag, pp. 117–193.
15. Geman, D. & Geman, S. (1987). Relaxation and annealing with constraints. Complex Systems Technical Report 35, Division of Applied Mathematics, Brown University, Providence, RI.
16. Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6): 721–741.
17. Gidas, B. (1995). Metropolis-type Monte Carlo simulation algorithms and simulated annealing. In J.L. Snell (ed.), *Topics in contemporary probability and its applications*. Probability and Stochastics Series. Boca Raton, FL: CRC Press, pp. 159–232.
18. Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematics of Operations Research* 13(2): 311–329.
19. Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1): 97–109.
20. Holley, R. & Stroock, D. (1988). Simulated annealing via Sobolev inequalities. *Communications in Mathematical Physics* 115: 553–569.
21. Ingrassia, S. (1993). Geometric approaches to the estimation of the spectral gap of reversible Markov chains. *Combinatorics, Probability and Computing* 2: 301–323.

22. Ingrassia, S. (1994). On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds. *The Annals of Applied Probability* 4(2): 347–389.
23. Jerrum, M. (1992). Large cliques elude the Metropolis process. *Random Structures and Algorithms* 3(4): 347–359.
24. Jerrum, M. (1995). A very simple algorithm for estimating the number of *k*-colorings of a low-degree graph. *Random Structures and Algorithms* 7(2): 157–165.
25. Jerrum, M. & Sinclair, A. (1989). Approximating the permanent. *SIAM Journal on Computing* 18(6): 1149–1178.
26. Jerrum, M. & Sinclair, A. (1996). The Markov chain Monte Carlo method: An approach to approximate counting and integration. In D.S. Hochbaum (ed.), *Approximation algorithms for NP-hard problems*. Boston: PWS, pp. 482–520.
27. Jerrum, M. & Sorkin, G.B. (1998). The Metropolis algorithm for graph bisection. *Discrete Applied Mathematics* 82: 155–175.
28. Kirkpatrick, S., Gelatt, C.D., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* 220(4598): 671–680.
29. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., & Teller, A.H. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6): 1087–1092.
30. Mitra, D., Sangiovanni-Vincentelli, F.R., & Sangiovanni-Vincentelli, A. (1985). Convergence and finite-time behavior of simulated annealing. In *Proceedings of the 24th Conference on Decision and Control*, pp. 761–767.
31. Peskun, P.H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* 60(3): 607–612.
32. Robini, M.C., Rastello, T., & Magnin, I.E. (1999). Simulated annealing, acceleration techniques and image restoration. *IEEE Transactions on Image Processing* 8(10): 1374–1387.
33. Sasaki, G.H. & Hajek, B. (1988). The time complexity of maximum matching by simulated annealing. *Journal of the Association for Computing Machinery* 35(2): 387–403.
34. Sinclair, A. (1992). *Improved bounds for mixing rates of Markov chains and multicommodity flow*. Lecture Notes in Computer Sciences, Vol. 583. Berlin: Springer-Verlag, pp. 474–487.
35. Sinclair, A. & Jerrum, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation* 82: 93–133.
36. Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* 22(4): 1701–1762.
37. Winkler, G. (1990). An ergodic L^2-theorem for simulated annealing in Bayesian image reconstruction. *Journal of Applied Probability* 28: 779–791.
38. Yao, J. (2000). On constrained simulation and optimization by Metropolis chains. *Statistics and Probability Letters* 46: 187–193.