J. Appl. Prob. **37**, 15–28 (2000)
Printed in Israel

GEOMETRIC INEQUALITIES FOR THE EIGENVALUES OF CONCENTRATED MARKOV CHAINS

OLIVIER FRANÇOIS,* *LMC/IMAG*

Abstract

This article describes new estimates for the second largest eigenvalue in absolute value of reversible and ergodic Markov chains on finite state spaces. These estimates apply when the stationary distribution assigns a probability higher than 0.702 to some given state of the chain. Geometric tools are used. The bounds mainly involve the isoperimetric constant of the chain, and hence generalize famous results obtained for the second eigenvalue. Comparison estimates are also established, using the isoperimetric constant of a reference chain. These results apply to the Metropolis–Hastings algorithm in order to solve minimization problems, when the probability of obtaining the solution from the algorithm can be chosen beforehand. For these dynamics, robust bounds are obtained at moderate levels of concentration.

Keywords: Reversible Markov chain; eigenvalues; isoperimetric constant; Metropolis–Hastings dynamics; minimization

AMS 1991 Subject Classification: Primary 60J10; 65C05
Secondary 15A42

1. Introduction

During the past few years, Markov chain Monte Carlo methods have proved to be useful in seeking the absolute minimum of a function H defined on a finite set E. In these methods, an ergodic reversible Markov chain is simulated, for which the stationary distribution tends to concentrate on the absolute minimum. However, Monte Carlo algorithms return approximate solutions, and the quality of the approximation depends on the number N of steps performed by the chain.

Most often, the stationary distribution is a Gibbs distribution, which depends on a positive parameter T called *temperature*:

$$\forall i \in E, \quad \pi_T(i) = \frac{\exp(-H(i)/T)}{Z_T}, \tag{1}$$

where

$$Z_T = \sum_{i \in E} \exp(-H(i)/T). \tag{2}$$

In this article, the function H is assumed to be non-negative and minimal at a unique point i_* in E such that $H(i_*) = 0$. Thus, one has

$$\pi_T(i_*) \to 1 \quad \text{as} \quad T \to 0. \tag{3}$$

Received 20 August 1997; revision received 13 April 1999.
* Postal address: LMC/IMAG, BP 53, 38041 Grenoble cedex 9, France. Email address: olivier.francois@imag.fr.

Hence, simulating π_T at low temperatures enables us to solve the minimization problem associated with H with high probability. Let α be a number in the interval $(0, 1)$, say $\alpha > \frac{1}{2}$, and assume that T can be fixed so that

$$\pi_T(i_*) \geq \alpha. \tag{4}$$

Denote by X_N the state which is returned at the end of the algorithm. For 'large' N, the probability $\mathrm{P}(X_N = i_*)$ is correctly approximated by $\pi_T(i_*)$. Then, one has

$$\mathrm{P}(X_N = i_*) \geq \alpha. \tag{5}$$

Therefore, the number α may be viewed as a level of the confidence that a user can have in the Monte Carlo minimization procedure.

In order to make the method efficient, the distribution of the Markov chain at the final step must be close to the stationary distribution π_T. Then, the main issue consists of finding the number N of steps needed to reach the stationary distribution.

It is well known that the rate of convergence towards the stationary distribution is controlled by the spectral gap of the chain (the second largest eigenvalue in absolute value). This article gives new estimates on the spectral gap of a reversible Markov chain under the hypothesis that Equation (4) is satisfied for α greater than an explicit value (close to 0.701, see Section 2). The approach is geometric. Intuitively, fast convergence is expected when the chain moves quickly to the subsets having large probability under the stationary distribution. This situation corresponds to large values of geometric quantities called *isoperimetric constants*. References [16, 21] shed light on the role played by these quantities on the control of the second eigenvalue of the chain. This article emphasizes the control which is exerted on the whole spectrum by the isoperimetric constant. It is organized as follows. The main results are stated in Section 2. The proofs are given in Section 3. Among Markov chain Monte Carlo methods, Metropolis–Hastings dynamics are very popular [12, 18]. In Section 4, our bounds are applied to the Metropolis–Hastings dynamics, yielding robust convergence estimates. The paper is concluded by a short discussion, in which the results are compared to those obtained in [15].

2. Background and presentation of the main results

2.1. Previous results

This article considers reversible ergodic Markov chains defined on the finite set

$$E = \{1, \dots, n\}, \quad n \geq 2. \tag{6}$$

Let π denote the (common) stationary distribution of these chains. The subscript T is used when $\pi = \pi_T$ is a Gibbs distribution at temperature T. The transition matrix is denoted by $P = (p(i, j))_{i,j=1\dots n}$, or P_T when the stationary distribution is π_T. Also denote

$$\forall i, j \in E, \quad a(i, j) = \pi(i)p(i, j). \tag{7}$$

Reversibility induces that

$$\forall i, j \in E, \quad a(i, j) = a(j, i). \tag{8}$$

The eigenvalues of P are real, and can be ordered as follows

$$-1 \leq \lambda_n \leq \cdots \leq \lambda_2 < \lambda_1 = 1. \tag{9}$$

Denote

$$\rho(P) = \max\{|\lambda_n|, \lambda_2\}. \tag{10}$$

It is well known that $\rho(P)$ controls the convergence of the chain towards its stationary distribution. For instance, such a control is expressed in the following result (see, for example, [6, 5, 14]):

$$\forall i \in E, \quad ||P^k(i,.) - \pi(.)||_{TV}^2 \leq \frac{1-\pi(i)}{4\pi(i)}\rho(P)^{2k}, \tag{11}$$

where $P^k(i,.)$ is the conditional distribution of the chain at step k starting from $i \in E$ and $||\,.\,||_{TV}$ denotes the total variation norm.

The *isoperimetric constant* of a subset $S \subset E$ is defined as

$$\phi(S) = \frac{\sum_{i\in S,\, j\notin S} a(i,j)}{\pi(S)}, \tag{12}$$

where $\pi(S) = \sum_{i\in S}\pi(i)$. This quantity represents the conditional probability under stationarity that the chain exits from the set S in a single step given that it starts in S. The *symmetric isoperimetric constant* of S is defined as

$$\bar{\phi}(S) = \frac{\phi(S)}{1-\pi(S)}. \tag{13}$$

The *global symmetric isoperimetric constant* is equal to

$$\bar{\phi} = \min_{S\subset E} \bar{\phi}(S), \tag{14}$$

where the minimum runs over the proper subsets of E ($S \neq \varnothing$ and $S \neq E$). In the literature about Markov chains, the most studied eigenvalue is λ_2. Lawler and Sokal [16] and Sinclair and Jerrum [21] have given a bound on the second eigenvalue in terms of the isoperimetric constant. Lawler and Sokal's bound is as follows

$$\lambda_2 \leq 1 - \frac{\bar{\phi}^2}{8}. \tag{15}$$

This bound parallels a previous result of Cheeger in Riemannian geometry [3]. To apply this result to convergence issues, many authors consider the modified transition matrix

$$P' = \theta I + (1-\theta)P, \quad 0 \leq \theta < 1 \tag{16}$$

which is still reversible. If θ is carefully chosen, the eigenvalues of P' are non-negative, and $\rho(P') = \lambda_2$. To avoid the computation of λ_n, some authors recommend $\theta = \frac{1}{2}$. However, this choice seems inefficient with regard to practical implementations (the dynamics slow down). A clever choice of θ may demand some knowledge about λ_n. Although it might be interesting to deal with this issue, we will not develop it further in this paper.

Diaconis and Stroock [6] should be mentioned here. This reference gives nice and useful bounds on the second and last eigenvalues by using geometric methods. Their technique has been inspired by Poincaré's inequalities (see for example [19]). However, a drawback of their approach is that two different bounds must be compared in order to obtain an inequality for $\rho(P)$. The comparison often involves several geometric quantities which are difficult to estimate (see [15]). The aim of the present paper is to use the isoperimetric constant to bound $|\lambda_n|$ as well as λ_2, in order to avoid the heavy computations sometimes required by Poincaré's inequalities.

In [7], a result related to those presented in this article has been proved. Assuming that there exists an $i_* \in E$ such that $\pi(i_*) > \frac{1}{2}$, one has

$$\rho(P)^2 \leq 1 - \left(2 - \frac{1}{\pi(i_*)}\right)^2 \phi^2. \tag{17}$$

In this equation, the constant ϕ has been defined as

$$\phi = \min_{S \subset E} \phi(S) \tag{18}$$

where the minimum runs over proper subsets of E. However, [7] also emphasizes that the bound is not accurate, even when $\pi(i_*)$ is close to 1. Its application to the Metropolis–Hastings dynamics leads to very rough estimates for the convergence rate. In addition, it has been observed that $\bar{\phi}$ would yield better estimates.

2.2. Main results

Our first result can be stated as follows. Let

$$v_* = (19/27 + \sqrt{33}/9)^{1/3} + \frac{4}{9}\left(\frac{1}{19/27 + \sqrt{33}/9}\right)^{1/3} - \frac{2}{3} \tag{19}$$

and let $i_* \in E$ be such that

$$\pi(i_*) > v_*^2. \tag{20}$$

Let $\lambda < 1$ be an eigenvalue of P. Then the following bound holds (Theorem 3.1):

$$\lambda^2 \leq 1 - K_*^2 \bar{\phi}^2, \tag{21}$$

with

$$K_* = 1 - \sqrt{1 - \pi(i_*)}(1 + \sqrt{\pi(i_*)}). \tag{22}$$

This result can be easily translated into a bound on the convergence rate of the chain towards its stationary distribution (Corollary 3.1). Note that the value v_*^2 is approximately equal to 0.701, and hence defines a suitable level of confidence for implementing a Monte Carlo minimization method.

The second result presented in this paper generalizes the former to the comparison of two Markov chains. A generic idea in comparison is to bound the eigenvalues of a chain by using the isoperimetric constant of another chain for which explicit calculations can be done. Let P_2 denote the transition matrix of the reference chain while P_1 denotes the transition matrix of the chain of interest. Let $i_* \in E$ be as in Equation (20) and $\lambda < 1$ be an eigenvalue of P_1. Then, we obtain the following bound (Theorem 3.2):

$$\lambda^2 \leq 1 - K_*^2 \bar{\phi}_2^2 / A^2, \tag{23}$$

where $\bar{\phi}_2$ is the isoperimetric constant associated with the transition matrix P_2. In this bound, A is a new geometric constant computed from P_1, P_2 and a fixed subset of finite trajectories for P_1.

The precise statement and proof of Theorems 3.1 and 3.2 will be given in Section 3. Theorem 3.1 is inspired by Cheeger's result, whereas Theorem 3.2 uses both Cheeger's and Poincaré's techniques.

Theorems 3.1 and 3.2 are appropriate to study the *Metropolis–Hastings dynamics*. These dynamics are frequently used in the context of minimization by simulated annealing [2, 11, 15, 5]. They correspond to the reversible transition matrix P_T which is defined as follows. Let Q be an arbitrary symmetric aperiodic and irreducible transition matrix on E, such that $q(i,i) = 0$ for all $i \in E$. Let

$$\forall i \neq j, \quad p_T(i,j) = \begin{cases} q(i,j) F\left(\dfrac{\pi_T(i)}{\pi_T(j)}\right) & \text{if } \pi_T(j) \geq \pi_T(i) \\[2ex] q(i,j) \dfrac{\pi_T(j)}{\pi_T(i)} F\left(\dfrac{\pi_T(j)}{\pi_T(i)}\right) & \text{otherwise,} \end{cases} \tag{24}$$

and

$$p_T(i,i) = 1 - \sum_{j \neq i} p_T(i,j), \tag{25}$$

where F is an arbitrary function such that $0 < F(x) \leq 1$ for $0 < x \leq 1$. (The standard Metropolis dynamics is obtained for $F \equiv 1$.) Let α be such that

$$v_*^2 < \alpha < 1, \tag{26}$$

and

$$T \leq \frac{\min_{i \neq i_*} H(i)}{\log(\alpha(n-1)/(1-\alpha))}. \tag{27}$$

Our third result involves a constant m called the *least total elevation gain* [13] (a definition will be given in Section 4). Let $\lambda_T < 1$ be an eigenvalue of the transition matrix P_T. Then, we have (Theorem 4.1)

$$\lambda_T^2 \leq 1 - K^2 \mathrm{e}^{-2m/T}, \tag{28}$$

with

$$K = \frac{1 - \sqrt{1-\alpha}(1 + \sqrt{\alpha})}{\alpha b_\Gamma} \min_x F(x) \min_{(k \to \ell)} q(k,\ell)$$

where b_Γ is a geometric constant to be defined in Section 4, $\min_x F(x)$ runs over all possible values of the ratio $x = \pi_T(j)/\pi_T(i)$, and $(k \to \ell)$ is an edge of the transition graph.

In comparison with the results obtained from Poincaré's inequalities [15], the main advantage of this approach is its robustness with respect to

$$\delta = \min_{\substack{(i \to j); \\ H(i) \neq H(j)}} |H(i) - H(j)|. \tag{29}$$

When T is fixed and δ is small, the estimates given by [15] may become very inaccurate. In contrast, the previous bound may be applied with small regard to the constant δ.

3. Geometric inequalities for the eigenvalues of a reversible Markov chain

3.1. Cheeger-like estimates

As usual, the space of functions defined on E is endowed with a Hilbertian structure. This space is denoted by $L^2(\pi)$ and

$$\forall f, g \in L^2(\pi), \quad \langle f, g\rangle = \sum_{i=1}^{n} \pi(i) f(i) g(i). \tag{30}$$

The transition matrix P is reversible if and only if P is self-adjoint as an operator of $L^2(\pi)$. In this section, the following result is proved.

Theorem 3.1. *Let P be the transition matrix of a reversible Markov chain on E and $i_* \in E$ be such that $\pi(i_*) > v_*^2$, where v_* is given by Equation (19). Let $\lambda < 1$ be an eigenvalue of P. Then we have*

$$1 - \lambda^2 \geq K_*^2 \bar{\phi}^2, \tag{31}$$

where $K_* = 1 - \sqrt{1 - \pi(i_*)}(1 + \sqrt{\pi(i_*)})$.

Corollary 3.1. *Let P be the transition matrix of an ergodic reversible Markov chain on E and $i_* \in E$ be such that $\pi(i_*) > v_*^2$. Then we have*

$$\forall i \in E, \quad ||P^k(i, .) - \pi(.)||_{TV}^2 \leq \frac{1 - \pi(i)}{4\pi(i)}(1 - K_*^2 \bar{\phi}^2)^k. \tag{32}$$

Proof. This is an obvious consequence of Theorem 3.1 and Equation (11).

The proof of Theorem 3.1 starts with a lemma on the eigenfunctions of the chain.

Lemma 3.1. *Let P be the transition matrix of a time-reversible Markov chain. Let $\lambda < 1$ be an eigenvalue of P and f an associated eigenfunction satisfying $\langle f, f\rangle = 1$. Then we have, for all $i \in E$,*

$$\sum_{j=1}^{n} \pi(j)|f(j)| \leq \sqrt{1 - \pi(i)}(1 + \sqrt{\pi(i)}). \tag{33}$$

Proof. First, f is orthogonal to the constant 1. Thus,

$$\pi(i) f(i) = -\sum_{j \neq i} \pi(j) f(j). \tag{34}$$

By elevating to the square and applying Cauchy–Schwarz's inequality, we obtain

$$\begin{aligned} f^2(i) = \frac{1}{\pi^2(i)}\Big(\sum_{j\neq i} \pi(j) f(j)\Big)^2 &\leq \frac{1}{\pi^2(i)}\Big(\sum_{j\neq i} \pi(j) f^2(j)\Big)\Big(\sum_{j\neq i} \pi(j)\Big) \\ &= \frac{1 - \pi(i)}{\pi^2(i)}(1 - \pi(i) f^2(i)). \end{aligned} \tag{35}$$

Therefore

$$f^2(i) \leq \frac{1}{\pi(i)} - 1, \quad \forall i \in E. \tag{36}$$

Again making use of the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\sum_{j=1}^{n} \pi(j)|f(j)| &= \sum_{j\neq i}^{n} \pi(j)|f(j)| + \pi(i)|f(i)| \\ &\leq \Big(\sum_{j\neq i}^{n} \pi(j)\Big)^{1/2} \Big(\sum_{j\neq i}^{n} \pi(j) f^2(j)\Big)^{1/2} + \pi(i)|f(i)|. \end{aligned}\tag{37}$$

To conclude, it follows from Equation (36) that

$$\begin{aligned}\sum_{j=1}^{n} \pi(j)|f(j)| &\leq \sqrt{1-\pi(i)} + \pi(i)\sqrt{\frac{1}{\pi(i)} - 1} \\ &= \sqrt{1-\pi(i)}(1+\sqrt{\pi(i)}\,). \end{aligned}\tag{38}$$

A lemma from probability theory is also needed.

Lemma 3.2. *Let X and Y be two i.i.d. random variables such that $E[X] = 0$ and $\mathrm{var}(X) = 1$. Then*

$$E[|X^2 - Y^2|] \geq 2(1 - E[|X|]). \tag{39}$$

Proof. See [16], Proposition 2.2 (Equation (2.28)), p. 563.

The proof of Theorem 3.1 can now be given.

Proof of Theorem 3.1. Let f be an eigenfunction associated with the eigenvalue $\lambda < 1$ and satisfying $\langle f, f\rangle = 1$. Reordering the elements of E, we assume that

$$f^2(1) \leq f^2(2) \leq \cdots \leq f^2(n). \tag{40}$$

For all real $t \geq 0$, we denote

$$U_f(t) = \sum_{i<j} a(i,j)\mathbf{1}_{(f^2(i), f^2(j))}(t) \tag{41}$$

and

$$V_f(t) = \sum_{i<j} \pi(i)\pi(j)\mathbf{1}_{(f^2(i), f^2(j))}(t), \tag{42}$$

where $\mathbf{1}_{(a,b)}$ denotes the indicator function of the interval (a, b), $a \leq b$. According to [19], the Dirichlet form associated with P is given by

$$\mathcal{E}(f, f) = \sum_{i<j} a(i,j)(f(i) - f(j))^2. \tag{43}$$

By Cauchy–Schwarz's inequality, we obtain

$$\left(\sum_{i<j} a(i,j)|f^2(i) - f^2(j)|\right)^2 \le \left(\frac{1}{2}\sum_{i,j=1}^{n} a(i,j)(f(i)+f(j))^2\right)\mathcal{E}(f,f)$$
$$= 1 - \lambda^2. \tag{44}$$

On the left-hand side, we have

$$\begin{aligned}\sum_{i<j} a(i,j)|f^2(i) - f^2(j)| &= \int_0^{f^2(n)} U_f(t)\,\mathrm{d}t \\ &= \int_{f^2(1)}^{f^2(n)} \frac{U_f(t)}{V_f(t)} V_f(t)\,\mathrm{d}t \\ &\ge \bar{\phi} \sum_{i<j} \pi(i)\pi(j)(f^2(j) - f^2(i)) \\ &= \frac{\bar{\phi}}{2}\sum_{i,j=1}^{n} \pi(i)\pi(j)|f^2(j) - f^2(i)|. \end{aligned} \tag{45}$$

Applying Lemma 3.2 yields

$$\sum_{i<j} a(i,j)|f^2(i) - f^2(j)| \ge \bar{\phi}\left(1 - \sum_{j=1}^{n} \pi(j)|f(j)|\right). \tag{46}$$

Before applying Lemma 3.1, note that

$$\sqrt{1-\pi(i_*)}(1+\sqrt{\pi(i_*)}) < 1 \tag{47}$$

if and only if

$$\pi(i_*) > v_*^2. \tag{48}$$

Finally,

$$\sqrt{1-\lambda^2} \ge \bar{\phi}(1 - \sqrt{1-\pi(i_*)}(1+\sqrt{\pi(i_*)})). \tag{49}$$

It can be seen from the special case of two-state transition matrices that the bound given in Theorem 3.1 is not sharp, but differs from the true result by the power two. For instance, consider the Metropolis independence base chain, whose transition matrix M is given by

$$\forall i \ne j, m(i,j) = \begin{cases} 1/n & \text{if} \quad \pi(i) < \pi(j), \\ \pi(j)/n\pi(i) & \text{otherwise} \end{cases} \tag{50}$$

$$m(i,i) = 1 - \sum_{j \ne i} m(i,j), \tag{51}$$

where we assume that $\pi(1) > 0.702$. This chain is a particular case of the Metropolis algorithm, and it will play a significant role in the proof of our third result. The spectrum of this chain can be entirely described [5]. The constant $\rho(P)$ is given by

$$\rho(P) = 1 - \frac{1}{n\pi(1)}. \tag{52}$$

The isoperimetric constant $\bar{\phi}$ is close to $1/n$ if $\pi(1)$ is close to 1. The bound obtained in Theorem 3.1 behaves as $(1 - 1/n^2)^{1/2}$. This does not give the correct order of convergence rate. However, this drawback is well known in the Cheeger-like approach. A more significant example will be studied in Section 4. In this perspective, the comparison results presented in the next section will be useful.

3.2. Comparison of two chains

In this section, the Cheeger-like approach is merged with Poincaré's method used in [6, 4], in order to give new estimates for the eigenvalues of reversible Markov chains. In what follows, P_2 denotes the transition matrix of a reference chain while P_1 denotes the transition matrix of the chain of interest. Both chains are reversible with respect to the same probability distribution π. For all $i, j \in E$, denote by γ_{ij} a path of P_1 from i to j without loops (irreducibility warrants the existence of such a path), that is, a sequence $i_0 = i, i_k, \ldots, i_r = j$ such that $p_1(i_k, i_{k+1}) > 0$ and $i_k \notin \{i_0, \ldots, i_{k-1}\}$. An edge of the transition graph between k and ℓ, $k \neq \ell$ is denoted by $(k \to \ell)$, still with respect to P_1. Let Γ be an arbitrary set of paths of P_1 consisting of one path γ_{ij} for all $i < j$ in E. Assume that Γ is symmetric. If $\gamma_{ij} \in \Gamma$, then Γ also contains γ_{ji}, which is obtained by reversing the sequence γ_{ij}. The following result is proved.

Theorem 3.2. *Let P_1, P_2 be the transition matrices of reversible Markov chains on E and let $i_* \in E$ be such that $\pi(i_*) > v_*^2$, where v_* is as in Equation (19). Let $\lambda < 1$ be an eigenvalue of P_1. Then we have*

$$1 - \lambda^2 \geq K_*^2 \bar{\phi_2}^2 / A^2 \tag{53}$$

where

$$A = \max_{(k \to \ell)} \sum_{\gamma_{ij} \ni (k \to \ell)} \frac{\pi(i) p_2(i, j)}{\pi(k) p_1(k, \ell)} \tag{54}$$

and $\bar{\phi_2}$ is the isoperimetric constant associated with the transition matrix P_2.

Proof. Let f be an eigenfunction of P_1 associated with the eigenvalue $\lambda < 1$ and satisfying $\langle f, f \rangle = 1$. Reordering the elements of E as before, we assume that

$$f^2(1) \leq f^2(2) \leq \cdots \leq f^2(n). \tag{55}$$

Denote

$$\forall i, j \in E, \quad a_s(i, j) = \pi(i) p_s(i, j), \quad s = 1, 2. \tag{56}$$

Following the same lines as in the proof of Theorem 3.1, we obtain

$$\bar{\phi_2}(1 - \sqrt{1 - \pi(i_*)}(1 + \sqrt{\pi(i_*)})) \leq \sum_{i<j} a_2(i, j) |f^2(i) - f^2(j)|. \tag{57}$$

Introducing the set of paths Γ, and using the triangle inequality

$$\begin{aligned}\sum_{i<j} a_2(i,j)|f^2(i) - f^2(j)| &= \sum_{i<j} a_2(i,j)\left|\sum_{(k\to\ell)\in\gamma_{ij}} f^2(\ell) - f^2(k)\right| \\ &\leq \sum_{i<j} a_2(i,j) \sum_{(k\to\ell)\in\gamma_{ij}} \frac{a_1(k,\ell)}{a_1(k,\ell)}|f^2(\ell) - f^2(k)| \\ &\leq \frac{1}{2}\sum_{k,\ell}\left(\sum_{\gamma_{ij}\ni(k\to\ell)} \frac{a_2(i,j)}{a_1(k,\ell)}\right) a_1(k,\ell)|f^2(\ell) - f^2(k)| \\ &\leq A(1-\lambda^2)^{1/2}. \end{aligned} \tag{58}$$

An obvious way to obtain a new bound on the spectrum of the chain P_1 is to choose

$$\forall i, j \in E, \quad p_2(i,j) = \pi(j). \tag{59}$$

This gives rise to the following result.

Theorem 3.3. *Let P be the transition matrix of a reversible Markov chain on E and let $i_* \in E$ be such that $\pi(i_*) > v_*^2$. Then we have*

$$\rho(P)^2 \leq 1 - K_*^2/\eta^2, \tag{60}$$

where

$$\eta = \max_{(k\to\ell)} \sum_{\gamma_{ij}\ni(k\to\ell)} \frac{\pi(i)\pi(j)}{\pi(k)p(k,\ell)}. \tag{61}$$

Proof. Check that $A = \eta$ and $\bar{\phi}_2 = 1$ when P_2 is defined as in Equation (59).

Since

$$\frac{1}{\eta} \leq \bar{\phi}, \tag{62}$$

the previous bound is always less accurate than (31). However, η may be easier to compute than ϕ, and such a bound may sometimes be useful.

4. Application to the Metropolis–Hastings dynamics

This section provides new results for the Metropolis–Hastings dynamics, built upon the geometric inequalities established in Section 3. The transition matrix $P = P_T$ is defined as in Equation (25) and $\rho_T = \rho(P_T)$. The graph associated with P_T has E as the set of vertices, and the set of edges is given by the pairs $\{i, j\}$ such that $q(i,j) > 0$.

The Metropolis chain has been studied in great detail [15, 5, 10, 9]. At low temperatures, Ingrassia [15] has applied Poincaré's inequalities, following the method initiated by [13]. In these references, the convergence rate of the Metropolis dynamics is expressed in terms of a parameter m called the *least total elevation gain.* For each path γ_{ij} between i and j (with respect to P_T), the *elevation* is defined as

$$\mathrm{elev}(\gamma_{ij}) = \max_k \{H(i_k)\}, \tag{63}$$

where the maximum runs over all vertices in γ_{ij}. Let H_{ij} be the lowest possible elevation between i and j over all self-avoiding paths γ_{ij} from i to j. Then, the least total elevation gain is

$$m = \max_{i,j\in E} \{H_{ij} - H(i) - H(j)\}. \tag{64}$$

Asymptotic results for Markov chains with rare transitions [8, 1] show that

$$\rho_T = \lambda_2(T) \tag{65}$$

for small T, and

$$1 - \lambda_2(T) \sim C\mathrm{e}^{-m/T} \quad \text{as } T \to 0, \tag{66}$$

where C is a positive constant. The results obtained in this section necessitate the use of the Comparison Theorem 3.2. To apply this theorem, the set of admissible paths Γ is defined as follows. As in [13], γ_{ij} is an admissible path ($\gamma_{ij} \in \Gamma$) if

$$\mathrm{elev}(\gamma_{ij}) - H(i) - H(j) \leq m. \tag{67}$$

The maximal number of paths which contain a fixed edge e is denoted by

$$b_\Gamma = \max_{e=(k\to\ell)} \#\{\gamma \in \Gamma \text{ s.t. } \gamma \ni e\}. \tag{68}$$

To start with, a lower bound on ρ_T can be given at low temperatures. According to the variational formula for eigenvalues, one has, for each proper subset $S \subset E$,

$$\rho_T \geq 1 - \bar{\phi}(S). \tag{69}$$

For a proof of this result, see for example [6]. According to Equation (69), a lower bound on ρ_T can be obtained by considering the subset $S \subset E$ defined below (actually, the same as [15], p. 357). Let i_0 and j_0 be two elements in E such that

$$H_{i_0 j_0} - H(i_0) - H(j_0) = m \tag{70}$$

and $H(i_0) \leq H(j_0)$. Consider the subset

$$S = \{i \in E \text{ s.t. } H_{i_0 i} < H_{i_0 j_0}\}. \tag{71}$$

Then, $i_0 \in S$ and $j_0 \notin S$. When T is small enough, one has

$$\rho_T \geq 1 - \max_{(k\to\ell)} q(k,\ell) \max_x F(x) \frac{n^2}{4}\mathrm{e}^{-m/T}, \tag{72}$$

where $x = \pi_T(j)/\pi_T(i)$, $i, j \in E$.

As upper estimates are concerned, the next result establishes a robust bound on the spectral gap.

Theorem 4.1. *Let* $1 \geq \alpha > v_*^2$, *and*

$$T \leq \frac{\min_{i \neq i_*} H(i)}{\log(\alpha(n-1)/(1-\alpha))}. \tag{73}$$

Then the constant ρ_T *associated with the transition matrix* P_T *given in Equation (25) satisfies*

$$\rho_T^2 \leq 1 - K^2 e^{-2m/T}, \tag{74}$$

with

$$K = \frac{1 - \sqrt{1-\alpha}(1+\sqrt{\alpha})}{\alpha b_\Gamma} \min_x F(x) \min_{(k \to \ell)} q(k, \ell), \tag{75}$$

where $\min_x F(x)$ *runs over the values of the ratio* $x = \pi_T(j)/\pi_T(i)$.

Proof. In general, the symmetric isoperimetric constant $\bar{\phi}$ cannot be evaluated explicitly. The bound is obtained by comparison with another chain (the independence base chain) according to Theorem 3.2. First of all, let

$$T \leq \frac{\min_{i \neq i_*} H(i)}{\log(\alpha(n-1)/(1-\alpha))}. \tag{76}$$

Then

$$\log((1-\alpha)/\alpha(n-1)) \geq -\frac{\min_{i \neq i_*} H(i)}{T}. \tag{77}$$

Hence we have

$$\frac{1}{\alpha} \geq 1 + (n-1)\exp\left(-\frac{\min_{i \neq i_*} H(i)}{T}\right) \geq Z_T \tag{78}$$

and

$$\pi_T(i_*) > v_*^2. \tag{79}$$

To apply Theorem 3.2, we choose as reference the Metropolis independence base chain M_T. Thus, we need to estimate the quantity

$$\frac{1}{\pi_T(k) p_T(k, \ell)} \sum_{\gamma_{ij} \ni (k \to \ell)} \pi_T(i) m_T(i, j). \tag{80}$$

For all $k \neq \ell$, one has

$$\pi_T(k) p_T(k, \ell) = q(k, \ell) \min\{\pi_T(k), \pi_T(\ell)\} F\left(\frac{\min\{\pi_T(k), \pi_T(\ell)\}}{\max\{\pi_T(k), \pi_T(\ell)\}}\right). \tag{81}$$

Letting $c = \max_{(k \to \ell)} q(k, \ell)^{-1}$,

$$A \leq \frac{c}{n} \max_{(k \to \ell)} \sum_{\gamma_{ij} \ni (k \to \ell)} \frac{\min\{\pi_T(i), \pi_T(j)\}}{\min\{\pi_T(k), \pi_T(\ell)\}} \left[F\left(\frac{\min\{\pi_T(k), \pi_T(\ell)\}}{\max\{\pi_T(k), \pi_T(\ell)\}}\right)\right]^{-1} \tag{82}$$

$$\leq c' \exp(m/T), \tag{83}$$

where

$$c' = b_\Gamma (n \min_x F(x) \min_{(k \to \ell)} q(k, \ell))^{-1}. \tag{84}$$

According to Equations (52) and (69), one has

$$\bar{\phi}_{M_T} \geq 1 - \rho(M_T) = \frac{1}{n\pi_T(i_*)}. \tag{85}$$

The result is obtained by applying Theorem 3.2.

5. Discussion

This section discusses the meaning of the confidence level introduced in Equation (4), and provides comparisons with other results concerning the Metropolis dynamics [15].

In this article, the main results have been established under the assumption that Equation (19) is satisfied. When π_T is a Gibbs distribution, it is crucial to determine the temperature below which this condition holds. The condition (27) is universal, and corresponds to the worst situation ($H(i) = h > 0$, for $i \neq i_*$). The upper bound in (27) may be improved when specific minimization problems are considered, but this issue is beyond the scope of this paper.

In general, the confidence level α can be chosen higher than 0.701. It does not seem necessary to choose this level very high (although a reflex would induce $\alpha = 0.95$). Indeed, it is well established that ergodic Markov chains satisfy large deviation bounds of Chernoff's type for the probability

$$\mathrm{P}_\pi\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \pi(f) > t\right),$$

where f is a function defined on E, and X_k is the state of the chain at step k. The results in [17, 19] give quantitative bounds for this probability. According to these results, the number of visits to i_* before step N can be estimated under stationarity by $N\pi(i_*)$, to the extent that N is large compared to $(1 - \lambda_2)^{-1}$. In this situation, a user can easily identify the absolute minimum. In view of specific applications (such as hard combinatorial problems), it would be useful to obtain analogous results at lower confidence levels (e.g. 0.5), but this issue deserves further work.

Next we turn to the comparison with the results obtained in [15]. Theorem 4.1 gives a convergence rate towards equilibrium which is roughly $1 - K'\mathrm{e}^{-2m/T}$ (for some $K' > 0$) as T goes to 0. However, the true order is $1 - C\mathrm{e}^{-m/T}$, for some $C > 0$. In this perspective, the results established by [15] regarding the Metropolis dynamics are more powerful. But, the approach of [15] has a weakness. Recall that

$$\delta = \min_{\substack{(i \to j);\\ H(i) \neq H(j)}} |H(i) - H(j)|. \tag{86}$$

In [15], the temperature below which the inequality

$$\rho_T \leq 1 - K''\mathrm{e}^{-m/T} \tag{87}$$

holds is

$$T_* = \min\left[m\left(\log\frac{n(A+B)}{2b_\Gamma\gamma_\Gamma}\right)^{-1}, \frac{\delta}{\log 2}\right]. \tag{88}$$

The definitions of the quantities γ_Γ, A and B are quite long. The interested reader may refer to [15], p. 359. For instance, let δ be very small and $T \geq T_*$. Then inequality (87) cannot be used, and the estimation of ρ_T obtained by [15] may be inaccurate.

In contrast, the constant δ does not significantly affect the probability $\pi_T(i_*)$. Applying Theorem 4.1 to the Metropolis chain leads to the following bound ($\alpha > v_*^2$):

$$\rho_T^2 \leq 1 - \left(q_* \frac{1 - \sqrt{1-\alpha}(1+\sqrt{\alpha})}{\alpha b_\Gamma} \right)^2 \mathrm{e}^{-2m/T}, \tag{89}$$

with $q_* = \min_{(k \to \ell)} q(k, \ell)$. Clearly, this bound improves on [15] when δ is small. In addition, this inequality avoids the computation of γ_Γ, A and B, and is thus simpler than [15].

Acknowledgement

I am grateful to Jean Diebolt for his careful reading of the manuscript. This work is dedicated to Hubert Pesce.

References

[1] AZENCOTT, R. (1988). Simulated annealing. *Sém. Bourbaki* **697**, 161–175.
[2] BERTSIMAS, D. AND TSITSIKLIS, J. (1993). Simulated annealing. *Statist. Sci.* **8**, 10–15.
[3] CHEEGER, J. (1970). A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in Analysis*, ed. R. C. Gunning. Princeton University Press, Princeton, NJ, pp. 195–199.
[4] DIACONIS, P. AND SALOFF-COSTE, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Prob.* **3**, 696–730.
[5] DIACONIS, P. AND SALOFF-COSTE, L. (1998). What do we know about the Metropolis algorithm? *J. Comp. Sci. Syst.* **57**, 20–36.
[6] DIACONIS, P. AND STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1**, 36–61.
[7] FRANÇOIS, O. (1999). On the spectral gap of a time reversible Markov chain. *Prob. Eng. Inform. Sci.* **13**, 95–101.
[8] FREIDLIN, M. I. AND WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.
[9] FRIGESSI, A., MARTINELLI, F. AND STANDER, J. (1997). Computational complexity of Markov chain Monte Carlo methods. *Biometrika* **84**, 1–18.
[10] 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. B* **55**, 205–220.
[11] HAJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Operat. Res.* **13**, 311–329.
[12] HASTINGS, W. K. (1970). Monte Carlo sampling using Markov chains and their application. *Biometrika* **57**, 77–109.
[13] HOLLEY, R. AND STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Commun. Math. Phys.* **115**, 553–569.
[14] INGRASSIA, S. (1993). Geometric approaches to the estimation of the spectral gap of time reversible Markov chains. *Combin. Prob. Comput.* **2**, 301–323.
[15] INGRASSIA, S. (1994). On the rate of convergence of the Metropolis algorithm and the Gibbs sampler by geometric bounds. *Ann. Appl. Prob.* **4**, 347–389.
[16] LAWLER, G. F. AND SOKAL, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309**, 557–580.
[17] LEZAUD, P. (1996). Chernoff bounds for finite Markov chains. Ph.D. Thesis, University of Toulouse.
[18] METROPOLIS, M., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A. AND TELLER, E. (1953). Equation of state calculation by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.
[19] SALOFF-COSTE, L. (1997). *Lectures on Finite Markov Chains*. Lect. Notes Math. (Saint-Flour) **1665**, Springer, Berlin, pp. 301–413.
[20] SINCLAIR, A. (1991). Improved bounds for mixing rates of Markov chains and multicommodity flows. *Combin. Prob. Comput.* **1**, 351–370.
[21] SINCLAIR, A. AND JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82**, 93–133.