

ABOUT RELAXATION TIME OF FINITE GENERALIZED METROPOLIS ALGORITHMS

BY L. MICLO

Instituto Nacional de Matemática Pura e Aplicada

In 1999 Catoni determined the critical rate H_3 for the relaxation time of generalized Metropolis algorithms, models for which the speed of convergence to equilibrium can be strongly influenced by the effects of a possible almost periodicity. We recover this result with the help of Dobrushin's coefficient and give characterizations of H_3 in terms of other ergodic constants. In particular, we prove that it also governs the large deviation behavior of the singular gap for a sufficiently large but finite number of iterations of the underlying kernel at low temperature.

1. Introduction. In recent lecture notes [2], Catoni considered a new critical exponent H_3 for finite generalized Metropolis algorithms. To present his result, we begin by introducing its setting and recalling some elementary large deviation facts.

On a finite set S containing at least two points, we are given a family $(P_\beta)_{\beta \geq 0}$ of $S \times S$ stochastic matrices, indexed by a nonnegative real parameter β which plays the role of an inverse temperature. We assume that the entries of these matrices satisfy large deviation principles at low temperature, which means that the following limits exist:

$$(1) \quad \forall x, y \in S, \quad V(x, y) \stackrel{\text{def.}}{=} - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(P_\beta(x, y)).$$

A priori these quantities belong to $\bar{\mathbb{R}}_+ \stackrel{\text{def.}}{=} \mathbb{R}_+ \sqcup \{+\infty\}$, but we furthermore make the hypothesis that the rate matrix $V \stackrel{\text{def.}}{=} (V(x, y))_{x, y \in S}$ is irreducible, in the sense that for all $x \neq y \in S$ there exists a finite sequence $x = x_0, x_1, \dots, x_n = y$ of elements of S such that, for any $0 \leq i < n$, $V(x_i, x_{i+1}) < +\infty$. Consequently, for $\beta \geq 0$ large enough, the Markov kernel P_β will also be irreducible, in the traditional meaning. In fact, up to a shift in the indexation by $\beta \geq 0$, there is no loss of generality in assuming (as we will do from now on) that we are always assured of

$$\forall x, y \in S, \quad P_\beta(x, y) > 0 \iff V(x, y) < +\infty$$

Received February 2001; revised January 2002.

AMS 2000 subject classifications. Primary 60J10; secondary 65C40, 49K45, 37A25, 15A18.

Key words and phrases. Generalized Metropolis algorithm at low temperature, critical rate for relaxation times, Dobrushin's coefficient and coupling, spectral gaps and singular values, classical or modified logarithmic Sobolev inequalities, delaying effect for ergodic constants, simulated annealing.

because we will only be interested in asymptotical behavior for β large. As is well known, this ensures that for any $\beta \geq 0$ there is a unique invariant probability μ_β relative to P_β . Using representations in terms of subtrees (cf., e.g., [7]), it is not very difficult to deduce a large deviation principle for it: there is a function $U : S \rightarrow \mathbb{R}_+$ (which is sometimes called the virtual potential associated with V) such that

$$\forall x \in S, \quad U(x) = - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\mu_\beta(x)).$$

Other important objects in this context are the mean exit times. More precisely, for given $x \in S$ and $\beta \geq 0$, let us designate by $\mathbb{P}_{x,\beta}$ the law of a homogeneous Markov chain $(X_n)_{n \geq 0}$ whose kernel of transition probabilities is P_β and which starts from the initial point x . Such processes are called generalized Metropolis algorithms by Catoni [2], the classical situation corresponding to

$$\forall \beta \geq 0, \forall x, y \in S, \quad P_\beta(x, y) \stackrel{\text{def.}}{=} P(x, y) \exp(-\beta(U(y) - U(x))_+),$$

where P is a reversible and irreducible Markovian kernel and $U : S \rightarrow \mathbb{R}$ is a given function (to be minimized). Let $\emptyset \neq A \neq S$ be a proper subset of S ; we write

$$T_A = \inf\{n \geq 0 : X_n \notin A\} \leq +\infty$$

but remark that, under any of the $\mathbb{P}_{x,\beta}$, this random variable is a.s. finite. Again by taking into account subtree representations of the following quantities, it can be shown (cf. [7] or [2]) that the limits

$$\forall x \in A, \quad h(x, A) \stackrel{\text{def.}}{=} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\mathbb{E}_{x,\beta}[T_A])$$

are well defined and obviously nonnegative. Then the exit height of A is naturally defined by

$$H(A) = \max_{x \in A} h(x, A)$$

and indeed it only depends on V .

Let us also consider the tensorized family $(P_\beta^{\otimes 2})_{\beta \geq 0}$: for any $\beta \geq 0$, $P_\beta^{\otimes 2}$ is the Kronecker product of P_β with itself, that is, the Markovian kernel on S^2 defined by

$$\forall (x_1, x_2), (y_1, y_2) \in S^2, \quad P_\beta^{\otimes 2}((x_1, x_2), (y_1, y_2)) = P_\beta(x_1, y_1) P_\beta(x_2, y_2)$$

and which corresponds to independent evolutions on the factors of $S \times S$. Clearly, this family obeys a large deviation principle with respect to the rate function $V \oplus V$ given by

$$\forall (x_1, x_2), (y_1, y_2) \in S^2, \quad V \oplus V((x_1, x_2), (y_1, y_2)) = V(x_1, y_1) + V(x_2, y_2).$$

However, the irreducibility is not preserved by this manipulation. In fact, $V \oplus V$ is irreducible if and only if V is aperiodic in an obvious sense. Nevertheless, the

irreducibility assumption was not relevant for the above definition of the height notion H , so we can apply it to $V \oplus V$ for proper subsets of S^2 , except that it can take the value $+\infty$. We are now in position to define the constant H_3 introduced by Catoni in [2]:

$$(2) \quad H_3 = H(S^2 \setminus \Delta),$$

where Δ is the diagonal of S^2 ,

$$\Delta = \{(x, x) \in S^2 : x \in S\}.$$

Its main interest is that it gives the critical exponential rate for a logarithmic uniform relaxation property:

THEOREM 1.1. *For any $\eta \geq 0$, consider the quantity*

$$l(\eta) = \limsup_{\beta \rightarrow +\infty} \sup_{x, y \in S} |\beta^{-1} \ln(\mathbb{P}_{x, \beta}[X_{\lfloor \exp(\eta\beta) \rfloor} = y]) - U(y)|,$$

where $\lfloor \cdot \rfloor$ denotes the integer part. Then we are assured of

$$\begin{aligned} \forall \eta > H_3, \quad l(\eta) &= 0, \\ \forall \eta < H_3, \quad l(\eta) &> 0. \end{aligned}$$

This result is due to Catoni [2], whose approach is based on large deviation considerations in the spirit of the work of Freidlin and Wentzell [7]; see also [1, 3, 17].

Our purpose here is to present a more direct proof via Dobrushin's coefficient and to discuss other characterizations of H_3 in terms of ergodic constants. This will lead us to wonder about the possibility of deducing their behavior for a large number of iterations from some fixed finite ones, asymptotically at low temperature. As we will see, this absence of a strong delaying effect is fulfilled by singular values even in essentially irreversible situations, while this is not necessarily true for the Dobrushin's coefficient. This phenomenon is technically important for the study of related simulated annealing algorithms and shows the advantage brought by a spectral approach.

The paper is organized as follows: in the next section, we will see the link between H_3 and the time of relaxation in the sense of Dobrushin's coefficient, which relation will be extended to spectral gap and other ergodic constants in Section 3. In Section 4, the above-mentioned difference of behavior in fixed times between Dobrushin's coefficient and spectral gap will be developed, followed by a discussion about the consequences for generalized simulated annealing.

2. Dobrushin’s coefficient. We will present another description of H_3 which enables us to derive straightforwardly Theorem 1.1.

More precisely, let us recall the definition of Dobrushin’s coefficient $a(P)$ associated with a Markovian kernel P , say on S :

$$a(P) = 1 - \frac{1}{2} \sup_{x,y \in S} \|P(x, \cdot) - P(y, \cdot)\|,$$

where $\|\cdot\|$ designates the total variation norm; for any signed measure m on S , $\|m\| := \sum_{x \in S} |m(x)|$. Then, for any given $0 < \varepsilon < 1$ and $\beta \geq 0$, we can consider the (deterministic) relaxation time

$$\tau_a(\beta) \stackrel{\text{def.}}{=} \inf\{n \geq 0 : a(P_\beta^n) \geq \varepsilon\} \leq +\infty.$$

Our main objective in this section is to show, always under the assumptions of the Introduction, the following proposition.

PROPOSITION 2.1. *Independently on the above choice of $0 < \varepsilon < 1$, we have*

$$H_3 = \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_a(\beta)).$$

To go in this direction, we begin by discussing the general definition of the height functional H . The next result can be deduced from Proposition 4.19 of [2] and in fact its proof is very similar.

LEMMA 2.2. *Let A be a proper subset of S and let $0 < \tilde{\varepsilon} < 1$ be fixed. We define*

$$t_A(\beta) \stackrel{\text{def.}}{=} \inf \left\{ t \geq 0 : \sup_{x \in A} \mathbb{P}_{x,\beta}[T_A > t] \leq \tilde{\varepsilon} \right\}.$$

Then we are assured of

$$H(A) = \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(t_A(\beta)).$$

The underlying idea is that if $x \in A$ is “the starting point from which it is the most difficult to get out of A ,” then, under $\mathbb{P}_{x,\beta}$, T_A is asymptotically an exponential time [whose parameter is of order $\exp(H(A)\beta)$]. This unpredictability property of the (conveniently renormalized) worse exit time can be made rigorous (cf. [10] or [12]), but we do not need to go so far here.

PROOF. Using the Markov property, we get that, for any $x \in A$, $\beta \geq 0$ and $t, n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}_{x,\beta}[T_A > nt] &= \mathbb{E}_{x,\beta}[\mathbb{1}_{\{T_A > (n-1)t\}} \mathbb{P}_{X_{(n-1)t},\beta}[T_A > t]] \\ &\leq \mathbb{P}_{x,\beta}[T_A > (n-1)t] \sup_{y \in A} \mathbb{P}_{y,\beta}[T_A > t] \\ &\leq \left(\sup_{y \in A} \mathbb{P}_{y,\beta}[T_A > t] \right)^n. \end{aligned}$$

Thus, by considering $t = t_A(\beta)$, we obtain

$$\begin{aligned} \mathbb{E}_{x,\beta}[T_A] &\leq \mathbb{E}_{x,\beta}\left[t_A(\beta) \sum_{n \geq 0} \mathbb{1}_{T_A > nt_A(\beta)}\right] \\ &= t_A(\beta) \sum_{n \geq 0} \mathbb{P}_{x,\beta}[T_A > nt_A(\beta)] \\ &\leq t_A(\beta) \sum_{n \geq 0} \left(\sup_{y \in A} \mathbb{P}_{y,\beta}[T_A > t_A(\beta)]\right)^n \\ &\leq t_A(\beta) \frac{1}{(1 - \tilde{\varepsilon})}, \end{aligned}$$

where we can next take the supremum over $x \in A$. On the other hand, we also have

$$\begin{aligned} \tilde{\varepsilon} &< \sup_{x \in A} \mathbb{P}_{x,\beta}[T_A \geq t_A(\beta)] \\ &\leq \frac{1}{t_A(\beta)} \sup_{x \in A} \mathbb{E}_{x,\beta}[T_A]. \end{aligned}$$

These bounds and the definition of $H(A)$ lead immediatly to the expected result. \square

We will apply this result to the particular proper subset $S^2 \setminus \Delta$ in the context of the tensorized family of Markov kernels, but we need first to recall some notions from coupling theory. For $\beta \geq 0$ and $(x, y) \in S^2$, let $\mathbb{P}_{(x,y),\beta}^{\otimes 2}$ be the law of a Markov chain $(X_n, \tilde{X}_n)_{n \geq 0}$, where $(X_n)_{n \geq 0}$ and $(\tilde{X}_n)_{n \geq 0}$ are independent and of respective distributions $\mathbb{P}_{x,\beta}$ and $\mathbb{P}_{y,\beta}$. We define

$$T_\Delta \stackrel{\text{def.}}{=} \inf\{n \geq 0 : X_n = \tilde{X}_n\}$$

and consider the sequence $(Y_n)_{n \geq 0}$ given by

$$\forall n \geq 0, \quad Y_n = \begin{cases} \tilde{X}_n, & \text{if } n \leq T_\Delta, \\ X_n, & \text{if } n \geq T_\Delta. \end{cases}$$

As is well known, $(Y_n)_{n \geq 0}$ admits $\mathbb{P}_{y,\beta}$ for law and $(X_n, Y_n)_{n \geq 0}$ is called a simple coupled chain.

Another basic computation is that, for any Markovian kernel P on S , we have

$$a(P) = \inf_{x,y \in S} \sum_{z \in S} P(x, z) \wedge P(y, z).$$

Now these statements easily imply the next result, from which Proposition 2.1 follows at once:

LEMMA 2.3. For any $\beta \geq 0$ and $n \in \mathbb{N}$, we have

$$(3) \quad \frac{a(P_\beta^n)^2}{\text{card}(S)} \leq \inf_{(x,y) \notin \Delta} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[T_\Delta \leq n] \leq a(P_\beta^n).$$

Consequently, we are assured of the inequalities

$$\begin{aligned} \forall \beta \geq 0, \quad \tau_a(\beta) &\geq t_{S^2 \setminus \Delta}(\beta) && \text{with } \tilde{\varepsilon} = 1 - \frac{\varepsilon^2}{\text{card}(S) + 1}, \\ \forall \beta \geq 0, \quad t_{S^2 \setminus \Delta}(\beta) &\geq \tau_a(\beta) && \text{with } \varepsilon = 1 - \tilde{\varepsilon}. \end{aligned}$$

PROOF. To get the second bound of (3), it is sufficient to consider the minimum in $x \neq y \in S$ in the inequality

$$\begin{aligned} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[T_\Delta \leq n] &= \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = Y_n] \\ &= \sum_{z \in S} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = z = Y_n] \\ &\leq \sum_{z \in S} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = z] \wedge \mathbb{P}_{(x,y),\beta}^{\otimes 2}[Y_n = z] \\ &= \sum_{z \in S} P_\beta^n(x, z) \wedge P_\beta^n(y, z). \end{aligned}$$

For the reverse minoration, we do the same thing, but with respect to the bounds

$$\begin{aligned} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = Y_n] &\geq \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = \tilde{X}_n] \\ &= \sum_{z \in S} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = z = \tilde{X}_n] \\ &= \sum_{z \in S} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = z] \mathbb{P}_{(x,y),\beta}^{\otimes 2}[\tilde{X}_n = z] \\ &\geq \sum_{z \in S} (\mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = z] \wedge \mathbb{P}_{(x,y),\beta}^{\otimes 2}[\tilde{X}_n = z])^2 \\ &\geq \frac{1}{\text{card}(S)} \left(\sum_{z \in S} \mathbb{P}_{(x,y),\beta}^{\otimes 2}[X_n = z] \wedge \mathbb{P}_{(x,y),\beta}^{\otimes 2}[\tilde{X}_n = z] \right)^2. \quad \square \end{aligned}$$

To finish this section, we will show how Theorem 1.1 can be deduced without difficulty from Proposition 2.1. So let $\eta > 0$ be given; for $\beta \geq 0$, we note $n_\beta \stackrel{\text{def.}}{=} \lfloor \exp((H_3 + \eta)\beta) \rfloor$ and $m_\beta \stackrel{\text{def.}}{=} \lfloor n_\beta / \lfloor \exp(\eta\beta/2) \rfloor \rfloor$. We want to see that

$$(4) \quad \lim_{\beta \rightarrow +\infty} \sup_{x,y \in S} |\beta^{-1} \ln(\mathbb{P}_{x,\beta}[X_{n_\beta} = y]) - U(y)| = 0.$$

To prove this convergence, let us recall some basic properties of Dobrushin's coefficient associated with a Markovian kernel P on S : for any probabilities ν and μ on S , we are assured of

$$\|\nu P - \mu P\| \leq (1 - a(P))\|\nu - \mu\|$$

and taking into account that if in the lhs we consider the supremum over all the possible choices of μ and ν , then we get an equality, one is easily convinced that, for any $n, m \in \mathbb{N}$,

$$1 - a(P^{nm}) \leq (1 - a(P^n))^m.$$

Coming back to our given family of kernels and using Proposition 2.1 and the definition of m_β , it appears that, for β sufficiently large,

$$a(P_\beta^{m_\beta}) \geq \frac{1}{2}.$$

Thus we obtain that, for such β and for any $x \in S$,

$$\begin{aligned} \|\delta_x P_\beta^{n_\beta} - \mu_\beta\| &= \|\delta_x P_\beta^{n_\beta} - \mu_\beta P_\beta^{n_\beta}\| \\ &\leq \left(\frac{1}{2}\right)^{\lfloor \exp(\eta\beta/2) \rfloor}, \end{aligned}$$

which goes to zero extremely fast for large β (exponential of exponential); in particular we have, for any $y \in S$, $\lim_{\beta \rightarrow +\infty} \mu_\beta^{-1}(y) 2^{-\lfloor \exp(\eta\beta/2) \rfloor} = 0$. Since we can also write

$$\|\delta_x P_\beta^{n_\beta} - \mu_\beta\| = \sum_{y \in S} \left| \frac{\mathbb{P}_{x,\beta}[X_{n_\beta} = y]}{\mu_\beta(y)} - 1 \right| \mu_\beta(y)$$

we get that, for any $y \in S$,

$$\lim_{\beta \rightarrow +\infty} \frac{\mathbb{P}_{x,\beta}[X_{n_\beta} = y]}{\mu_\beta(y)} = 1,$$

from which (4) follows, via the definition of the quasipotential U .

Now to see that H_3 is really the critical exponent for that relaxation property, let $\eta > 0$ be fixed and this time we let $n_\beta \stackrel{\text{def.}}{=} \lfloor \exp((H_3 - \eta)\beta) \rfloor$ and $m_\beta \stackrel{\text{def.}}{=} \lfloor \exp(\eta\beta/2) \rfloor$. We have to show that there exist $x, y \in S$ such that

$$\limsup_{\beta \rightarrow +\infty} |\beta^{-1} \ln(\mathbb{P}_{x,\beta}[X_{n_\beta} = y]) - U(y)| > 0.$$

Indeed, if it were not true, we could find a $\beta_0 \geq 0$ such that, for all $\beta \geq \beta_0$ and all $x, y \in S$,

$$\frac{\mathbb{P}_{x,\beta}[X_{n_\beta} = y]}{\mu_\beta(y)} \geq \exp(-\eta\beta/2),$$

a bound which implies that

$$a(P^{n\beta}) \geq \exp(-\eta\beta/2).$$

Thus we would obtain that

$$a(P^{m\beta}) \geq 1 - (1 - \exp(-\eta\beta/2))^{\lfloor \exp(\eta\beta/2) \rfloor},$$

which is larger than $\varepsilon \stackrel{\text{def.}}{=} 1/10$ for $\beta \geq \beta_1$ with $\beta_1 \geq \beta_0$ large enough. One could then derive that, for $\beta \geq \beta_1$, $\tau_a(\beta) \leq m\beta$ and passing to the limit for large β , conclude the contradiction that $H_3 \leq H_3 - \eta/2$.

REMARK 2.4. In fact the above approach enables one to see that H_3 is also the critical exponent for a stronger convergence:

$$\begin{aligned} \forall \eta > H_3, \quad \limsup_{\beta \rightarrow +\infty} \sup_{x, y \in S} \left| \frac{\mathbb{P}_{x, \beta}[X_{\lfloor \exp(\eta\beta) \rfloor} = y]}{\mu_\beta(y)} - 1 \right| &= 0, \\ \forall \eta < H_3, \quad \limsup_{\beta \rightarrow +\infty} \sup_{x, y \in S} \left| \frac{\mathbb{P}_{x, \beta}[X_{\lfloor \exp(\eta\beta) \rfloor} = y]}{\mu_\beta(y)} - 1 \right| &> 0. \end{aligned}$$

However, due to the weak assumption on the family $(P_\beta)_{\beta \geq 0}$, one cannot expect precise estimations on μ_β or $\tau_a(\beta)$, only large deviation results are available for them (whose rates are respectively given by U and H_3). For further reference, also remark that this characterization remains true if in the above left-hand side [or in the definition of $l(\eta)$], we rather consider a particular increasing sequence $(\beta_k)_{k \geq 0}$ of positive real numbers satisfying $\lim_{k \rightarrow \infty} \beta_k = +\infty$ instead of all large $\beta \geq 0$. Indeed, the above limsup can be replaced by liminf, for instance.

3. Spectral gaps and other ergodic constants. The previous analysis of Dobrushin’s coefficient can be extended to some other quantifiers of convergence to equilibrium speed, in particular due to the finiteness of the state space.

We begin by giving a general result. Let $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}$ be a strictly convex function such that $\varphi(1) = 0$. For fixed $\beta \geq 0$, we define a functional $H_{\varphi, \beta}$ on \mathcal{P} the set of probabilities on S by

$$\forall m \in \mathcal{P}, \quad H_{\varphi, \beta}(m) \stackrel{\text{def.}}{=} \sum_{x \in S} \varphi\left(\frac{m(x)}{\mu_\beta(x)}\right) \mu_\beta(x).$$

In some sense, this quantity measures a “distance” to μ_β ; at least an application of Jensen’s inequality shows readily that $H_{\varphi, \beta}(m) > 0$ for $m \neq \mu_\beta$. If P is a Markov kernel leaving μ_β invariant, let $\alpha_{\varphi, \beta}(P)$ be its relative contraction coefficient:

$$\alpha_{\varphi, \beta}(P) \stackrel{\text{def.}}{=} \inf_{m \in \mathcal{P} \setminus \{\mu_\beta\}} \frac{H_{\varphi, \beta}(m) - H_{\varphi, \beta}(mP)}{H_{\varphi, \beta}(m)} \geq 0.$$

As before, we consider the order of the number of iterations of P_β which will produce a sizeable effect with respect to this ergodic coefficient; that is, for a fixed $0 < \varepsilon < 1$, we let

$$\forall \beta \geq 0, \quad \tau_\varphi(\beta) \stackrel{\text{def.}}{=} \inf\{n \geq 0 : \alpha_{\varphi, \beta}(P_\beta^n) \geq \varepsilon\} \leq +\infty.$$

Taking into account the general lower bound (cf. [4] or [5])

$$\alpha_{\varphi, \beta}(P) \geq a(P)$$

for all Markovian kernels P over S for which μ_β is invariant, it already appears that

$$\limsup_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_\varphi(\beta)) \leq H_3.$$

The next result gives a sufficient condition on φ for this bound to be an equality with a true limit.

PROPOSITION 3.1. *Assume that, for all $\eta > 0$, $\varphi(x) \ll \exp(x^\eta)$ for large $x \geq 0$. Then we are assured of*

$$\lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_\varphi(\beta)) = H_3.$$

PROOF. Let $(\beta_k)_{k \geq 0}$ be an increasing sequence of positive reals diverging to $+\infty$, such that

$$\lim_{k \rightarrow \infty} \beta_k^{-1} \ln(\tau_\varphi(\beta_k)) = \gamma_0 \stackrel{\text{def.}}{=} \liminf_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_\varphi(\beta)).$$

Let $\gamma > \gamma_0$ be arbitrary; we let $n_k = \lfloor \exp(\gamma \beta_k) \rfloor$ for $k \geq 0$. Due to the remark at the end of the previous section, it is sufficient to show that

$$\forall x, y \in S, \quad \lim_{k \rightarrow \infty} \frac{P_{\beta_k}^{n_k}(x, y)}{\mu_{\beta_k}(y)} = 1.$$

First, we note that up to the addition to φ of a linear function of type $\mathbb{R}_+ \ni x \mapsto a(x - 1)$, with $a \in \mathbb{R}$, which does not alter the definition of $H_{\varphi, \beta}$, we can assume that φ is increasing (respectively decreasing) on $[1, +\infty[$ (resp. $[0, 1]$), and in particular positive on $\mathbb{R}_+ \setminus \{1\}$. Thus, we obtain

$$\begin{aligned} \sup_{m \in \mathcal{P}} H_{\varphi, \beta}(m) &\leq \sup_{x \in S} \varphi(1/\mu_\beta(x)) \\ &\leq \varphi(\exp(K\beta)) \end{aligned}$$

for a sufficiently large constant $K > 0$, uniformly over all $\beta \geq 1$.

Then we reconsider arguments similar to the ones used in the first part of the deduction of Theorem 1.1 from Proposition 2.1. Let $z := (\gamma - \gamma_0)/4 > 0$; we have, for all sufficiently large $k \geq 0$,

$$\begin{aligned} \sup_{m \in \mathcal{P}} H_{\varphi, \beta_k}(m P_{\beta_k}^{n_k}) &\leq [1 - \alpha_{\varphi, \beta_k}(P_{\beta_k}^{\lfloor \exp(\beta_k(\gamma_0+z)) \rfloor})]^{\exp(2z\beta_k)} \sup_{m \in \mathcal{P}} H_{\varphi, \beta_k}(m) \\ &\leq (1 - \varepsilon)^{\exp(2z\beta_k)} \varphi(\exp(K\beta)) \\ &\leq \exp(\ln(1 - \varepsilon) \exp(2z\beta_k)) \exp(\exp(z\beta_k)) \\ &\leq \exp(-\exp(z\beta_k)), \end{aligned}$$

where in the last line but one, we have used the hypothesis made on φ with $\eta = z/K$. Again the strength of this convergence to zero of the rhs enables us to conclude that, for all $x, y \in S$,

$$\lim_{k \rightarrow \infty} \varphi\left(\frac{P_{\beta_k}^{n_k}(x, y)}{\mu_{\beta_k}(y)}\right) = 0,$$

which can be rewritten as the expected behavior. \square

Certainly the more interesting special case corresponds to the quadratic mapping

$$\varphi_1 : \mathbb{R}_+ \ni x \mapsto (x - 1)^2$$

since for any Markovian kernel P leaving μ_β invariant [in the sequel the set of such matrices will be designated by $\mathcal{I}(\mu_\beta)$], $\alpha_{\varphi_1, \beta}(P)$ is just $\lambda(\mu_\beta, P P^*)$, the largest singular value of the restriction of P to the orthogonal complement of $\text{Vect}(\mathbb{1})$ in the Hilbert space $\mathbb{L}^2(\mu_\beta)$, that is, the largest eigenvalue on this subspace of $P P^*$, where P^* is the adjoint operator of P with respect to the natural scalar product of $\mathbb{L}^2(\mu_\beta)$ (cf. [9]).

Related notions are the classical and modified logarithmic Sobolev constants defined respectively by

$$\begin{aligned} \tilde{l}(\mu_\beta, P P^*) &\stackrel{\text{def.}}{=} \inf_{f \in \mathbb{L}^2(\mu_\beta) \setminus \text{Vect}(\mathbb{1})} \frac{\mu_\beta[f(\text{Id} - P P^*)(f)]}{\text{Ent}(f^2, \mu_\beta)}, \\ l(\mu_\beta, P P^*) &\stackrel{\text{def.}}{=} \inf_{f \in \mathbb{L}^2(\mu_\beta) \setminus \text{Vect}(\mathbb{1})} \frac{\mu_\beta[f^2(\text{Id} - P P^*)(\ln(f^2))]}{\text{Ent}(f^2, \mu_\beta)}, \end{aligned}$$

where the entropy of a function $f \in \mathbb{L}^2(\mu_\beta)$ is given by

$$\text{Ent}(f^2, \mu_\beta) \stackrel{\text{def.}}{=} \int f^2 \ln(f^2 / \mu_\beta(f^2)) d\mu_\beta.$$

As usual, it is natural to consider the associated relaxation times, always for sufficiently small fixed $\varepsilon > 0$ and any $\beta \geq 0$,

$$\begin{aligned} \tau_{\tilde{l}}(\beta) &\stackrel{\text{def.}}{=} \inf\{n \geq 0 : \tilde{l}(\mu_\beta, P_\beta^n P_\beta^{n*}) \geq \varepsilon\}, \\ \tau_l(\beta) &\stackrel{\text{def.}}{=} \inf\{n \geq 0 : l(\mu_\beta, P_\beta^n P_\beta^{n*}) \geq \varepsilon\}. \end{aligned}$$

Due to the following general bounds, valid for any kernel $P \in \mathcal{L}(\mu_\beta)$ (cf. [5, 15]),

$$\begin{aligned} \frac{2 - 4 \min_{x \in S} \mu_\beta(x)}{\ln(1 / \min_{x \in S} \mu_\beta(x) - 1)} \lambda(\mu_\beta, P P^*) &\leq 2\tilde{l}(\mu_\beta, P P^*) \\ &\leq l(\mu_\beta, P P^*) \leq 2\lambda(\mu_\beta, P P^*) \end{aligned}$$

[in case $\min_{x \in S} \mu_\beta(x) = 1/2$, the first factor above must be understood as 1, and note that for β large it is always at least of order $1/\beta$], and taking into account that, for any given $\eta > 0$ and $K > 0$, we have, for sufficiently large $\beta \geq 0$,

$$\left(1 - \frac{\varepsilon}{K\beta}\right)^{\exp(\eta\beta)} \geq (1 - \varepsilon)^{\exp(\eta\beta/2)},$$

it is easy to see that the same convergences as before are taking place:

$$\begin{aligned} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_{\tilde{l}}(\beta)) &= H_3, \\ \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_l(\beta)) &= H_3. \end{aligned}$$

The latter could also have been deduced directly from Proposition 3.1, since it has been shown that there exists a universal constant $0 < \rho < 1$ such that, for any $\beta \geq 0$ and any kernel $P \in \mathcal{L}(\mu_\beta)$,

$$\rho l(\mu_\beta, P P^*) \leq \alpha_{\varphi_2, \beta}(P) \leq l(\mu_\beta, P P^*)$$

with $\varphi_2 : \mathbb{R}_+ \ni x \mapsto x \ln(x)$ (see [5]).

Finally, let us note that the isoperimetric constants associated with our family will also have the same limiting behavior: they are defined for any $\beta \geq 0$ and kernel $P \in \mathcal{L}(\mu_\beta)$ by

$$i(\mu_\beta, P P^*) \stackrel{\text{def.}}{=} \inf_{A \subset S : 0 < \mu_\beta(A) \leq 1/2} \frac{\mu_\beta(\mathbb{1}_{A^c} P P^*(\mathbb{1}_A))}{\mu_\beta(A)}.$$

Due to the general bounds (cf. [14])

$$(5) \quad \forall \beta \geq 0, \quad \frac{i(\mu_\beta, P P^*)}{\text{card}(S) - 1} \leq \lambda(\mu_\beta, P P^*) \leq 2i(\mu_\beta, P P^*),$$

we obtain, without much surprise,

$$\lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\tau_i(\beta)) = H_3$$

with, for sufficiently small fixed $\varepsilon > 0$,

$$\tau_i(\beta) \stackrel{\text{def.}}{=} \inf\{n \geq 0 : i(\mu_\beta, P_\beta^n P_\beta^{n*}) \geq \varepsilon\}.$$

Thus it appears that in this context of finite generalized Metropolis algorithms, the relaxation times $\tau_a(\beta)$, $\tau_i(\beta)$, $\tau_l(\beta)$, $\tau_r(\beta)$ and $\tau_\varphi(\beta)$, with φ satisfying the condition of Proposition 3.1, are all of the same order $\exp(\beta H_3)$ asymptotically at large inverse temperatures $\beta > 0$, that is, independently of “reasonable” ways of measuring the distance to the equilibrium probability μ_β .

4. Absence of delaying effect. The objective of this section is to relate the asymptotical behavior at low temperature of the singular values for a large number of iterations to fixed finite ones.

First, let us recall some basic large deviation results for spectral quantities. A family of Markovian kernels being given as before, for any $\beta \geq 0$, we associate a spectral gap with P_β by the formula

$$\lambda(\mu_\beta, P_\beta) \stackrel{\text{def.}}{=} \inf_{f \in \mathbb{L}^2(\mu_\beta) \setminus \text{Vect}(\mathbb{1})} \frac{\mu_\beta[f(\text{Id} - P_\beta)(f)]}{\mu_\beta[(f - \mu_\beta(f))^2]}.$$

Strictly speaking, it is the gap between 1 and the rest of the spectrum of the autoadjoint operator $(P_\beta + P_\beta^*)/2$ (sometimes called the additive symmetrization of P_β , to distinguish it from the multiplicative one $P_\beta P_\beta^*$) in $\mathbb{L}^2(\mu_\beta)$. Nevertheless, as below P_β will mainly be replaced by “reversibilized” kernels, we will keep this denomination, which in this case is justified. Then it is well known that, under the hypotheses of the Introduction, the next limit exists:

$$H_2 \stackrel{\text{def.}}{=} - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(\mu_\beta, P_\beta)) \geq 0.$$

Indeed, this behavior can be deduced at once from the bounds (5) where P_β replaces PP^* , and the irreducibility of V implies the finiteness of the limit in that situation. An interesting feature of this quantity is that it admits an explicit description (cf. [8], whose computations can be extended to the present setting; see also, for instance, [14]): let $C_{x,y}$ denote the set of “paths” going from x to y , that is, the finite sequences $p = (p_i)_{0 \leq i \leq n}$ of elements of S satisfying $p_0 = x$ and $p_n = y$. The elevation of such a path p is

$$e(p) = \max_{0 \leq i < n} U(p_i) + V(p_i, p_{i+1}).$$

Then the previous constant H_2 is given by

$$H_2 = \max_{x,y \in S} H(x, y) - U(x) - U(y),$$

where the difficulty $H(x, y)$ of going from x to y is defined by

$$\forall x, y \in S, \quad H(x, y) \stackrel{\text{def.}}{=} \min_{p \in C_{x,y}} e(p)$$

(note that all these objects depend only on V). The important property of H being symmetrical in its two arguments was first noticed by Trouvé [16] and is linked to the fact that, in the above definition of elevation, it is not necessary to consider “nonoriented paths.”

However, as we have already alluded to, the kernels of interest in defining natural Dirichlet forms in a discrete time setting are rather the $P_\beta^n P_\beta^{n*}$, for $n \geq 1$. Thus we are led to consider the limit

$$\forall n \in \mathbb{N}, \quad h(n) \stackrel{\text{def.}}{=} - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(\mu_\beta, P_\beta^n P_\beta^{n*})) \leq +\infty.$$

These quantities are easily seen to be nonincreasing in $n \geq 0$, since for $\beta \geq 0$ fixed, $\mathbb{N} \ni n \mapsto \lambda(\mu_\beta, P_\beta^n P_\beta^{n*})$ is nondecreasing, and in fact a stationarity property is taking place:

LEMMA 4.1. *There exists $n_0 \geq 0$ such that, for all $n \geq n_0$, we have*

$$h(n) = h(n_0).$$

In the sequel the rhs will then also be designated by $h(\infty)$.

PROOF. This result is quite clear in the degenerate situation where, for all $n \geq 0$, $h(n) = +\infty$. Thus it is sufficient to consider the case where there exists $n_1 \geq 1$ such that $h(n_1) < +\infty$.

Taking into account the monotonicity and the nonnegativity of the sequence $(h(n))_{n \geq n_1}$, the lemma follows at once if we prove that $\{h(n) : n \geq n_1\}$ is discrete. A priori, the latter set is included in the group generated on \mathbb{R} by the real numbers $V(w, w')$, $w, w' \in S$, which unfortunately can be dense. So, to get round this little difficulty, we define, for $n \geq 0$ and $x, y \in S$, the rate function

$$W_n(x, y) \stackrel{\text{def.}}{=} - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(P_\beta^n P_\beta^{n*}(x, y)) \leq +\infty,$$

which can be rewritten as

$$\begin{aligned} W_n(x, y) &= \min_{z \in S, (p_i)_{0 \leq i \leq n} \in C_{x,z}, (q_i)_{0 \leq i \leq n} \in C_{y,z}} (V(x, p_1) + V(p_1, p_2) + \cdots + V(p_{n-1}, z) \\ &\quad + U(y) + V(y, q_1) + V(q_1, q_2) + \cdots + V(q_{n-1}, z) - U(z)). \end{aligned}$$

Due to the above representations, for any $n \geq 0$, there exist $x, y, v, w \in S$ such that $h(n) = W_n(x, y) + U(x) - U(v) - U(w)$. This leads us to consider E the additive semigroup on \mathbb{R}_+ generated by the values $V(w, w')$, $w, w' \in S$, and E' the finite set $\{U(w_1) + U(w_2) - U(w_3) - U(w_4) - U(w_5) : \forall 1 \leq i \leq 5, w_i \in S\}$. Indeed, for any $n \geq n_1$, there exist $e \in E$ and $e' \in E'$ such that $h(n) = e + e'$, which implies in particular that

$$e \leq K \stackrel{\text{def.}}{=} h(n_1) + \max_{e'' \in E'} |e''|.$$

Noting that the set $E \cap [0, K]$ is also finite, we see that only a finite number of values are permitted for $h(n)$, $n \geq n_1$. \square

We already observe that there is an immediate relation between $h(\infty)$ and H_3 , which has corresponded to a well-known procedure of estimation since the article of Fill [6]:

LEMMA 4.2. *We are assured of*

$$H_3 \leq h(\infty).$$

PROOF. The case $h(\infty) = +\infty$ being obvious, we can restrict the study to the situation $h(\infty) < +\infty$. Let $n_0 \geq 1$ be such that $h(\infty) = h(n_0)$. Using the following general inequality for $k \in \mathbb{N}^*$,

$$1 - \lambda(\mu_\beta, P_\beta^{kn_0} P_\beta^{*kn_0}) \leq (1 - \lambda(\mu_\beta, P_\beta^{n_0} P_\beta^{*n_0}))^k$$

[which can be deduced by iteration from the bound $1 - \lambda(\mu_\beta, PQQ^*P^*) \leq (1 - \lambda(\mu_\beta, PP^*))(1 - \lambda(\mu_\beta, QQ^*))$, valid for all $P, Q \in \mathcal{L}(\mu_\beta)$], it is easy to see that, for any sufficiently small fixed $\varepsilon > 0$,

$$\begin{aligned} \tau_\lambda(\beta) &\stackrel{\text{def.}}{=} \inf\{n \geq 0 : \lambda(\mu_\beta, P_\beta^n P_\beta^{*n}) \geq \varepsilon\} \\ &\leq n_0 \inf\{k \geq 0 : \lambda(\mu_\beta, P_\beta^{kn_0} P_\beta^{*kn_0}) \geq \varepsilon\} \\ &\leq n_0 \inf\{k \geq 0 : 1 - (1 - \lambda(\mu_\beta, P_\beta^{n_0} P_\beta^{*n_0}))^k \geq \varepsilon\} \\ &= n_0 \inf\{k \geq 0 : (1 - \lambda(\mu_\beta, P_\beta^{n_0} P_\beta^{*n_0}))^k \leq 1 - \varepsilon\} \\ &= n_0 \left\lceil \frac{\ln(1 - \varepsilon)}{\ln(1 - \lambda(\mu_\beta, P_\beta^{n_0} P_\beta^{*n_0}))} \right\rceil, \end{aligned}$$

where, for all $x \in \mathbb{R}$, $\lceil x \rceil \stackrel{\text{def.}}{=} -\lfloor -x \rfloor$, from which we deduce without difficulty the above bound. \square

If the inequality $H_3 < h(\infty)$ were to be satisfied by some examples, it could be interpreted as a “delaying effect,” since it means that the rate of convergence to equilibrium will eventually be stronger than what can be expected by observation of the relaxation at any fixed finite time (more specifically, we should call this property “strong” delaying, to distinguish it from its weak form just meaning that the smaller $n_0 \geq 1$ for which Lemma 4.1 is satisfied can be strictly larger than 1). Nevertheless, we will now show our main result, saying that this phenomenon cannot occur:

THEOREM 4.3. *Indeed, we always have*

$$H_3 = h(\infty).$$

The proof of the inequality $h(\infty) \leq H_3$ is not so straightforward as that of its reverse in Lemma 4.2; in fact it will be based on an easy extension of a result proved in [13] which we now recall. Let \mathcal{L} be the set of local minima associated with V , which are the points $x \in S$ such that

$$\forall y \in S, \quad U(y) < U(x) \implies H(x, y) > U(x).$$

We introduce an equivalence relation \sim on \mathcal{L} by

$$(6) \quad \forall x, y \in \mathcal{L}, \quad x \sim y \iff H(x, y) = U(x) = U(y)$$

and let $\mathcal{L}_1, \dots, \mathcal{L}_K$ be the associated equivalence classes. Another way to describe them is to say that they are the biconnected [with respect to the directed graph whose oriented edges are the arrows $x \rightarrow y$ between vertices $x, y \in S$ satisfying $V(x, y) = 0$] components of the bottoms of the wells corresponding to the energy landscape defined by (U, V) (which only depends on V). We have shown in [13] that if, for any $1 \leq k \leq K$, there exists at least one $x \in \mathcal{L}_k$ such that $V(x, x) = 0$, then with the previous notation we can take $n_0 \leq \text{card}(S) - 1$ and $h(\infty) = H_2$. The main idea was to use in some sense nonincreasing and nondecreasing paths to connect the sets $\mathcal{L}_1, \dots, \mathcal{L}_K$, the points x satisfying $V(x, x) = 0$ enabling us to manage the length of the paths.

Returning to the general situation, we denote by V_n , for $n \in \mathbb{N}$, the rate matrix associated with the family of Markovian kernels $(P_\beta^n)_{\beta \geq 0}$. We remark that the subset \mathcal{L} remains the same for all the V_n , $n \geq 1$ (this is not necessarily true for the $\mathcal{L}_1, \dots, \mathcal{L}_K$, which can be divided into a finer partition of \mathcal{L}). Furthermore, noticing that, for any given $x \in \mathcal{L}$, there exists a number $n_x \in \mathbb{N}^*$ such that $V_{n_x}(x, x) = 0$, we see that we can find $l \in \mathbb{N}^*$ (for instance, it is possible to take $l = \prod_{x \in \mathcal{L}} n_x$) for which V_l will satisfy

$$(7) \quad \forall x \in \mathcal{L}, \quad V_l(x, x) = 0.$$

According to the above reminder, we thus obtain

$$\begin{aligned} h(\infty) &= - \lim_{n \rightarrow +\infty} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(\mu_\beta, P_\beta^{ln} P_\beta^{ln*})) \\ &= - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(\mu_\beta, P_\beta^l)) \\ &\stackrel{\text{def.}}{=} H_2(l). \end{aligned}$$

We will also designate by $H_3(n)$, for $n \geq 1$, the constant defined as in the Introduction, but with respect to the family of kernels $(P_\beta^n)_{\beta \geq 0}$. Theorem 4.3 will then follow from the next two results:

LEMMA 4.4. *For any $n \geq 1$, we have $H_2(n) \leq H_3(n)$.*

This bound was deduced by Catoni [2] from his characterization of H_2 and H_3 as critical exponents, but we will present here an alternative proof.

PROOF OF LEMMA 4.4. We note at once that we can restrict ourselves to treat the case $n = 1$. For that purpose, we first come back to the notion of exit height $H(A)$, for proper subsets A of S , to recall a way to compute H_3 . Having resort to either large deviation techniques (cf., e.g., [2]) or martingale and spectral arguments (see [12]), it is possible to show that

$$H(A) = \max_{x \in A} \min_{y \notin A} H(x, y) - U(x).$$

To apply this explicit formulation to the tensorized kernels and $A = \Delta$, we begin by assuming that V is aperiodic. Then $V \oplus V$ is irreducible and its associated quasipotential is $U \oplus U : S^2 \ni (x, y) \mapsto U(x) + U(y)$ (use, e.g., the fact that under our hypotheses, for $\beta \geq 0$, $\mu_\beta^{\otimes 2}$ will necessarily be the unique invariant probability of $P_\beta^{\otimes 2}$). Thus finally it appears that

$$(8) \quad \begin{aligned} H_3 = & \max_{x \neq y \in S} \min_{z \in S, n \in \mathbb{N}^*} \min_{p=(p_i)_{0 \leq i \leq n} \in C_{x,z}, q=(q_i)_{0 \leq i \leq n} \in C_{y,z}} \max_{0 \leq i < n} U(p_i) + U(q_i) \\ & + V(p_i, p_{i+1}) + V(q_i, q_{i+1}) - U(x) - U(y). \end{aligned}$$

This equality remains true in case V is periodic, since then both sides are $+\infty$. How awful it may seem that this formula admits at least the advantage of showing the following monotonicity of H_3 : if $V^{(1)}$ and $V^{(2)}$ are two irreducible rate matrices admitting the same function U for quasipotential and if $V^{(1)} \leq V^{(2)}$ (entry by entry), then with obvious notation $H_3^{(1)} \leq H_3^{(2)}$. In particular, if \tilde{H}_3 is the constant associated with the family of kernels $((P_\beta + P_\beta^*)/2)_{\beta \geq 0}$, then we have $\tilde{H}_3 \leq H_3$ and it is sufficient to prove that $H_2 = \tilde{H}_2 \leq \tilde{H}_3$. Meanwhile, $(P_\beta + P_\beta^*)/2$ being autoadjoint in $\mathbb{L}^2(\mu_\beta)$, let f be a left-eigenvector associated with $1 - \lambda(\mu_\beta, P_\beta)$, that is, satisfying

$$f \frac{P_\beta + P_\beta^*}{2} = (1 - \lambda(\mu_\beta, P_\beta)) f.$$

We can suppose furthermore that $\sup_{x \in S} |f(x)| = 1$ and consider then the probability $m_0 \stackrel{\text{def.}}{=} (1 + f)\mu_\beta$ [one would have noticed that $1 + f \geq 0$ and that $\mu_\beta(f) = 0$]. We have

$$m_0 \left(\frac{P_\beta + P_\beta^*}{2} \right)^{[\lambda(\mu_\beta, P_\beta)^{-1}]} = [1 + (1 - \lambda(\mu_\beta, P_\beta))^{[\lambda(\mu_\beta, P_\beta)^{-1}]} f] \mu_\beta.$$

Now assume that $H_2 > 0$ (otherwise the bound $H_2 \leq \tilde{H}_3$ is trivial), in order that $\lim_{\beta \rightarrow +\infty} \lambda(\mu_\beta, P_\beta) = 0$; we deduce from the previous equality that if we start with m_0 as initial distribution,

$$\forall y \in S, \quad \lim_{\beta \rightarrow +\infty} \frac{\mathbb{P}_{m_0, \beta}[X_{[\lambda(\mu_\beta, P_\beta)^{-1}]} = y]}{\mu_\beta(y)} = 1 + \frac{f(y)}{e},$$

and taking into account Remark 2.4, we obtain $H_2 \leq \tilde{H}_3$ (note also that the equality $\tilde{H}_2 = \tilde{H}_3$ is not always satisfied, due to the possible 2-periodicity of the kernels $(P_\beta + P_\beta^*)/2$, for $\beta \geq 0$). \square

To conclude, we make the following observation.

LEMMA 4.5. *For any $n \geq 1$, we are assured of $H_3(n) = H_3$.*

PROOF. As we will give an entirely probabilist argument, let us consider again $(X_n, Y_n)_{n \geq 0}$ a Markov chain of distribution $\mathbb{P}_{(x,y),\beta}^{\otimes 2}$ for $(x, y) \in S^2$ and $\beta \geq 0$. For fixed $n \geq 1$, we consider the stopping time

$$T_\Delta^{(n)} \stackrel{\text{def.}}{=} \inf\{m \geq 0 : X_{nm} = Y_{nm}\}$$

because $H_3(n)$ can be deduced from its expectations

$$H_3(n) = \sup_{x \neq y \in S} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln \left(\mathbb{E}_{(x,y),\beta}^{\otimes 2} [T_\Delta^{(n)}] \right).$$

However, we note that $nT_\Delta^{(n)} \geq T_\Delta^{(1)}$; thus the bound $H_3 = H_3(1) \leq H_3(n)$ follows at once.

For the reverse inequality, we will rather use the description of H_3 presented in Section 2. Taking into account that, for any $x \in S$ and $\beta \geq 0$, there exists at least one point $y_\beta \in S$ such that $P_\beta(x, y_\beta) \geq 1/\text{card}(S)$, we deduce without difficulty that

$$\forall x \in S, \forall \beta \geq 0, \quad \mathbb{P}_{(x,x),\beta}^{\otimes 2} [(X_n, Y_n) \in \Delta] \geq \frac{1}{\text{card}(S)^{2n}}.$$

Using the Markov property in a traditional way, it appears that, for all $t \in \mathbb{N}$, all $\beta \geq 0$ and all $(x, y) \in S^2$,

$$\frac{1}{\text{card}(S)^{2n}} \mathbb{P}_{(x,y),\beta}^{\otimes 2} [T_\Delta^{(1)} \leq t] \leq \mathbb{P}_{(x,y),\beta}^{\otimes 2} [T_\Delta^{(n)} \leq \lfloor t/n \rfloor + 1].$$

This leads us to introduce, for $0 < \varepsilon < 1$,

$$t_\beta^{(n)}(\varepsilon) \stackrel{\text{def.}}{=} \inf \left\{ t \in \mathbb{N} : \inf_{(x,y) \in S^2 \setminus \Delta} \mathbb{P}_{(x,y),\beta}^{\otimes 2} [T_\Delta^{(n)} \leq t] \geq \varepsilon \right\}$$

since we then have

$$\forall \beta \geq 0, \quad t_\beta^{(n)} \left(\frac{\varepsilon}{\text{card}(S)^{2n}} \right) \leq \left\lfloor \frac{t_\beta^{(1)}(\varepsilon)}{n} \right\rfloor + 1.$$

The bound $H_3 \geq H_3(n)$ is now a consequence of Lemma 2.2, which can be rewritten to say that, for all $n \geq 1$ and $0 < \varepsilon < 1$,

$$H_3(n) = \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(t_\beta^{(n)}(\varepsilon)). \quad \square$$

Let us indicate that there are situations where Theorem 4.3 can be deduced in a simpler way. We begin by assuming furthermore that, for any $\beta \geq 0$, the kernel P_β is reversible with respect to μ_β . In this setting, for all $n \geq 1$, $1 - \lambda(\mu_\beta, P_\beta^n P_\beta^{*n})$ is the largest eigenvalue of the restriction of P_β^{2n} on $\mathbb{L}^2(\mu_\beta) \setminus \text{Vect}(\mathbb{1})$. The nonnegativity of the spectrum of $P_\beta^2 = P_\beta P_\beta^*$ enables us then to see that, for all $n \geq 1$,

$$\lambda(\mu_\beta, P_\beta^n P_\beta^{*n}) = 1 - (1 - \lambda(\mu_\beta, P_\beta^2))^n.$$

Let $0 < \varepsilon < 1$ and $\beta \geq 0$ be given; it appears that

$$\begin{aligned} \tau_\lambda(\beta) &= \inf\{n \geq 0 : (1 - \lambda(\mu_\beta, P_\beta^2))^n \leq 1 - \varepsilon\} \\ &= \left\lceil \frac{\ln(1 - \varepsilon)}{\ln(1 - \lambda(\mu_\beta, P_\beta^2))} \right\rceil. \end{aligned}$$

By considering asymptotic behavior for $\beta \geq 0$ large [special care is needed to treat the cases $h(1) = 0$ and $h(1) = +\infty$] in this equality, we get $H_3 = h(1)$ and consequently $H_3 \geq h(\infty)$. Thus Lemma 4.2 shows that, in symmetrical situations, we are assured of the identities $H_3 = h(1) = h(\infty)$. This approach can be extended to deal with the weak reversible cases, which verify the hypothesis that

$$\forall x, y \in S, \quad U(x) + V(x, y) = U(y) + V(y, x) \leq +\infty.$$

Indeed, if V satisfies that condition, we can find a new associated family of kernels $(\widehat{P}_\beta)_{\beta \geq 0}$ such that, for any $\beta \geq 0$, \widehat{P}_β is reversible with respect to its invariant probability $\widehat{\mu}_\beta$. For instance, we can consider, for all $\beta \geq 0$,

$$\begin{aligned} \forall x, y \in S, \quad \widehat{P}_\beta(x, y) &= \frac{\exp(-\beta V(x, y))}{\sum_{z \in S} \exp(-\beta V(x, z))}, \\ (9) \quad \forall x \in S, \quad \widehat{\mu}_\beta(x) &= \frac{\sum_{y \in S} \exp(-\beta V(x, y))}{\sum_{y, z \in S} \exp(-\beta V(y, z)) \exp(-\beta U(y))} \\ &\quad \times \exp(-\beta U(x)). \end{aligned}$$

Then we obtain $\widehat{H}_2 = \widehat{H}_3$ and we conclude by using the fact that these quantities only depend on V , namely $H_2 = \widehat{H}_2$ and $H_3 = \widehat{H}_3$. More generally, equations (9) show that an $S \times S$ matrix whose entries belong to $\bar{\mathbb{R}}_+$ is a rate matrix (i.e., there exists a corresponding family of Markov kernels) if and only if any of its rows contains a null entry.

Opposite to the previous considerations, we note that a delaying effect can occur for Dobrushin's coefficient. More precisely, it is easy to be convinced that, for any $n \geq 1$, the following limit exists:

$$\bar{h}(n) = - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(a(P_\beta^n)) \in [0, +\infty].$$

Furthermore, by reasoning quite similar to that of the beginning of this section, we see that the sequence $(\bar{h}(n))_{n \geq 1}$ is decreasing and stationary: there exists $n_1 \in \mathbb{N}^*$ such that, for all $n \geq n_1$, $\bar{h}(n) = \bar{h}(n_1) \stackrel{\text{def.}}{=} \bar{h}(\infty)$. It also appears that $H_3 \leq \bar{h}(\infty)$. However, the equality is no longer enforced, as the next counterexamples show, for $N \geq 3$: on the state space $\{1, 2, \dots, N\}$, we consider the rate matrix V given by

$$\forall 1 \leq x, y \leq N, \quad V(x, y) = \begin{cases} 0, & \text{if } x = y, \\ 1, & \text{if } |x - y| = 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

We compute that the associated quasipotential U is identically null, that $H_3 = 1$ and that $\bar{h}(\infty) = \lceil N/2 \rceil$, proving there is no possibility of bounding $\bar{h}(\infty)$ from above in terms of H_3 , uniformly in the cardinal of the state space S .

Nevertheless, due to the relations we have mentioned in Section 3 between the spectral gap, the isoperimetric constant and the (classic or modified) logarithmic Sobolev coefficient, we see that there is no delay for the latter ones. A heuristic reason for that property is that their definitions use the corresponding invariant probability and the presence of such a time asymptotic object really helps.

To end this section, here are a few supplementary observations about related extensions:

REMARK 4.6. (a) The previous considerations show that the following characterization also holds:

$$H_3 = \max_{n \geq 1} H_2(n) \geq H_2(1) = H_2.$$

Certainly the simplest example for which $H_3 > H_2$ is given by $S = \{0, 1\}$ and

$$V = \begin{pmatrix} 1 & 0 \\ 0 & +\infty \end{pmatrix},$$

for which $H_2 = 0$ and $H_3 = 1$. In particular, contrary to the analogous property for H_3 proved in Lemma 4.5, we are not always assured of $H_2(n) = H_2$ for all $n \geq 1$. Let us recall a probabilist interpretation of $H_2(n)$, for $n \geq 1$: choose an arbitrary global minimum x_0 [i.e., a point of S satisfying $U(x_0) = 0$] and define

$$T_{x_0}^{(n)} = \inf\{m \geq 0 : X_{nm} = x_0\};$$

we have (cf. [12] or [2])

$$H_2(n) = \max_{x \in S} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\mathbb{E}_{x, \beta}[T_{x_0}^{(n)}]).$$

The occurrence of $\max_{n \geq 1} H_2(n) > H_2$ is then seen to be permitted by the eventual “nonstability” of x_0 , that is, the possibility that $V(x_0, x_0) > 0$ [since the proof of Lemma 4.5 can be reused to imply that $H_2(n) = H_2$ for all $n \geq 1$, if $V(x_0, x_0) = 0$].

However, the subset \mathcal{M} of all the global minima is asymptotically stable, in the sense that

$$\forall x \in \mathcal{M}, \quad \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(P_\beta(x, \mathcal{M})) = 0;$$

thus we obtain that, for all $n \geq 1$, $H_1(n) = H_1 \stackrel{\text{def.}}{=} H_1(1)$, where

$$\forall n \geq 1, \quad H_1(n) = \max_{x \in S} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\mathbb{E}_{x, \beta}[T_{\mathcal{M}}^{(n)}])$$

with

$$T_{\mathcal{M}}^{(n)} = \inf\{m \geq 0 : X_{nm} \in \mathcal{M}\}.$$

The critical roles of the three constants H_1 , H_2 and H_3 are described by Catoni in [2] for generalized Metropolis algorithms; see also [11] for a spectral interpretation of H_1 (given in some continuous time special cases, but that could be extended to the present setting).

(b) We strongly believe that the picture given here could be made even more precise by the introduction of (low temperature) asymptotical periodicity classes, at least when $H_2 < H_3$. The period should be the larger common divisor of the natural numbers $n \geq 1$ satisfying $H_2(n) = H_3$. In particular this would allow us to show that, in Lemma 4.1, we can choose $n_0 \leq \text{card}(S) - 1$.

(c) The existence of such a finite $n_0 \in \mathbb{N}^*$ satisfying

$$(10) \quad H_3 = - \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln(\lambda(\mu_\beta, P_\beta^{n_0} P_\beta^{n_0*}))$$

is technically important for a spectral treatment of simulated annealing. In this context, we need to make another assumption on the variations of the kernel P_β with respect to β , which is not a consequence of (1) nor implies it:

(H) There exists a constant $K > 0$ such that, for all $\beta, \beta' \geq 0$ (or sufficiently large),

$$\forall x, y \in S, \quad \left| \ln \left(\frac{P_\beta(x, y)}{P_{\beta'}(x, y)} \right) \right| \leq K |\beta - \beta'|$$

(with the special convention that $0/0 = 1$), which in particular implies, via the representation in terms of subtrees (cf. [7]), that

$$\forall x \in S, \quad \left| \ln \left(\frac{\mu_\beta(x)}{\mu_{\beta'}(x)} \right) \right| \leq K (\text{card}(S) - 1) |\beta - \beta'|.$$

Now let $(X_n)_{n \geq 0}$ be a time-inhomogeneous Markov chain starting from an initial distribution m_0 and whose kernel of transition probabilities at any time $n \in \mathbb{N}$ is $P_{\beta(n)}$, where, for example, $\beta(n) = k^{-1} \ln(n + 1)$ with $k > H_3$. Using the approach presented in [13], it can be shown that, for large n , the distribution of X_n will be

closed to $\mu_{\beta(n)}$, in total variation norm or relative entropy sense. Thus X_n will eventually concentrate on the global minima of U , giving to them weights close to those granted by $\mu_{\beta(n)}$ (which itself may not converge under the assumptions of the Introduction). Due to the delaying effect for Dobrushin’s coefficients, we could not reach this precision by only using them. Indeed, we would have to play with very large times, which are especially difficult to handle in inhomogeneous settings. On the contrary, (10) is more easy to perturb. This puts forward the advantage of considering singular values or similar ergodic constants. The constant H_3 is critical for the above behavior, which means that this property will no longer be satisfied for every initial probability m_0 if $0 < k < H_3$. This can be seen for instance by applying the estimation of exit times presented in [12] to simulated annealing algorithms associated with the tensorized kernels $(P_\beta^{\otimes 2})_{\beta \geq 0}$ and by taking into account the fact that in the maximum appearing in (8) [or hidden in (2)], we can choose $(x, y) \notin \Delta$ such that $x, y \in S$ are both global minima of U . However, again, the development of the above point (b) would help the understanding of this behavior.

(d) Let us give an example of application for the strong ergodicity in law (namely the convergence in law for large time to a probability not depending on the initial distribution) for generalized simulated annealing with a cost function depending (a little) on time. On an oriented irreducible finite graph, we are given for any edge (x, y) a C^1 mapping $\mathbb{R}_+ \ni \beta \mapsto U_\beta(x, y) \in \mathbb{R}$ [if (x, y) is not an edge of the graph, $U_\beta(x, y) \equiv -\infty$]. We assume the existence of $U(x, y) := \lim_{\beta \rightarrow +\infty} U_\beta(x, y)$ in \mathbb{R} for any edge (x, y) and that $\limsup_{\beta \rightarrow +\infty} \beta |\partial_\beta U_\beta(x, y)| < +\infty$. Then we consider, for any $\beta \geq 0$, the Markovian kernel P_β defined by

$$\forall x, y \in S, \quad P_\beta(x, y) = \frac{\exp(\beta U_\beta(x, y))}{\sum_{z \in S} \exp(\beta U_\beta(x, z))}.$$

It is immediate to verify that (1) and (H) are satisfied. Let $X = (X_n)_{n \in \mathbb{N}}$ be an inhomogeneous Markov chain associated as above with an evolution of the inverse of the temperature given by $\beta_n = k^{-1} \ln(1 + n)$, for all $n \geq 0$. As a consequence of the previous considerations, H_3 is the infimum of the values k such that X is strongly ergodic in law [and the limit distribution can be written in terms of the $U(x, y)$, where (x, y) is an edge].

(e) For $n \geq 1$, we define

$$H_2^{(2)}(n) = \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln[\lambda(\mu_\beta^{\otimes 2}, (P_\beta^n)^{\otimes 2})],$$

$$h^{(2)}(n) = \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln[\lambda(\mu_\beta^{\otimes 2}, (P_\beta^n)^{\otimes 2} (P_\beta^{n*})^{\otimes 2})].$$

Let x_0 be a fixed global minimum; writing

$$\forall n \geq 1, \quad T_{(x_0, x_0)}^{(n)} = \inf\{m \geq 0 : (X_{nm}, Y_{nm}) = (x_0, x_0)\}$$

the former quantities also admit the interpretation

$$H_2^{(2)}(n) = \max_{(x,y) \in S^2} \lim_{\beta \rightarrow +\infty} \beta^{-1} \ln \left(\mathbb{E}_{(x,y),\beta}^{\otimes 2} [T_{(x_0,x_0)}^{(n)}] \right).$$

Thus it appears that, for all $n \geq 1$,

$$H_3 \leq H_2^{(2)}(1) \leq H_2^{(2)}(n).$$

Note, however, that the subset of local minima for $U \oplus U$ is just the product of the set of local minima for U with itself; thus if $l \geq 1$ is such that (7) is satisfied, then we also obtain

$$H_2^{(2)}(l) = h^{(2)}(l) = h(l) = H_2(l),$$

the middle equality coming from a general property of singular values under the Kronecker product (cf. Theorem 4.2.15 of [9]), saying that, for all $n \geq 1$ and all $\beta \geq 0$, $\lambda(\mu_\beta^{\otimes 2}, (P_\beta^n)^{\otimes 2} (P_\beta^{n*})^{\otimes 2}) = \lambda(\mu_\beta, P_\beta^n P_\beta^{n*})$. Taking into account that our previous proofs indeed show the equality $H_2(l) = H_3$, we conclude that

$$H_2^{(2)}(1) = H_3$$

and consequently that, for all $n \geq 1$,

$$H_2^{(2)}(n) = H_3(n) = H_3.$$

Notice that this new characterization of H_3 (or that of Theorem 4.3) seems not so trivial to work out directly from its formulation in terms of elevations of paths given in (8).

(f) Let $\mathcal{M}_1, \dots, \mathcal{M}_\kappa$ be the classes of equivalence for the restriction to the subset of global minima \mathcal{M} of the relation introduced in (6) (in fact, for $1 \leq i \leq \kappa$, there exists $1 \leq k_i \leq K$ such that $\mathcal{M}_i = \mathcal{L}_{k_i}$). The conjectured simulated annealing interpretation of the constants H_2 and H_1 is that they should be critical respectively for the next behaviors, according to the position of the parameter k in front of the previous inverse logarithmic evolution of the temperature with respect to these constants [always under (H)]:

- (i) for all initial distributions and for all $1 \leq i \leq \kappa$, asymptotically when $n \geq 0$ goes to infinity, the probability that X_n belongs to \mathcal{M}_i is close to the weight $\mu_{\beta(n)}(\mathcal{M}_i)$;
- (ii) the probability that X_n belongs to \mathcal{M} converges to 1 for large $n \geq 0$.

This latter characterization of H_1 was proved by Trouvé [17] under a hypothesis a little stronger than just a large deviation behavior for the entries of the Markov kernels $(P_\beta)_{\beta \geq 0}$, but already, in the case of a general cost function V , generalizing results obtained by Catoni [1] in the reversible situation. The corresponding property of H_2 for generalized Metropolis algorithms, meaning that, for all $\eta > 0$,

$$\sup_{x \in S, 1 \leq i \leq K} \lim_{\beta \rightarrow +\infty} |\mathbb{P}_{\beta,x} [X_{\lfloor \exp(\beta(H_2+\eta)) \rfloor} \in \mathcal{M}_i] - \mu_\beta(\mathcal{M}_i)| = 0$$

and

$$\sup_{x \in S, 1 \leq i \leq K} \limsup_{\beta \rightarrow +\infty} |\mathbb{P}_{\beta, x}[X_{[\exp(\beta(H_2 - \eta))]} \in \mathcal{M}_i] - \mu_\beta(\mathcal{M}_i)| > 0$$

would also be new. It should come from the precise behavior of conveniently renormalized exit times which enables us to agglutinate the points of each \mathcal{M}_i , for $1 \leq i \leq \kappa$ (see [12]).

Acknowledgments. I am very grateful to the referee and to an Associate Editor, whose careful readings pointed out and corrected several mistakes in a first version of this paper. I also thank the Instituto Nacional de Matemática Pura e Aplicada, CNPq, Brazil, where this work was done, for its hospitality and support.

REFERENCES

- [1] CATONI, O. (1992). Rough large deviation estimates for simulated annealing: application to exponential schedules. *Ann. Probab.* **20** 1109–1146.
- [2] CATONI, O. (1999). Simulated annealing algorithms and Markov chains with rare transitions. *Séminaire de Probabilités XXXIII. Lecture Notes in Math.* **1709** 69–119. Springer, Berlin.
- [3] CATONI, O. and CERF, R. (1995–1997). The exit path of a Markov chain with rare transitions. *ESAIM Probab. Statist.* **1** 95–144 (electronic).
- [4] COHEN, J. E., IWASA, Y., RĂUȚU, G., RUSKAI, M. B., SENETA, E. and ZBĂGANU, G. (1993). Relative entropy under mappings by stochastic matrices. *Linear Algebra Appl.* **179** 211–235.
- [5] DEL MORAL, P., LEDOUX, M. and MICLO, L. (2000). About supercontraction properties of Markov kernels. Preprint 2001-01. Laboratoire de Statistique et Probabilités, Univ. Paul Sabatier, Toulouse.
- [6] FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.
- [7] FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, Berlin.
- [8] HOLLEY, R. and STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 553–569.
- [9] HORN, R. A. and JOHNSON, C. R. (1991). *Topics in Matrix Analysis*. Cambridge Univ. Press.
- [10] MATHIEU, P. (1995). Spectra, exit times and long time asymptotics in the zero-white-noise limit. *Stochastics Stochastics Rep.* **55** 1–20.
- [11] MICLO, L. (1995). Comportement de spectres d’opérateurs de Schrödinger à basse température. *Bull. Sci. Math.* **119** 529–553.
- [12] MICLO, L. (1996). Sur les problèmes de sortie discrets inhomogènes. *Ann. Appl. Probab.* **6** 1112–1156.
- [13] MICLO, L. (1997). Remarques sur l’hypercontractivité et l’évolution de l’entropie pour des chaînes de Markov finies. *Séminaire de Probabilités XXXI. Lecture Notes in Math.* **1655** 136–167. Springer, Berlin.
- [14] MICLO, L. (1998). Une variante de l’inégalité de Cheeger pour les chaînes de Markov finies. *ESAIM Probab. Statist.* **2** 1–21.
- [15] SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. *Lectures on Probability Theory and Statistics. Ecole d’Eté de Probabilités de Saint-Flour XXVI. Lecture Notes in Math.* **1665**. Springer, Berlin.

- [16] TROUVÉ, A. (1996). Cycle decompositions and simulated annealing. *SIAM J. Control Optim.* **34** 966–986.
- [17] TROUVÉ, A. (1996). Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms. *Ann. Inst. H. Poincaré Probab. Statist.* **32** 299–348.

LABORATOIRE DE STATISTIQUE ET PROBABILITÉS, UMR 5583
CNRS AND UNIVERSITÉ PAUL SABATIER
118, ROUTE DE NARBONNE
31062 TOULOUSE CEDEX 4
FRANCE
E-MAIL: miclo@cict.fr