The Annals of Applied Probability
1994, Vol. 4, No. 4, 1012–1056

L^2 CONVERGENCE OF TIME NONHOMOGENEOUS MARKOV PROCESSES: I. SPECTRAL ESTIMATES

BY JEAN-DOMINIQUE DEUSCHEL AND CHRISTIAN MAZZA[1]

ETH Zurich and University of Fribourg

We study the convergence of nonsymmetric annealing processes, extending the classical Dirichlet form approach to a broad class of Markov chains with exponentially vanishing transition functions. We show that both the true and symmetrized spectral gaps are logarithmically equivalent, and give robust estimates for the gap using geometric methods.

Contents

1. Introduction. Let $U: \Omega \to R$ be given function on a finite set Ω. Simulated annealing is a Monte Carlo method which enables one to locate the set

$$U_{\min} \equiv \left\{ x \in \Omega : U(x) = \min_{y \in \Omega} U(y) \right\}$$

Received November 1992; revised March 1994.
[1]Work supported by the Swiss National Science Foundation.
AMS 1991 *subject classifications.* Primary 60J27; secondary 60F10, 93E25, 15A18, 60J60
Key words and phrases. Dirichlet forms, first hitting time, geometric bounds, L^2 convergence, Metropolis, nonsymmetric Markov chains, spectral gap, ultrametricity.

of minimizing elements of U. In probabilistic terms, the algorithm consists of a time inhomogeneous Markov chain $\{X_t: t \geq 0\}$ generated by irreducible transition matrices $L_{\beta(t)} = \{q_{\beta(t)}(x, y),\ x, y \in \Omega\}$ depending on a parameter $\beta: \mathbb{R}^+ \to \mathbb{R}^+$. The transition matrices have the property that for fixed $t \in \mathbb{R}^+$ the corresponding invariant distribution $\pi_{\beta(t)}$ concentrates on the set $U_{\min}$ for large $\beta(t)$. The increasing schedule β is then chosen to make the distribution of the chain converge to $\pi_{\beta(t)}$ as $t \to \infty$. Since the early developments of Metropolis, there exists an extensive mathematical literature investigating the convergence of the algorithm; see Hajek (1988), Chiang and Chow (1988a, b), Holley and Stroock (1989) and its various applications. One of the main interests is to estimate the rate of convergence in terms of the structure of the set Ω and the profile of the function U. In particular, Holley and Stroock have proposed a universal L^2 method based on Dirichlet forms and spectral gap estimates. This approach has been implemented by Goetze (1992) yielding explicit constants for symmetric annealing of the Metropolis type. The aim of this paper is to extend the method of these two works to nonsymmetric situations, more precisely to the Freidlin–Wentzell family $\mathscr{L}$ of transition matrices $L_\beta = \{q_\beta(x, y),\ x, y \in \Omega\}$, $\beta \geq 0$ with exponentially vanishing coefficients:

$$(1.1)\quad q_\beta(x, y) \asymp \exp(-\beta V(x \to y)) \quad \text{as } \beta \to \infty \text{ with } V(x \to y) \geq 0.$$

The broad family $\mathscr{L}$ contains various dynamics used for simulations in several contexts like statistical mechanics (Glauber dynamics), image processing (sequential and parallel Gibbs samplers), neural computing (Boltzmann machines, evolutionary algorithms) and optimization (simulated annealing).

Our starting point will be the following general convergence result for nonhomogeneous, time continuous Markov chains. Assume that the invariant distribution π_β is constant on $U_{\min}$ and concentrates exponentially fast on $U_{\min}$:

$$(1.2)\qquad \pi_\beta(U^c_{\min}) \leq Ae^{-\beta B},$$

for some $A = A(U^c_{\min})$, $B = B(U^c_{\min}) > 0$, and that

$$(1.3)\qquad \left|\frac{d}{d\beta}\log \pi_\beta(x)\right| \leq M, \qquad x \in \Omega,$$

for some $M > 0$. Next, let $\tilde{C}(\beta)$ be the spectral gap of the symmetrized operator $\tilde{L}_\beta = (L_\beta + L^*_\beta)/2$, where L^*_β denotes the π_β-adjoint of L_β, and let $m, K \in \mathbb{R}^+$ be such that

$$(1.4)\qquad \tilde{C}(\beta) \geq Ke^{-m\beta}.$$

Then, choosing β of the form

$$\beta(t) = \frac{1}{m}\log(1 + \rho t) \quad \text{where} \quad \rho = \frac{2mK}{3M},$$

yields

$$P_x(X(t) \notin U_{\min}) \le 5A(1+\rho t)^{-B/m} \tag{1.5}$$
$$+ A^{1/2}\big(\pi_0(x)^{-1} - 1\big)^{1/2}(1+\rho t)^{-(B+2M)/(2m)}.$$

Here P_x denotes the law of the Markov chain starting at $x \in \Omega$. The symmetric Metropolis process (see Example 3.1.9) is a generic example where (1.2) and (1.3) hold.

In the discrete time case, Azencott (1988) obtains a very similar result under stricter conditions, using the Jordan decomposition. The Dirichlet form method is much simpler and gives explicit constants. On the other hand, Azencott's result is apparently better since it is formulated in terms of the true spectral gap, $C(\beta)$, of L_β, which is in general larger than $\tilde{C}(\beta)$.

We prove in our main result, Theorem 3.2.11, that for transition matrices of the Freidlin–Wentzell form (1.1) both the true spectral gap, $C(\beta)$, and the symmetrized spectral gap, $\tilde{C}(\beta)$, are logarithmically equivalent:

$$\lim_{\beta\to\infty} \frac{1}{\beta}\log(C(\beta)) = \lim_{\beta\to\infty} \frac{1}{\beta}\log\big(\tilde{C}(\beta)\big) = -m. \tag{1.6}$$

Our proof is based on Wentzell and Freidlin's graph method, which also gives an adequate expression for the invariant distribution π_β and easily verifies assumptions (1.2) and (1.3). However, this method is not suitable for estimating the constant K in (1.4) and thus we apply a geometric Poincaré technique, as in Diaconis and Stroock (1991) and Goetze (1992). We show the asymptotic equivalence of both techniques in Theorem 4.1.13 and give a robust estimate of K in terms of the spectral gap of an associate Metropolis chain using a filling method. We also derive some related spectral estimates for the first hitting time of $U_{\min}$ using a collapsed chain technique.

The argument, based on the underlying ultrametric structure, sheds a new light on the geometric bounds. In particular, we illustrate the interplay between getting the optimal rate m and the biggest K in (1.4). Moreover, the spectral estimates give some information concerning polynomial time simulations with Metropolis dynamics on large combinatorial sets Ω_d, for which $|\Omega_d|$ grows exponentially fast in $d \in \mathbb{N}$.

The paper is organized as follows: In Section 2 we derive some properties of the spectral gap and present the general convergence result (1.5) for time nonhomogeneous Markov chains. Section 3 introduces the Wentzell–Freidlin class $\mathscr{L}$ and proves (1.6). In Section 4, we derive the geometric estimates for the spectral gap and give some examples related to Metropolis chains in Section 5. In Section 6, we show the logarithmic equivalence for the symmetrized and true spectral gaps of diffusion on a compact manifold with small noise. Finally, the Appendix contains the proof of the convergence result (1.5).

2. L^2 convergence.

2.1. *Preliminaries.* In this section we introduce the basic notation and recall a few elementary facts about spectral gaps.

Let $L\colon \Omega\times\Omega\to\mathbb{R}$, $L=\{q(x,y),\ x,y\in\Omega\}$, be a transition function on a finite set Ω:

$$q(x,y)\ge 0 \quad \text{for } x\ne y \quad \text{and} \quad q(x,x)=-\sum_{y\ne x} q(x,y).$$

We will assume that L is *irreducible* and denote by π its unique invariant distribution: $\sum_x \pi(x)q(x,y)=0$, $y\in\Omega$. For $g,f\in L^2(\pi)$, we write

$$\langle f,g\rangle_\pi := \sum_{x\in\Omega} f(x)g(x)\pi(x) \quad \text{and} \quad \|f\|_\pi := \|f\|_{L^2(\pi)}.$$

By irreducibility $\pi(x)>0$, for $x\in\Omega$, and we can define L^*, the π dual of L:

$$\text{(2.1.1)}\qquad \begin{aligned} q^*(x,y) &:= \frac{q(y,x)\pi(y)}{\pi(x)} \quad \text{for } x\ne y \quad \text{and}\\ q^*(x,x) &= -\sum_{y\ne x} q^*(x,y). \end{aligned}$$

Define the symmetrized operator $\tilde L := (L+L^*)/2$. Then,

$$\langle f,Lg\rangle_\pi=\langle L^*f,g\rangle_\pi,\qquad \langle f,\tilde Lg\rangle_\pi=\langle \tilde Lf,g\rangle_\pi,\qquad f,g\in L^2(\pi).$$

Let $\{0=\lambda_1,\lambda_2,\dots,\lambda_{|\Omega|}\}\subset\mathbb{C}$ and $\{0=\tilde\lambda_1,\tilde\lambda_2,\dots,\tilde\lambda_{|\Omega|}\}\subset\mathbb{R}$ be the eigenvalues, counted with their algebraic multiplicities, of L (or L^*) and $\tilde L$. Define C and $\tilde C\in\mathbb{R}^+$,

$$C := -\max\{\Re(\lambda_i);\, i\ne 1\},\qquad \tilde C := -\max\{\tilde\lambda_i;\, i\ne 1\},$$

to be the *spectral gaps* of L and $\tilde L$. Next let $\mathscr{E}_\pi$ be the Dirichlet form associated with $\tilde L$:

$$\text{(2.1.2)}\qquad \mathscr{E}_\pi(\varphi,\varphi) := -\langle \tilde L\varphi,\varphi\rangle_\pi=\tfrac12\sum_{x,y}Q(x,y)\{\varphi(y)-\varphi(x)\}^2,$$

where

$$Q(x,y)=Q(y,x):=\tfrac12\big(\pi(x)q(x,y)+\pi(y)q(y,x)\big).$$

Then the spectral gap $\tilde C$ of $\tilde L$ is given by variational formula

$$\text{(2.1.3)}\qquad \tilde C=\inf\{\mathscr{E}_\pi(\varphi,\varphi);\ \|\varphi\|_\pi=1,\ \langle\varphi,1\rangle_\pi=0\}.$$

2.1.4. LEMMA. *We have*

$$\tilde C\le C.$$

PROOF. Without loss of generality, we may assume that $C=-\Re(\lambda_2)$ and $\tilde C=-\tilde\lambda_2$. Let φ be a eigenvector of L such that $L\varphi=\lambda_2\varphi$ and $\|\varphi\|_\pi=1$. Since $\lambda_2\ne 0$, we get $\langle 1,\Re\varphi\rangle_\pi=\langle 1,\Im\varphi\rangle_\pi=0$. Thus

$$\begin{aligned} -2\Re(\lambda_2) &= \langle\overline{\varphi},(-L)\varphi\rangle_\pi+\langle\varphi,(-L)\overline{\varphi}\rangle_\pi=2\langle\varphi,(-\tilde L)\overline{\varphi}\rangle_\pi\\ &=2\big(\langle\Re\varphi,(-\tilde L)\Re\varphi\rangle_\pi+\langle\Im\varphi,(-\tilde L)\Im\varphi\rangle_\pi\big)\ge -2\tilde\lambda_2. \qquad\square \end{aligned}$$

The spectral gap $\tilde{C}$ controls the rate at which a Markov chain converges to its invariant distribution. More precisely, let ν be a given probability distribution on Ω and let $f(x) = (\nu(x)/\pi(x))$. Set

$$f_t(x) = e^{tL^*}f(x), \qquad \tilde{f}_t = e^{t\tilde{L}}f(x).$$

Then $\nu_t(x) = f_t(x)\pi(x)$, respectively, $\tilde{\nu}_t(x) = \tilde{f}_t(x)\pi(x)$, is the distribution of a continuous-time Markov chain with initial distribution ν generated by L, respectively, by $\tilde{L}$, and

$$\|\nu_t - \pi\|_{\mathrm{var}} = \|f_t - 1\|_{L^1(\pi)} \le \|f_t - 1\|_\pi,$$

where $\|\cdot\|_{\mathrm{var}}$ is the variational distance between the probability measures ν_t and π [see Diaconis and Stroock (1992) and Fill (1991)]. We have

$$\begin{aligned}\|f_t - 1\|_\pi^2 &= \langle f_t - 1, f_t - 1\rangle_\pi = \langle (f-1)_t, (f-1)_t\rangle_\pi \\ &= \langle (f-1), e^{tL}\cdot e^{tL^*}(f-1)\rangle_\pi\end{aligned}$$

and

$$\begin{aligned}\|\tilde{f}_t - 1\|_\pi^2 &= \langle \tilde{f}_t - 1, \tilde{f}_t - 1\rangle_\pi = \langle (\widetilde{f-1})_t, (\widetilde{f-1})_t\rangle_\pi \\ &= \langle (f-1), c^{t\tilde{L}}\cdot e^{t\tilde{L}}(f-1)\rangle_\pi = \langle (f-1), e^{t(L+L^*)}(f-1)\rangle_\pi.\end{aligned}$$

In particular, we have the following lemma.

2.1.5. LEMMA. *Assume that L is normal; that is, L and L^* commute. Then $C = \tilde{C}$ with*

$$\|f_t - 1\|_\pi^2 = \|\tilde{f}_t - 1\|_\pi^2.$$

PROOF. Simply note that, for normal operators, L and L^* have the same eigenvectors and $e^{tL}\cdot e^{tL^*} = e^{t(L+L^*)}$. □

In general, we have

$$\begin{aligned}\frac{d}{dt}\|f_t - 1\|_\pi^2 &= 2\langle L^*(f-1)_t, (f-1)_t\rangle_\pi = 2\langle \tilde{L}(f-1)_t, (f-1)_t\rangle_\pi \\ &= -2\mathscr{E}_\pi((f-1)_t, (f-1)_t) \le -2\tilde{C}\|f_t - 1\|_\pi^2.\end{aligned}$$

Thus

$$\|f_t - 1\|_\pi \le e^{-\tilde{C}t}\|f - 1\|_\pi.$$

The same argument shows

$$\|\tilde{f}_t - 1\|_\pi \le e^{-\tilde{C}t}\|f - 1\|_\pi.$$

However, although by the spectral decomposition theorem for $\tilde{L}$,

$$\inf\left\{-\lim_{t\to\infty}\frac{1}{t}\log\|\tilde{f}_t - 1\|_\pi; \|f - 1\|_\pi \ne 0\right\} = \tilde{C},$$

it may well be that by the Jordan decomposition of (the not necessarily diagonalizable) L,

$$\inf\left\{ -\lim_{t\to\infty} \frac{1}{t}\log\|f_t - 1\|_\pi ; \|f-1\|_\pi \neq 0 \right\} = C > \tilde{C}.$$

2.1.6. EXAMPLE (Random walks on the the torus). This is a generic situation where L and L^* commute and, therefore, $C = \tilde{C}$. Let Ω be a finite Abelian group, for example, the N-torus. Assume that L is the generating matrix of a random walk on Ω:

$$q(x,y) = q(x-y), \qquad x \neq y,$$

with invariant distribution $\pi(x) \equiv 1/|\Omega|$. Then

$$q^*(x,y) = q(y,x) = q(y-x)$$

and, therefore,

$$\begin{aligned} L^* \cdot L(x,y) &= \sum_{\omega\in\Omega} q^*(x,\omega)q(\omega,y) = \sum_{\omega\in\Omega} q(\omega - x)q(\omega - y) \\ &= \sum_{\omega\in\Omega} q(\omega + y - x)q(\omega) = \sum_{\omega\in\Omega} q(y-\omega)q(x-\omega) \\ &= \sum_{\omega\in\Omega} q(x,\omega)q^*(\omega,y) = L\cdot L^*(x,y). \end{aligned}$$

Thus L and L^* commute.

Finally, let us introduce the notion of *collapsed chain*, which we will use in the next section as a tool for estimating the probability of first hitting times. Let $\Omega_* \subset \Omega$ be such that $\Omega_* \neq \varnothing$ and $\Omega_* \neq \Omega$. Set $\Omega^* := \Omega \setminus \Omega_*$.

2.1.7. DEFINITION. Let $L[\Omega_*]$ be the transition function on $\overline{\Omega} := \Omega^* \cup \{\delta\}$, obtained from L by collapsing the elements of Ω_* into one element δ, defined by

$$\begin{aligned} q[\Omega_*](x,y) &= q(x,y) \quad \text{for } x,y \in \Omega^*, \\ q[\Omega_*](x,\delta) &:= \sum_{y\in\Omega_*} q(x,y), \qquad x\in\Omega^*, \\ q[\Omega_*](\delta,x) &:= \pi[\Omega_*](\delta)^{-1} \sum_{y\in\Omega_*} \pi(y)q(y,x), \qquad x \in \Omega^*, \\ q[\Omega_*](\delta,\delta) &:= -\sum_{x\in\Omega^*} q[\Omega_*](\delta,x), \end{aligned}$$

where $\pi[\Omega_*](\cdot)$, the invariant measure associated with $L[\Omega_*]$, is given by

$$\pi[\Omega_*](x) = \pi(x), \quad \text{for } x\in\Omega^*, \quad \text{and} \quad \pi[\Omega_*](\delta) = \sum_{y\in\Omega_*} \pi(y).$$

We denote by $\mathscr{E}[\Omega_*]_{\pi[\Omega_*]}$ the corresponding Dirichlet form and by $\tilde{C}[\Omega_*]$ its spectral gap:

$$\tilde{C}[\Omega_*] := \inf\{\mathscr{E}[\Omega_*]_{\pi[\Omega_*]}(\varphi,\varphi); \langle\varphi,\varphi\rangle_{\pi[\Omega_*]} = 1, \langle\varphi,1\rangle_{\pi[\Omega_*]} = 0\}$$
$$= \inf\{\mathscr{E}_\pi(\varphi,\varphi); \varphi|\Omega_* \text{ is constant}, \langle\varphi,\varphi\rangle_\pi = 1, \langle\varphi,1\rangle_\pi = 0\} \geq \tilde{C}.$$

2.2. *L^2 convergence results for time nonhomogeneous Markov chains.* We give some general results about the convergence of nonsymmetric Markov chains using the ideas of Holley and Stroock (1989) and Goetze (1992).

Let $L.: \mathbb{R}^+ \times \Omega \times \Omega \to \mathbb{R}$, $L_\beta = \{q_\beta(x,y),\ x, y \in \Omega\}$, be a one-parameter family of transition functions of Ω. We will assume that L_β is *irreducible* and denote by π_β the unique invariant probability distribution, by L_β^* the π_β dual of L_β [see (2.1.1)], by $\tilde{L}_\beta := (L_\beta + L_\beta^*)/2$ the symmetrized operator, by $\mathscr{E}_{\pi_\beta}$ the associated Dirichlet form [see (2.1.2)] and by $\tilde{C}(\beta)$ the spectral gap associated with $\tilde{L}_\beta$. Next suppose that $\beta(\cdot) \in C^1(0,+\infty)$ with $\dot\beta(t) = (d/dt)\beta(t) > 0$, $\forall\ t > 0$, and consider $P_{s,t} = (p_{s,t}(x,y),\ x, y \in \Omega)$, $0 \leq s \leq t < \infty$, the Markov semigroup generated by $L_{\beta(t)}$, that is, the solution to the forward equation

$$\frac{d}{dt}(P_{s,t}\varphi)(x) = \big(P_{s,t}[L_{\beta(t)}\varphi]\big)(x),$$
$$(P_{s,s}\varphi)(x) = \varphi(x).$$

Let $P_{s,t}^* = (p_{s,t}^*(x,y),\ x, y \in \Omega)$ denote the $\pi_{\beta(t)}$ adjoint of $P_{s,t}$:

$$p_{s,t}^*(x,y) := \frac{p_{s,t}(y,x)\pi_{\beta(t)}(y)}{\pi_{\beta(t)}(x)}.$$

For simplicity, we write $P_t = P_{0,t}$ and $P_t^* = P_{0,t}^*$. For an initial probability ν_0, let ν_t be the distribution of the Markov chain at time $t \geq 0$. Introduce the densities

$$f_t(x) = \frac{\nu_t(x)}{\pi_{\beta(t)}(x)}, \qquad t \geq 0.$$

Now, we make our *main assumptions*: There exist positive constants K, m and $M \in (0,+\infty)$ such that

$$\tilde{C}(\beta) \geq Ke^{-\beta m} \tag{2.2.1}$$

and

$$\left|\frac{d}{d\beta}\log\pi_\beta(x)\right| \leq M \qquad \forall\ x \in \Omega. \tag{2.2.2}$$

2.2.3. LEMMA. *Assume* (2.2.1) *and* (2.2.2), *and choose* β *such that*

$$\beta(t) = \frac{1}{m}\log(1+\rho t), \qquad t \geq 0 \textit{ where } \rho = \frac{2mK}{3M}. \tag{2.2.4}$$

Then

$$\|f_t - 1\|_{\pi_{\beta(t)}} < 1 + \exp(-M\beta(t))\|f_0 - 1\|_{\pi_{\beta(0)}}. \tag{2.2.5}$$

Next, let $\Omega^* \subset \Omega$ *with* $\Omega_* = \Omega \setminus \Omega^* \neq \varnothing$, *and assume that*

$$\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}} \text{ is constant on } \Omega_*. \tag{2.2.6}$$

Then

$$\|f_t - 1\|_{\pi_{\beta(t)}} \leq 4\pi_{\beta(t)}(\Omega^*)^{1/2} + \exp(-M\beta(t))\|f_0 - 1\|_{\pi_{\beta(0)}}. \tag{2.2.7}$$

The proof is given in the Appendix.

2.2.8. COROLLARY. *Let* $A = A(\Omega^*) \geq 0$ *and* $B = B(\Omega^*) \in (0, M)$ *be such that*

$$\pi_{\beta(t)}(\Omega^*) \leq A\exp(-\beta(t)B). \tag{2.2.9}$$

Assume (2.2.1) *and* (2.2.2), *and choose* $\beta(\cdot)$ *as in* (2.2.4). *Then, for each* $x \in \Omega$,

$$\begin{aligned} &P_x(X_t \notin \Omega_*) \\ &\quad\leq \pi_{\beta(t)}(\Omega^*)^{1/2}\Big(1 + \exp(-M\beta(t))\big(\pi_{\beta(0)}(x)^{-1} - 1\big)^{1/2}\Big) + \pi_{\beta(t)}(\Omega^*) \\ &\quad\leq 2A^{1/2}(1+\rho t)^{-B/2m} + \big(\pi_{\beta(0)}(x)^{-1} - 1\big)^{1/2} A^{1/2}(1+\rho t)^{-(B+2M)/(2m)}. \end{aligned}$$

If (2.2.6) *is also satisfied, then*

$$\begin{aligned} &P_x(X_t \notin \Omega_*) \\ &\quad\leq \pi_{\beta(t)}(\Omega^*)^{1/2}\Big(5\pi_{\beta(t)}(\Omega^*)^{1/2} + \exp(-M\beta(t))\big(\pi_{\beta(0)}(x)^{-1} - 1\big)^{1/2}\Big) \\ &\quad\leq 5A(1+\rho t)^{-B/m} + \big(\pi_{\beta(0)}(x)^{-1} - 1\big)^{1/2} A^{1/2}(1+\rho t)^{-(B+2M)/2m}. \end{aligned}$$

PROOF. Simply note that $\nu_0 = \delta_x$. Then

$$\begin{aligned} \nu_t(\Omega^*) &= \pi_{\beta(t)}(\Omega^*) + \int_{\Omega^*} (f_t - 1)\, d\pi_{\beta(t)} \\ &\leq \pi_{\beta(t)}(\Omega^*) + \pi_{\beta(t)}(\Omega^*)^{1/2}\|(f_t - 1)\|_{\pi_{\beta(t)}} \end{aligned}$$

with $\|f_0 - 1\|_{\pi_{\beta(0)}} = (\pi_{\beta(0)}(x)^{-1} - 1)^{1/2}$. □

If (2.2.6) is not satisfied, we can modify the original Markov chain as in Goetze (1992), and get a similar result for the probability of the first hitting time of Ω_*. Consider the collapsed chain $L_\beta[\Omega_*]$ (see Definition 2.1.7). We denote by $\mathscr{E}[\Omega_*]_{\pi_\beta[\Omega_*]}$ the corresponding Dirichlet form and by $\tilde{C}[\Omega_*](\beta)$ its spectral gap. Thus $\tilde{C}[\Omega_*](\beta) \geq \tilde{C}(\beta)$, and under (2.2.1) we have

$$\tilde{C}[\Omega_*](\beta) \geq \overline{K}\exp(-\overline{m}\beta), \tag{2.2.10}$$

for some $\overline{m} \leq m$ and $\overline{K} > 0$. Let $\bar{\nu}_0$ be the initial probability distribution of the Markov chain generated by $L[\Omega_*]_{\beta(t)}$ and let $\bar{\nu}_t$ be its distribution at time $t > 0$. As above we write

$$\bar{f}_t(x) = \frac{\bar{\nu}_t(x)}{\pi_{\beta(t)}[\Omega_*](x)}, \qquad t > 0.$$

Note that (2.2.2) and (2.2.9) imply

$$\left|\frac{\dot{\pi}_{\beta(t)}[\Omega_*]}{\pi_{\beta(t)}[\Omega_*]}(x)\right| \leq \overline{M}\dot{\beta}(t), \qquad x \in \Omega_* \cup \{\delta\}, \tag{2.2.11}$$

and

$$\pi_\beta[\Omega_*](\Omega^*) = \pi_\beta(\Omega^*) \leq A\exp(-\beta B),$$

with $\overline{M} \leq M$. *Let* $\{X_t: t \geq 0\}$ ($\{\overline{X}_t: t \geq 0\}$) be the Markov process generated by $L_{\beta(t)}$ ($L_{\beta(t)}[\Omega_*]$). Set

$$\tau := \inf\{t \geq 0: X_t \notin \Omega^*\} \quad \text{and} \quad \bar{\tau} := \inf\{t \geq 0: \ \overline{X}_t \notin \Omega^*\}.$$

Then for any ν_0 concentrated on Ω^*,

$$P_{\nu_0}(\tau > t) = P_{\nu_0}(\bar{\tau} > t) \leq \inf_{0 \leq s \leq t} P_{\nu_0}\big(\overline{X}_s \in \Omega^*\big) \leq \bar{\nu}_t(\Omega^*).$$

Thus we get the following proposition.

2.2.12. PROPOSITION. *Assume* (2.2.2), (2.2.9) *and* (2.2.10), *and choose* $\beta(\cdot)$ *of the form*

$$\beta(t) = \frac{1}{\overline{m}}\log(1 + \bar{\rho}t) \quad \text{where} \quad \bar{\rho} := \frac{2\overline{m}\overline{K}}{3\overline{M}}.$$

Then for each $x \in \Omega^*$,

$$\begin{aligned}
P_x(\tau > t) &\leq P_x\big(\overline{X}_t \in \Omega^*\big)\\
&\leq \pi_{\beta(t)}(\Omega^*)^{1/2}\Big(5\pi_{\beta(t)}(\Omega^*)^{1/2} + \exp\big(-\overline{M}\beta(t)\big)\big(\pi_{\beta(0)}(x)^{-1} - 1\big)^{1/2}\Big)\\
&\leq 5A(1 + \bar{\rho}t)^{-B/\overline{m}} + \big(\pi_{\beta(0)}(x)^{-1} - 1\big)^{1/2} A^{1/2}(1 + \bar{\rho}t)^{-(B + 2\overline{M})/(2\overline{m})}
\end{aligned}$$

2.2.13. REMARK. From the above we see that the speed of convergence is of the order $O(t^{-B/m})$ or $O(t^{-B/(2m)})$, depending whether (2.2.6) is satisfied or not, and $O(t^{-B/\overline{m}})$ for the first hitting time. It thus depends on two constants: m (or $\overline{m}$) and $B = B(\Omega^*)$. The first constant m, linked to the spectral gap of the process, gives the logarithmic scale at which β tends to infinity, whereas the second constant B provides information about the rate at which the invariant distribution concentrates on Ω_*.

3. Definition and basic properties of $\mathscr{L}$.

3.1. *The family* $\mathscr{L}$. In this section we introduce the family $\mathscr{L}^1$ of *exponentially vanishing* Markov chains and show that it satisfies assumption (2.2.2). Let $\mathscr{C}$ be the family of all functions $\varphi: \mathbb{R}^+ \to \mathbb{R}^+$ with bounded logarithm and let $\mathscr{C}^1$ be the set of $\varphi \in \mathscr{C}$ with bounded $(d/d\beta)\log(\varphi(\beta))$.

3.1.1. DEFINITION. Let $\mathscr{L}$ be the family of irreducible transitions functions $L_\beta: \Omega \times \Omega \to \mathbb{R}$, $\beta \geq 0$, of the form

$$q_\beta(x, y) = \exp(-\beta V(x \to y))\Lambda_{x,y}(\beta), \qquad x \neq y, \tag{3.1.2}$$

where $0 \leq V(x \to y) \leq +\infty$ and $\Lambda_{x,y}(\cdot) \in \mathscr{C}$, $\forall\ x \neq y \in \Omega$. If $\Lambda_{x,y} \in \mathscr{C}^1$, $\forall$ $x, y \in \Omega$, we write $L_\beta \in \mathscr{L}^1$. We adopt the convention that $V(x \to y) = +\infty$ when $q_\beta(x, y) \equiv 0$. Given $L_\beta \in \mathscr{L}$, we associate a connected oriented graph

$$\mathscr{G} = (\Omega, E) \tag{3.1.3}$$

with node set Ω and edge set E, which compose the underlying combinatorial structure supporting the Markov chain: $e := (e^- \to e^+) \in E$ iff $q_\beta(e^-, e^+) > 0$ for some $\beta \geq 0$ [or $V(e^- \to e^+) < +\infty$ in this setting].

This family contains various stochastic algorithms which are used in contexts like image processing [Geman (1984), and Geman and Geman (1988)], neural computing [Aarts and Korst (1989)] or optimization [Azencott (1992), Goetze (1992), Mazza (1992) and Trouvé (1992, 1993)]. First let us characterize the invariant measure π_β associated with $L_\beta \in \mathscr{L}$, using the Freidlin–Wentzell framework [Freidlin and Wentzell (1984) and Wentzell (1972)].

3.1.4. DEFINITION. Let $W \subset \Omega$. A directed graph g with vertex set Ω, consisting of a family of arrows $(x \to y)$, $x, y \in \Omega$, is called a W-graph if it satisfies the following conditions: (1) $\forall\ x \in \Omega - W$, g contains a unique arrow starting at x; (2) g contains no cycles; (3) if $x \in W$, g contains no arrow starting at x. Define $G(W)$ to be the set of all W-graphs and set $G := \bigcup_{W \subset \Omega} G(W)$.

Here is the *matrix tree theorem* of Bott and Mayberry (1954):

3.1.5. LEMMA. *Let $L = (q(x, y);\ x, y \in \Omega)$ be the transition function associated with an irreducible Markov chain on a finite set Ω. Let π be its invariant measure. Then $\pi(x) = (\sum_{y \in \Omega} R(y))^{-1} R(x)$, where $R(x) = \sum_{g \in G\{x\}} L(g)$ and $L(g) = \prod_{(m \to n) \in g} q(m, n)$.*

[For the proof, see, e.g., Freidlin and Wentzell (1984).] In view of (3.1.5), let us introduce the weight function $V: G \to \mathbb{R} \cup \{+\infty\}$, defined by

$$V(g) := \sum_{(x \to y) \in g} V(x \to y). \tag{3.1.6}$$

3.1.7. DEFINITION. Let $V_{\cdot} : \Omega \to \mathbb{R}$ be the function

$$V_x := \min_{g \in G\{x\}} V(g),$$

which is bounded by irreducibility. Moreover, set

$$V^1 := \min_x V_x$$

and

$$V_{\min} := \{x \in \Omega; V_x = V^1\}.$$

3.1.8. LEMMA. *Let $L_\beta \in \mathscr{L}$ and let π_β be its invariant measure. Then*

$$\pi_\beta(x) = \exp(-\beta(V_x - V^1))\Lambda_x(\beta),$$

where $\Lambda_x(\cdot) \in \mathscr{C}$, $\forall\, x \in \Omega$. Moreover, if $V_{\min}^c \neq \varnothing$, then

$$\pi_\beta(V_{\min}^c) \le A(V_{\min}^c)\exp(-\beta B(V_{\min}^c)),$$

where

$$B(V_{\min}^c) = \min_{x \in V_{\min}^c} V_x - V^1 \quad \text{and} \quad A(V_{\min}^c) = \sup_\beta \sum_{x \in V_{\min}^c} \Lambda_x(\beta).$$

Finally, if $L_\beta \in \mathscr{L}^1$, then $\Lambda_x(\cdot) \in \mathscr{C}^1$, $\forall\, x \in \Omega$, and

$$\left|\frac{d}{d\beta}\log(\pi_\beta(x))\right| < M \qquad \forall\, x \in \Omega,$$

with

$$M \le \sup_{x, \beta}\left|-(V_x - V^1) + \frac{d}{d\beta}\log \Lambda_x(\beta)\right|.$$

PROOF. By Lemma 3.1.5 each $R_\beta(x)$ is a sum of products of transition probabilities

$$\begin{aligned} R_\beta(x) &= \sum_{g \in G\{x\}} \exp(-\beta V(g)) \prod_{(m \to n) \in g} \Lambda_{m,n}(\beta) \\ &= \exp(-\beta V_x)\tilde{\Lambda}_x(\beta), \end{aligned}$$

for some $\tilde{\Lambda}_x(\cdot) \in \mathscr{C}$ and, therefore, $\sum_y R_y(\beta) = \exp(-\beta V^1)\Lambda(\beta)$, for some $\Lambda(\cdot) \in \mathscr{C}$. Thus $\pi_\beta(\cdot)$ takes the form $\pi_\beta(x) = \exp(-\beta(V_x - V^1))\Lambda_x(\beta)$, with $\Lambda_x := \tilde{\Lambda}_x \Lambda^{-1} \in \mathscr{C}$. Clearly the same argument applies for $L_\beta \in \mathscr{L}^1$. □

3.1.9. EXAMPLE (The Metropolis chain). Let $q_0(x, y)$ be a given transition function on a finite set Ω and let π_0 be a probability measure on Ω. Suppose that π_0 is L_0-symmetric, that is,

$$Q_0(x, y) = \pi_0(x) q_0(x, y) = Q_0(y, x).$$

Next, given $U: \Omega \to \mathbb{R}$, define the Metropolis chain

$$(3.1.10)\quad q_\beta(x, y) := \begin{cases} q_0(x, y)\exp\left(-\beta(U(y) - U(x))^+\right), & x \neq y, \\ -\Sigma_{y \neq x} q_\beta(x, y), & x = y, \end{cases}$$

with invariant measure π_β of the form

$$\pi_\beta(x) = \exp(-\beta U(x))\pi_0(x)Z(\beta)^{-1},$$

where $Z(\beta) := \Sigma_{y \in \Omega}\pi_0(y)\exp(-\beta U(y))$. Thus $L_\beta \in \mathscr{L}^1$ is symmetric with

$$Q_\beta(x, y) = \pi_\beta(x)q_\beta(x, y) = \exp(-\beta(U(x) \vee U(y)))Z(\beta)^{-1} = Q_\beta(y, x).$$

Here (2.2.2) is satisfied with $M = \max_{x \in \Omega} U(x) - \min_{x \in \Omega} U(x)$ and π_β is constant on $U_{\min}$. Inequality (2.2.9) holds with

$$A(U_{\min}^c) = \pi_0(U_{\min})^{-1} - 1 \quad \text{and} \quad B(U_{\min}^c) = \min_{x \in U_{\min}^c} U(x) - \min_{x \in U_{\min}} U(x).$$

3.2. *Asymptotic behavior of the spectral gap.* We show that $C(\beta)$ and $\tilde{C}(\beta)$ are logarithmically equivalent.

3.2.1. LEMMA. *Let $L_\beta \in \mathscr{L}$. Then*:

(i) *$L_\beta^* \in \mathscr{L}$ and $\tilde{L}_\beta \in \mathscr{L}$.*

(ii) *$V^*(x \to y) = V(y \to x) + V_y - V_x$, where $V^*(\cdot \to \cdot)$ is the weight function associated with L_β^*.*

(iii) *$\tilde{V}(x \to y) = \min\{V(x \to y), V^*(x \to y)\}$.*

PROOF. As we have seen in Lemma 3.1.8, $\pi_\beta(x) = \exp(-\beta(V_x - V^1))\Lambda_x(\beta)$, where $\Lambda_x(\cdot) \in \mathscr{C}$. Thus we can write

$$(3.2.2)\quad \begin{aligned} q_\beta^*(x, y) &= \pi_\beta(x)^{-1}\pi_\beta(y)q_\beta(y, x) \\ &= \exp\left(-\beta\left(V_y + V(y \to x) - V_x\right)\right)\Lambda_{x,y}^*(\beta), \end{aligned}$$

where $\Lambda_{x,y}^*(\cdot) \in \mathscr{C}$ (see the proof of Lemma 3.1.8). It remains only to see that $V_y + V(y \to x) - V_x \geq 0$, $\forall\, x \neq y \in \Omega$. Obviously this is the case if $V(y \to x) = +\infty$. Thus assume that $V(y \to x) < \infty$ and let $g \in G\{y\}$ be such that $V(g) = V_y$ (which is finite by irreducibility). Let $(x \to s(x))$ be the unique arrow of g which starts at x. Thus $V(x \to s(x)) < +\infty$. Consider the graph $\hat{g} := g \setminus (x \to s(x)) \cup (y \to x) \in G\{x\}$, with $V(\hat{g}) < +\infty$ since $V(y \to x) < +\infty$. Therefore,

$$\begin{aligned} V(\hat{g}) &= V(g) - V(x \to s(x)) + V(y \to x) \\ &= V_y + V(y \to x) - V(x \to s(x)) \geq V_x \end{aligned}$$

and thus $L_\beta^* \in \mathscr{L}$ with weight function given by $V^*(x \to y) := V(y \to x) + V_y - V_x$. Furthermore $\tilde{L}_\beta \in \mathscr{L}$ since by definition $\tilde{L}_\beta := (L_\beta + \tilde{L}_\beta)/2$. $\square$

Wentzell (1972) gives a theorem which yields spectral estimates for stochastic matrices with exponential vanishing transition probabilities. For each $L_\beta \in \mathscr{L}$ we can associate a stochastic matrix P_β in the following way. Consider a diagonal element of L_β,

$$q_\beta(x,x) = -\sum_{y\neq x} q_\beta(x,y) = -\sum_{y\neq x} \exp(-\beta V(x\to y))\Lambda_{x,y}(\beta),$$

and let $B = |\Omega|\sup_\beta \sup_{x,y} \Lambda_{x,y}(\beta)$. Then $\Sigma_{y\neq x} q_\beta(x,y) \leq B$, $\forall\ \beta \in (0,\infty)$. For $x \neq y$, set $P_\beta(x,y) = q_\beta(x,y)B^{-1}$. From the above we have $P_\beta(x,x) := 1 - \Sigma_{y\neq x} P_\beta(x,y) \geq 0$, $\forall\ \beta \in (0,\infty)$, and it follows that the matrix

$$P_\beta := B^{-1}L_\beta + Id \tag{3.2.3}$$

is stochastic for all $\beta \in (0,\infty)$. Moreover, $P_\beta(x,y) = \exp(-\beta V(x\to y))\overline{\Lambda}_{xy}(\beta)$, where $\overline{\Lambda}_{xy}(\cdot) \in \mathscr{C}$, and thus has the same weight function as L_β.

3.2.4. Definition. Let G^i, $1 \leq i \leq |\Omega|$, be the set of all W-graphs, where W runs through all possible i-element subsets of Ω. Define the constants

$$V^i := \min_{g\in G^i} V(g), \qquad 1 \leq i \leq |\Omega| - 1, V^{|\Omega|} := 0,$$

where $V(\cdot)$ has been defined in (3.1.6).

3.2.5. Theorem [Wentzell (1972)]. *Let $\mu_1(\beta) = 0$ and $\mu_2(\beta),\dots,\mu_{|\Omega|}(\beta)$ denote the eigenvalues of the matrix $P_\beta - Id$, arranged in order of decreasing real part. Then, for $i = 2,\dots,|\Omega|$,*

$$\Re(-\mu_i(\beta)) \asymp \exp(-(V^{i-1} - V^i)\beta), \tag{3.2.6}$$

where $\asymp$ means that $\lim_{\beta\to\infty} \beta^{-1}\log(-\Re(\mu_i(\beta))) = -(V^{i-1} - V^i)$, $\forall\ i \geq 2$. *Moreover, the polygonal arc through the points (i, V^i) is convex downward*:

$$V^1 - V^2 \geq V^2 - V^3 \geq \cdots \geq V^{|\Omega|-1} - V^{|\Omega|}. \tag{3.2.7}$$

Assume that $\lambda(\beta)$ is some eigenvalue of $L_\beta \in \mathscr{L}$. Then $B^{-1}\lambda(\beta)$ [see (3.2.3)] is an eigenvalue of $P_\beta - Id$. Let $\lambda_1(\beta) = 0, \lambda_2(\beta),\dots$ denote the eigenvalues of L_β arranged in order of decreasing real parts. Then, according to (3.2.6),

$$\Re(-\lambda_i(\beta)) \asymp \exp(-\beta(V^{i-1} - V^i)), \qquad i = 2,\dots,|\Omega|. \tag{3.2.8}$$

Let $C(\beta)$ [resp. $\tilde{C}(\beta)$] be the spectral gap associated with L_β (resp. $\tilde{L}_\beta$). Then, by (3.2.8),

$$\lim_{\beta\to\infty} \beta^{-1}\log(C(\beta)) = -(V^1 - V^2),$$

for the family of weights $V(x\to y)$ associated with L_β and

$$\lim_{\beta\to\infty} \beta^{-1}\log(\tilde{C}(\beta)) = -(\tilde{V}^1 - \tilde{V}^2),$$

for the family of weights $\tilde{V}(x \to y)$ associated with $\tilde{L}_\beta$. Thus, using Lemma 2.1.4, we have

$$\tilde{V}^1 - \tilde{V}^2 \geq V^1 - V^2. \tag{3.2.9}$$

In the sequel, we will see that

$$V^1 - V^2 = \tilde{V}^1 - \tilde{V}^2, \tag{3.2.10}$$

for any element L_β of $\mathscr{L}$. With the graph theoretic interpretation, (3.2.10) is related to the so-called *directed spanning tree problem* [see Lawler (1976)], which is well known in the context of combinatorial optimization.

3.2.11. THEOREM. *Let $L_\beta \in \mathscr{L}$. Then*

$$\tilde{V}^i = V^i, \qquad i = 1, \dots, |\Omega|$$

and, therefore,

$$\Re(-\lambda_i(\beta)) \asymp -\tilde{\lambda}_i(\beta), \qquad i = 2, \dots, |\Omega|.$$

The proof proceeds in several steps. Let L be an irreducible transition function on Ω and let $\Omega_* \subset \Omega$. The following lemma deals with the collapsed chain $L[\Omega_*]$ on $\overline{\Omega} = \Omega^* \cup \{\delta\}$ (see Definition 2.1.7).

3.2.12. LEMMA. *Let L be an irreducible transition function on Ω and let $\tilde{L}$ be its symmetrized version. Let $\Omega_* \subset \Omega$ be such that $\Omega_* \neq \emptyset, \Omega$. Then*

$$\tilde{L}[\Omega_*] = \widetilde{L[\Omega_*]},$$

where $\widetilde{L[\Omega_*]}$ is the symmetrization of $L[\Omega_*]$ on $\Omega^* \cup \{\delta\}$.

Recall first that

$$\tilde{q}(x,y) := \tfrac{1}{2}\big(q(x,y) + \pi(x)^{-1}\pi(y)q(y,x)\big)$$

and

$$\widetilde{q[\Omega_*]}(x,y) := \tfrac{1}{2}\big(q[\Omega_*](x,y) + \pi[\Omega_*](x)^{-1}\pi[\Omega_*](y)q[\Omega_*](y,x)\big).$$

For x and $y \in \Omega^*$ the statement is obvious. Next, remark that $\tilde{q}[\Omega_*](x,\delta) := \sum_{y \in \Omega_*} \tilde{q}(x,y)$ for $x \in \Omega^*$. Hence,

$$\begin{aligned}
\widetilde{q[\Omega_*]}(x,\delta) &= \tfrac{1}{2}\big(q[\Omega_*](x,\delta) + \pi[\Omega_*](x)^{-1}\pi[\Omega_*](\delta)q[\Omega_*](\delta,x)\big)\\
&= \tfrac{1}{2}\Bigg(\sum_{y\in\Omega_*} q(x,y)\\
&\qquad + \pi(x)^{-1}\pi[\Omega_*](\delta)\pi[\Omega_*](\delta)^{-1}\sum_{y\in\Omega_*}\pi(y)q(y,x)\Bigg)\\
&= \sum_{y\in\Omega_*}\tfrac{1}{2}\big(q(x,y) + \pi(x)^{-1}\pi(y)q(y,x)\big)\\
&= \sum_{y\in\Omega_*}\tilde{q}(x,y)
\end{aligned}$$

and thus $\tilde{q}[\Omega_*](x,\delta) = \widetilde{q[\Omega_*]}(x,\delta)$. Finally, let us consider $\tilde{q}[\Omega_*](\delta, x)$ and $\widetilde{q[\Omega_*]}(\delta, x)$:

$$\begin{aligned}\tilde{q}[\Omega_*](\delta,x) &= \tilde{\pi}[\Omega_*](\delta)^{-1} \sum_{y\in\Omega_*} \tilde{\pi}(y)\tilde{q}(y,x)\\ &= \pi[\Omega_*](\delta)^{-1}\tfrac{1}{2} \sum_{y\in\Omega_*} (\pi(y)q(y,x) + \pi(x)q(x,y)),\end{aligned}$$

since L and $\tilde{L}$ have the same invariant measure $\tilde{\pi} \equiv \pi$. Moreover,

$$\begin{aligned}\widetilde{q[\Omega_*]}(\delta,x) &= \frac{1}{2}\big(q[\Omega_*](\delta,x) + \pi[\Omega_*](\delta)^{-1}\pi[\Omega_*](x)q[\Omega_*](x,\delta)\big)\\ &= \frac{1}{2}\bigg(\pi[\Omega_*](\delta)^{-1} \sum_{y\in\Omega_*} \pi(y)q(y,x)\\ &\qquad + \pi[\Omega_*](\delta)^{-1}\pi(x) \sum_{y\in\Omega_*} q(x,y)\bigg)\\ &= \frac{1}{2\pi[\Omega_*](\delta)} \sum_{y\in\Omega_*} (\pi(y)q(y,x) + \pi(x)q(x,y)).\end{aligned}$$

The following lemma is an easy consequence of the definitions.

3.2.13. LEMMA. *Let $L_\beta \in \mathscr{L}$. Let $\Omega_* \subset \Omega$ be such that $\Omega_* \neq \emptyset, \Omega$. Define $V[\Omega_*]$ to be the weight function associated with $L_\beta[\Omega_*]$ on $\Omega^* \cup \{\delta\}$. Then we have*

$$\begin{aligned}V[\Omega_*](x\to y) &= V(x\to y) \quad \text{for } x,y\in\Omega^*,\\ V[\Omega_*](x\to\delta) &= \min_{p\in\Omega_*} V(x\to p), \qquad x\in\Omega^*,\\ V[\Omega_*](\delta\to x) &= \min_{p\in\Omega_*}\big(V_p + V(p\to x)\big) - \min\nolimits_{p\in\Omega_*}V_p, \qquad x\in\Omega^*.\end{aligned}$$

3.2.14. LEMMA. *Let $L_\beta \in \mathscr{L}$, let Ω_* with $|\Omega_*| = i$ and assume that $V^i = V(g)$ for some $g \in G(\Omega_*)$. Then*

$$V[\Omega_*]^1 := \min_{p\in\Omega^*\cup\{\delta\}} V[\Omega_*]_p = V[\Omega_*]_\delta = V^i,$$

where $V[\Omega_*]_.$ is the function (3.1.7) *associated with $L_\beta[\Omega_*]$.*

PROOF. First let us check that $V[\Omega_*]_\delta = V^i$. Let $\hat{g}$ be an arbitrary element of $G[\Omega_*]\{\delta\}$, the set of $\{\delta\}$-graphs with vertices in $\overline{\Omega}$. Then,

$$\begin{aligned}V[\Omega_*](\hat{g}) &= \sum_{\substack{(m\to n)\in\hat{g}\\ m,n\in\Omega^*}} V[\Omega_*](m\to n) + \sum_{(m\to\delta)\in\hat{g}} V[\Omega_*](m\to\delta)\\ &= \sum_{\substack{(m\to n)\in\hat{g}\\ m,n\in\Omega^*}} V(m\to n) + \sum_{(m\to\delta)\in\hat{g}} \min_{p\in\Omega_*} V(m\to p).\end{aligned}$$

Thus for each $\hat{g} \in G[\Omega_*]\{\delta\}$ we can associate a graph $\hat{g}' \in G(\Omega_*)$ such that

$$V[\Omega_*](\hat{g}) = V(\hat{g}') \geq \min_{g \in G(\Omega_*)} V(g) = V^i. \tag{3.2.15}$$

By hypothesis there exists $g \in G(\Omega_*)$ such that $V(g) = V^i$. Set

$$\hat{g} := \left(g \Bigg\backslash \left(\bigcup_{\substack{(m \to p) \in g \\ m \in \Omega^*, p \in \Omega_*}} (m \to p) \right) \right) \bigcup_{\substack{(m \to p) \in g \\ m \in \Omega^*, p \in \Omega_*}} (m \to \delta).$$

Thus we have

$$V[\Omega_*](\hat{g}) = V^i$$

because $V(g) = V^i$ for $g \in G(\Omega_*)$ implies that $V(m \to p) = \min_{q \in \Omega_*} V(m \to q)$, $\forall\ (m \to p) \in g,\ p \in \Omega_*$. It follows that $V[\Omega_*]_\delta := \min_{\hat{g} \in G\{\delta\}} V[\Omega_*](\hat{g}) = V^i$.

It remains to see that $V[\Omega_*]_\delta = V[\Omega_*]^1$. Let $\omega \in \Omega^*$ and take some $\hat{g} \in G[\Omega_*]\{\omega\}$. As above $V[\Omega_*](m \to n) = V(m \to n)\ \forall\ m, n \in \Omega^*$. Let $(\delta \to \omega')$ be the unique arrow of $\hat{g}$ which starts at δ. By definition $V[\Omega_*](\delta \to \omega') = \min_{p \in \Omega_*}(V_p + V(p \to \omega')) - \min_{p \in \Omega_*} V_p$. Let p_1 be any element of Ω_* which achieves the minimum $\min_{p \in \Omega_*}(V_p + V(p \to \omega'))$. For any arrow $(m \to \delta) \in \hat{g}$ we associate an arrow $(m \to p(m))$ for $p(m) \in \Omega_*$ such that $V(m \to p(m)) = \min_{p \in \Omega_*} V(m \to p)$. In this way we get a graph $\hat{g}' \in G(\Omega_* \setminus \{p_1\} \cup \{\omega\})$ defined by

$$\hat{g}' = \left(\hat{g} \Bigg\backslash \left(\bigcup_{(m \to \delta) \in \hat{g}} (m \to \delta) \cup (\delta \to \omega') \right) \right)$$
$$\bigcup_{(m \to \delta) \in \hat{g}} (m \to p(m)) \cup (p_1 \to \omega'),$$

which satisfies

$$\begin{aligned} V[\Omega_*](\hat{g}) &= \sum_{(m \to \delta) \in \hat{g}} V(m \to p(m)) \\ &\quad + \sum_{\substack{(m \to n) \in \hat{g} \\ m, n \in \Omega^*}} V(m \to n) + V_{p_1} + V(p_1 \to \omega') - \min_{p \in \Omega_*} V_p \\ &\geq \sum_{(m \to \delta) \in \hat{g}} V(m \to p(m)) + \sum_{\substack{(m \to n) \in \hat{g} \\ m, n \in \Omega^*}} V(m \to n) + V(p_1 \to \omega') \\ &= V(\tilde{g}') \geq V^i. \end{aligned}$$

The statement is proved since we have already shown that $V[\Omega_*]_\delta = V^i$. □

PROOF OF THEOREM 3.2.11. The proof proceeds by induction on the size N of Ω. Let us first check the statement for $N = 2$. Assume that $\Omega = \{x, y\}$ and

PROOF OF THEOREM 3.2.11. The proof proceeds by induction on the size N of Ω. Let us first check the statement for $N = 2$. Assume that $\Omega = \{x, y\}$ and $V(x \to y) < +\infty$, $V(y \to x) < +\infty$ (by irreducibility). In any case $V_x = V(y \to x)$ and $V_y = V(x \to y)$, and, therefore,

$$\begin{aligned}\tilde{V}(x \to y) &:= \min\{V(x \to y), V(y \to x) + V_y - V_x\}\\ &= \min\{V(x \to y), V(y \to x) + (V(x \to y) - V(y \to x))\}\\ &= V(x \to y),\end{aligned}$$

and thus $\tilde{V}^1 := \min\{\tilde{V}_x, \tilde{V}_y\} = \min\{V_x, V_y\} = V^1$. Next assume that $\tilde{V}^i = V^i$, $1 \le i \le N$, for any $L_\beta \in \mathscr{L}$ on Ω with $|\Omega| \le N$. Take some L_β on a set Ω with $|\Omega| = N + 1$. Let Ω_* with $|\Omega_*| = i$, $2 \le i \le N + 1$, be such that there exists $g \in G(\Omega_*)$ with $\tilde{V}(g) = \tilde{V}^i$. By Lemmas 3.2.14 and 3.2.12 we have

$$\tilde{V}^i = \tilde{V}[\Omega_*]^1 = \widetilde{V[\Omega_*]}^1 = V[\Omega_*]^1, \tag{3.2.16}$$

where the last equality is a consequence of the induction hypothesis, since, as $2 \le i \le N + 1$, $\bar{\Omega} = \Omega^* \cup \{\delta\}$ has at most N elements. As we have seen in Lemma 3.2.14, $\tilde{V}[\Omega_*]^1 = \widetilde{V[\Omega_*]}^1$ is realized for a graph $g \in G[\Omega_*]\{\delta\}$. As $L_\beta[\Omega_*]$ and $\widetilde{L_\beta[\Omega_*]}$ have the same invariant measure, it follows that $V[\Omega_*]^1 = V[\Omega_*](\hat{g})$ for some $\hat{g} \in G[\Omega_*]\{\delta\}$. Working as in the proof of Lemma 3.2.14 we can find a graph $\hat{g}' \in G(\Omega_*)$ such that $V[\Omega_*](\hat{g}) = V(\hat{g}') \ge V^i$ [see (3.2.15)], and it follows from (3.2.16) that $\tilde{V}^i \ge V^i$, which implies that $\tilde{V}^i = V^i$ since by the definition of the symmetrized operator $\tilde{V}(m \to n) \le V(m \to n)$, $\forall\, m, n \in \Omega$. It remains to see that $\tilde{V}^1 = V^1$: by (3.2.9) we know that in any case $\tilde{V}^1 - \tilde{V}^2 \ge V^1 - V^2$, $\forall\ L_\beta \in \mathscr{L}$; therefore, $\tilde{V}^1 = V^1$ as $\tilde{V}^1 \le V^1$ by definition. □

3.3. *Processes drifted by a potential function.* The weight function $\tilde{V}$ associated with $\tilde{L}_\beta$ has some interesting properties which will be useful for the estimation of the spectral gap.

3.3.1. DEFINITION. We say that a function $\Psi\colon \Omega \to \mathbb{R}$ is a potential function for the family $(V(x \to y))_{x, y \in \Omega}$ if

$$V(x \to y) - V(y \to x) = \Psi(y) - \Psi(x) \qquad \forall\, x, y \in \Omega,$$

when $V(x \to y) < +\infty$.

Note that by definition a potential function Ψ is bounded. For example the Metropolis chain has the potential function $\Psi \equiv U$; cf. Example 3.1.9.

3.3.2. LEMMA. *Assume that $L_\beta \in \mathscr{L}$ has a potential function Ψ. Then*

$$V(x \to y) < +\infty \quad \Leftrightarrow \quad V(y \to x) < +\infty.$$

3.3.3. LEMMA. *Let $L_\beta \in \mathscr{L}$ and let $\tilde{L}_\beta \in \mathscr{L}$ be its symmetrized version. Then $\tilde{L}_\beta$ has*

$$V_\cdot\colon \quad \Omega \to \mathbb{R}$$

(*see Definition* 3.1.7) *as a potential function.*

PROOF. First assume that $\tilde{V}(x \to y) < +\infty$. Thus, as $V_.$ is finite by irreducibility, either $V(x \to y) < +\infty$ or $V(y \to x) < +\infty$ and then $\tilde{V}(y \to x) = \min\{V(x \to y); V(y \to x) + V_y - V_x\}$ is finite. Therefore, we can write

$$\begin{aligned}\tilde{V}(x \to y) - \tilde{V}(y \to x) &= \min\{V(x \to y), V(y \to x) + V_y - V_x\} \\ &\quad - \min\{V(y \to x), V(x \to y) + V_x - V_y\} \\ &= V_y - V_x. \qquad \square\end{aligned}$$

3.3.4. LEMMA. *Assume that $L_\beta \in \mathscr{L}$ has a potential function Ψ. Then*

$$V_x - V_y = \Psi(x) - \Psi(y) \qquad \forall\, x, y \in \Omega,$$

$$\tilde{V}(x \to y) = V(x \to y) \qquad \forall\, x, y \in \Omega,$$

$$\tilde{V}_x = V_x \quad \forall\, x \in \Omega.$$

PROOF. Let $x, y \in \Omega$ and consider $G\{x\}$ and $G\{y\}$. Let $g \in G\{x\}$ and let γ_{yx} be the unique geodesic of g which takes y to x and let γ_{xy} be the reversal of γ_{yx}. Let $F(\cdot)$: $G\{x\} \to G\{y\}$ be the mapping given by $F(g) := (g \setminus \gamma_{yx}) \cup \gamma_{xy} \in G\{y\}$. As the function $V(\hat{g}) = \sum_{(m \to n) \in \hat{g}} V(m \to n)$ is additive, we get

$$V(F(g)) = V(g) - V(\gamma_{yx}) + V(\gamma_{xy}) = V(g) + \Psi(y) - \Psi(x)$$

since $V(\cdot, \cdot)$ has Ψ as potential function. Thus,

$$\begin{aligned}V_x &= \min_{g \in G\{x\}} V(g) = \min_{g \in G\{x\}} (V(F(g)) + \Psi(x) - \Psi(y)) \\ &= \min_{g \in G\{x\}} V(F(g)) + \Psi(x) - \Psi(y) \\ &= \min_{g \in G\{y\}} V(g) + \Psi(x) - \Psi(y) \\ &= V_y + \Psi(x) - \Psi(y),\end{aligned}$$

since $F(\cdot)$ is *onto*. Thus $V_x - V_y = \Psi(x) - \Psi(y)\ \forall\, x, y \in \Omega$ and the first part of the lemma is proved. By hypothesis,

$$V(x \to y) - V(y \to x) = \Psi(y) - \Psi(x),$$

so that

$$V(y \to x) = V(x \to y) + \Psi(x) - \Psi(y) = V(x \to y) + V_x - V_y$$

and it follows that

$$V^*(x \to y) := V(y \to x) + V_y - V_x = V(x \to y)$$

and thus

$$\tilde{V}(x \to y) = V(x \to y) \qquad \forall\, x, y \in \Omega. \qquad \square$$

3.3.5. REMARK. Assume that $L_\beta \in \mathscr{L}$ has a potential function $\Psi(\cdot)$. The preceding lemma implies that we can without loss of generality take $\Psi \equiv V_.$ as its potential function.

Let T be a maximal connected spanning tree and let $x \in \Omega$. Define T_x to be the directed spanning tree of $G\{x\}$ obtained by pointing T at x.

3.3.6. LEMMA. *Assume that $L_\beta \in \mathscr{L}$ has a potential function Ψ. There exists a maximal connected spanning tree T, such that*

$$V_x = V(T_x), \qquad \forall\, x \in \Omega.$$

PROOF. Let T be any directed spanning tree for G such that $V(T_x) = V_x$. Then for $x \neq y$, $T_y = (T_x \setminus t_{yx}) \cup t_{xy}$, where t_{yx} is the unique geodesic of T_x taking y and t_{xy} as its reversal. Then

$$V(T_y) = V(T_x) - V(t_{yx}) + V(t_{xy}) = V_x - (V_x - V_y) = V_y,$$

since by Lemma 3.3.4, $V_.$ can be taken to be the potential function which exists for L_β by hypothesis. □

3.3.7. EXAMPLE. The following processes have potential functions:

(i) Processes running on trees.
(ii) Symmetric processes.

First assume that L_β is running on a tree T, that is $\mathscr{G} = (\Omega, E)$ is a tree and that $\{V(e)\}$ runs over directed edges. In this case, $G\{x\} = \{T_x\}$, $\forall\, x \in \Omega$ [see (3.3.6)] and obviously $V_x = V(T_x)$ for all $x \in \Omega$. Therefore, $V_.$ is a potential function for the family $\{V(e)\}_{e \in E}$.

Next, consider a symmetric $L_\beta \in \mathscr{L}$, that is, $\tilde{L}_\beta \equiv L_\beta$. Then, $\tilde{V}(x \to y) \equiv V(x \to y)$ and thus the result follows from Lemma 3.3.3.

4. Spectral estimates for $\mathscr{L}$.

4.1. *Geometric bounds and ultrametrics.* In the previous section, we have seen that $\mathscr{L}^1$-chains satisfy (2.2.2). In order to apply Corollary 2.2.8 and Proposition 2.2.12, we must check that $\tilde{C}(\beta) \geq K \exp(-\beta m)$ for some $K > 0$ and $m \geq 0$. Moreover, we should try to have m as small as possible and K as large as possible. As pointed out by Goetze (1992), this is a major problem. In this section we will verify (2.2.1) using the Poincaré geometric method, as in Diaconis and Stroock (1991), Sinclair (1991) and Goetze (1992), and show that our estimate is consistent with Wentzell's (3.2.8).

In the remaining text, we assume *symmetry*, that is, $\tilde{L}_\beta \equiv L_\beta$ and $\tilde{C}(\beta) \equiv C(\beta)$.

4.1.1. DEFINITION. The main ingredient for bounding $C(\beta)$ is the choice of a collection of oriented paths $\Gamma \equiv \{g_{xy} \in \mathscr{G} = (\Omega, E);\ x, y \in \Omega\}$, connecting each $x \in \Omega$ with each $y \in \Omega$. The paths may have repeated nodes but a given edge appears at most once in a given path.

Next, let $|g_{xy}|$ be the length of g_{xy} and set

$$|g_{xy}|_{Q_\beta} = \sum_{e \in g_{xy}} \frac{1}{Q_\beta(e)},$$

where, for an oriented edge $e = (e^- \to e^+) \in g_{xy}$, we write

$$Q_\beta(e) = Q_\beta(e^-, e^+) = \tfrac{1}{2}\big(\pi_\beta(e^-)q_\beta(e^-, e^+) + \pi_\beta(e^+)q_\beta(e^+, e^-)\big)$$
$$= Q_\beta(-e).$$

Consider the Poincaré bound [see Diaconis and Stroock (1991)]

$$C(\beta) \ge \frac{1}{\kappa_\beta^{(1)}(\Gamma)}, \tag{4.1.2}$$

where

$$\kappa_\beta^{(1)}(\Gamma) := \sup_{e \in \mathscr{G}(\Gamma)} \sum_{g_{xy} \ni e} |g_{xy}|_{Q_\beta} \pi_\beta(x)\pi_\beta(y)$$

and $\mathscr{G}(\Gamma) := \{e \in \mathscr{G};\ \exists\ g_{xy} \ni e,\ g_{xy} \in \Gamma\}$.

4.1.3. DEFINITION. Let $L_\beta \in \mathscr{L}$ be symmetric. Set

$$\mathscr{O}(e) := V_{e^-} + V(e^- \to e^+).$$

Note that, by Lemma 3.2.1 and symmetry, $\mathscr{O}(\cdot)$ is orientation independent; that is,

$$\mathscr{O}(e) = \mathscr{O}(-e),$$

where $-e := (e^+ \to e^-)$. Let C_{xy} be the set of all the paths of $\mathscr{G}$ taking x to y. For $\gamma \in C_{xy}$, let $\mathrm{Evel}(\gamma) := \max_{e \in \gamma} \mathscr{O}(e)$. For any collection of paths Γ set

$$m^{(1)}(\Gamma) := \max_{x \ne y}\big(\mathrm{Evel}(g_{xy}) + V^1 - V_x - V_y\big), \qquad m^{(1)} = \inf_\Gamma m^{(1)}(\Gamma).$$

For $x, y \in \Omega$, define

$$F_{x,y}(\beta) := \Lambda_x(\beta)\Lambda_y(\beta) \sum_{e \in g_{xy}} \big(\Lambda_{e^-}(\beta)\Lambda_{e^-,e^+}(\beta)\big)^{-1},$$

with $F_{x,y} \in \mathscr{C}$, $x, y \in \Omega$ [see (3.1.2) and Lemma 3.1.8].

4.1.4. LEMMA. *Let $L_\beta \in \mathscr{L}$ be symmetric. Then, for all collections of paths Γ, we have*

$$C(\beta) \ge \exp(-\beta m^{(1)}(\Gamma))\Bigg(\max_{e \in \mathscr{G}(\Gamma)} \sum_{g_{xy} \ni e} F_{x,y}(\beta)\Bigg)^{-1}. \tag{4.1.5}$$

PROOF. By (4.1.2), it suffices to show

$$\kappa_\beta^{(1)}(\Gamma) \le \exp(\beta m^{(1)}(\Gamma)) \max_{e \in \mathscr{G}(\Gamma)} \sum_{g_{xy} \ni e} F_{x,y}(\beta).$$

By definition, $\kappa_\beta^{(1)}(\Gamma) = \max_{e \in \mathscr{G}(\Gamma)} \sum_{g_{xy} \ni e} \pi_\beta(x)\pi_\beta(y)|g_{xy}|_{Q_\beta}$, where

$$|g_{xy}|_{Q_\beta} = \sum_{\hat{e} \in g_{xy}} \frac{1}{\pi_\beta(\hat{e}^-)L_\beta(\hat{e}^-, \hat{e}^+)}.$$

Using Lemma 3.1.8 we obtain

$$\kappa_\beta^{(1)} = \max_{e\in\mathscr{G}(\Gamma)} \sum_{g_{xy}\ni e} \exp\bigl(-\beta\bigl(V_x + V_y - 2V^1\bigr)\bigr)\Lambda_x(\beta)\Lambda_y(\beta)|g_{xy}|_{Q_\beta}$$

and

$$\begin{aligned}|g_{xy}|_{Q_\beta} &= \sum_{\hat e\in g_{xy}} \exp\bigl(-\beta\bigl(V^1 - V_{\hat e^-}\bigr)\bigr)\Lambda_{\hat e^-}(\beta)^{-1}\exp(\beta V(\hat e^-\to \hat e^+))\Lambda_{\hat e^-,\hat e^+}(\beta)^{-1}\\ &= \sum_{\hat e\in g_{xy}} \exp\bigl(-\beta\bigl(V^1 - V_{\hat e^-} - V(\hat e^-\to\hat e^+)\bigr)\bigr)\bigl(\Lambda_{\hat e^-}(\beta)\Lambda_{\hat e^-,\hat e^+}(\beta)\bigr)^{-1}\end{aligned}$$

Therefore,

$$\begin{aligned}\kappa_\beta^{(1)}(\Gamma) &\le \max_{e\in\mathscr{G}(\Gamma)} \sum_{g_{xy}\ni e}\biggl[\exp\bigl(-\beta\bigl(V_x + V_y - 2V^1\bigr)\bigr)\\ &\qquad\times\exp\Bigl(-\beta\Bigl(-\max_{\hat e\in g_{xy}}\bigl(V_{\hat e^-} + V(\hat e^-\to\hat e^+)\bigr)\Bigr)\Bigr)\\ &\qquad\times\exp(-\beta V^1)F_{x,y}(\beta)\biggr]\\ &\le \exp(\beta m^{(1)}(\Gamma))\max_{e\in\mathscr{G}(\Gamma)}\sum_{g_{xy}\ni e} F_{x,y}(\beta).\end{aligned}$$

□

4.1.6. EXAMPLE (Metropolis chains). Consider the general symmetric Metropolis chain (3.1.10), which has $U(\cdot)$ as a potential function. By Lemma 3.3.4 there exists a constant $D \in \mathbb{R}$ such that $V \equiv U(\cdot) + D$. Therefore, $\mathscr{O}(e) \equiv U(e^-) \vee U(e^+) + D$ and thus

$$\begin{aligned}\mathrm{Evel}(g_{xy}) &= \max_{e\in g_{xy}}\bigl(U(e^-) + (U(e^+) - U(e^-))^+\bigr) + D\\ &= \max_{e\in g_{xy}}\max(U(e^-), U(e^+)) + D\\ &= \max\bigl\{U(\tilde x);\ \tilde x \text{ is a node of } g_{xy}\bigr\} + D\end{aligned}$$

and it follows that

$$m^{(1)}(\Gamma) \equiv \max_{x\neq y}\bigl(U_* + \max_{e\in g_{xy}}(U(e)) - U(x) - U(y)\bigr),$$

where $U_* := \min_x U(x)$ and $U(e) \equiv U(e^-) \vee U(e^+)$.

Ultrametrics and hierarchies. Consider the distance $\rho(\cdot,\cdot)$ on Ω given by

$$\rho(x,y) := \min_{\gamma\in C_{xy}} \mathrm{Evel}(\gamma), \qquad x,y\in\Omega, \tag{4.1.7}$$

with $\rho(x,y) = \rho(y,x)$ by symmetry (see Definition 4.1.3), which satisfies the so-called *ultrametric inequality*

$$\rho(x,y) \le \rho(x,w) \vee \rho(w,y), \qquad x,y,w\in\Omega. \tag{4.1.8}$$

Given a finite set Ω and an ultrametric ρ on Ω, that is, a so-called ultrametric set (Ω,ρ), we can associate an *indexed hierarchy* (see Figure 1) on Ω [see Roux (1985) or Rammal, Toulouse and Virasoro (1986)], which yields a representation of (Ω,ρ) in a simple way: The elements of Ω are the

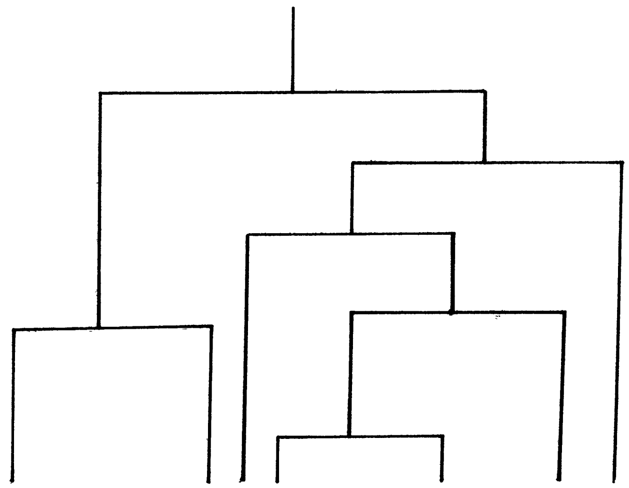


FIG. 1. *The indexed hierarchy.*

leaves of a tree T, where the distance ρ between nodes is an increasing function of the height of the closest common ancestor.

4.1.9. DEFINITION. Set

$$D(x) := \max_{y \neq x}\left(\rho(x,y) + V^1 - V_x - V_y\right), \qquad \Delta := \max_{x \in \Omega} D(x),$$
$$V_{\min} := \left\{x \in \Omega; V_x = V^1\right\}, \qquad R := \max_{x, y \in V_{\min}} \rho(x,y),$$
$$\hat{\rho}(x) := \min_{y \in V_{\min}} \rho(x,y), \qquad \rho(x) := \max_{y \in V_{\min}} \rho(x,y),$$
$$\rho_{\min} := \left\{x \in \Omega;\, \rho(x) = \min_{y \in \Omega} \rho(y)\right\}.$$

If $V_{\min} = \{x_0\}$, we set $\rho(x_0) := V_{x_0} = V^1$ and $R := V^1$. Moreover, let $\mathcal{S}^*$ be the set of collections of paths $\Gamma = \{g_{xy}\}$ such that

$$\text{Evel}(g_{xy}) \equiv \rho(x,y), \qquad x \neq y. \tag{4.1.10}$$

4.1.11. LEMMA. *Let $L_\beta \in \mathcal{L}$ be symmetric. Then, for all collections of paths Γ, we have*

$$m^{(1)}(\Gamma) \geq \Delta. \tag{4.1.12}$$

Let $\Gamma \in \mathcal{S}^*$. Then,

$$m^{(1)}(\Gamma) = m^{(1)} = \Delta.$$

Moreover,

$$\rho(x) \equiv R \vee \hat{\rho}(x) \quad \textit{and} \quad \Delta = \max_{x \in \Omega}\left(\rho(x) - V_x\right).$$

PROOF. The first two statements are direct consequences of the definitions. Assume next that $|V_{\min}| \geq 2$. Cut T at each node n which has height R, where T is the tree of the indexed hierarchy associated with the ultrametric set (Ω, ρ). The deleted portions of the tree form a collection of subtrees of T, and one of them, say T_1, rooted at n, contains $V_{\min}$, and $V_{\min}$ is not contained in any one single branch emanating from n (if this is the case, necessarily $R = V^1$). Define $\tilde{\rho}_{\min}$ to be the subset of Ω which consists of the leaves of T_1. The above considerations show that $\rho(x) \equiv \rho(x, y) > R$, $\forall\ y \in V_{\min}$, $\forall\ x \in \Omega \setminus \tilde{\rho}_{\min}$. Therefore, $\rho(x) = \hat{\rho}(x)$, $\forall\ x \in \Omega \setminus \tilde{\rho}_{\min}$ and $\rho(x) = R$, $\forall\ x \in \tilde{\rho}_{\min}$. Thus $\tilde{\rho}_{\min} = \rho_{\min}$. Finally, let us prove the last statement. Trivially $\Delta \geq \max_x(\rho(x) - V_x)$. Next let $x, y \in \Omega$ and $z \in V_{\min}$. Then by (4.1.8),

$$\begin{aligned}\rho(x, y) - V_x - V_y + V^1 &\leq (\rho(x, z) - V_x) \vee (\rho(y, z) - V_y)\\ &\leq (\rho(x) - V_x) \vee (\rho(y) - V_y)\\ &\leq \max_{x \in \Omega}(\rho(x) - V_x). \qquad \square\end{aligned}$$

Recall that the aim is to obtain an inequality of the form $C(\beta) \geq K\exp(-\beta m)$ with m as small as possible. Using (3.2.6) we get that, in general, $m \geq V^1 - V^2$. From the above lemma we know that Lemma 4.1.4 yields at best $m^{(1)} = \Delta$. The following theorem shows that $V^1 - V^2$ can be realized by $\kappa_\beta^{(1)}(\Gamma)$ for some well chosen collection of paths Γ.

4.1.13. THEOREM. *Let $L_\beta \in \mathscr{L}$ be symmetric. Then*

$$V^1 - V^2 = \Delta. \tag{4.1.14}$$

PROOF. We only need to show that $\Delta \leq V^1 - V^2$. Set for convenience $\Xi(\gamma) := \mathrm{Evel}(\gamma) + V^1 - V_x - V_y$, $x, y \in \Omega$, $\gamma \in C_{xy}$. Let $(x, y) \in \Omega^2$ be such that $D(y) = \Delta$ and $D(y) = \Xi(f)$ for some $f \in C_{xy}$. Let $h \in G\{x\}$ be such that $V(g) = V_x$ and let h_{yx} be the unique geodesic of h which takes y to x. Assume that h_{yx} has the normal form $h_{yx} \equiv (e_0, \dots, e_{k-1}, e, e_{k+1}, \dots, e_n)$, where e is the arc of h_{yx} which realizes $\mathrm{Evel}(h_{yx})$. Therefore, we have

$$\Delta = \Xi(f) \leq \Xi(h_{yx}) = V^1 + \mathrm{Evel}(h_{yx}) - V_x - V_y. \tag{4.1.15}$$

Set $\hat{h} := h - h_{yx} \cup (e_{k+1}, \dots, e_n) \cup (-e_{k-1}, \dots, -e_0) \in G\{x, y\}$. Then we have

$$\begin{aligned}V(\hat{h}) &= V(h) - V(h_{yx}) + V((e_{k+1}, \dots, e_n)) + V((-e_{k-1}, \dots, -e_0))\\ &= V(h) - V(e) - V((e_0, \dots, e_{k-1})) + V((-e_{k-1}, \dots, -e_0))\\ &= V(h) - V(e) - V_{e^-} + V_y\\ &= V(h) - \mathscr{O}(e) + V_y,\end{aligned}$$

since $e_{k-1}^+ = e^-$ and $e_0^- = y$. However, because $\hat{h} \in G\{x, y\}$, we have by (4.1.15)

$$(4.1.16)\qquad \begin{aligned} V^2 &\le V(\hat{h}) = V(h) - \mathscr{O}(e) + V_y \\ &\le V(h) + V^1 - V_x - \Delta \\ &= V^1 - \Delta, \end{aligned}$$

since $h \in G\{x\}$ is such that $V(h) = V_x$. □

In view of Lemmas 2.1.4, 3.1.8, and 4.1.4 and Theorems 3.2.11 and 4.1.13, we get the following theorem.

4.1.17. THEOREM. *Let L_β be any (not necessarily symmetric) operator of the Freidlin–Wentzell type $\mathscr{L}^1$. Then both* (2.2.1) *and* (2.2.2) *are satisfied with*

$$m = \Delta = V^1 - V^2, \qquad K = \max_{\Gamma \in \mathscr{S}^*} \left(\max_{e \in \mathscr{G}(\Gamma)} \sum_{g_{xy} \ni e} F_{x,y}(\beta) \right)^{-1}$$

and

$$M \le \sup_{x,\beta} \left| -\left(V_x - V^1\right) + \frac{d}{d\beta} \log \Lambda_x(\beta) \right|.$$

We conclude this section with some related estimates and show their asymptotic equivalence.

Let $\Gamma_{x_0} \equiv \{g_{xx_0} \in \mathscr{G};\ x \in \Omega\}$ be a collection of paths connecting each $x \in \Omega$ with $x_0 \in \Omega$. The paths may have repeated nodes, but a given edge appears at most once in a given path. Define

$$\kappa_\beta^{(2)}(\Gamma) := \sup_{e \in \mathscr{G}(\Gamma)} \frac{1}{Q_\beta(e)} \sum_{g_{xy} \ni e} |g_{xy}| \pi_\beta(x) \pi_\beta(y),$$

$$\kappa_\beta^{(3)}(\Gamma_{x_0}) := 2 \sup_{e \in \mathscr{G}(\Gamma_{x_0})} \frac{1}{Q_\beta(e)} \sum_{g_{xx_0} \ni e} |g_{xx_0}| \pi_\beta(x),$$

where $\mathscr{G}(\Gamma_{x_0}) := \{e \in \mathscr{G};\ \exists\, g_{xx_0} \ni e,\ g_{xx_0} \in \Gamma_{x_0}\}$. Note that $\kappa_\beta^{(2)}(\Gamma)$ is Sinclair's constant and $\kappa_\beta^{(3)}(\Gamma_{x_0})$ is Goetze's. Set

$$m^{(2)}(\Gamma) \equiv \sup_{e \in \mathscr{G}} \sup_{g_{xy} \ni e} \left(\mathscr{O}(e) - V_x - V_y + V^1\right), \qquad m^{(2)} := \inf_\Gamma m^{(2)}(\Gamma),$$

and

$$m^{(3)}(\Gamma_{x_0}) \equiv \sup_{e \in \mathscr{G}(\Gamma_{x_0})} \sup_{g_{xx_0} \ni e} \left(\mathscr{O}(e) - V_x\right), \qquad m^{(3)} := \inf_{x_0 \in \Omega} \inf_{\Gamma_{x_0}} m^{(3)}(\Gamma_{x_0}).$$

As in Lemma 4.1.4 we have

$$C(\beta) \ge \frac{1}{\kappa_\beta^{(2)}(\Gamma)}, \qquad C(\beta) \ge \frac{1}{\kappa_\beta^{(3)}(\Gamma_{x_0})},$$

$$\kappa_\beta^{(2)}(\Gamma) \le F^{(2)}(\Gamma) \exp(\beta m^{(2)}(\Gamma))$$

and

$$\kappa_\beta^{(3)}(\Gamma_{x_0}) \le F^{(3)}(\Gamma_{x_0})\exp\big(\beta m^{(3)}(\Gamma_{x_0})\big),$$

where $F^{(2)}(\Gamma)$ and $F^{(3)}(\Gamma_{x_0})$ are positive constants. The following lemma shows that the value $V^1 - V^2$ can be attained by the estimates $\kappa_\beta^{(2)}(\Gamma)$ and $\kappa_\beta^{(3)}(\Gamma_{x_0})$, for well chosen collections of paths.

4.1.18. LEMMA. *Let $L_\beta \in \mathscr{L}$ be symmetric. Then*

$$\Delta = m^{(1)} = m^{(2)} = m^{(3)}.$$

PROOF. To see the first equality choose Γ as in (4.1.10) and use the definition of $m^{(1)}$. Next, for $e \in g_{xy}$ we have $\operatorname{Evel}(g_{xy}) \ge \mathscr{O}(e)$. Thus, $m^{(1)}(\Gamma) \ge m^{(2)}(\Gamma)$. Also setting $y = x_0$ in $m^{(2)}(\Gamma)$ we get that $m^{(2)}(\Gamma) \ge m^{(3)}(\Gamma_{x_0})$ if $V_{x_0} = V^1$, and thus $m^{(2)} \ge m^{(3)}$. It remains to see that $m^{(1)} \le m^{(3)}$. Recall that [see (4.1.14)] $m^{(1)} = \Delta = V^1 - V^2$. Then use the inequality of Lemma 4.1.4 and (3.2.6), by applying $-\beta^{-1}\log(\cdot)$ to both sides and taking the limit as $\beta \to \infty$. □

4.2. *The spectral gap of the collapsed chain and first hitting times.* In this section we describe the asymptotics of the spectral gap of the collapsed chain using the terminology of Section 2.2. The following theorem establishes a link between the speed at which a symmetric $L_\beta \in \mathscr{L}$ concentrates on $V_{\min}$ and the distribution of the first hitting time of $V_{\min}$ (see Corollary 2.2.8 and Proposition 2.2.12).

4.2.1. THEOREM. *Let $L_\beta \in \mathscr{L}$ be symmetric and consider the symmetric operator $\overline{L}_\beta := L_\beta[V_{\min}]$ (cf. Definition 2.1.7). Let $C(\beta)$ [resp. $\overline{C}(\beta)$] be the spectral gap associated with L_β (resp. $\overline{L}_\beta$). Then $\overline{L}_\beta$ is symmetric with*

$$\overline{\Delta} = V[V_{\min}]^1 - V[V_{\min}]^2 = -\lim_{\beta\to\infty} \beta^{-1}\log\big(\overline{C}(\beta)\big)$$

given by

$$\overline{\Delta} = \max_{x \in V_{\min}^c} (\hat{\rho}(x) - V_x).$$

Moreover,

$$\Delta = \overline{\Delta} \vee (R - V^1).$$

PROOF. Let us first check that $\overline{\Delta} = \max_{x \in V_{\min}^c}(\hat{\rho}(x) - V_x)$. Remark that $L_\beta[V_{\min}]$ is symmetric since $\widetilde{L_\beta[V_{\min}]} = \tilde{L}_\beta[V_{\min}] = L_\beta[V_{\min}]$ by Lemma 3.2.12. By definition, $L_\beta[V_{\min}]$ acts on $\overline{\Omega} := V_{\min}^c \cup \{\delta\}$, where δ is any element of $V_{\min}$. Thus we can use the setup of Lemma 4.1.11 to obtain that $\overline{\Delta} = \max_{x \in \overline{\Omega}}(\bar{\rho}(x) - \overline{V}_x)$. Recall that

$$\overline{V}(x \to y) := V[V_{\min}](x \to y) = V(x \to y)$$

if $x, y \in V_{\min}^c$, $\overline{V}(x \to \delta) := \min_{y \in V_{\min}} V(x \to y)$ and

$$\begin{aligned}\overline{V}(\delta \to x) &:= \min_{y \in V_{\min}}\big(V_y + V(y \to x)\big) - \min_{y \in V_{\min}} V_y\\ &= \min_{y \in V_{\min}} V(y \to x).\end{aligned}$$

Our basic observation here is that the weight function $\overline{V}$ has a potential function since $L_\beta[V_{\min}]$ is symmetric. By Lemmas 3.3.3 and 3.3.4,

$$\begin{aligned}\overline{V}(x \to y) - \overline{V}(y \to x) &= V(x \to y) - V(y \to x)\\ &= V_y - V_x \quad \text{for } x, y \in V_{\min}^c,\end{aligned}$$

and therefore $\overline{V}_y - \overline{V}_x = V_y - V_x$, $x, y \in V_{\min}^c$, where $\overline{V}_x \equiv V[V_{\min}]_x$. It follows that there exists a constant D such that $\overline{V}_. \equiv V_. + D$ on $V_{\min}^c$. Let us show that $\overline{V}_{\min} := V[V_{\min}]_{\min} = \{\delta\}$. Since $\overline{V}$ has a potential function, $\overline{V}_x - \overline{V}_\delta = \overline{V}(\gamma_{\delta x}) - \overline{V}(-\gamma_{\delta x})$, where $\gamma_{\delta x} \in C_{\delta x}$ is simple, and $-\gamma_{\delta x}$ is the reversal of $\gamma_{\delta x}$. Suppose that $\gamma_{\delta x}$ has the normal form $(e_0, \dots, e_n)$, where $e_0^- = \delta$, $e_0^+ = y \neq \delta$ and $e_n^+ = x$. Let $\overline{\gamma}_{yx}$ be the segment of $\gamma_{\delta x}$ which takes y to x, and let $-\overline{\gamma}_{yx}$ be its reversal. Then

$$\begin{aligned}\overline{V}_x - \overline{V}_\delta &= \overline{V}(\gamma_{\delta x}) - \overline{V}(-\gamma_{\delta x})\\ &= \overline{V}(\gamma_{yx}) - \overline{V}(-\gamma_{yx}) + \overline{V}(\delta \to y) - \overline{V}(y \to \delta)\\ &= V(\gamma_{yx}) - V(-\gamma_{yx}) + \overline{V}(\delta \to y) - \overline{V}(y \to \delta)\\ &= V_x - V_y + \min_{p \in V_{\min}} V(p \to y) - \min_{p \in V_{\min}} V(y \to p)\\ &= V_x - V_y + \min_{p \in V_{\min}} \big(V(y \to p) + V_y - V_p\big) - \min_{p \in V_{\min}} V(y \to p)\\ &= V_x - V^1 > 0\end{aligned}$$

since $x \notin V_{\min}$ and, therefore, we have just proved that $\overline{V}_{\min} = \{\delta\}$, that is, $\overline{V}_x > \overline{V}_\delta$, $\forall\ x \in V_{\min}^c$. Next let us recall some definitions: $\overline{\rho}(x, \delta) := \min_{\gamma \in C_{x\delta}} \overline{\text{Evel}}(\gamma)$, $\overline{\text{Evel}}(\gamma) := \max_{e \in \gamma} \overline{\mathscr{O}}(e)$ and $\overline{\mathscr{O}}(e) := \overline{V}_{e^-} + \overline{V}(e^- \to e^+)$. We have $\overline{\rho}(x) = \overline{\hat{\rho}}(x)$, $\overline{R} = \overline{V}_\delta$, since $|\overline{V}_{\min}| = 1$. Observe that it is sufficient to consider simple paths γ of $C_{x\delta}$ [by symmetry $\overline{\text{Evel}}(\gamma) = \max_{e \in \gamma} \overline{\overline{\mathscr{O}}}(e)$, with $\overline{\overline{\mathscr{O}}}(e) \equiv \overline{\mathscr{O}}(-e)$]. Suppose that γ has the normal form $\gamma = (e_0, \dots, e_n)$, with $e_0^- = x$ and $e_n^+ = \delta$. Then

$$\overline{\text{Evel}}(\gamma) = \max_{e \in \gamma} \overline{\mathscr{O}}(e) = \Big(\max_{e \in \gamma \setminus e_n} \overline{\mathscr{O}}(e)\Big) \vee \overline{\mathscr{O}}(e_n) = \Big(\max_{e \in \gamma \setminus e_n} \mathscr{O}(e)\Big) \vee \overline{\mathscr{O}}(e_n).$$

On the other hand,

$$\overline{\mathscr{O}}(e_n) = \overline{V}_{e_n^-} + \overline{V}(e_n^- \to e_n^+) = V_{e_n^-} + D + \min_{y \in V_{\min}} V(e_n^- \to y).$$

Therefore

$$\overline{\text{Evel}}(\gamma) = D + \Big(\max_{e \in \gamma \setminus e_n} \mathscr{O}(e) \vee \big(\min_{y \in V_{\min}} \mathscr{O}(e_n^- \to y)\big)\Big)$$

and thus

$$\begin{aligned}\overline{\rho}(x, \delta) &= \min_{\gamma \in C_{x\delta}} \overline{\text{Evel}}(\gamma) = D + \min_{\gamma \in C_{x\delta}} \Big(\max_{e \in \gamma \setminus e_n} \mathscr{O}(e) \vee \Big(\min_{y \in V_{\min}} \mathscr{O}(e_n^- \to y)\Big)\Big)\\ &= D + \min_{y \in V_{\min}} \min_{\gamma' \in C_{xy}} \text{Evel}(\gamma') = D + \min_{y \in V_{\min}} \rho(x, y) = D + \hat{\rho}(x)\\ &= \big(\overline{V}_x - V_x\big) + \hat{\rho}(x).\end{aligned}$$

Finally,

$$\overline{\Delta} = \max_{x\in\overline{\Omega}}\left(\bar{\rho}(x) - \overline{V}_x\right) = \max_{x\in V_{\min}^c}\left(\bar{\rho}(x) - \overline{V}_x\right) = \max_{x\in V_{\min}^c}\left(\bar{\rho}(x,\delta) - \overline{V}_x\right)$$

$$= \max_{x\in V_{\min}}\left(\hat{\rho}(x) + \left(\overline{V}_x - V_x\right) - \overline{V}_x\right) = \max_{x\in V_{\min}^c}\left(\hat{\rho}(x) - V_x\right),$$

as required.

Next let us show that $\Delta = \overline{\Delta} \vee (R - V^1)$. By definition

$$\begin{aligned}\Delta &= \max_{x\in\Omega} D(x) = \max_{x\in\Omega}\left(\rho(x) - V_x\right)\\ &= \max_{x\in\rho_{\min}}\left(\rho(x) - V_x\right) \vee \max_{x\in\rho_{\min}^c}\left(\rho(x) - V_x\right)\\ &= \max_{x\in\rho_{\min}}\left(R - V_x\right) \vee \max_{x\in\rho_{\min}^c}\left(\hat{\rho}(x) - V_x\right)\\ &= (R - V^1) \vee \max_{x\in\rho_{\min}^c}\left(\hat{\rho}(x) - V_x\right),\end{aligned}$$

since $V_{\min} \subset \rho_{\min}$. Therefore, we must check that

$$(R - V^1) \vee \max_{x\in\rho_{\min}^c}\left(\hat{\rho}(x) - V_x\right) = (R - V^1) \vee \max_{x\in V_{\min}^c}\left(\hat{\rho}(x) - V_x\right),$$

where $\rho_{\min}^c \subset V_{\min}^c$. Let $x \in V_{\min}^c \cap \rho_{\min}$. Then by Lemma 4.1.11, $\hat{\rho}(x) \le \rho(x) \equiv R$ and, therefore, $\max_{x\in V_{\min}^c\cap\rho_{\min}}(\hat{\rho}(x) - V_x) < (R - V^1)$, since $V_x > V^1$, $\forall x \in V_{\min}^c$. □

4.2.2. REMARK. Take a symmetric Metropolis chain, as in, Example 3.1.9. The assertion $\Delta = \overline{\Delta} \vee (R - V^1)$ corresponds to a result of Chiang and Chow (1988b).

4.3. *The filling method.* We give an alternative approach to (4.1.4). We start with a general result. Let $L_\beta \in \mathscr{L}$ and π_β be given and assume symmetry, that is,

$$Q_\beta(x,y) = \pi_\beta(x) q_\beta(x,y) \equiv Q_\beta(y,x).$$

Consider an arbitrary weight function $W(x \to y)$ with potential function $\Psi(\cdot)$ on Ω (W does not need to be nonnegative!). Define the new transition

$$q'_\beta(x,y) := \exp(-\beta W(x\to y)) q_\beta(x,y), \quad \text{for } x \ne y,$$

with invariant distribution

$$\pi'_\beta(x) \equiv \exp(-\beta\Psi(x))\pi_\beta(x) Z(\beta)^{-1},$$

where $Z(\beta) := \sum_{x\in\Omega}\exp(-\beta\Psi(x))\pi_\beta(x)$. Note that $L'_\beta \in \mathscr{L}$ if and only if $V(x\to y) + W(x\to y) \ge 0$, $\forall\, x, y \in \Omega$. We have

$$\begin{aligned}Q'_\beta(x,y) &:= \pi'_\beta(x) q'_\beta(x,y) = \exp(-\beta(\Psi(x) + W(x\to y))) Z(\beta)^{-1} Q_\beta(x,y)\\ &= \exp(-\beta\mathscr{O}'(x\to y)) Z(\beta)^{-1} Q_\beta(x,y) = Q'_\beta(y,x),\end{aligned}$$

where $\mathscr{O}'(x \to y) := \Psi(x) + W(x \to y)$. Let $C(\beta)$ and $C'(\beta)$ be the spectral gaps associated with L_β and L'_β.

4.3.1. LEMMA. *Let $L_\beta \in \mathscr{L}$ be symmetric. Then*

$$C(\beta) \geq \exp\big(-\beta(M - m_Q)\big)C'(\beta),$$

where

$$M := \max_{x \in \Omega} \Psi(x) \quad \text{and} \quad m_Q := \min_{e:\, Q_0(e) > 0} \mathscr{O}'(e).$$

PROOF. For any edge $e \in E$ set $\varphi(e) := \rho(e^+) - \varphi(e^-)$. Then

$$\begin{aligned}
\mathscr{E}_{\pi_\beta}(\varphi, \varphi) &= \tfrac{1}{2} \sum_{e \in E} \varphi(e)^2 Q_\beta(e) \\
&\geq \tfrac{1}{2} \sum_{e \in E} \varphi(e)^2 Q'_\beta(e) Z(\beta) \exp(\beta m_Q) \\
&\geq C'(\beta) \|\varphi - \langle 1, \varphi \rangle_{\pi'_\beta}\|^2_{\pi'_\beta} Z(\beta) \exp(\beta m_Q) \\
&\geq C'(\beta) \|\varphi - \langle 1, \varphi \rangle_{\pi'_\beta}\|^2_{\pi_\beta} \exp\big(-\beta(M - m_Q)\big) \\
&\geq C'(\beta) \|\varphi - \langle 1, \varphi \rangle_{\pi_\beta}\|^2_{\pi_\beta} \exp\big(-\beta(M - m_Q)\big). \qquad \square
\end{aligned}$$

Consider the new potential

$$\Psi(x) := \rho(x) - V_x,$$

where $\rho(x) := \max_{y \in V_{\min}} \rho(x, y)$ (see Definition 4.1.7) and the transitions

$$W(x \to y) := (\rho(y) - \rho(x))^+ - V(x \to y).$$

Clearly W has potential $\Psi(\cdot)$ with

$$\mathscr{O}'(x \to y) = \rho(x) \vee \rho(y) - \mathscr{O}(x \to y).$$

Note that $\mathscr{O}' \geq 0$, so $m_Q \geq 0$, and by construction

$$M = \max_{x \in \Omega} \Psi(x) = \Delta;$$

see Lemma 4.1.11. This yields

$$C(\beta) \geq \exp(-\beta\Delta)C'(\beta),$$

where $C'(\beta)$ is the spectral gap associated with the Metropolized chain $L'_\beta \in \mathscr{L}$ defined by $q'_\beta(x, y) = \Lambda_{x,y}(\beta)\exp(-\beta(\rho(y) - \rho(x))^+)$. As we will see later, the inequality permits a better understanding of the estimation procedure and *reduces the problem to the estimation of the spectral gap of a Metropolis* chain. In the sequel, we will prove the following theorem.

4.3.2. THEOREM. *Let $L_\beta \in \mathscr{L}$ be symmetric and let $\Gamma \in \mathscr{S}^*$. Then*

$$(m')^{(1)}(\Gamma) = (m')^{(1)} = 0$$

and

$$C(\beta) \geq \exp(-\beta\Delta)C'(\beta)$$

with

$$\inf_{\beta>0} C'(\beta) \geq \max_{\Gamma \in \mathscr{S}^*} \left(\max_{e \in \mathscr{G}(\Gamma)} \sum_{g_{xy} \ni e} F_{x,y}(\beta) \right)^{-1} > 0.$$

For an illustration of the application of Theorem 4.3.2, see Examples 5.1.6 and 5.1.7. The proof proceeds in two steps.

4.3.3. LEMMA. *Let $\Gamma \in \mathscr{S}^*$. Then $\hat{\rho}(\cdot)$ is decreasing along any path $g_{xs(x)}$ of Γ such that $\hat{\rho}(x) = \rho(x, s(x))$, $s(x) \in V_{\min}$.*

PROOF. Take some $x \in \Omega$ and let $s(x) \in V_{\min}$ be such that $\hat{\rho}(x) = \rho(x, s(x))$. Choose any node y of $g_{xs(x)}$. Define $\bar{g}_{ys(x)}$ to be the oriented segment of $g_{xs(x)}$ which takes y to $s(x)$. By assumption $\rho(p,q) = \text{Evel}(g_{pq}) = \min_{\gamma \in C_{pq}} \text{Evel}(\gamma)$, $\forall\ p, q \in \Omega$, and, therefore, $\text{Evel}(\bar{g}_{ys(x)}) \geq \text{Evel}(g_{ys(x)})$. Then we have

$$\begin{aligned}\hat{\rho}(y) &:= \min_{p \in V_{\min}} \text{Evel}(g_{yp}) \leq \text{Evel}(\bar{g}_{ys(x)}) \leq \text{Evel}(g_{ys(x)}) \leq \text{Evel}(g_{xs(x)}) \\ &= \hat{\rho}(x),\end{aligned}$$

since

$$\text{Evel}(\bar{g}_{ys(x)}) = \max\{\mathscr{O}(e); e \in \bar{g}_{ys(x)}\} \leq \max\{\mathscr{O}(e); e \in g_{xs(x)}\}. \qquad \square$$

4.3.4. LEMMA. *Assume that $\Gamma \in \mathscr{S}^*$ and let $\mathscr{G}_\rho := (\rho_{\min}, E_\rho)$ be the subgraph of $\mathscr{G} = (\Omega, E)$ with node set $\rho_{\min}$ and edge set $E_\rho := \{e \in E;\ e^- \in \rho_{\min}, e^+ \in \rho_{\min}\}$. Then $\mathscr{G}_\rho$ is connected.*

PROOF. For $x \in \rho_{\min}$ [i.e., $\rho(x) = R$], let $s(x)$ be the element of $V_{\min}$ such that $\hat{\rho}(x) = \text{Evel}(g_{xs(x)})$. By Lemma 4.3.3 we know that $\hat{\rho}(p) \leq \hat{\rho}(x) \leq R$, for all nodes p of $g_{xs(x)}$, and thus that $g_{xs(x)} \subset \rho_{\min}$ and $g_{ys(y)} \subset \rho_{\min}$, for $x, y \in \rho_{\min}$. It follows that the construction of a path $\gamma_{xy} \in C_{xy}$, $\gamma_{xy} \subset \rho_{\min}$ will be achieved as soon as we can find $\gamma \in C_{pq}$, $\gamma \subset \rho_{\min}$, $\forall\ p, q \in V_{\min}$. Take any node r of g_{pq} and let $\bar{g}_{pr}$ be the oriented segment of g_{pq} which takes p to r. Thus if $-\bar{g}_{pr}$ is the reversal of $\bar{g}_{pr}$, we have

$$\begin{aligned}\text{Evel}(g_{pq}) = \max\{\mathscr{O}(e); e \in g_{pq}\} &\geq \max\{\mathscr{O}(e); e \in \bar{g}_{pr}\} \\ &= \max\{\mathscr{O}(e); e \in -\bar{g}_{pr} \in C_{rp}\} \\ &\geq \text{Evel}(g_{rp}) \\ &\geq \min_{t \in V_{\min}} \text{Evel}(g_{rt}) = \hat{\rho}(r),\end{aligned}$$

since $\mathscr{O}(e) \equiv \mathscr{O}(-e)$. Thus if r were not in $\rho_{\min}$, that is, if $\hat{\rho}(r) > R$, we would have $\rho(p,q) = \text{Evel}(g_{pq}) > R$, which is a contradiction. $\square$

PROOF OF THEOREM 4.3.2. As $\mathscr{G}_\rho$ is connected, let T_{x_0} be any directed spanning tree of $\mathscr{G}_\rho$, pointing at x_0. By definition $V_\rho(x \to y) = (\rho(y) - \rho(x))^+$ and, therefore, $V_\rho(T_{x_0}) = 0$. Consider the following algorithm:

(i) Take any $x_1 \in \rho^c_{\min}$ and consider a path $g_{x_1 s(x_1)}$ such that $\hat{\rho}(x_1) = \operatorname{Evel}(g_{x_1 s(x_1)})$, with as usual $s(x_1) \in V_{\min}$. Stop it as soon as it meets the spanning tree T_{x_0}, at some node x_2. Let $\bar{g}_{x_1 x_2}$ be the oriented segment of $g_{x_1 s(x_1)}$ which takes x_1 to x_2. We have $V_\rho(\bar{g}_{x_1 x_2}) = 0$ since by Lemma 4.3.3, $\rho(\cdot) = \hat{\rho}(\cdot)$ is decreasing along $g_{x_1 s(x_1)}$.

(ii) Take some $x_3 \in \rho^c_{\min} \setminus \bar{g}_{x_1 x_2}$ and proceed as above to get a segment $\bar{g}_{x_2 x_4}$ with $V_\rho(\bar{g}_{x_3 x_4}) = 0$, with the difference that the path is stopped at the first index at which it meets $T_{x_0} \cup \bar{g}_{x_1 x_2}$.

(iii) Repeat the above operation until $\rho^c_{\min} \setminus \bar{g}_{x_1 x_2} \setminus \bar{g}_{x_2 x_4} \setminus \cdots \setminus \bar{g}_{x_k x_{k+1}} = \varnothing$, for some index $k \in \mathbb{N}$.

Set $T_\rho := T_{x_0} \cup \bar{g}_{x_1 x_2} \cup \cdots \cup \bar{g}_{x_k x_{k+1}} \in G(\{x_0\})$, which satisfies, by construction, $V_\rho(T_\rho) = 0$. Therefore, $V^1_\rho = V^2_\rho = 0$ and $(m')^{(1)} = 0$. □

4.4. *Spectral properties of subchains of a given chain.* In this section, we diverge slightly from the symmetric setting of the previous sections. Let us start with a simple remark. Consider two irreducible transition functions $L' = (q'(x, y);\ x, y \in \Omega)$ and $L = (q(x, y);\ x, y \in \Omega)$, with invariant distributions π' and π. Suppose that $L' \leq L$ in the sense that $q'(x, y) \leq q(x, y)$, $x \neq y$. Then $\pi \equiv \pi'$ implies that $\tilde{C}' \leq \tilde{C}$, where $\tilde{C}'$ and $\tilde{C}$ are the spectral gaps of the symmetrized versions of L' and L. This follows from the variational formula (2.1.3) and the fact that $L' \leq L$ implies $L'^* \leq L^*$ when $\pi' \equiv \pi$. Next consider two transition functions L_β and $L'_\beta \in \mathscr{L}$. We write

$$q'_\beta(x, y) \prec q_\beta(x, y) \quad \text{if} \quad \lim_{\beta \to \infty} \beta^{-1} \log q'_\beta(x, y) \leq \lim_{\beta \to \infty} \beta^{-1} \log q_\beta(x, y)$$

and

$$L'_\beta \prec L_\beta,$$

if the above inequality holds for all $x \neq y \in \Omega$. In terms of the corresponding weight functions V and V', this simply means $V'(x \to y) \geq V(x \to y)$, $x \neq y \in \Omega$. If $L'_\beta \prec L_\beta$ and $L_\beta \prec L'_\beta$, then $L_\beta \asymp L'_\beta$.

4.4.1. LEMMA. *Let L_β and L'_β be two transition kernels of $\mathscr{L}$ such that $L'_\beta \prec L_\beta$. If the corresponding invariant distributions satisfy $\pi'_\beta \asymp \pi_\beta$, then*

$$C'(\beta) \prec C(\beta).$$

PROOF. Since $\widetilde{C'}(\beta) \asymp C'(\beta)$ and $\tilde{C}(\beta) \asymp C(\beta)$ by Theorem 3.2.11, it is enough to prove that

$$\widetilde{C'}(\beta) \prec \tilde{C}(\beta). \tag{4.4.2}$$

First note that $\pi'_\beta \asymp \pi_\beta$ implies the existence of a constant $k \in \mathbb{R}$ with $V'_\cdot \equiv V_\cdot + k$. From this we have

$$\begin{aligned}\widetilde{V'}(e) &= \min\{V'(e); V'(-e) + V'_{e^+} - V'_{e^-}\} = \min\{V'(e); V'(-e) + V_{e^+} - V_{e^-}\}\\ &\geq \min\{V(e); V(-e) + V_{e^+} - V_{e^-}\} = \tilde{V}(e),\end{aligned}$$

that is, $\widetilde{L'_\beta} \prec \tilde{L}_\beta$. As both $\widetilde{L'_\beta}$ and $\tilde{L}_\beta$ are symmetric, we can use the setup of Section 4.1. In particular, we have

$$\begin{aligned}\mathscr{O}'(e) &= V'_{e^-} + \widetilde{V'}(e^- \to e^+) = V_{e^-} + k + \widetilde{V'}(e^- \to e^+)\\ &\geq V_{e^-} + k + \tilde{V}(e^- \to e^+) = \mathscr{O}(e) + k.\end{aligned}$$

This yields $\rho'(x, y) \geq \rho(x, y) + k$, $x \neq y$, and

$$\begin{aligned}\Delta' &= \max_{x \neq y}\left(\rho'(x, y) + V'^1 - V'_x - V'_y\right)\\ &\geq \max_{x \neq y}\left(\rho(x, y) + k + V^1 + k - V_x - k - V_y - k\right) = \Delta.\end{aligned}$$

Now (4.4.2) follows from Theorem 4.1.14. □

Let $L_\beta \in \mathscr{L}$ be a transition function on Ω and let $\mathscr{G} = (\Omega, E)$ be its associated graph. For any subset $E^0 \subset E$ of edges, consider the *subchain* L^0_β of L_β given by $q^0_\beta(e) \equiv q_\beta(e) I_{E^0}(e)$, where I_{E^0} is the indicator function. As is easily seen, $L^0_\beta \in \mathscr{L}$ if and only if the graph $\mathscr{G}^0 := (\Omega, E^0)$ is connected, that is, the chain L^0_β is irreducible. Let us denote by π_β^0 its invariant measure and $C^0(\beta)$ its spectral gap. Let $V^0(e) := V(e)$, $e \in E^0$, and $V^0(e) = +\infty$, $e \notin E^0$, be the corresponding weight function. A natural question in this setting is to compare the asymptotics of π_β^0 and π_β and of $C^0(\beta)$ and $C(\beta)$. One knows that, in general, $\operatorname{supp}(\pi_\infty) \neq \operatorname{supp}(\pi_\infty^0)$ [see Mazza (1992) and Trouvé (1992)]. We can view L^0_β as a deletion operation on $E \setminus E^0$. In contrast to the collapsing operation (see Lemma 2.1.5), the deletion operation does not necessarily commute with the symmetrization. More precisely we have the following lemma.

4.4.3. LEMMA. *The following are equivalent*:

$$\widetilde{L^0_\beta} \asymp \left(\tilde{L}_\beta\right)^0 \tag{4.4.4}$$

and

$$E^0 = -E^0 := \{-e; e \in E^0\} \quad \text{and} \quad \pi_\beta^0 \asymp \pi_\beta. \tag{4.4.5}$$

PROOF. Assume (4.4.4) and take $e \in E \setminus E^0$. Then $(\tilde{V})^0(e) = \infty$ and

$$\widetilde{V^0}(e) = \min\{V^0(e), V^0(-e) + V^0_{e^+} - V^0_{e^-}\} = \min\{\infty, V^0(-e) + V^0_{e^+} - V^0_{e^-}\}.$$

If both lines agree by assumption, then $\widetilde{V^0}(e) = \infty$ and $V^0(-e) = \infty$. Therefore, $-e \in E \setminus E^0$. This implies $E^0 = -E^0$. Next take $e \in E^0$. Since $-e \in E^0$, we have

$$\widetilde{V^0}(e) - \widetilde{V^0}(-e) = V^0_{e^+} - V^0_{e^-}$$

and

$$(\tilde{V})^0(e) - (\tilde{V})^0(-e) = \tilde{V}(e) - \tilde{V}(-e) = V_{e^+} - V_{e^-};$$

see Lemma 3.3.3. Thus $V^0_{e^+} - V^0_{e^-} = V_{e^+} - V_{e^-}$ for each $e \in E^0$. As $\mathscr{G}^0$ is connected by assumption, there is a constant $k \in \mathbb{R}$ such that $V^0_\cdot \equiv V_\cdot + k$ and, therefore, $\pi^0_\beta \asymp \pi_\beta$.

Next assume (4.4.5). Then, for $e, -e \in E^0$,

$$\begin{aligned}\widetilde{V^0}(e) &= \min\{V^0(e), V^0(-e) + V^0_{e^+} - V^0_{e^-}\}\\ &= \min\{V(e), V(-e) + V_{e^+} - V_{e^-}\} = (\tilde{V})^0(e),\end{aligned}$$

whereas for $e, -e \in E \setminus E^0$,

$$\widetilde{V^0}(e) = \min\{V^0(e), V^0(-e) + V^0_{e^+} - V^0_{e^-}\} = \infty = (\tilde{V})^0(e). \qquad \square$$

4.4.6. THEOREM. *Let $L^0_\beta \in \mathscr{L}$ be a subchain of L_β with spectral gap $C^0(\beta)$ and invariant distribution π^0_β. If $\widetilde{L^0_\beta} \asymp (\tilde{L}_\beta)^0$, then*

$$\pi^0_\beta \asymp \pi_\beta \quad \textit{and} \quad C^0(\beta) \prec C(\beta). \tag{4.4.7}$$

In particular, (4.4.7) *holds when both L^0_β and L_β have potentials.*

PROOF. In view of the above lemmas, we only have to show that the last line implies (4.4.4). If both L^0_β and L_β have potentials, then $\widetilde{L^0_\beta} \asymp L^0_\beta$ and $\tilde{L}_\beta \asymp L_\beta$. Thus $(\tilde{L}_\beta)^0 \asymp L^0_\beta \asymp \widetilde{L^0_\beta}$. $\square$

4.4.8. EXAMPLE.

(i) L^0_β and L_β are symmetric.
(ii) L_β has a potential function and $\mathscr{G}^0$ is a tree.

See Example 3.3.7.

4.4.9. REMARK. When one of the two chains $L_\beta \in \mathscr{L}$ and $L^0_\beta \in \mathscr{L}$ does not have a potential function, both the situations $\Delta < \Delta^0$ and $\Delta > \Delta^0$ are possible. An example of this situation is the case of sequential and parallel annealing on Gibbsian fields [Deuschel and Mazza (1994)].

5. Examples.

5.1. *Illustration with Metropolis chains.* Let us see what happens with the spectral estimate (4.1.5) for symmetric Metropolis chains. Define $\mathscr{L}_0$ to be the subfamily of $\mathscr{L}$ containing the transition matrices of the form

$$q_\beta(x, y) = q_0(x, y)\exp(-\beta V(x \to y)) \quad \text{for } x \neq y,$$

with q_0 irreducible on Ω and π_0-symmetric for a given probability measure π_0. For the transition kernels of this family we have

$$\pi_\beta(x) \equiv \exp(-\beta V_x)\pi_0(x)Z(\beta)^{-1},$$

where $Z(\beta) := \sum_{y\in\Omega}\exp(-\beta V_y)\pi_0(y)$. The estimates of Section 4 have the closed form

$$C(\beta) \ge \kappa_\beta^{(i)}(\Gamma)^{-1} \ge \exp(\beta V^1)Z(\beta)\kappa_0^{(i)}(\Gamma)^{-1}\exp(-\beta m^{(i)}(\Gamma))$$
$$\ge \exp(-\beta m^{(i)}(\Gamma))\pi_0(V_{\min})\kappa_0^{(i)}(\Gamma)^{-1}, \qquad i = 1,2,$$

and

$$C(\beta) \ge \kappa_\beta^{(3)}(\Gamma_{x_0})^{-1} \ge \exp\big(-\beta m^{(3)}(\Gamma_{x_0})\big)\kappa_0^{(3)}(\Gamma_{x_0})^{-1}. \qquad (5.1.1)$$

For $\Gamma \in \mathscr{S}^*$, the filling method yields

$$C(\beta) \ge \exp(-\beta\Delta)C'(\beta) \ge \exp(-\beta\Delta)(\kappa')_\beta^{(1)}(\Gamma)^{-1}$$
$$\ge \exp(-\beta\Delta)\pi_0(\rho_{\min})\kappa_0^{(1)}(\Gamma)^{-1},$$

since $\pi_0' \equiv \pi_0$ and $\kappa_0^{(1)}(\Gamma) \equiv (\kappa')_0^{(1)}(\Gamma)$. Note that $V_{\min} \subset \rho_{\min}$ and, therefore, $\pi_0(\rho_{\min})$ could be much bigger than $\pi_0(V_{\min})$. In particular if $\rho_{\min} = \Omega$, that is, $\rho(\cdot) \equiv R$, then $q_\beta' \equiv q_0$ and $C'(\beta) = C'(0) = C(0)$, $\forall\, \beta \ge 0$, and, therefore,

$$C(\beta) \ge \exp(-\beta\Delta)C(0).$$

Inequality (5.1.1) decomposes in three parts:

1. Set $\beta = 0$. Note that $\kappa_0^{(1)}(\Gamma)$ is a bound for the second largest eigenvalue of L_0. Define Γ_1 to be the collection of paths which best approximate the spectral gap $C(0)$.
2. As we have seen, $m^{(1)}(\Gamma) \ge \Delta = V^1 - V^2$, $\forall\ \Gamma$. Choose Γ_2 such that $m^{(1)}(\Gamma_2) = \Delta$ [see (4.1.10)].
3. The extra factor $\exp(-\beta V^1)Z(\beta)^{-1}$ does not depend of Γ and can be taken as a systematic bound.

For an arbitrary collection Γ, we have $m^{(1)}(\Gamma) \ge m^{(1)}(\Gamma_2)$ and $\kappa_0^{(1)}(\Gamma) \ge \kappa_0^{(1)}(\Gamma_1)$, and, therefore, $m^{(1)}(\Gamma_1) \ge m^{(1)}(\Gamma_2)$ and $\kappa_0^{(1)}(\Gamma_2) \ge \kappa_0^{(1)}(\Gamma_1)$. This illustrates the competition between the speed of diffusion of the basic chain L_0 on Ω and the difficulty for the chain to reach the barrier Δ.

Note that for a given potential U, the Metropolis chain L_β with spectral gap $C(\beta)$ gives the *fastest chain*, that is, among all $L_\beta' \in \mathscr{L}_0$ with potential U and spectral gap $C'(\beta)$, we have

$$q_\beta'(x,y) \le q_0(x,y)\exp\big(-\beta(U(y)-U(x))^+\big) = q_\beta(x,y), \qquad x \ne y \in \Omega,$$

and, therefore, $C'(\beta) \le C(\beta)$.

In the following example we consider only Metropolis chains.

5.1.2. Example (A "Metropolized" Ehrenfest urn). Let $\Omega := \{0,1,\dots,d-1,d\}$ and consider the so-called "distance chain" $q_0(x,y)$ on Ω given by the transition probabilities $q_0(x, x+1) := 1 - x/d$ and $q_0(x, x-1) := x/d$, which is π_0-symmetric where $\pi_0(x) := 2^{-d}\binom{d}{x}$. The second largest eigenvalue is $2/d$ and the Poincaré bound is $\kappa_0^{(1)} = d/2$ [see Diaconis and Stroock (1991)]. Such a chain is a model of the classic Ehrenfest urn. Our "Metropolized" urn will be the chain obtained by taking $U(x) \equiv x$ with $U_* = 0$. Set for convenience $z := \exp(-\beta)$. With this choice, the chain becomes $q_\beta(x, x+1) = q_0(x, x+1)z$ and $q_\beta(x, x-1) = q_0(x, x-1)$ and has

$$\pi_\beta(x) = z^x \pi_0(x) \Big/ \sum_{y=0}^{d} z^y \pi_0(y)$$

as its invariant measure. As the graph $\mathscr{G} = (\Omega, E)$ is a tree, there is a unique collection of paths Γ for which $m^{(1)}(\Gamma) = \Delta = 0$ since $U(\cdot)$ is increasing, namely, the set of geodesics. Therefore, (5.1.1) becomes $\kappa_0^{(1)}/\sum_{y=0}^{d} z^y \pi_0(y)$. Then we have

$$\kappa_\beta^{(1)} \le \kappa_0^{(1)} \Big/ \sum_{y=0}^{d} z^y 2^{-d}\binom{d}{y} = \kappa_0^{(1)} 2^d \Big/ \sum_{y=0}^{d} z^y \binom{d}{y} \tag{5.1.3}$$
$$= \kappa_0^{(1)} 2^d/(1+z)^d.$$

Thus when $z < 1$ is fixed, the bound is exponential in d. It would be of interest to have the exact value of $C(\beta)$ to see the loss in the estimation procedure. Now let us see what happens with Goetze's estimate $\kappa_\beta^{(3)}$ for this example. As $\mathscr{G}$ is a tree, we take Γ_{x_0} as the directed spanning tree $\mathscr{G}_{x_0}$ obtained by pointing $\mathscr{G}$ at location x_0, $\forall\ x_0 \in \Omega$. By (5.1.1), $\kappa_\beta^{(3)} \le \exp(\beta m^{(3)}(\Gamma_{x_0}))\kappa_0^{(3)}(\Gamma_{x_0})$, $\forall\ x_0 \in \Omega$. By Lemma 4.1.18 we have $m^{(3)} = m^{(1)} = 0$ since $m^{(1)} = \Delta = 0$. Concerning $m^{(3)}(\Gamma_{x_0})$, it is not hard to check that $m^{(3)}(\Gamma_0) = 0$ and $m^{(3)}(\Gamma_{x_0}) > 0$, $x_0 \ne 0$, and, therefore, as $\beta \gg 1$, the optimal choice is $x_0 = 0$. By definition, $\kappa_0^{(3)}(\Gamma_0) = 2\sup_{e \in \mathscr{G}(\Gamma_0)} Q_0^{-1}(e)\sum_{g_{x0} \ni e}|g_{x0}|\pi_0(x)$. Set $e = (j \to j-1)$, where $1 \le j \le d$. Then we have

$$Q_0^{-1}(e) \sum_{g_{x0} \ni e} |g_{x0}|\pi_0(x) = \frac{1}{\binom{d}{j}(j/d)} \sum_{x \ge j} x\binom{d}{x} = \frac{1}{\binom{d-1}{j-1}} \sum_{x \ge j} x\binom{d}{x}$$

which is maximized at $j = 1$, and thus we get

$$\kappa_0^{(3)}(\Gamma_0) = \sum_{x=1}^{d} x\binom{d}{x} = 2^d \sum_{x=0}^{d} x 2^{-d}\binom{d}{x} = 2^d(d/2) = d2^{d-1}.$$

Consider the limiting case $\beta \to \infty$ or $z = 0$: The matrix L_β converges to the triangular matrix L_∞ which has $(0, -(x/d)\colon 0 < x \le d)$ as diagonal and thus as spectrum, so that $C(\infty) = d^{-1}$, which is far from (1) $(\kappa_0^{(1)} 2^d/(1+z)^d)^{-1} \to (d2^{d-1})^{-1}$ as $z \to 0$ and (2) $(\kappa_0^{(3)}(\Gamma_0))^{-1} \equiv (2^{d-1}d)^{-1}$.

Now let us see what happens with the Goetze estimate $\kappa_\beta^{(3)}(\Gamma_0)$ as $\beta \to \infty$, without using (5.1.1). Any arrow e of $\mathscr{G}(\Gamma_0)$ has the form $e = (j \to j-1)$, $1 \le j \le d$, and, therefore,

$$Q_\beta^{-1}(e) \sum_{g_{x0} \ni e} |g_{x0}| \pi_\beta(x) = \frac{1}{z^j Q_0(j)} \sum_{x \ge j} xz^x \pi_0(x) = \frac{1}{\binom{d-1}{j-1}} \sum_{x \ge j} xz^{x-j} \binom{d}{x}$$

$$\le \frac{d}{dz}(1+z)^d$$

(as $0 < z < 1$, the maximum of the above expression is obtained for $j = 1$). Therefore, $\kappa_\beta^{(3)}(\Gamma_0) \equiv 2d(1+z)^{d-1}$, for $\beta \ge \beta_0(d)$. Thus $\lim_{\beta \to \infty} \kappa_\beta^{(3)}(\Gamma_0) = 2d$, which equals $\lim_{\beta \to \infty}(2C(\beta))^{-1}$ and is more than acceptable and shows that the upper bound of (5.1.1) achieves the "barrier" $\Delta = m^{(1)} = m^{(3)} = 0$, but is poor as a function of $d \in \mathbb{N}$.

5.1.4. REMARK. Lemma 4.1.18 and (5.1.1) achieve the barrier Δ in the limit $\beta \to \infty$. Being general, they do not give information on $C(\beta)$ as a function of the problem size $d \in \mathbb{N}$ (see Example 5.1.2). In his work, Goetze shows that it is possible to give examples with more accurate estimates for $\min_{x_0 \in \Omega} \kappa_\beta^{(3)}(\Gamma_{x_0})$.

5.1.5. EXAMPLE (Collapsed symmetric Metropolis chains and hitting times). Take a symmetric Metropolis chain L_β on Ω, with cost function $U(\cdot)\colon \Omega \to \mathbb{R}$. Using Corollary 2.2.8 and the results of this section, we see that the smallest m that satisfies (2.2.1) is given by $\Delta = V^1 - V^2$. Consider Proposition 2.2.12, which gives a bound for the distribution of the first hitting time of $U_{\min}$. The smallest $\overline{m}$ for which the analogue of (2.2.1) holds is $\overline{\Delta}$ (see, for example, Remark 4.2.2). The quantities Δ and $\overline{\Delta}$ are linked by the relation $\Delta = \overline{\Delta} \vee (R - U_*)$, where $U_* := \min_{x \in \Omega} U(x)$, since $U(\cdot)$ is a potential function. It is not hard to check that

$$R = \sup_{x, y \in U_{\min}} \inf\Big\{h \ge 0; \exists\, \gamma \in C_{xy} \text{ with } q_0(\gamma) > 0 \text{ and } \max_{e \in \gamma} U(e) = h\Big\},$$

where we set $F = U_*$ if $|U_{\min}| = 1$. In the same vein we have

$$\max_{x \in U_{\min}^c} (\hat{\rho}(x) - U(x)) = \max_{x \in U_{\min}^c} h(x),$$

where

$$h(y) := \inf\big\{h \ge 0; \exists\, x \in \Omega \text{ with } U(x) < U(y)$$
$$\text{and } \gamma \in C_{yx} \text{ such that } U(p) \le U(y) + h\ \forall\, p \in \gamma\big\},$$

is the *depth* associated with y, with $h(y) = 0$ on the set of global minima.

Therefore,

$$V^1 - V^2 = \Delta = (R - U_*) \vee \max_{x \in U_{\min}^c} h(x)$$

[cf. Chiang and Chow (1988a)].

In the special case of a symmetric Metropolis chain, one can consider various hitting times. Assume that $M := \max_{x \in \Omega} U(x) - \min_{x \in \Omega} U(x) > 0$ and choose a number U between $\max_{x \in \Omega} U(x)$ and $\min_{x \in \Omega} U(x)$. Moreover, assume that there exists $y \in \Omega$ such that $U(y) = U$. Set $\Omega_* := \{x \in \Omega; U(x) \leq U\}$ and consider the first hitting time $\tau := \inf\{t \geq 0; U(X(t)) \leq U\}$. Let $C[\Omega_*](\beta)$ be the spectral gap associated with $L_\beta[\Omega_*]$, with $C[\Omega_*](\beta) \geq \kappa \exp(-\beta \overline{m})$. Remark first that by Lemma 3.2.12, $\widetilde{L_\beta[\Omega_*]} \equiv \tilde{L}_\beta[\Omega_*] = L_\beta[\Omega_*]$, since by assumption L_β is symmetric and, therefore, $L_\beta[\Omega_*]$ is symmetric. It follows that we can use the setup of this section. Consider the weight function $V[\Omega_*]$. Use Lemma 3.2.13 to get that:

(i) $V[\Omega_*](x \to y) = (U(y) - U(x))^+$ for $x, y \in \Omega^*$ such that $(x \to y) \in \mathscr{G}$ [see (3.1.3)],

(ii) $V[\Omega_*](x \to \delta) = 0$ if there exists $y \in \Omega_*$ with $(x \to y) \in \mathscr{G}$,

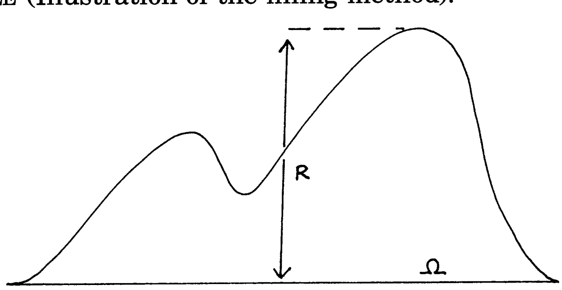
(iii) $V[\Omega_*](\delta \to x) = \min_{p \in \Omega_*}(U(p) + (U(x) - U(p))^+) - \min_{p \in \Omega_*} U(p)$, which is equal to $U(x) - U_*$ if there exists $y \in \Omega_*$ such that $(y \to x) \in \mathscr{G}$.

Thus we can define a new Metropolis chain on $\Omega^* \cup \{\delta\}$, with cost function $\overline{U}(x) := U(x)$, $x \in \Omega^*$, and $\overline{U}(\delta) := U_*$. By Theorem 4.1.13, we have

$$\overline{m} = V[\Omega_*]^1 - V[\Omega_*]^2 = \max_y \overline{h}(y) = \max_{y \in \Omega^*} h(y)$$

(since there is only one global minimum), where $h(y)$ is the height of y [see Hajek (1988) and Holley and Stroock (1989)], and thus $\overline{m} = \max_{y \in \Omega^*} h(y)$; compare Goetze (1992).

5.1.6. EXAMPLE (Illustration of the filling method).



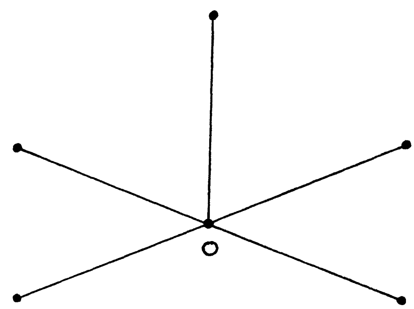
Take, for example, a symmetric Metropolis chain on Ω, with cost function $U(\cdot)$, such that $R = \max_{x \in \Omega} U(x)$ (see Lemma 4.1.11). In this case, $\rho(\cdot) \equiv R$ and, therefore, $C(\beta) \geq \exp(-\beta(R - U_*))C(0)$, where $C(0)$ is the spectral gap associated with the basic chain L_0. Take, for example, $\Omega_d := Z_2^d$, the space of binary vectors of length $d \in \mathbb{N}$, with some cost function $U_d(\cdot)$ and assume that $R_d = \max_{x \in Z_2^d} U_d(x)$. Then the filling method gives $C(\beta) \geq \exp(-\beta(R_d - U_*))d^{-1}$, where $C(0)$ has been estimated by $\kappa_0^{(1)}$ [see Diaconis and Stroock

(1991)], whereas (5.1.1) yields

$$C(\beta) \geq \exp(-\beta m^{(1)}(\Gamma))\kappa_0^{(1)}(\Gamma)^{-1}\frac{|U_{\min}|}{2^d},$$

with $m^{(1)}(\Gamma) \geq (R_d - U_*)$ and $\kappa_0^{(1)} \leq d$, $\forall\ \Gamma$, $\forall\ d \in \mathbb{N}$.

5.1.7. EXAMPLE (The star).



Here $\mathscr{G} = (\Omega, E)$ is a tree with a central node O and $n \in \mathbb{N}$ outside nodes. Let L_0 be the transition function associated with the simple random walk on $\mathscr{G}$. Then L_0 has 0, -1 and -2 as eigenvalues with -1 having multiplicity $n - 1$. To avoid periodicity, put loops at each node, with the same holding rate $\vartheta - 1$, for $0 \leq \vartheta \leq 1$. This yields a transition function L_ϑ given by

$$q_\vartheta(x, O) \equiv 1 - \vartheta, \qquad q_\vartheta(O, x) \equiv n^{-1}(1 - \vartheta), \qquad x \neq O,$$

and L_ϑ has 0, $\vartheta - 1$ and $2(\vartheta - 1)$ as eigenvalues with the above multiplicites. Diaconis and Stroock's (1991) estimate $\kappa^{(1)}$ equals

$$\kappa^{(1)} = \frac{3n-2}{2n}\frac{1}{1-\vartheta}, \qquad C \geq (1-\vartheta)\frac{2n}{3n-2}. \tag{5.1.8}$$

Let $(V(e))_e$ be the weight family given by $V(e) = +\infty$ for $e \notin E$, and $V(e) \equiv V > 0$ for $e \in E$, and consider the transition function $L_\beta \in \mathscr{L}$ given by $q_\beta(e) \equiv q_0(e)\exp(-\beta V(e))$. Also, setting $\vartheta \equiv 1 - \exp(-\beta V)$, we see that $L_\vartheta \equiv L_\beta$ and, therefore, we can use the filling method for the spectral gap of L_ϑ. It is easy to see that for this star $V_x \equiv nV$, $\mathscr{O}(e) \equiv (n + 1)V$, $\rho(x, y) \equiv (n + 1)V$, $x \neq y$, $\rho(x) \equiv \hat{\rho}(x) \equiv R$ and $R = V(n + 1)$ (see Definition 4.1.9). The filling method yields

$$C(\beta) \geq \exp(-\beta(R - nV))C'(\beta) \equiv \exp(-\beta V)C'(\beta),$$

where $C'(\beta)$ is the spectral gap of L_0 since $q'_\beta(e) \equiv q_0(e)\exp(-\beta(\rho(e^+) - \rho(e^-))^+) \equiv q_0(e)$. Therefore,

$$C(\beta) \geq \exp(-\beta V)C(0) \equiv \vartheta - 1, \tag{5.1.9}$$

which equals the spectral gap. Note that the filling method reduces the estimation procedure to the estimation of $C(0)$.

6. Diffusions on compact manifolds.

6.1. *Spectral estimates for diffusions with small noise.* Let $\mathbf{M}$ be a compact, connected, N-dimensional C^∞-Riemannian manifold. Denote by $T(\mathbf{M})$ the tangent bundle over $\mathbf{M}$ and by $\Gamma(T(\mathbf{M}))$ the space of smooth sections. For $X, Y \in T(\mathbf{M})$ we will denote by $(X|Y)$ the inner product of X and Y and by $|X| = (X|X)^{1/2}$, the length of X. For given B_β, $B \in \Gamma(T(\mathbf{M}))$, $\beta > 0$, such that uniformly on $\mathbf{M}$,

$$\lim_{\beta\to\infty} B_\beta = B, \tag{6.1.1}$$

consider the elliptic operator L_β: $C^\infty(\mathbf{M}) \to \mathbf{C}^\infty(\mathbf{M})$,

$$L_\beta \varphi = \beta\big(B_\beta|\nabla\varphi\big) + \tfrac{1}{2}\,\Delta\varphi = \beta\big(B_\beta|\nabla\varphi\big) + \tfrac{1}{2}\nabla\cdot\nabla\varphi,$$

where we use ∇, $\nabla\cdot$ and Δ to denote, respectively, the gradient, divergence and Laplace–Beltrami operator on $\mathbf{M}$. Since the manifold is compact, there exists a unique normalized L_β-invariant distribution π_β such that

$$\langle 1, L_\beta\varphi\rangle_{\pi_\beta} = 0 \quad \text{for all } \varphi \in C^\infty(\mathbf{M}).$$

Moreover, π_β has a smooth density $\rho_\beta > 0$ with respect to the Riemannian measure π_0 which solves the equation

$$-\beta\nabla\cdot\big(\rho_\beta B_\beta\big) + \tfrac{1}{2}\Delta\rho_\beta = 0. \tag{6.1.2}$$

Next let L_β^* and $\tilde{L}_\beta = (L_\beta + L_\beta^*)/2$ be, respectively, the π_β-adjoint and symmetrization of L_β:

$$L_\beta^*\varphi = \beta\big(B_\rho^*|\nabla\varphi\big) + \frac{1}{2}\Delta\varphi \quad \text{with} \quad B_\beta^* = -B_\beta + \frac{1}{\beta}\nabla\log\rho_\beta \tag{6.1.3}$$

and

$$\tilde{L}_\beta\varphi = \beta\big(\tilde{B}_\beta|\nabla\varphi\big) + \frac{1}{2}\,\Delta\varphi = \frac{1}{2\rho_\beta}\,\nabla\cdot\big(\rho_\beta\nabla\varphi\big) \quad \text{with} \quad \tilde{B}_\beta = \frac{1}{2\beta}\nabla\log\rho_\beta.$$

Both L_β and $\tilde{L}_\beta$ have discrete $L^2(\pi_\beta)$-spectra $\{0 = \lambda_1(\beta), \lambda_2(\beta), \dots\}$, and $\{0 = \tilde{\lambda}_1(\beta), \tilde{\lambda}_2(\beta), \dots\}$ with $\Re(\lambda_i(\beta)) < 0$, $i \neq 1$, and $\lambda_i(\beta) < 0$, $i \neq 1$. Let $C(\beta) = -\sup\{\Re(\lambda_i(\beta))\colon i \neq 1\}$ and $\tilde{C}(\beta) = -\sup\{\lambda_i(\beta)\colon i \neq 1\}$ be the spectral gaps associated with L_β and $\tilde{L}_\beta$.

The aim of this section is to show that $C(\beta)$ and $\tilde{C}(\beta)$ are logarithmically equivalent. We will express the asymptotics of $C(\beta)$ and $\tilde{C}(\beta)$ in terms of Markov chains, following the ideas of Section 6 of Freidlin and Wentzell

(1984). Their monograph deals with operators of the form

$$\hat{L}_\varepsilon \varphi = (B_\varepsilon | \nabla\varphi) + \frac{\varepsilon^2}{2}\Delta,$$

where $\varepsilon \searrow 0$. The translation of their results to our situation follows from $\hat{L}_\varepsilon = (1/\beta)L_\beta$, where $\beta = 1/\varepsilon^2$. In particular, the invariant distributions are the same and for the spectrum we have $\hat{\lambda}_i(\varepsilon) = (1/\beta)\lambda_i(\beta)$.

The main tool of Freidlin and Wentzell is the theory of large deviations based on the action functional S_T: $C^1([0,T]; \mathbf{M}) \to \mathbb{R}^+$,

$$S_T(\varphi) = \tfrac{1}{2}\int_0^T |\dot{\varphi} - B(\varphi_t)|^2\, dt,$$

and the weight function V: $\mathbf{M} \times \mathbf{M} \to \mathbb{R}^+$,

$$V(x,y) = \inf\{S_T(\varphi): \varphi_0 = x,\ \varphi_T = y,\ T \ge 0\}.$$

We will be working under Assumption A of Freidlin and Wentzell (1984):

ASSUMPTION A. There exist a finite number of conpacta $K_1, K_2, \dots, K_n \subseteq \mathbf{M}$ such that:

1. For any two points $x, y \in K_i$, $V(x,y) = V(y,x) = 0$.
2. If $x \in K_i$ and $y \notin K_i$, then $V(x,y) \neq 0$.
3. Every ω-limit point of the dynamical system $\dot{x}_t = B(x_t)$ is contained in one of the K_i.

Next introduce the weight function V on $\Omega \times \Omega$ with $\Omega = \{1, \dots, n\}$:

$$V(i \to j) = \inf\left\{S_T(\varphi): \varphi_0 \in K_i,\ \varphi_T \in K_j,\ \varphi_t \in \mathbf{M} \Big\backslash \bigcup_{k \neq i,j} K_k \text{ for } 0 < t < T\right\}.$$

If no such function exists, write $V(i \to j) = \infty$. Consider the irreducible Markov chain on E associated with the weights V and define the corresponding V: $\Omega \to \mathbb{R}^+$, V^1 and V^2 as in Section 3. Finally, define W: $\mathbf{M} \to \mathbb{R}^+$ by

$$W(x) = \min\{V_i + V(K_i, x): i = 1, \dots, n\},$$

where $V(K_i, x) = \min\{V(y,x): y \in K_i\}$. The following theorem, proved in Freidlin and Wentzell [(1984), Theorems 6.4.3 and 6.7.4], show that the asymptotic behavior of the diffusion is determined by the Markov chain.

6.1.4. THEOREM. *Let Assumption A hold. Let $\gamma > 0$. Then for any sufficiently small neighborhood $G(x)$ of $x \in \mathbf{M}$, there exists $\beta_0 > 0$ such that for $\beta \ge \beta_0$ we have*

$$\exp\{-\beta(W(x) - V^1 + \gamma)\} \le \pi_\beta(G(x)) \le \exp\{-\beta(W(x) - V^1 - \gamma)\}. \tag{6.1.5}$$

Moreover,

$$\lim_{\beta \to \infty} \left[-\frac{1}{\beta} \log C(\beta) \right] = V^1 - V^2.$$

The main result of this section is the following theorem.

6.1.6. THEOREM. *Let Assumption A hold and suppose that there is* $B^* \in \Gamma(T(\mathbf{M}))$ *such that uniformly on* $\mathbf{M}$,

$$\lim_{\beta \to \infty} B_\beta^* = B^* \quad \textit{and} \quad \lim_{\beta \to \infty} \frac{1}{\beta} \nabla \cdot \left(B_\beta^* - B_\beta \right) = 0. \tag{6.1.7}$$

Then

$$\lim_{\beta \to \infty} \left[-\frac{1}{\beta} \log \tilde{C}(\beta) \right] = \lim_{\beta \to \infty} \left[-\frac{1}{\beta} \log C(\beta) \right] = V^1 - V^2. \tag{6.1.8}$$

We believe that the hypothesis (6.1.7) should be true in the general setting of smooth compact manifolds. In particular, by (6.1.3) and (6.1.5), B^* should have the form $B^* = -B - \nabla W$. Also one expects that

$$\nabla \cdot \left(B_\beta^* - B_\beta \right) = -2 \nabla B_\beta + \frac{1}{\beta} \Delta \log \rho_\beta$$

remains bounded as $\beta \to \infty$, which of course yields the second limit in (6.1.7). Note that $C(\beta) \ge \tilde{C}(\beta)$ and, in general, $C(\beta) > \tilde{C}(\beta)$; see Hwang, Hwang-Ma and Sheu (1992). In view of Theorem 3.2.11, a similar statement holds for the real part of the first n eigenvalues of L_β and $\tilde{L}_\beta$. It would be very interesting to study the asymptotics of the next eigenvalues.

In the symmetric case where $\pi_\beta(dx) = (1/Z(\beta))e^{-\beta U(x)}\pi_0(dx)$ and $B_\beta = B_\beta^* = -\frac{1}{2}\nabla U$ for a given $U \in C^\infty(\mathbf{M})$, Holley, Kusuoka and Stroock (1989) have derived a Poincaré estimate for $C(\beta)$. They show that $m = -\lim_{\beta \to \infty} (1/\beta)\log C(\beta)$ can be expressed in terms of the potential U and a suitable collection of paths on the manifold. Theorems 4.1.13 and 6.1.6 extend this result to a much broader class of diffusions. However, the derivation of more precise estimates of the type

$$\tilde{C}(\beta) \ge K_{\mathbf{M}}(\beta \vee 1)^{-5N+2} e^{-\beta m},$$

[cf. Holley, Kusuoka and Stroock (1989), Theorem 1.14] is well beyond the scope of this section.

The main step in the proof of (6.1.8) is the following lemma.

6.1.9. LEMMA. *Let Assumption* A *and* (6.1.7) *hold. Let* $S_T^*(\varphi) = \frac{1}{2}\int_0^T |\dot{\varphi}_t - B^*(\varphi_t)|^2 \, dt$ *be the action functional associated with* L_β^*. *Consider* $\varphi \in C([0,T]; \mathbf{M})$ *such that* $\varphi_0 = x$ *and* $\varphi_T = y$, *and set* $\varphi_t^* = \varphi_{T-t}$, $0 \le t \le T$. *Then*

$$S_T^*(\varphi^*) = S_T(\varphi) + W(x) - W(y).$$

PROOF. By (6.1.1), (6.1.7) and the trivial fact that $\dot{\varphi}_t^* = -\dot{\varphi}_{T-t}$ we have

$$\begin{aligned}
S_T^*(\varphi^*) &= o(1) + \tfrac{1}{2}\int_0^T |\dot{\varphi}_t + B_\beta^*(\varphi_t)|^2\, dt \\
&= o(1) + \tfrac{1}{2}\int_0^T |\dot{\varphi}_t - B_\beta(\varphi_t) + B_\beta^*(\varphi_t) + B_\beta(\varphi_t)|^2\, dt \\
&= o(1) + S_T(\varphi) + \int_0^T \big(\dot{\varphi}_t | B_\beta^*(\varphi_t) + B_\beta(\varphi_t)\big)\, dt \\
&\quad + \tfrac{1}{2}\int_0^T \big(B^*(\varphi_t) - B_\beta(\varphi_t) | B_\beta^*(\varphi_t) + B_\beta(\varphi_t)\big)\, dt.
\end{aligned}$$

In view of (6.1.3) and (6.1.5) we have

$$\begin{aligned}
\int_0^T \big(\dot{\varphi}_t | B_\beta^*(\varphi_t) + B_\beta(\varphi_t)\big)\, dt &= \frac{1}{\beta}\int_0^T \big(\dot{\varphi}_t | \nabla \log \rho_\beta(\varphi_t)\big)\, dt \\
&= \frac{1}{\beta}\big(\log(\rho_\beta(y)) - \log(\rho_\beta(x))\big) \\
&= W(x) - W(y) + o(1).
\end{aligned}$$

Next note that (6.1.2) implies

$$-\nabla \cdot \left(B_\beta - \frac{1}{2\beta}\nabla \log \rho_\beta \right) = \left(B_\beta - \frac{1}{2\beta}\nabla \log \rho_\beta \middle| \nabla \log \rho_\beta \right),$$

but this together with (6.1.3) and (6.1.6) shows

$$\big(B_\beta^* - B_\beta | B_\beta^* + B_\beta\big) = -\frac{1}{\beta}\nabla \cdot \big(B_\beta^* - B_\beta\big) = o(1)$$

and we get the result from the above. □

PROOF OF THEOREM 6.1.6. From Lemma 6.1.9 we see that if $V^*(i \to j)$ are the weights associated with S^*, then

$$V^*(i \to j) = V(j \to i) + V_j - V_i.$$

In other words, the Markov chain associated with the asymptotics L_β^* is the adjoint of the chain associated with L_β. Now (6.1.8) follows from Theorems 6.1.4 and 3.2.11. □

6.1.10. EXAMPLE (Perturbation of the symmetric case). Let $U \in C^\infty(\mathbf{M})$ be given and let $C \in \Gamma(T(\mathbf{M}))$ satisfy

$$(6.1.11) \qquad \nabla \cdot C \equiv 0 \quad \text{and} \quad (C|\nabla U) \equiv 0.$$

For instance, let A be a co-closed second order differential form with vanishing exterior derivative $dA \equiv 0$. Then C defined by

$$(C|\nabla f) = A(\nabla U, \nabla f), \qquad f \in C^\infty(\mathbf{M}),$$

satisfies (6.1.11). In particular, if $\mathbf{M} = T_N$, the N-dimensional torus equipped with the usual Euclidian metric, we can choose $C = A\nabla U$, where A is a skew-symmetric matrix with constant entries.

Next choose

$$B_\beta \equiv B = -\tfrac{1}{2}\nabla U + C.$$

Then $\pi_\beta(dx) = e^{-\beta U(x)} Z(\beta)^{-1}\pi_0(dx)$ and

$$B_\beta^* \equiv B^* = -\tfrac{1}{2}\nabla U - C. \qquad \square$$

APPENDIX

A.1. Proof of the results of Section 2.2.

A.1.1. LEMMA. *Assume that $t \to L_{\beta(t)}$ is differentiable. Then $t \to \pi_{\beta(t)}$ is differentiable with*

$$\dot{f}_t(x) = \frac{d}{dt} f_t(x) = L^*_{\beta(t)} f_t(x) - \frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}(x) f_t(x), \qquad x \in \Omega.$$

PROOF. Let us first check that $t \to \pi_{\beta(t)}$ is differentiable. Let $\dot{\pi}_{\beta(t)}$ be the solution of

$$\sum_x \dot{\pi}_{\beta(t)}(x) q_{\beta(t)}(x, y) = -\sum_x \pi_{\beta(t)}(x) \dot{q}_{\beta(t)}(x, y), \qquad y \in \Omega,$$

$$\sum_x \dot{\pi}_{\beta(t)}(x) = 0.$$

This system has a unique solution since $\sum_y q_{\beta(t)}(x, y) = 0$ implies $\sum_y \dot{q}_{\beta(t)}(x, y) = 0$ and $L_{\beta(t)}$ has rank $(|\Omega| - 1)$. From this one gets

$$\pi_{\beta(t+\varepsilon)}(x) - \pi_{\beta(t)}(x) = \varepsilon \dot{\pi}_{\beta(t)}(x) + o(\varepsilon)$$

and, therefore, $t \to \pi_{\beta(t)}$ is differentiable. Next let us verify that $f_t = P_t^*(\nu_0/\pi_{\beta(t)})$. Take a bounded measurable h. Then

$$\int_\Omega h \cdot f_t \, d\pi_{\beta(t)} = \int_\Omega h \, d\nu_t = \int_\Omega P_t h \, d\nu_0 = \int_\Omega P_t h \cdot \frac{\nu_0}{\pi_{\beta(t)}} \, d\pi_{\beta(t)}$$

$$= \int_\Omega h \cdot P_t^*\left(\frac{\nu_0}{\pi_{\beta(t)}}\right) d\pi_{\beta(t)}.$$

Next note that

$$\frac{d}{dt}\int_\Omega h \, d\nu_t = \int_\Omega P_t[L_{\beta(t)} h] \, d\nu_0$$

$$= \int_\Omega P_t[L_{\beta(t)} h] \cdot \frac{\nu_0}{\pi_{\beta(t)}} \, d\pi_{\beta(t)} = \int_\Omega h \cdot L^*_{\beta(t)}\left[P_t^*\left(\frac{\nu_0}{\pi_{\beta(t)}}\right)\right] d\pi_{\beta(t)}$$

$$= \int_\Omega h \cdot L^*_{\beta(t)} f_t \, d\pi_{\beta(t)}.$$

On the other hand, we have

$$\frac{d}{dt}\int_\Omega h\,d\nu_t = \frac{d}{dt}\int_\Omega h\cdot f_t\,d\pi_{\beta(t)} = \int_\Omega h\cdot \dot{f}_t d\pi_{\beta(t)} + \int_\Omega h\cdot f_t \frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}$$

and, therefore, $\dot{f}_t = L^*_{\beta(t)} f_t - (\dot{\pi}_{\beta(t)}/\pi_{\beta(t)}) f_t$. □

A.1.2. LEMMA.

$$\frac{d}{dt}\int_\Omega (f_t-1)^2\,d\pi_{\beta(t)} = -2\mathscr{E}_{\pi_{\beta(t)}}(f_t, f_t) - \int_\Omega (f_t-1)^2 \frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)} - 2\int_\Omega (f_t-1)\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}.$$

PROOF. Using Lemma A.1.1 we see that

$$\begin{aligned}\frac{d}{dt}\int_\Omega (f_t-1)^2\,d\pi_{\beta(t)} &= \int_\Omega 2(f_t-1)\dot{f}_t\,d\pi_{\beta(t)} + \int_\Omega (f_t-1)^2\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}\\ &= 2\int_\Omega (f_t-1)L^*_{\beta(t)}f_t\,d\pi_{\beta(t)} - 2\int_\Omega (f_t-1)f_t\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}\\ &\quad + \int_\Omega (f_t-1)^2\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}\\ &= -2\mathscr{E}_{\pi_{\beta(t)}}(f_t f_t) - \int_\Omega (f_t-1)^2\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}\\ &\quad - 2\int_\Omega (f_t-1)\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\,d\pi_{\beta(t)}.\end{aligned}$$

□

PROOF OF LEMMA 2.2.3. Note that (2.2.4) implies

$$K\exp(-m\beta(t)) - \frac{M}{2}\dot{\beta}(t) = M\dot{\beta}(t). \tag{A.1.3}$$

Set $u_t := \|f_t - 1\|^2_{\pi_{\beta(t)}}$. Then, by Lemma A.1.2 and (2.2.1), (2.2.2), (A.1.3) and the Cauchy–Schwarz inequality, we get

$$\begin{aligned}\dot{u}_t &\le -2\left(K\exp(-m\beta(t)) - \frac{M}{2}\dot{\beta}(t)\right)u_t - 2\int_\Omega (f_t-1)\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}d\pi_{\beta(t)}\\ &\le -2M\dot{\beta}(t)u_t + 2M\dot{\beta}(t)u_t^{1/2}.\end{aligned}$$

Thus if $w_t := u_t^{1/2}$, we have

$$\dot{w}_t \le -M\dot{\beta}(t)w_t + M\dot{\beta}(t),$$

which implies (2.2.5). Next assume (2.2.6) and let $\delta \in \Omega_*$. Since

$$0 = \int_\Omega (f_t - 1)\, d\pi_{\beta(t)} = \int_{\Omega^*} (f_t - 1)\, d\pi_{\beta(t)} + \int_{\Omega_*} (f_t - 1)\, d\pi_{\beta(t)},$$

we have

$$\begin{aligned} -\int_\Omega (f_t - 1)\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}\, d\pi_{\beta(t)} &= -\int_{\Omega_*} (f_t - 1)\left(\frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}} - \frac{\dot{\pi}_{\beta(t)}}{\pi_{\beta(t)}}(\delta)\right) d\pi_{\beta(t)} \\ &\le 2M\dot{\beta}(t)\pi_{\beta(t)}(\Omega^*)^{1/2} u_t^{1/2}, \end{aligned}$$

where we have used (2.2.2) and the Cauchy–Schwarz inequality. Remembering that $w_t = u_t^{1/2}$, we get from the above that

$$\begin{aligned} \dot{w}_t &\le -\left(Ke^{-m\beta(t)} - \frac{M}{2}\dot{\beta}(t)\right) w_t + 2M\dot{\beta}(t)\pi_{\beta(t)}(\Omega^*)^{1/2} \\ &= -M\dot{\beta}(t) w_t + 2M\dot{\beta}(t)\pi_{\beta(t)}(\Omega^*)^{1/2}. \end{aligned}$$

Note that by (2.2.2), for $0 \le s \le t$,

$$\begin{aligned} \left|\log(\pi_{\beta(s)}(\Omega^*)) - \log(\pi_{\beta(t)}(\Omega^*))\right| &= \left|\int_s^t \left(\frac{\int_{\Omega^*}(\dot{\pi}_{\beta(u)}/\pi_{\beta(u)})\, d\pi_{\beta(u)}}{\int_{\Omega^*} d\pi_{\beta(u)}}\right) du\right| \\ &\le M(\beta(t) - \beta(s)). \end{aligned}$$

This yields

$$\begin{aligned} w_t &\le e^{-M\beta(t)} w_0 + 2M\pi_{\beta(t)}(\Omega^*)^{1/2} \int_0^t \exp\big[-M(\beta(t) - \beta(s)) \\ &\qquad + \tfrac{1}{2}\log(\pi_{\beta(s)}(\Omega^*)) - \tfrac{1}{2}\log(\pi_{\beta(t)}(\Omega^*))\big]\, \dot{\beta}(s)\, ds \\ &\le e^{-M\beta(t)} w_0 + 2M\pi_{\beta(t)}(\Omega^*)^{1/2} \int_0^t \exp\left[-\frac{M}{2}(\beta(t) - \beta(s))\right] \dot{\beta}(s)\, ds \\ &\le e^{-M\beta(t)} w_0 + 4\pi_{\beta(t)}(\Omega^*)^{1/2}. \end{aligned}$$

□

Acknowledgments. We are grateful to Alain-Sol Sznitman and a referee for their valuable comments on a previous version of this paper.

REFERENCES

AARTS, E. and KORST, J. (1989). Simulated annealing and Boltzmann machines. In *A Stochastic Approach to Combinatorial Optimization and Neural Computing*. Wiley, New York.

AZENCOTT, R. (1988). *Simulated Annealing. Séminaire Bourbaki. Astérisque* **161–162** 223–237.

AZENCOTT, R. (1992). *Simulated Annealing: Parallelization Techniques*. Wiley, New York.

BOTT, R. and MAYBERRY, J. P. (1954). Matrices and trees. *Economics Activity Analysis*. Wiley, New York.

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

CHIANG, T. S. and CHOW, Y. (1988b). On the convergence rate of annealing processes. *SIAM J. Control Optim.* **26** 1455–1470.

DEUSCHEL, J. D. and MAZZA, C. (1994). L^2 convergence of time non-homogeneous Markov processes: II. Parallel and sequential annealing process on Gibbsian fields. Preprint.

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

FILL, J. (1991). Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with applications to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.

FREIDLIN, M. I. and WENTZELL, A. D., (1984). *Random Perturbations of Dynamical Systems.* Springer, New York.

GEMAN, D. (1988). Random fields and inverse problems in imaging. *Ecole d'Été de Probabilités de Saint-Flour XVIII. Lecture Notes in Math.* **1427**. Springer, Berlin.

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

GOETZE, F. (1992). Rate of convergence of simulated annealing processes. Unpublished manuscript.

HAJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329.

HOLLEY, R. and STROOCK, D. (1989). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 553–569.

HOLLEY, R., KUSUOKA, S. and STROOCK, D. (1989). Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83** 333–347.

HWANG, C. R., HWANG-MA, S. Y. and SHEU, S. J. (1992). Accelerating Gaussian diffusions. Preprint.

LAWLER, E. (1976). *Combinatorial Optimization: Network and Matroids*, Holt, Rinehart and Winston, New York.

MAZZA, C. (1992). Parallel simulated annealing. *Rand. Struct. Algorithms* **3** 139–148.

RAMMAL, R., TOULOUSE, G. and VIRASORO, M. A. (1986). Ultrametricity for physicists *Rev. Mod. Phys.* **58** 765–788.

ROUX, M. (1985). *Algorithmes de classification.* Masson, Paris.

SINCLAIR, A. (1991). Improved bounds for mixing rates of Markov chains and multicommodity flow. LFCS Report Series, Dept. Computer Science, Univ. Edinburgh.

TROUVÉ, A. (1992). Massive parallelization of simulated annealing: A mathematical study. In *Simulated Annealing: Parallelization Techniques* (R. Azencott, ed.). Wiley, New York.

TROUVÉ, A. (1993). Parallélisation massive du recuit simulé. Thèse, Univ. Paris II.

WENTZELL, A. (1972). On the asymptotics of eigenvalues of matrices with elements of order $\exp(-V_{ij}/(2\varepsilon^2))$. *Sov. Math. Dokl.* **13** 65–68.

FACHBEREICH MATHEMATIK
TECHNISCHE UNIVERSITÄT BERLIN
D-10623 BERLIN
GERMANY

INSTITUT DE MATHÉMATIQUES
UNIVERSITÉ DE FRIBOURG
CH-1700 FRIBOURG
SWITZERLAND