Journal of Statistical Physics, Vol. 51, Nos. 5/6, 1988

Absence of Mass Gap for a Class of Stochastic Contour Models

Alan D. Sokal[1] **and Lawrence E. Thomas**[2,3]

Received March 21, 1988

We study a class of Markovian stochastic processes in which the state space is a space of lattice contours and the elementary motions are local deformations. We show, under suitable hypotheses on the jump rates, that the infinitesimal generator has zero mass gap. This result covers (among others) the BFACF dynamics for fixed-endpoint self-avoiding walks and the Sterling–Greensite dynamics for fixed-boundary self-avoiding surfaces. Our models also mimic the Glauber dynamics for the low-temperature Ising model. The proofs are based on two new general principles: the minimum hitting-time argument and the mean (or mean-exponential) hitting-time argument.

KEY WORDS: Markov chain; Markov process; contour model; mass gap; dynamic critical phenomena; Monte Carlo; Glauber dynamics; self-avoiding walk.

1. INTRODUCTION

In this paper we study a class of Markovian stochastic processes (in either discrete or continuous time) in which the state space is a space of *lattice contours* γ (of variable length) and the elementary motions are *local deformations*. These stochastic processes are reversible (i.e., satisfy detailed balance) with respect to the probability measure $\pi_\beta(\gamma) = \text{const} \times e^{-\beta|\gamma|}$, where $|\gamma|$ is the length of γ. Our main concern is the rate at which the process approaches equilibrium, and in particular whether this approach occurs exponentially fast.

[1] Department of Physics, New York University, New York, New York 10003.

[2] Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802.

[3] Permanent address: Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903.

0022-4715/88/0600-0907$06.00/0

Reversibility implies that the time-evolution semigroups are self-adjoint in the Hilbert space $L^2(\pi_\beta)$, and our approach in this paper is to study the spectral properties of these semigroups. In the discrete-time case, the transition matrix is a contraction with an eigenvalue at 1; exponential convergence to equilibrium (in the L^2 sense) means that the rest of the spectrum is contained in some interval $[-\lambda, \lambda]$ with $\lambda < 1$. In the continuous-time case, the infinitesimal generator is positive-semidefinite with an eigenvalue at 0; exponential convergence to equilibrium (in the L^2 sense) means that the rest of the spectrum is bounded away from 0. In such cases we say that the transition matrix or infinitesimal generator has a nonzero *mass gap.*

In a previous paper[(1)] (see also ref. 2), the authors considered a superficially similar model[(3)] in which the state space consists of self-avoiding walks (of variable length) with *free* endpoints, and the elementary moves are creation and destruction of bonds at one of the endpoints. We showed[(1,2)] that this stochastic process exhibits exponential convergence to equilibrium, and we obtained two-sided bounds on the mass gap. By contrast, in this paper we show, under suitable hypotheses on the jump rates, that the contour models with local-deformation dynamics do *not* have exponential convergence to equilibrium—that is, they have *zero mass gap.*

At first we found this result to be extremely surprising. Indeed, in models of self-avoiding walks of *fixed* length N with local-deformation dynamics, the mass gap is believed to be of order $N^{-(2+2\nu)}$, where ν is the critical exponent for the mean size of self-avoiding walks[4]; this belief is based on a simple heuristic argument involving the diffusion of the center-of-mass of the SAW.[(7)] Since the models considered in this paper are the natural generalizations of local-deformation dynamics to the grand canonical (variable-N) ensemble, it is natural to conjecture that the mass gap is of order $\langle N \rangle^{-(2+2\nu)}$, where $\langle N \rangle$ is the average length of walks in the equilibrium ensemble. This conjecture is, however, based on the implicit assumption that transitions between walks of different lengths N are sufficiently fast (i.e., faster than equilibration of walks at a fixed length N), and it is this assumption that breaks down in the models considered here if the jump rates grow sufficiently slowly with N. Roughly speaking, in these processes there are rectangular configurations γ in which no immediate motions shorten the length of γ, and these configurations behave

[4] This statement requires some qualification. First, local-deformation dynamics for fixed-length SAWs is always nonergodic,[(4)] so the conjectured mass gap refers to the spctrum of the transition matrix restricted to some particular ergodic class (e.g., the ergodic class of a straight rod). Second, if the dynamics happens to obey spcial conservation laws, then the mass gap may be considerably smaller than $N^{-(2+2\nu)}$ (see refs. 5 and 6). For more discussion, see ref. 7.

in a metastable manner: it takes too many jumps (or too long a time) to reach a fixed shorter configuration. Here, "too many" is defined relative to a property of the invariant measure π_β. In Section 3 we abstract this idea into two general principles for proving upper bounds on the mass gap: the *minimum hitting-time argument* (Theorem 3.1) and the *mean (or mean-exponential) hitting-time argument* (Propositions 3.3–3.10). In Section 4 we then apply these ideas to the contour models.

Our models have two very different applications, depending on the interpretation of the contours:

(a) *Peierls contours.* If the contours are interpreted as Peierls contours separating regions of + and − spins in a d-dimensional Ising model, then our dynamics mimics the single-spin-flip Glauber dynamics[8,9] for the Ising model, in the low-temperature approximation in which contours are dilute and hence noninteracting. Here our parameter β is proportional to the inverse temperature of the corresponding Ising model, and our total jump rates $r(\gamma)$ should be taken to be proportional to the length of the contour. In an important recent paper, Huse and Fisher[10] argued heuristically that this contour model should have zero mass gap if (and only if) $1 < d < 3$. Unfortunately, our rigorous bounds are one power of n weaker than their heuristic estimates,[5] and so our theorems apply (for these jump rates) only for $1 < d < 2$. (In other words, for $d = 2$ we are just barely unable to treat jump rates growing as fast as the length of the contour.) See Section 4.3 for discussion.

(b) *Self-avoiding walks, surfaces, etc.* If the contours are interpreted as self-avoiding walks with fixed endpoints, then our dynamics includes, as a special case, the Berg–Foerster–Aragão de Carvalho–Caracciolo–Fröhlich (BFACF)[11-14,7] Monte Carlo algorithm for simulating a grand canonical ensemble of such walks. Here $e^{-\beta}$ plays the role of a bond activity (fugacity), and the jump rates should be taken to be bounded. Our results imply the vanishing of the mass gap for this Markov chain. Higher-dimensional versions can be interpreted as self-avoiding surfaces with fixed boundary, in which case the dynamics includes the Sterling–Greensite algorithm.[15-20]

2. MARKOV CHAINS AND MARKOVIAN JUMP PROCESSES

In this section we review the basic theory of Markov chains and Markovian jump processes from both an analytic and a probabilistic point

[5] Here n is the linear extension of an approximately spherical or cubical contour γ_n; thus, $|\gamma_n| \sim n^{d-1}$, and the volume enclosed by γ_n is of order n^d.

of view. (More information on Markov chains and processes can be found in refs. 21–22 and 22–27, respectively.)

Our primary point of view is analytic: starting from a transition probability kernel P or an infinitesimal generator kernel G, our aim is to study the spectrum of P or G acting as an operator on the space $L^2(\pi)$, where π is the invariant measure. In fact, our main theorems and their proofs are purely analytic; from a logical point of view, probability theory plays no role. However, the *motivation* for studying these particular operators arises from probability theory: we want to understand the rate of convergence to equilibrium for the stochastic process generated by P or G. Moreover, the *physical interpretation* of our theorems and proofs involves hitting times in this stochastic process. It is thus incumbent on us to construct this process, and to make the connection between the probabilistic and analytic approaches. In discrete time, this is relatively trivial. In continuous time, however, it is far from trivial, and for a good *physical* reason: if the jump rates are unbounded, there arises the possibility that the system could make infinitely many jumps in a finite time ("explosion"). Much of Section 2.2 is devoted, therefore, to giving sufficient conditions for the absence of explosion. (In Section 4.1 we verify these hypotheses for our stochastic contour models under extremely weak conditions on the transition rates.) These issues are nevertheless somewhat tangential to our main theme of L^2 spectral bounds, and so the reader who is willing to take on faith the existence of a stochastic process with the desired properties may skim lightly over Section 2.2.

We state our results for Markov chains and processes on a general (measurable) state space $(X, \mathscr{X})$ whenever it causes no complications to do so. However, in all the applications in this paper, X is countable (and of course $\mathscr{X}$ is the σ-field of all subsets of X). Therefore, the reader is welcome to imagine that all integrals are in fact sums, all kernels are matrices, and so on.

2.1. Discrete Time

Let us start by considering a positive-recurrent, discrete-time Markov chain with measurable state space $(X, \mathscr{X})$, transition probability kernel $P(x, dy)$, and invariant probability measure π. This means that if the system is in state x at time t, then its state at time $t+1$ will be chosen randomly according to the probability distribution $P(x, \cdot)$. The invariance of π means that

$$\int \pi(dx)\, P(x, dy) = \pi(dy) \tag{2.1}$$

In this situation it is not hard to show (e.g., using Hölder's inequality) that P induces a positivity-preserving linear contraction on $L^2(\pi)$ [and in fact on all the spaces $L^p(\pi)$] by

$$(Pf)(x) = \int P(x, dy)\, f(y) \tag{2.2}$$

The constant function $\mathbf{1}$ is an eigenvector of P (and of its adjoint P^*) with eigenvalue 1. The spectrum of $P \restriction \mathbf{1}^\perp$ thus lies in the unit disk, and the goal of this paper (along with refs. 1, 2, and 28) is to prove bounds on its location. In particular, we show that in certain circumstances there must be spectrum very near 1.

The Markov chain is said to be *reversible* (or to satisfy *detailed balance*) with respect to π in case

$$\pi(dx)\, P(x, dy) = \pi(dy)\, P(y, dx) \tag{2.3}$$

Equivalently, the chain is reversible if the operator P on $L^2(\pi)$ is self-adjoint. In this case the spectrum of $P \restriction \mathbf{1}^\perp$ lies in the interval $[-1, 1]$, and the *mass gap* (or L^2 spectral gap) is, by definition, the distance of this spectrum from the point 1. For convenience we introduce the operator $\tilde{P} \equiv I - P$ and discuss the spectrum of $\tilde{P} \restriction \mathbf{1}^\perp$ near 0. The mass gap is therefore

$$m \equiv \inf \operatorname{spec}(\tilde{P} \restriction \mathbf{1}^\perp) \tag{2.4}$$

Finally, it is not hard to show (ref. 29, Sections V-1 and V-2) that given any initial probability distribution α, there exists a (essentially unique) stochastic process $\{X_0, X_1, X_2,...\}$ satisfying

$$P(X_0 \in A) = \alpha(A) \tag{2.5a}$$

$$P(X_{n+1} \in A \mid X_0,..., X_n) = P(X_n, A) \tag{2.5b}$$

for all $A \in \mathscr{X}$ and all $n \geqslant 0$.

For future use, let Π be the expectation operator

$$(\Pi f)(x) = \int \pi(dy)\, f(y) \qquad \text{for all } x \tag{2.6}$$

Equivalently, Π is the orthogonal projection in $L^2(\pi)$ onto the constant functions.

2.2. Continuous Time

Next let us consider a positive-recurrent, continuous-time Markovian jump process with measurable state space $(X, \mathscr{X})$, transition rate kernel

$J(x, dy)$, and invariant probability measure π. The total transition rate out of state x,

$$r(x) \equiv J(x, X) \tag{2.7}$$

is assumed to be finite for all x, but it is not necessarily bounded. Thus, if the system is in state x at time t, it will wait a random time that is exponentially distributed with mean $1/r(x)$ and then jump to another state chosen randomly according to the probability distribution $J(x, \cdot)/r(x)$. The invariance of π means that

$$\int \pi(dx)\, J(x, dy) = r(y)\, \pi(dy) \tag{2.8}$$

i.e., the mean transition rate into the state y equals the mean transition rate out of it.

The infinitesimal generator G of this jump process is the signed kernel[6]

$$G(x, \cdot) \equiv r(x)\delta_x - J(x, \cdot) \tag{2.9}$$

If the transition rates are *bounded* (say, by some number M), then there is a unique semigroup of transition probability kernels $\{P_t\}_{t \geqslant 0}$ generated by G, namely

$$P_t \equiv \exp(-tG) \equiv \sum_{n=0}^{\infty} \frac{(-G)^n}{n!} \tag{2.10}$$

a series which is absolutely convergent in the Banach space of bounded signed kernels.[7] Moreover, G induces a bounded linear operator (of norm $\leqslant 2M$) on $L^2(\pi)$ [and in fact on all the spaces $L^p(\pi)$] by

$$(Gf)(x) = \int J(x, dy)\, [f(x) - f(y)] \tag{2.11}$$

[6] We follow the physicists' sign convention for G, which makes the associated operator positive-semidefinite in the reversible case [see (2.11) and (2.14)]. Mathematicians use the opposite sign convention, which makes G negative-semidefinite.

[7] Powers of kernels are defined exactly as powers of matrices: $G^0 \equiv I$, where I is the identity kernel $I(x, \cdot) = \delta_x$, and

$$G^n(x, B) \equiv \int \cdots \int G(x, dx_1)\, G(x_1, dx_2) \cdots G(x_{n-1}, B)$$

for $n \geqslant 1$. To see that P_t is indeed a positive kernel, write

$$P_t \equiv \exp(-tM) \exp[t(MI - G)] \equiv \exp(-tM) \sum_{n=0}^{\infty} \frac{(MI - G)^n}{n!}$$

(using properties of the exponential function that are valid in any Banach algebra) and note that $MI - G \geqslant 0$. It is trivial that $P_t(x, X) = 1$ for all x, since $G(x, X) = 0$ for all x. We learned this argument from ref. 27, p. 152, where it is attributed to John Kingman.

Finally, it can be shown (ref. 26, pp. 311–319) that given any initial probability distribution α, there exists a strong Markov process $\{X_t\}_{t \geqslant 0}$ whose sample paths are piecewise-constant and right-continuous with only isolated jumps, such that

$$P(X_0 \in A) = \alpha(A) \tag{2.12a}$$

$$P(X_u \in A \mid \{X_s\}_{0 \leqslant s \leqslant t}) = P_{u-t}(X_t, A) \tag{2.12b}$$

for all $A \in \mathscr{X}$ and all $u \geqslant t \geqslant 0$.

If, on the other hand, the transition rates are *unbounded*, then some subtleties arise.[8] From a purely analytic point of view, our goal is to define the operator G acting on some suitable space of functions. One approach is to try to make sense of (2.11) for functions f in some dense subspace of $L^p(\pi)$. We prefer, however, a simpler but more restricted approach based on quadratic forms. We assume that the jump process is reversible (with respect to π), i.e., that it satisfies

$$\pi(dx)\, J(x, dy) = \pi(dy)\, J(y, dx) \tag{2.13}$$

Then we can define the positive-semidefinite sesquilinear form

$$\mathscr{G}(f, g) = \int \pi(dx)\, J(x, dy)\, f(x)^* \; [g(x) - g(y)] \tag{2.14a}$$

$$= \frac{1}{2} \int \pi(x)\, J(x, dy)\; [f(x) - f(y)]^* \; [g(x) - g(y)] \tag{2.14b}$$

on the form domain

$$\mathscr{D}(\mathscr{G}) = \left\{ f \in L^2(\pi) \colon \int \pi(dx)\, J(x, dy) \; |f(x) - f(y)|^2 < \infty \right\} \tag{2.15}$$

Note that $\mathscr{D}(\mathscr{G})$ contains the set

$$\mathscr{D}_1(\mathscr{G}) = \left\{ f \in L^2(\pi) \colon \int \pi(dx)\, r(x) \; |f(x)|^2 < \infty \right\} \tag{2.16}$$

and so is dense in $L^2(\pi)$. We note the following additional facts about $\mathscr{G}$:

(a) $\mathscr{D}(\mathscr{G})$ is a linear subspace.

(b) $\mathscr{D}(\mathscr{G})$ contains the constant functions.

(c) $\mathscr{G}$ defined on $\mathscr{D}(\mathscr{G})$ is a *closed* form.

[8] The reader uninterested in these subtleties may skim lightly over the rest of Section 2.2.

(d) If $\Phi: \mathbf{R} \to \mathbf{R}$ has Lipschitz norm $\leqslant K < \infty$ [this means that $|\Phi(z) - \Phi(z')| \leqslant K|z - z'|$ for all z, z'] and $f \in \mathcal{Q}(\mathcal{G})$, then $\Phi \circ f \in \mathcal{Q}(\mathcal{G})$ and $\mathcal{G}(\Phi \circ f, \Phi \circ f) \leqslant K^2 \mathcal{G}(f, f)$.

(e) If $f, g \in \mathcal{Q}(\mathcal{G})$, then $f \vee g \equiv \max(f, g)$ and $f \wedge g \equiv \min(f, g)$ also belong to $\mathcal{Q}(\mathcal{G})$.

(f) If $f, g \in \mathcal{Q}(\mathcal{G}) \cap L^\infty(\pi)$, then $fg \in \mathcal{Q}(\mathcal{G})$ and

$$\mathcal{G}(fg, fg) \leqslant [\|f\|_\infty \mathcal{G}(g, g)^{1/2} + \|g\|_\infty \mathcal{G}(f, f)^{1/2}]^2$$

(g) If $f \in \mathcal{Q}(\mathcal{G})$ and $f_n \equiv ((-n) \vee f) \wedge n$, then $f_n \in \mathcal{Q}(\mathcal{G})$ and $\lim_{n \to \infty} \mathcal{G}(f_n - f, f_n - f) = 0$. In particular, $\mathcal{Q}(\mathcal{G}) \cap L^\infty(\pi)$ is a form core for $\mathcal{G}$.

(h) Under the additional hypothesis $\int r(x)\, \pi(dx) < \infty$, we have $\mathcal{Q}(\mathcal{G}) \supset \mathcal{Q}_1(\mathcal{G}) \supset L^\infty(\pi)$.

(i) Under the additional hypothesis $\int r(x)\, \pi(dx) < \infty$, the quadratic form $\mathcal{G}$ is *maximal Markovian*, i.e., there does not exist a proper extension of $\mathcal{G}$ that has property (d).[9]

Most of these statements are easy to prove; see ref. 30, pp. 13–14 and 25–26. The proof of (i) goes as follows: Let $\mathcal{G}'$ be an extension of $\mathcal{G}$ having property (d), and let $f \in \mathcal{Q}(\mathcal{G}')$. Then

$$f_n \equiv ((-n) \vee f)) \wedge n \in L^\infty(\pi) \subset \mathcal{Q}(\mathcal{G}) \subset \mathcal{Q}(\mathcal{G}')$$

and

$$\mathcal{G}(f_n, f_n) = \mathcal{G}'(f_n, f_n) \leqslant \mathcal{G}'(f, f) < \infty$$

by property (d) for $\mathcal{G}'$. But

$$\mathcal{G}(f_n, f_n) = \frac{1}{2} \int \pi(dx)\, J(x, dy)\, |f_n(x) - f_n(y)|^2$$

$$\to \frac{1}{2} \int \pi(dx)\, J(x, dy)\, |f(x) - f(y)|^2$$

by the monotone convergence theorem; hence $f \in \mathcal{Q}(\mathcal{G})$.

Since $\mathcal{G}$ is closed and semibounded, it follows (ref. 31, Section VI.2) that there is a unique self-adjoint operator G with dense domain $\mathcal{D}(G) \subset \mathcal{Q}(\mathcal{G})$ such that $\mathcal{G}(f, g) = (f, Gg)$ for all $f \in \mathcal{Q}(\mathcal{G})$ and $g \in \mathcal{D}(G)$. Moreover:

(a′) $\mathcal{D}(G)$ is a form core for $\mathcal{G}$.

[9] We do not know whether $\mathcal{G}$ is maximal Markovian if $\int r(x)\, \pi(dx) = \infty$.

(b′) $g \in \mathscr{D}(G)$ if and only if there exists $h \in L^2(\pi)$ such that $\mathscr{G}(f, g) = (f, h)$ holds for all f in some form core for $\mathscr{G}$; and in that case $Gg = h$.

(c′) G is positive-semidefinite, $\mathscr{D}(G^{1/2}) = \mathscr{Q}(\mathscr{G})$, and $\mathscr{G}(f, g) = (G^{1/2}f, G^{1/2}g)$ for all $f, g \in \mathscr{Q}(\mathscr{G})$.

(d′) A linear subspace is a form core for $\mathscr{G}$ if and only if it is an operator core for $G^{1/2}$.

It appears to be a difficult problem to describe exactly the operator domain $\mathscr{D}(G)$. However, it does follow from (b′) that the set

$$\mathscr{D}_1(G) = \left\{ f \in L^2(\pi) \colon \int \pi(dx) \left[\int J(x, dy) \, |f(x) - f(y)| \right]^2 < \infty \right\} \tag{2.17}$$

is contained in $\mathscr{D}(G)$, and that for $f \in \mathscr{D}_1(G)$ we have Gf given by (2.11) [that is, the right-hand side of (2.11) is absolutely convergent for π-a.e. x and equals $(Gf)(x)$]. Anyway, most of our work can be done using only the quadratic form $\mathscr{G}$.

Since G is positive-semidefinite, its spectrum lies in the half-line $[0, \infty)$. The constant function $\mathbf{1}$ belongs to $\mathscr{D}_1(G) \subset \mathscr{D}(G)$ and is an eigenvector of G with eigenvalue 0. The goal of this paper is to prove bounds on the spectrum of $G \restriction \mathbf{1}^\perp$ and in particular to show that in certain circumstances there must be spectrum very near 0. We define, therefore, the mass gap

$$m \equiv \inf \operatorname{spec}(G \restriction \mathbf{1}^\perp) \tag{2.18}$$

The main theorems of this paper are *upper bounds* on the mass gap.

From a purely analytic point of view, this is all we need: our main theorems are operator-theoretic statements about G (or $\mathscr{G}$). However, the *motivation* for studying this particular operator comes from probability theory: G is allegedly the infinitesimal generator of a Markovian jump process. We now turn, therefore, to the problem of constructing this process and ascertaining its properties. We temporarily drop the assumption of reversibility, as it is not needed.

It is, in fact, straightforward to construct the jump process $\{X_t\}_{t \geq 0}$, using the description in terms of exponentially distributed waiting times followed by jumps (see ref. 25, Section I.12 or ref. 26, pp. 362–369). But since the rate function r is unbounded, there arises the possibility that the system could make infinitely many jumps in a finite time ("explosion"), after which the evolution would be undetermined: in the so-called "minimal process," the system simply "dies," but there exists a continuum of other possibilities in which the system is "reborn" (see ref. 26, pp. 363–364 or ref. 27, Section 6.5). Corresponding to each such process is a semigroup $\{P_t\}_{t \geqslant 0}$ of kernels with generator G, which may, however, be substochastic rather than stochastic [i.e., $P_t(x, X) \leqslant 1$]; the minimal such semigroup, denoted $\{\bar{P}_t\}_{t \geqslant 0}$, is the one corresponding to the minimal process (see ref. 26, pp. 364–366 and ref. 27, Section 7.5).

The details of the construction are as follows: Define

$$\tilde{J}(x, dy) = \begin{cases} J(x, dy)/r(x) & \text{if } r(x) > 0 \\ \delta_x(dy) & \text{if } r(x) = 0 \end{cases} \tag{2.19}$$

Then $\tilde{J}$ is the transition probability kernel for a discrete-time Markov chain $\{\tilde{X}_0, \tilde{X}_1, \ldots\}$; we write E_x for expectations in this Markov chain with $\tilde{X}_0 \equiv x$. Next, let $\tau_0, \tau_1, \ldots$ be exponentially distributed random variables of mean $1/r(\tilde{X}_0)$, $1/r(\tilde{X}_1), \ldots$, which are conditionally independent given $\{\tilde{X}_0, \tilde{X}_1, \ldots\}$. Define $T_0 = 0$, $T_n = \sum_{i=0}^{n-1} \tau_i =$ time of nth jump, and $T_\infty = \sum_{i=0}^{\infty} \tau_i =$ time of first explosion (if any). Then we can define a continuous-time process (the "minimal process") by

$$X_t = \begin{cases} \tilde{X}_n & \text{if } T_n \leqslant t < T_{n+1} \\ \partial & \text{if } t \geqslant T_\infty \end{cases} \tag{2.20}$$

where $\partial \notin X$ is the "cemetery" state. It can be shown (ref. 25, Section I.12) that $\{X_t\}_{t \geqslant 0}$ is a strong Markov process. The minimal semigroup $\{\bar{P}_t\}_{t \geqslant 0}$ is defined by

$$(\bar{P}_t f)(x) = E_x[\chi(t < T_\infty)\, f(X_t)] \tag{2.21}$$

and the minimal resolvent $\{\bar{R}_\lambda\}_{\lambda \geqslant 0}$ by

$$\bar{R}_\lambda = \int_0^\infty dt\, e^{-\lambda t} \bar{P}_t \tag{2.22}$$

A crucial role will be played by the identities

$$E_x[e^{-\lambda T_1} f(\tilde{X}_1)] = \frac{r(x)}{r(x) + \lambda} \int \tilde{J}(x, dy)\, f(y)$$
$$\equiv \int \tilde{J}_\lambda(x, dy)\, f(y) \tag{2.23}$$

and more generally

$$E_x[e^{-\lambda T_n} f(\tilde{X}_n)] = \int \tilde{J}_\lambda^n(x, dy)\, f(y) \tag{2.24}$$

for $\lambda \geqslant 0$, which may be derived straightforwardly from the definition of the process. Here we have defined for each $\lambda \geqslant 0$ the subprobability kernel

$$\tilde{J}_\lambda(x, dy) = \frac{1}{r(x) + \lambda} J(x, dy) \tag{2.25}$$

which will play an important role in our development (see also ref. 32, and ref. 33, pp. 217–226).

The following proposition gives some necessary and sufficient conditions for the nonoccurrence of explosion:

Proposition 2.1. The following are equivalent (in order to save space, equivalent alternatives are shown in brackets):

(a) For every initial state x, the probability of having infinitely many jumps in a finite time interval is zero, i.e., $E_x[\chi(T_\infty < \infty)] = 0$ for all x.

(b) The minimal semigroup $\{\bar{P}_t\}_{t \geqslant 0}$ is stochastic, i.e., $\bar{P}_t(x, X) = 1$ for all t, x.

(c) There is at most one substochastic [or stochastic] semigroup $\{P_t\}_{t \geqslant 0}$ satisfying $P_t(x, A) = \chi_A(x) - tG(x, A) + o(t)$ as $t \to 0$ for all x, A.[10] (This unique semigroup is, of course, $\{\bar{P}_t\}$.)

(d) In the discrete-time Markov chain $\{\tilde{X}_0, \tilde{X}_1, \ldots\}$ with transition probability $\tilde{J}$, we have, for every initial state x,

$$\sum_{n=0}^{\infty} \frac{1}{r(\tilde{X}_n)} = \infty$$

with probability 1.

(e) For some [or all] $\lambda > 0$, we have $\lim_{n \to \infty} (\tilde{J}_\lambda^n \mathbf{1})(x) = 0$ for all x (not necessarily uniformly).

(f) For some [or all] $\lambda > 0$, the only bounded measurable [or nonnegative bounded measurable] solution to the equation $(\lambda + G) f = 0$ is the zero function.

Sketch of proof. (a)$\Leftrightarrow$(b) is immediate by definition of $\bar{P}_t$. (b)$\Rightarrow$(c) follows from the minimality of $\bar{P}_t$, while (c)$\Rightarrow$(a) follows by the construction of a continuum of distinct stochastic semigroups whenever (a) fails (see ref. 26, pp. 363–364 and ref. 27, Sections 6.5 and 7.5). (a)$\Leftrightarrow$(d) holds because the waiting times τ_n are (conditionally) independent

[10] Note that this asymptotic expansion is *not* required to be uniform in x. It does turn out, however, to be uniform in A for each fixed x (ref. 23, p. 330).

exponentially distributed random variables of mean $1/r(\tilde{X}_n)$, and it can be shown that $\sum_{n=0}^{\infty} \tau_n$ is finite or infinite (with probability 1) according as $\sum_{n=0}^{\infty} [1/r(\tilde{X}_n)]$ is (see ref. 27, pp. 153–154, Lemma 5.33). Finally, for each $\lambda > 0$ define $f_\lambda(x) \equiv E_x[\exp(-\lambda T_\infty)]$. By the dominated convergence theorem and (2.24), we have

$$f_\lambda(x) = \lim_{n \to \infty} E_x(e^{-\lambda T_n}) = \lim_{n \to \infty} (\tilde{J}_\lambda^n 1)(x)$$

This proves (a) $\Leftrightarrow$ (e). Now, clearly, $0 \leqslant f_\lambda \leqslant 1$, and it is not hard to show that $(\lambda + G) f_\lambda = 0$; this proves (f) $\Rightarrow$ (a). Conversely, it can be shown that for *any* function f satisfying $(\lambda + G) f = 0$ and $-1 \leqslant f \leqslant 1$, we have $-f_\lambda \leqslant f \leqslant f_\lambda$ (see ref. 26, pp. 367–368); this proves (a) $\Rightarrow$ (f). ∎

The next three propositions give simple sufficient (but *not* necessary) conditions for the nonoccurrence of explosion:

Proposition 2.2. Let the state space X be countable, and consider the discrete-time Markov chain with transition matrix $\tilde{J}$. If, in this chain, every state is recurrent, then the continuous-time process has no explosion. In particular, this occur if $r(x)\,\pi(x) > 0$ for all x and $\sum_x r(x)\,\pi(x) < \infty$.

Proof. If $\tilde{X}_0 = x$, then recurrence means that infinitely many of the $\tilde{X}_n$ equal x (with probability 1). This implies condition (d) of Proposition 2.1.

On the other hand, it is easy to check that $\tilde{\pi} \equiv r\pi$ is an invariant masure for the chain with transition matrix $\tilde{J}$ [this is equivalent to (2.8)]. If $\tilde{\pi}$ is a *finite* measure that gives nonzero weight to every state, then the state space consists only of positive-recurrent classes (ref. 22, Theorems I.7.1 and I.7.2), i.e., every state is positive-recurrent. ∎

Proposition 2.3. Let ψ be a nonnegative measurable function on the state space X, such that:

(a) r is bounded on each set $\{x: \psi(x) \leqslant K\}$.

(b) $\int J(x, dy)[\psi(y) - \psi(x)] \leqslant C\psi(x)$ for some constant C.

Then the continuous-time process has no explosion.

Heuristic explanation. Formally, hypothesis (b) implies that $(d/dt)\, E(\psi(X_t)) \leqslant CE(\psi(X_t))$, so that $E(\psi(X_t)) \leqslant e^{Ct}\psi(X_0)$. Morally, this means that the "Liapunov function" ψ cannot become infinite in a finite time—so, by hypothesis (a), neither can r—hence there is no explosion. However, more careful consideration shows that it is not enough to look only at expectations; instead, we will prove the stronger result that $\{e^{-Ct}\psi(X_t)\}_{t \geqslant 0}$ is a supermartingale, hence almost surely bounded. For technical resaons, it is convenient to work instead with the discrete-time Markov chain $\{(\tilde{X}_0, 0), (\tilde{X}_1, \tau_0), (\tilde{X}_2, \tau_1),...\}$ consisting of the states visited and their holding times.

Proof. Hypothesis (b) says that

$$E(\psi(\tilde{X}_{n+1}) \mid \tilde{X}_0,..., \tilde{X}_n, \tau_0,..., \tau_{n-1}) \leqslant \left(1 + \frac{C}{r(\tilde{X}_n)}\right) \psi(\tilde{X}_n) \tag{2.26}$$

On the other hand, we know that

$$E(e^{-\lambda \tau_n} \mid \tilde{X}_0,..., \tilde{X}_n, \tau_0,..., \tau_{n-1}) = \frac{r(\tilde{X}_n)}{r(\tilde{X}_n) + \lambda} \tag{2.27}$$

for all $\lambda \geqslant 0$. Moreover, $\tilde{X}_{n+1}$ and τ_n are independent conditionally on $\tilde{X}_0,..., \tilde{X}_n, \tau_0,..., \tau_{n-1}$. Taking $\lambda = C$, we conclude that

$$\left\{\psi(\tilde{X}_n) \exp\left[-C \sum_{i=0}^{n-1} \tau_i\right]\right\}_{n \geqslant 0}$$

is a supermartingale. Since $\psi(\tilde{X}_0) < \infty$, it follows by standard martingale theory (Ref. 34, Proposition II-2-7) that

$$\sup_{n \geqslant 0} \psi(\tilde{X}_n) \exp\left[-C \sum_{i=0}^{n-1} \tau_i\right] < \infty \quad \text{a.s.}$$

If $\sum_{i=0}^{\infty} \tau_i < \infty$, this means that $\sup_{n \geqslant 0} \psi(\tilde{X}_n) < \infty$, and hence, by hypothesis (a), that $\sup_{n \geqslant 0} r(\tilde{X}_n) < \infty$; but this would imply (as noted previously) that in fact $\sum_{i=0}^{\infty} \tau_i = \infty$ a.s., a contradiction. We conclude that $\sum_{i=0}^{\infty} \tau_i = \infty$ with probability 1, i.e., there is no explosion. ∎

Remark. This proposition was inspired by ref. 35, p. 263, Problem 4.11.15(c) and by ref. 36, Theorem 10.2.1. See also ref. 37.

Proposition 2.4. Assume that the state space can be partitioned as $X = \bigcup_{n=1}^{\infty} \hat{X}_n$, where $J(x, \hat{X}_n) = 0$ whenever $x \in \hat{X}_m$ and $n > m + 1$. Assume further that

$$\sup_{x \in \hat{X}_m} J(x, \hat{X}_{m+1}) \leqslant M_m$$

with

$$\sum_{m=1}^{\infty} \frac{1}{M_m} = \infty \tag{2.28}$$

Then the continuous-time process has no explosion.

Proof. This is a corollary of the preceding proposition: just take

$$\psi(x) = \prod_{i=1}^{m-1} \left(1 + \frac{C}{M_i}\right)$$

for $x \in \hat{X}_m$. Alternatively, it is easy to show directly that these hypotheses imply condition (d) of Proposition 2.1; the details are left to the reader. ∎

Example (Birth–death process). Let $X = \{0, 1, 2,...\}$ and let $J(x, x+1) = \lambda_x$, $J(x, x-1) = \mu_x$. Assume for simplicity that all $\lambda_x, \mu_x > 0$. Then the unique invariant measure is given by

$$\pi_x = \frac{\lambda_0 \lambda_1 \cdots \lambda_{x-1}}{\mu_1 \mu_2 \cdots \mu_x} \pi_0$$

(in fact, J reversible with respect to π). We have the following necessary and sufficient conditions (see ref. 26, pp. 368–369 and ref. 22, Section I.12):

- (a) No explosion $\Leftrightarrow \sum_{n=0}^{\infty} (1/\lambda_n \pi_n)(\pi_n + \pi_{n-1} + \cdots + \pi_0) = \infty$.
- (b) Recurrence of discrete-time chain $\Leftrightarrow \sum_{n=0}^{\infty} (1/\lambda_n \pi_n) = \infty$.
- (c) Positive-recurrence of discrete-time chain $\Leftrightarrow \sum_{n=0}^{\infty} \lambda_n \pi_n < \infty$.
- (d) Condition of Proposition 2.4 $\Leftrightarrow \sum_{n=0}^{\infty} (1/\lambda_n) = \infty$.

Clearly, (c) $\Rightarrow$ (b) $\Rightarrow$ (a) and (d) $\Rightarrow$ (a), as required by Propositions 2.2 and 2.4, although these implications are in general far from being sharp. [Note, however, that (a) $\Leftrightarrow$ (b) if the invariant measure π is finite.]

Further discussion of what happens in the event of "explosion" can be found in refs. 22, 27, 32, and 38–41. In this paper, however, we restrict attention to models in which there is *no* explosion.

The next step is to verify that the semigroup $\{P_t\}_{t \geqslant 0}$ has all the desired properties:

Proposition 2.5. Let f be a bounded measurable function, and fix $x \in X$. Then $(\bar{P}_t f)(x)$ is a continuous function of t for $t \geqslant 0$. In fact, $(\bar{P}_t f)(x)$ is a differentiable function of t and satisfies the backward Kolmogorov equation

$$\frac{d}{dt} (\bar{P}_t f)(x) = -(G\bar{P}_t f)(x) \tag{2.29}$$

where the action of G on bounded measurable functions is defined by (2.11).

Proof. See ref. 23, pp. 331–336. ∎

Proposition 2.6. Assume that there is no explosion, and let π be a probability measure satisfying $\int r(x)\,\pi(dx) < \infty$.

(a) Suppose that π satisfies the infinitesimal invariance condition (2.8). Then the (unique) semigroup $\{P_t\}_{t \geqslant 0}$ leaves π invariant, i.e.,

$$\int \pi(dx)\,(P_t f)(x) = \int \pi(dx)\,f(x) \tag{2.30}$$

for all $t \geqslant 0$ and all bounded measurable functions f.

(b) Suppose that π satisfies the infinitesimal reversibility condition (2.13). Then the (unique) semigroup $\{P_t\}_{t \geqslant 0}$ is symmetric with respect to π, i.e.,

$$\int \pi(dx)\,(P_t f)(x)\,g(x) = \int \pi(dx)\,f(x)\,(P_t g)(x) \tag{2.31}$$

for all $t \geqslant 0$ and all bounded measurable functions f, g.

Proof. (a) Let f be a bounded measurable function; then for the resolvent R_λ $(\lambda > 0)$ we compute

$$\begin{aligned}(R_\lambda f)(x) &= \int_0^\infty dt\,e^{-\lambda t} E_x[\chi(t < T_\infty)\,f(X_t)] \\ &= \sum_{n=0}^{\infty} E_x\left[\int_{T_n}^{T_{n+1}} dt\,e^{-\lambda t} f(\tilde{X}_n)\right] \\ &= \frac{1}{\lambda}\sum_{n=0}^{\infty} E_x[e^{-\lambda T_n}(1 - e^{-\lambda \tau_n})\,f(\tilde{X}_n)] \\ &= \sum_{n=0}^{\infty} \int \tilde{J}_\lambda^n(x, dy)\,\frac{1}{r(y)+\lambda}\,f(y)\end{aligned} \tag{3.32}$$

where in the last step we used the identity (2.24). Now multiply both sides by λ and integrate with respect to $\pi(dx)$. From (2.8) we easily derive the identities

$$\lambda\pi\tilde{J}_\lambda = r\pi - (r\pi)\tilde{J}_\lambda \tag{2.33}$$

and hence

$$\lambda\pi\tilde{J}_\lambda^n = (r\pi)\tilde{J}_\lambda^{n-1} - (r\pi)\tilde{J}_\lambda^n \tag{2.34}$$

for $n \geqslant 1$. Substituting (2.34) into (2.32) and telescoping the sum, we obtain

$$\lambda \int \pi(dx)(R_\lambda f)(x) = \int \pi(dy)\,f(y) - \lim_{N\to\infty} \int \pi(dx)\,r(x)\,\tilde{J}_\lambda^N(x, dy)\,\frac{1}{r(y)+\lambda}\,f(y) \tag{2.35}$$

Now, by the absence of explosion, we have $\lim_{N \to \infty} \tilde{J}_\lambda^N g = 0$ pointwise for every bounded measurable function g [see Proposition 2.1(e)]; and of course $\|\tilde{J}_\lambda^N g\|_\infty \leqslant \|g\|_\infty$. Since $r\pi$ is by hypothesis a finite measure, the final term in (2.33) vanishes by the dominated convergence theorem. By the uniqueness theorem for Laplace transforms, it follows that (2.30) holds for almost all $t \geqslant 0$. But then Proposition 2.5 and the dominated convergence theorem imply that (2.30) holds for all $t \geqslant 0$.

(b) By (2.32) we have

$$\int \pi(dx)(R_\lambda f)(x)\, g(x) = \sum_{n=0}^{\infty} \int \pi(dx_0)\, J(x_0, dx_1) \cdots J(x_{n-1}, dx_n)\, g(x_0)\, f(x_n) \prod_{i=0}^{n-1} \frac{1}{r(x_i) + \lambda} \tag{2.36}$$

On the other hand, by repeated use of the infinitesimal reversibility condition (2.13), we have

$$\pi(dx_0)\, J(x_0, dx_1) \cdots J(x_{n-1}, dx_n) = \pi(dx_n)\, J(x_n, dx_{n-1}) \cdots J(x_1, dx_0) \tag{2.37}$$

as measures on X^{n+1}. These two facts together imply (2.31). ∎

Remark. Proposition 2.6 can alternatively be proven by random time change starting from the continuous-time jump process with jump rates $\tilde{J}$ (S. R. S. Varadhan, private communication). Proposition 2.6(a) can alternatively be proven using the backward Kolmogorov equation (2.29) together with the uniform bound $|(P_t f)(x) - f(x)|/t \leqslant 2r(x)\|f\|_\infty$ (see ref. 23, pp. 331–332 for this bound). See also refs. 42 and 43 for related results.

From now on we assume that $\int r(x)\, \pi(dx) < \infty$. It follows from the invariance of π, just as in the discrete-time case, that P_t is contractive in the $L^p(\pi)$ norm, i.e.,

$$\|P_t f\|_{L^p(\pi)} \leqslant \|f\|_{L^p(\pi)} \tag{2.38}$$

for each $1 \leqslant p \leqslant \infty$ and all bounded measurable functions f. (In particular, $P_t f$ is π-null whenever f is.) Since the bounded measurable functions are dense in $L^p(\pi)$, the semigroup $\{P_t\}_{t \geqslant 0}$ extends by continuity to a contraction semigroup on $L^p(\pi)$; in fact, this extended semigroup is given by the obvious integral formula

$$(P_t f)(x) = \int P_t(x, dy)\, f(y) \tag{2.39}$$

where the integral is well-defined for π-a.e. x. Finally, Proposition 2.5 together with the dominated convergence theorem and an $\varepsilon/3$ argument imply that the semigroup $\{P_t\}_{t\geqslant 0}$ is strongly continuous on $L^p(\pi)$ for $1\leqslant p<\infty$ (though not necessarily for $p=\infty$). Thus, this semigroup has an infinitesimal generator G_p, with domain $\mathscr{D}(G_p)$ dense in $L^p(\pi)$.

We now restore the assumption of reversibility. It follows that the operator G_2 on $L^2(\pi)$ is self-adjoint and positive-semidefinite; we denote by $\mathscr{G}_2$ the corresponding closed sesquilinear form with $\mathscr{Q}(\mathscr{G}_2)=\mathscr{D}(G_2^{1/2})$. It is easy to see that $\mathscr{G}_2$ is Markovian, i.e., satisfies property (d) above. Now let f be a bounded measurable function. Then, by Proposition 2.5 and its proof (see ref. 23, pp. 331–332), the quantity $[(P_tf)(x)-f(x)]/t$ is bounded uniformly in t by $2r(x)\|f\|_\infty$, and as $t\to 0$ it converges pointwise to $-(Gf)(x)$. By the dominated convergence theorem [and using again the hypothesis $\int r(x)\ \pi(dx)<\infty$], we conclude that for bounded measurable functions f, g we have

$$\begin{aligned}\lim_{t\to\infty}\frac{(g,P_tf)_{L^2(\pi)}-(g,f)_{L^2(\pi)}}{t}&=-\int\pi(dx)\ g(x)\ (Gf)(x)\\&=-\mathscr{G}(g,f)\end{aligned}\tag{2.40}$$

In particular, by taking $g=f$ and using the spectral theorem, it follows [Ref. 30, p. 21, Lemma 1.3.4(i)] that $L^\infty(\pi)\subset\mathscr{Q}(\mathscr{G}_2)$ and $\mathscr{G}_2\upharpoonright L^\infty(\pi)=\mathscr{G}\upharpoonright L^\infty(\pi)$. On the other hand, $L^\infty(\pi)$ is a form core for $\mathscr{G}$ [fact (g) above], so it follows that $\mathscr{Q}(\mathscr{G}_2)\supset\mathscr{Q}(\mathscr{G})$ and $\mathscr{G}_2\upharpoonright\mathscr{Q}(\mathscr{G})=\mathscr{G}$—that is, $\mathscr{G}_2$ is an extension of $\mathscr{G}$. But since $\mathscr{G}$ is maximal Markovian [fact (i) above], it follows that $\mathscr{G}_2=\mathscr{G}$ and hence $G_2=G$.

This completes our analysis of the continuous-time Markovian jump process with generating kernel G.

Our final topic in this section concerns Dirichlet boundary conditions. Consider a reversible Markovian jump process with generating kernel G, as at the beginning of this subsection, and let A be a subset of the state space. We define G_A to be the generator for the process that evolves according to G as long as it stays within A, but is killed when it tries to jump outside A:

$$G_A(x,\cdot)\equiv r(x)\delta_x-J(x,\cdot\cap A)\tag{2.41}$$

This process can be constructed in precisely the same manner as the original process; the construction is unique provided that the original process has no explosion. Of course, the semigroup $\{(P_A)_t\}_{t\geqslant 0}$ is in general *substochastic*. Thus, if $\{X_t\}_{t\geqslant 0}$ is the right-continuous stochastic process whose generator is G, so that

$$E_x[f(X_t)]=(P_tf)(x)\tag{2.42}$$

then we have

$$E_x[\chi(\tau_{A^c} > t)\, f(X_t)] = ((P_A)_t f)(x) \tag{2.43}$$

where

$$\tau_{A^c} \equiv \inf\{t\colon X_t \in A^c\} \tag{2.44}$$

is the time of first exit from A.

If $\pi(A) > 0$, we can make sense of G_A as an operator on the Hilbert space $L^2(A, \pi)$. In fact, let $\mathscr{G}_A$ be the positive-semidefinite sesquilinear form

$$\mathscr{G}_A(f, g) = \mathscr{G}(\hat{f}, \hat{g}) \tag{2.45}$$

with form domain

$$\mathscr{D}(\mathscr{G}_A) = \{f \in L^2(A, \pi)\colon \hat{f} \in \mathscr{D}(\mathscr{G})\} \tag{2.46}$$

where

$$\hat{f}(x) \equiv \begin{cases} f(x), & x \in A \\ 0, & x \notin A \end{cases} \tag{2.47}$$

Note that $\mathscr{D}(\mathscr{G}_A)$ contains the set

$$\mathscr{D}_1(\mathscr{G}_A) = \left\{ f \in L^2(A, \pi)\colon \int_A \pi(dx)\; r(x)\, |f(x)|^2 < \infty \right\} \tag{2.48}$$

and so is dense in $L^2(A, \pi)$. Then $\mathscr{G}_A$ determines, as before, a positive-semidefinite, self-adjoint operator G_A acting on $L^2(A, \pi)$. Moreover, the semigroup $\{(P_A)_t\}_{t \geqslant 0}$ is a strongly continuous, self-adjoint contraction semigroup on $L^2(A, \pi)$, and its generator $G_{2,A}$ is precisely G_A.

Finally, note that the transition probability P of a discrete-time Markov chain can serve also as the transition rate kernel of a continuous-time jump process (the process that waits an exponentially distributed time of mean 1 and then jumps according to P). The generator G of this process is just $\tilde{P} \equiv I - P$. Thus, it suffices to state our results for the generators G of continuous-time jump processes; the analogous results for the operators $\tilde{P}$ associated with discrete-time Markov chains follow immediately as a special case.

3. UPPER BOUNDS ON THE MASS GAP FOR REVERSIBLE MARKOV CHAINS AND MARKOVIAN JUMP PROCESS

In this section we explain two general principles—the minimum hitting-time argument (Theorem 3.1) and the mean (or mean-exponential) hitting-time argument (Propositions 3.3–3.10)—which provide upper bounds on the mass gap (L^2 spectral gap) for the generators of reversible Markov chains and Markovian jump processes.

3.1. The Minimum Hitting-Time Argument

Let G be the generator of a reversible Markovian jump process with transition rates bounded by M, that is,

$$r(x) \leqslant M < \infty \qquad \text{for all } x \tag{3.1}$$

The boundedness of transition rates is, of course, a rather severe restriction. It does, nevertheless, cover many interesting cases, including all discrete-time Markov chains (for which $M = 1$). Clearly, G is a positive-semidefinite, self-adjoint operator on $L^2(\pi)$ with spectrum in the interval $[0, 2M]$.

For any subsets $A, B \subset X$, let T_{AB} be the minimum number of jumps in which the system can get from A to B with nonzero probability (for a π-nonnull set of points in A): that is,

$$T_{AB} \equiv \min \left\{ n \geqslant 0: \quad \int \pi(x)\, \chi_A(x)\, J^n(x, B) > 0 \right\} \tag{3.2}$$

We then have the following result:

Theorem 3.1. Let G be the generator of a reversible Markovian jump process with transition rates uniformly bounded by M, and let m be its mass gap. Then

$$m \leqslant \inf_{A, B \subset X} 2M \tanh^2 \left[\frac{1}{2(T_{AB} - 1)} \operatorname{arccosh} \left(\frac{[1 - \pi(A)][1 - \pi(B)]}{\pi(A)\, \pi(B)} \right)^{1/2} \right] \tag{3.3a}$$

$$\leqslant \inf_{A, B \subset X} \frac{M}{8(T_{AB} - 1)^2} \log^2 \frac{4}{\pi(A)\, \pi(B)} \tag{3.3b}$$

Proof. Let $A, B \subset X$, and let p_n be a polynomial of degree $n < T_{AB}$. Then

$$(\chi_A, p_n(G) \chi_B) = 0 \tag{3.4}$$

by hypothesis, and

$$p_n(G) \mathbf{1} = p_n(0) \mathbf{1} \tag{3.5a}$$

$$p_n(G)^* \mathbf{1} = p_n(0)^* \mathbf{1} \tag{3.5b}$$

since $G\mathbf{1} = G^*\mathbf{1} = 0$. Hence

$$(\chi_A - \pi(A)\,\mathbf{1},\, p_n(G)[\chi_B - \pi(B)\,\mathbf{1}]) = -\pi(A)\,\pi(B)\,p_n(0) \tag{3.6}$$

On the other hand,

$$\| p_n(G) \restriction \mathbf{1}^\perp \| \leqslant \sup_{z \in [m, 2M]} |p_n(z)| \tag{3.7}$$

and hence, by the Schwarz inequality,

$$\begin{aligned} &|(\chi_A - \pi(A)\mathbf{1},\, p_n(G)[\chi_B - \pi(B)\mathbf{1}])| \\ &\quad \leqslant \|\chi_A - \pi(A)\mathbf{1}\|\, \|\chi_B - \pi(B)\mathbf{1}\| \sup_{z \in [m, 2M]} |p_n(z)| \\ &\quad = \pi(A)^{1/2}[1 - \pi(A)]^{1/2}\, \pi(B)^{1/2}[1 - \pi(B)]^{1/2} \sup_{z \in [m, 2M]} |p_n(z)| \end{aligned} \tag{3.8}$$

An easy bound (which is good enough for our applications in Section 4) can be obtained by taking $p_n(z) = (1 - z/2M)^n$. But we can do better, by taking p_n to be the polynomial that minimizes $\sup_{z \in [m, 2M]} |p_n(z)/p_n(0)|$. That is, we take

$$p_n(z) = \text{const} \times T_n\left(\frac{2M + m - 2z}{2M - m}\right) \tag{3.9}$$

where

$$T_n(\zeta) = \begin{cases} \cos(n \arccos \zeta) & -1 \leqslant \zeta \leqslant 1 \\ \cosh(n \operatorname{arccosh} \zeta) & |\zeta| \geqslant 1 \end{cases} \tag{3.10}$$

is the nth Chebyshev polynomial. We then have

$$\sup_{z \in [m, 2M]} \left| \frac{p_n(z)}{p_n(0)} \right| = \frac{1}{\cosh\{n \operatorname{arccosh}[(2M + m)/(2M - m)]\}} \tag{3.11}$$

(see, e.g., ref. 44). Combining (3.6), (3.8), and (3.11) and letting $n = T_{AB} - 1$, we arrive after some algebra at (3.3a). The elementary inequalities $\tanh x \leqslant x$ and $\operatorname{arccosh} x \leqslant \log(2x)$ then yield (3.3b). ∎

Remark. The statement and proof of Theorem 3.1 are purely analytic: they make no reference to the underlying stochastic process.

Theorem 3.1 says that if the minimum number of jumps to get from A to B is large, and this is not justified by the rarity of A and/or B in the invariant distribution—here the "justifiable" time is of order $\log[1/\pi(A)\,\pi(B)]$—then the mass gap must be small.

Examples. 1. Symmetric random walk on $\{0,..., N\}$ with reflecting boundary conditions. The transition matrix is

$$P(0, 1) = 1 \tag{3.12a}$$

$$P(N, N-1) = 1 \tag{3.12b}$$

$$P(x, x-1) = P(x, x+1) = \tfrac{1}{2} \qquad \text{for} \quad 1 \leqslant x \leqslant N-1 \tag{3.12c}$$

The exact mass gap is $m = 1 - \cos(\pi/N) \approx \pi^2/2N^2$. If we take $A = \{x: x \leqslant i\}$ and $B = \{x: x \geqslant j\}$ with $i < j$, then Theorem 3.1 yields the bound

$$m \leqslant \frac{1}{8(j-i-1)^2} \log^2 \frac{4N^2}{(i+1/2)(N-j+1/2)} \tag{3.13}$$

Clearly, taking $i \approx \alpha N$, $j \approx (1-\alpha)N$ with $0 < \alpha < 1/2$ gives the optimal-order bound $m \leqslant \text{const} \times N^{-2}$. (However, the coefficient is not sharp.)

2. Random walk with inward drift on $\mathbf{Z}_+$, with elastic boundary conditions at 0. The transition matrix is

$$P(x, x+1) = p \tag{3.14a}$$

$$P(x, x-1) = 1-p \qquad \text{for} \quad x \geqslant 1 \tag{3.14b}$$

$$P(0, 0) = 1-p \tag{3.14c}$$

with $0 < p < 1/2$. The invariant measure is $\pi(x) = \text{const} \times [p/(1-p)]^x$, i.e., it has exponential decay with mean $\langle N \rangle = p/(1-2p)$, while the exact mass gap is $m = 1 - 2p^{1/2}(1-p)^{1/2}$ (see ref. 3, Appendix A or ref. 45). If we take $A = \{x: x \leqslant i\}$ and $B = \{x: x \geqslant j\}$ with $i < j$, then Theorem 3.1 yields the bound

$$m \leqslant \frac{1}{8(j-i-1)^2} \log^2 \frac{4}{\beta^j(1-\beta^{i+1})} \tag{3.15}$$

where $0 < \beta \equiv p/(1-p) < 1$. Taking $j \to \infty$ (at any fixed i), we get

$$m \leqslant \frac{1}{8} \log^2 \frac{1}{\beta} \tag{3.16}$$

As $p \to \frac{1}{2}$ (i.e., as $\langle N \rangle \to \infty$), both this upper bound and the exact mass gap behave as $2(\frac{1}{2}-p)^2 + O[(\frac{1}{2}-p)^4]$, so even the coefficient is sharp!

In the problems treated in this paper, we will not need the full strength of Theorem 3.1, but only the following obvious corollary:

Corollary 3.2. Let G be the generator of a reversible Markovian jump process with transition rates bounded by $M < \infty$, and let m be its mass gap. If

$$\inf_{A, B \subset X} \frac{\log[1/\pi(A)\, \pi(B)]}{T_{AB}} = 0$$

then $m = 0$.

3.2. The Mean (or Mean-Exponential) Hitting-Time Argument

Let G be the generator of a reversible Markovian jump process with state space X (in this subsection we are *not* assuming that the transition rates are bounded), and let A be a subset of X with $\pi(A) > 0$. Recall that G_A is the generator obtained from G by imposing Dirichlet boundary conditions on the complement of A. More precisely, G_A is the positive-semidefinite, self-adjoint operator on the space $L^2(A, \pi)$ defined via the quadratic form

$$(f, G_A f)_{L^2(A,\pi)} = (\hat{f}, G\hat{f})_{L^2(X,\pi)} \tag{3.17}$$

where

$$\hat{f}(x) \equiv \begin{cases} f(x), & x \in A \\ 0, & x \notin A \end{cases} \tag{3.18}$$

In probabilistic terms, G_A is the generator of the process that evolves according to G as long as it stays within A, but is killed when it tries to jump outside A. That is, if $\{X_t\}_{t \geqslant 0}$ is the right-continuous stochastic process whose generator is G, so that

$$E_\alpha[f(X_t)] = (d\alpha/d\pi, e^{-tG} f)_{L^2(\pi)} \tag{3.19}$$

for any initial distribution α whose Radon–Nikodým derivative $d\alpha/d\pi$ lies in $L^2(\pi)$, then

$$E_\alpha[\chi(\tau_{A^c} > t)\, f(X_t)] = (d\alpha/d\pi, e^{-tG_A} f)_{L^2(A,\pi)} \tag{3.20}$$

where

$$\tau_{A^c} \equiv \inf\{t\colon X_t \in A^c\} \tag{3.21}$$

is the time of first exit from A.

Finally, we define

$$m_A \equiv \inf \operatorname{spec}(G_A) \tag{3.22}$$

In view of (3.20), m_A measures the exponential rate at which probability is guaranteed to leak out of the set A.

Proposition 3.3. For any set A with $0 < \pi(A) < 1$, we have

$$m_A \geqslant m\pi(A^c) \tag{3.23}$$

Proof. Let $f \in L^2(\pi)$ be supported in A, with $\|f\| = 1$. Now decompose $f = a\mathbf{1} + f^\perp$ with $f^\perp$ orthogonal to $\mathbf{1}$ in $L^2(X, \pi)$. Then, by the Schwarz inequality,

$$|a| = |(f, \mathbf{1})| = |(f, \chi_A)| \leqslant \pi(A)^{1/2} \tag{3.24}$$

so that $\|f^\perp\| \geqslant \pi(A^c)^{1/2}$. Hence

$$(f, G_A f) = (f, Gf) = (f^\perp, Gf^\perp) \geqslant m\,\|f^\perp\|^2 \geqslant m\pi(A^c) \tag{3.25}$$

Since $f \in L^2(A, \pi)$ is arbitrary, this proves the proposition. ∎

Remarks. 1. This proposition was also proven by Maitre and Musy (ref. 46, Proposition 4, bound α_1) in an entirely different context (convergence proofs for the multigrid iteration in numerical analysis). A slightly weaker bound is implicit in the work of Carmona and Klein (ref. 47, proof of Theorem 1).

2. If A^c is a one-point set, then the reverse inequality $m \geqslant m_A$ holds as a consequence of the min-max theorem.[48] More generally, the inequality

$$m \geqslant \inf_{A:0<\pi(A)<1} \max(m_A, m_{A^c})$$

holds; see ref. 2, Proposition 3.3.

Proposition 3.3 shows that m_A (the exponential decay rate of the killing time τ_{A^c}) cannot be too small if neither $\pi(A^c)$ nor the mass gap m is too small. Alternatively, it shows that if m_A *is* small and $\pi(A^c)$ is not too small, then the mass gap m must be small.

The next step is to get an upper bound on m_A. From a purely analytic point of view we have the following result:

Proposition 3.4. Let ε and c be real numbers, and let h be a function (not identically zero) in $\mathcal{Q}(\mathcal{G})$ satisfying[11]

[11] Strictly speaking, hypothesis (3.26a) should be interpreted as saying that $\mathcal{G}(\phi, h) \leqslant (\phi, \varepsilon h + c\chi_A)$ for all functions ϕ in $\mathcal{Q}(\mathcal{G})$ that are nonnegative on A and vanishing on A^c. Of course, if h happens to lie in $\mathcal{D}(G)$, then this statement is equivalent to (3.26a) interpreted in the pointwise π-a.e. sense.

$$(G-\varepsilon)h \leqslant c \qquad \text{on } A \tag{3.26a}$$

$$h \geqslant 0 \qquad \text{on } A \tag{3.26b}$$

$$h = 0 \qquad \text{on } A^c \tag{3.26c}$$

Then

$$m_A \leqslant \varepsilon + c\frac{(\chi_A, h)}{(h, h)} \tag{3.27a}$$

$$\leqslant \varepsilon + c\frac{\pi(A)^{1/2}}{\|h\|_{L^2(\pi)}} \qquad (\text{if } c \geqslant 0) \tag{3.27b}$$

Proof. Use h as a Rayleigh–Ritz trial function for the quadratic form $\mathscr{G}_A$:

$$\frac{\mathscr{G}_A(h, h)}{(h, h)} = \frac{\mathscr{G}(h, h)}{(h, h)} \leqslant \frac{(h, \varepsilon h + c\chi_A)}{(h, h)} \tag{3.28}$$

This proves (3.27a). Then (3.27b) follows by the Schwarz inequality $(\chi_A, h) \leqslant \|\chi_A\| \, \|h\| = \pi(A)^{1/2} \|h\|$. ∎

There is a variant of Proposition 3.4 in which h need not vanish identically on A^c:

Proposition 3.5. Let $\varepsilon \geqslant 0$ and $c \in \mathbf{R}$, and let h be a function in $\mathscr{Q}(\mathscr{G})$ satisfying (see footnote 11)

$$(G-\varepsilon)h \leqslant c \qquad \text{on } A \tag{3.29a}$$

$$h \geqslant 0 \qquad \text{on } A \tag{3.29b}$$

$$h \leqslant 0 \qquad \text{on } A^c \tag{3.29c}$$

and such that $h_+ \equiv \max(h, 0)$ is not identically zero. Then h_+ satisfies the hypotheses of Proposition 3.4, so that

$$m_A \leqslant \varepsilon + c\frac{(\chi_A, h_+)}{(h_+, h_+)} \tag{3.30a}$$

$$\leqslant \varepsilon + c\frac{\pi(A)^{1/2}}{\|h_+\|_{L^2(\pi)}} \qquad (\text{if } c \geqslant 0) \tag{3.30b}$$

Proof. By Lemma 3.6 below, we have $h_+ \in \mathscr{Q}(\mathscr{G})$ and

$$\mathscr{G}(\phi, h_+) \leqslant \mathscr{G}(\phi, h) \leqslant (\phi, \varepsilon h + c\chi_A) = (\phi, \varepsilon h_+ + c\chi_A) \tag{3.31}$$

for any nonnegative function $\phi \in \mathscr{Q}(\mathscr{G})$ that vanishes on A^c. Thus, the function h_+ satisfies the hypotheses of Proposition 3.4. ∎

Lemma 3.6. Let $h \in \mathscr{Q}(\mathscr{G})$. Then $h_+ \in \mathscr{Q}(\mathscr{G})$, and

$$\mathscr{G}(\phi, h_+) \leqslant \mathscr{G}(\phi, h) \tag{3.32}$$

for all nonnegative $\phi \in \mathscr{Q}(\mathscr{G})$ that are supported on the set where $h \geqslant 0$.

Proof. We already know that $h_+ \in \mathscr{Q}(\mathscr{G})$ [see Section 2.2, fact (d) or (e)]. Now let $\phi \in \mathscr{Q}(\mathscr{G})$ be supported on the set where $h \geqslant 0$. Then

$$\begin{aligned}\mathscr{G}(\phi, h_+) &\equiv \int \pi(dx)\, J(x, dy)\, \phi(x)\, [h_+(x) - h_+(y)] \\ &= \int \pi(dx)\, J(x, dy)\, \phi(x)\, [h(x) - h_+(y)] \\ &\leqslant \int \pi(dx)\, J(x, dy)\, \phi(x)\, [h(x) - h(y)] \\ &\equiv \mathscr{G}(\phi, h) \quad \blacksquare\end{aligned} \tag{3.33}$$

Remark. Lemma 3.6 is a weakened version of a limiting case of Lemma 3.8 below.

The probabilistic interpretation of Propositions 3.4 and 3.5 is given by the following:

Proposition 3.7.

(a) Fix $c \geqslant 0$. The function

$$\bar{h}_\varepsilon(x) \equiv \begin{cases} (c/\varepsilon)\, E_x[\exp(\varepsilon \tau_{A^c}) - 1] & \text{if} \quad \varepsilon \neq 0 \\ c E_x(\tau_{A^c}) & \text{if} \quad \varepsilon = 0 \end{cases} \tag{3.34}$$

is nonnegative on A and vanishing on A^c; if it is everywhere finite, then it satisfies (3.29a) as an equality.

(b) Let h be any function in $\mathscr{D}(G)$ that satisfies (3.29a) and (3.29c). Then $h(x) \leqslant \bar{h}_\varepsilon(x)$ for all x.

Proof. (a) Consider the jump process starting at $x \in A$, and condition on the first jump. By the strong Markov property, we have

$$E_x[\exp(\varepsilon \tau_{A^c})] \equiv \begin{cases} \dfrac{r(x)}{r(x) - \varepsilon} \displaystyle\int \tilde{J}(x, dy)\, E_y[\exp(\varepsilon \tau_{A^c})] & \text{if} \quad \varepsilon < r(x) \\ +\infty & \text{if} \quad \varepsilon \geqslant r(x) \end{cases} \tag{3.35}$$

where $\tilde{J}(x, dy) \equiv J(x, dy)/r(x)$. After some algebra this gives equality in (3.29a) for the function $\bar{h}_\varepsilon$ ($\varepsilon \neq 0$). A similar computation handles the case $\varepsilon = 0$.

(b) Let h be any function in the domain of G. Then the random function

$$M(t) = e^{\varepsilon t} h(X_t) + \int_0^t e^{\varepsilon s}((G-\varepsilon)h)(X_s)\, ds \tag{3.36}$$

is a martingale, so by optional stopping we have

$$h(x) = E_x[\exp(\varepsilon \tau_{A^c})\, h(X_{\tau_{A^c}})] + E_x \int_0^{\tau_{A^c}} e^{\varepsilon s}((G-\varepsilon)h)(X_s)\, ds \tag{3.37}$$

If h satisfies (3.29a) and (3.29c), the right-hand side is

$$\leqslant E_x \int_0^{\tau_{A^c}} c e^{\varepsilon s}\, ds \tag{3.38a}$$

$$= \bar{h}_\varepsilon(x) \quad \blacksquare \tag{3.38b}$$

Thus, for any fixed ε, the optimal choice of h in (3.27b)/(3.30b) is the mean-exponential hitting time function $\bar{h}_\varepsilon$. On the other hand, any function h satisfying (3.29a) and (3.29c) serves as a lower bound on the true hitting-time function $\bar{h}_\varepsilon$. We refer to such a function h as a *Liapunov function.*

The physical meaning of Propositions 3.3–3.5 may thus be summarized as follows: If, for a large (in π-measure) set of states $x \in A$, the mean (or mean-exponential) time to hit A^c is large, then the Dirichlet spectral gap m_A must be small; and if A^c is not too small (again in π-measure), then the mass gap m must itself be small.

The final step in our argument is to notice that given a Liapunov function h for one value of ε, we can form Liapunov functions for values $\varepsilon' \geqslant \varepsilon$ and then make an optimal choice of ε'. The key fact is the following:

Lemma 3.8. Let I be an interval of the real line, and let $\Phi\colon I \to \mathbf{R}$ be convex and once differentiable with a first derivative that is bounded and globally Lipschitz. Let $h \in \mathcal{Q}(\mathcal{G})$ with $\operatorname{Ran} h \subset I$. Then $\Phi \circ h \in \mathcal{Q}(\mathcal{G})$, $\Phi' \circ h \in \mathcal{Q}(\mathcal{G}) \cap L^\infty(\pi)$, and

$$\mathcal{G}(\phi, \Phi \circ h) \leqslant \mathcal{G}((\Phi' \circ h) \cdot \phi, h) \tag{3.39}$$

for all nonnegative $\phi \in \mathcal{Q}(\mathcal{G}) \cap L^\infty(\pi)$. If, in addition, $h \in \mathcal{D}(G)$ and $\Phi \circ h \in \mathcal{D}(G)$, then

$$G(\Phi \circ h) \leqslant (\Phi' \circ h) \cdot Gh \tag{3.40}$$

pointwise π-a.e. [We apologize for all the technical complications; the key inequality is (3.40).]

Proof. The hypotheses on Φ imply that $\Phi \circ h \in \mathscr{Q}(\mathscr{G})$, $\Phi' \circ h \in \mathscr{Q}(\mathscr{G}) \cap L^{\infty}(\pi)$, and hence that $(\Phi' \circ h) \cdot \phi \in \mathscr{Q}(\mathscr{G})$ [see Section 2.2, facts (d) and (f)]. Then

$$\begin{aligned}
&\int J(x, dy)\ [\Phi(h(x)) - \Phi(h(y))] \\
&\quad = r(x) \left[\Phi(h(x)) - \int \tilde{J}(x, dy)\, \Phi(h(y)) \right] \\
&\quad \leqslant r(x) \left[\Phi(h(x)) - \Phi\left(\int \tilde{J}(x, dy)\, h(y) \right) \right] \\
&\quad \leqslant r(x)\, \Phi'(h(x)) \left[h(x) - \int \tilde{J}(x, dy)\, h(y) \right] \\
&\quad = \Phi'(h(x)) \int J(x, dy)[h(x) - h(y)]
\end{aligned} \tag{3.41}$$

where we used first Jensen's inequality on the probability measure $\tilde{J}(x, \cdot) \equiv J(x, \cdot)/r(x)$ and then the convexity inequality $\Phi(a) - \Phi(b) \leqslant \Phi'(a)(a-b)$. Now multiply this by $\phi(x)$ and integrate with respect to $\pi(dx)$; this proves (3.39).

Now if (3.40) were false, we could take ϕ to be nonnegative, not identically zero, and supported on the set where (3.40) fails; there exist such functions in $\mathscr{Q}_1(\mathscr{G}) \cap L^{\infty}(\pi)$ and hence in $\mathscr{Q}(\mathscr{G}) \cap L^{\infty}(\pi)$. ∎

Remarks. 1. If $h \in \mathscr{D}_1(G)$, then $\Phi \circ h \in \mathscr{D}_1(G) \subset \mathscr{D}(G)$.

2. The inequality (3.40) is true quite generally for generators of Markov semigroups; see ref. 42.

Now define, for $\varepsilon' \geqslant \varepsilon \geqslant 0$ and $c, c' \geqslant 0$, the function $\Phi_{\varepsilon,\varepsilon'}: \mathbf{R}_+ \to \mathbf{R}_+$ by

$$\Phi_{\varepsilon,\varepsilon'}(z) = \begin{cases} \dfrac{c'}{\varepsilon'} \left[\left(1 + \dfrac{\varepsilon}{c} z \right)^{\varepsilon'/\varepsilon} - 1 \right] & \text{if} \quad \varepsilon' \geqslant \varepsilon > 0 \\ \dfrac{c'}{\varepsilon'} \left[\exp\left(\dfrac{\varepsilon'}{c} z \right) - 1 \right] & \text{if} \quad \varepsilon' > \varepsilon = 0 \\ \dfrac{c'}{c} z & \text{if} \quad \varepsilon' = \varepsilon = 0 \end{cases} \tag{3.42}$$

(Strictly speaking, we should call this function $\Phi_{\varepsilon,\varepsilon',c,c'}$, but for simplicity we suppress c, c' from the notation.) This function is convex and infinitely differentiable, and satisfies the differential equation

$$\frac{d\Phi_{\varepsilon,\varepsilon'}}{dz} = \frac{\varepsilon'\Phi_{\varepsilon,\varepsilon'} + c}{\varepsilon z + c} \tag{3.43}$$

with the initial condition $\Phi_{\varepsilon,\varepsilon'}(0) = 0$. Unfortunately, $\Phi_{\varepsilon,\varepsilon'}$ does not satisfy the technical hypotheses of Lemma 3.8, because its derivative is not globally bounded. So we define cutoff versions of $\Phi_{\varepsilon,\varepsilon'}$ which grow linearly for $z > M$:

$$\Phi_{\varepsilon,\varepsilon',M}(z) = \begin{cases} \Phi_{\varepsilon,\varepsilon'}(z) & \text{for} \quad 0 \leqslant z \leqslant M \\ \Phi_{\varepsilon,\varepsilon'}(M) + (z - M)\,\Phi'_{\varepsilon,\varepsilon'}(M) & \text{for} \quad z \geqslant M \end{cases} \tag{3.44}$$

The function $\Phi_{\varepsilon,\varepsilon',M}$ satisfies all the hypotheses of Lemma 3.8 as well as the differential inequality

$$\frac{d\Phi_{\varepsilon,\varepsilon',M}}{dz} \leqslant \frac{\varepsilon'\Phi_{\varepsilon,\varepsilon',M} + c}{\varepsilon z + c} \tag{3.45}$$

We therefore conclude:

Corollary 3.9. Let $\varepsilon' \geqslant \varepsilon \geqslant 0$ and $c, c', M \geqslant 0$. Let $h \in \mathscr{Q}(\mathscr{G})$ satisfy (see footnote 11)

$$(G - \varepsilon)h \leqslant c \qquad \text{on } A \tag{3.46a}$$

$$h \geqslant 0 \qquad \text{on } A \tag{3.46b}$$

$$h = 0 \qquad \text{on } A^c \tag{3.46c}$$

Then $\Phi_{\varepsilon,\varepsilon',M} \circ h \in \mathscr{Q}(\mathscr{G})$ and (see footnote 11)

$$(G - \varepsilon')(\Phi_{\varepsilon,\varepsilon',M} \circ h) \leqslant c' \qquad \text{on } A \tag{3.47a}$$

$$\Phi_{\varepsilon,\varepsilon',M} \circ h \geqslant 0 \qquad \text{on } A \tag{3.47b}$$

$$\Phi_{\varepsilon,\varepsilon',M} \circ h = 0 \qquad \text{on } A^c \tag{3.47c}$$

Corollary 3.9 has a nice probabilistic interpretation: Let h be the function $\bar{h}_\varepsilon$ defined in (3.35). Then Corollary 3.9 and Proposition 3.7(b) imply that

$$\bar{h}_{\varepsilon'}(x) \geqslant \Phi_{\varepsilon,\varepsilon'}(\bar{h}_\varepsilon(x)) = \lim_{M \to \infty} \Phi_{\varepsilon,\varepsilon',M}(\bar{h}_\varepsilon(x)) \tag{3.48}$$

But this is precisely the content of Jensen's inequality applied to hitting times:

$$\bar{h}_{\varepsilon'}(x) \equiv E_x(\Phi_{\varepsilon,\varepsilon'}(F_\varepsilon(\tau_{A^c}))) \geqslant \Phi_{\varepsilon,\varepsilon'}(E_x(F_\varepsilon(\tau_{A^c}))) \equiv \Phi_{\varepsilon,\varepsilon'}(\bar{h}_\varepsilon(x)) \tag{3.49}$$

where

$$F_\varepsilon(z) \equiv \begin{cases} (c/\varepsilon)(e^{\varepsilon z} - 1) & \text{if } \varepsilon \neq 0 \\ cz & \text{if } \varepsilon = 0 \end{cases} \tag{3.50}$$

Thus, Lemma 3.8 and Corollary 3.9 are the analytic expressions of the fact that G generates a Markov semigroup, so that the associated hitting times satisfy Jensen's inequality.

In summary: Given a function h satisfying the hypotheses of Proposition 3.4 for some particular value of ε, we can form functions $\Phi_{\varepsilon,\varepsilon',M} \circ h$ that satisfy the hypotheses of Proposition 3.4 for values $\varepsilon' \geqslant \varepsilon$. For each fixed ε' we first let $M \to \infty$ (using the monotone convergence theorem); we then optimize over ε'. The result is:

Proposition 3.10. Let $\varepsilon, c, c' \geqslant 0$ and let h be a function (not identically zero) in $\mathscr{D}(\mathscr{G})$ satisfying (3.46a)–(3.46c). Then

$$m_A \leqslant \inf_{\varepsilon' \geqslant \varepsilon} \left[\varepsilon' + c' \frac{\pi(A)^{1/2}}{\|\Phi_{\varepsilon,\varepsilon'} \circ h\|_{L^2(\pi)}} \right] \tag{3.51}$$

where

$$\|\Phi_{\varepsilon,\varepsilon'} \circ h\|_{L^2(\pi)} \equiv +\infty \qquad \text{if} \quad \Phi_{\varepsilon,\varepsilon'} \circ h \notin L^2(\pi)$$

Corollary 3.11. Let $\varepsilon \geqslant 0$ and let h be a function (not identically zero) in $\mathscr{D}(\mathscr{G})$ satisfying (3.46a)–(3.46c) for some $c < \infty$. Define

$$\varepsilon_{\max}(h) \equiv \sup\{\varepsilon' : \Phi_{\varepsilon,\varepsilon'} \circ h \in L^2(\pi)\} \tag{3.52}$$

Then

$$m_A \leqslant \varepsilon_{\max}(h) \tag{3.53}$$

Example. Random walk with inward drift on $\mathbf{Z}_+$, with elastic boundary conditions at 0. The transition matrix is given by (3.14), the invariant measure is $\pi(x) = \text{const} \times [p/(1-p)]^x$, and the exact mass gap is $m = 1 - 2p^{1/2}(1-p)^{1/2}$. Now fix $j \geqslant 0$, and let $A = \{x : x > j\}$. Then the Liapunov function $h(x) = (x-j)_+$ satisfies (3.46a)–(3.46c) with $\varepsilon = 0$ and $c = 1 - 2p$. Thus, by Corollary 3.11 it follows that

$$m_A \leqslant \varepsilon_{\max}(h) = \frac{1-2p}{2} \log \frac{1-p}{p}$$

Taking $j \to \infty$ and using Proposition 3.3, we conclude that

$$m \leqslant \frac{1-2p}{2} \log \frac{1-p}{p}$$

as well. Note that as $p \to 1/2$, both this upper bound and the exact mass gap behave as $2(\frac{1}{2}-p)^2 + O[(\frac{1}{2}-p)^4]$, so even the coefficient is sharp!

We conclude with a remark which is important in applications. Suppose that we have a function $\tilde{h}$ satisfying $(G-\varepsilon)\tilde{h} \leqslant c$ on A, *but which is not necessarily nonpositive on* A^c. We then let k be the function

$$k(x) = E_x[\exp(\varepsilon\tau_{A^c})\,\tilde{h}_+(X_{\tau_{A^c}})] \tag{3.54}$$

which solves the boundary-value problem

$$(G-\varepsilon)k = 0 \qquad \text{on } A \tag{3.55a}$$

$$k = \tilde{h}_+ \qquad \text{on } A^c \tag{3.55b}$$

Then $h \equiv \tilde{h} - k$ satisfies all the hypotheses of Proposition 3.5, and so can be used as a Liapunov function. We refer to $\tilde{h}$ as a *pre-Liapunov function*. It is then necessary to obtain upper bounds on k. If $m_A > \varepsilon$ and $\chi_{A^c}\tilde{h}_+ \in \mathscr{D}(G)$, then k can be written as

$$k = (G_A - \varepsilon)^{-1} I_A G(\chi_{A^c}\tilde{h}_+) \tag{3.56}$$

where I_A is the operator of multiplication by χ_A; in particular,

$$\|k\|_{L^2(\pi)} \leqslant (m_A - \varepsilon)^{-1} \|I_A G(\chi_{A^c}\tilde{h}_+)\|_{L^2(\pi)} \tag{3.57}$$

4. STOCHASTIC CONTOUR MODELS

In Section 4.1 we define our stochastic contour models and verify that the continuous-time dynamics is well-defined (i.e., without "explosion"). In Section 4.2 we show that if the jump rates grow sufficiently slowly, then *the mass gap is zero*. In Section 4.3 we compare our results with the heuristic predictions of Huse and Fisher.[10]

4.1. Definitions

Let X be the space of all simple closed contours γ (of arbitrary length) in $\mathbf{Z}^2 \subset \mathbf{R}^2$ which bound an area containing an origin fixed at $(\frac{1}{2}, \frac{1}{2}) \in \mathbf{R}^2$. For convenience, we assume that X also contains the null contour $\gamma = \varnothing$. Thus, a nonnull contour γ is made up of unit-length segments between integer lattice points; and since γ is simple, a lattice point on γ has precisely two such segments hitting it. We denote by $|\gamma|$ the length of γ. Let a_n be the number of contours of length n; then it is not hard to show[49] that

$$\mu \equiv \lim_{\substack{n \to \infty \\ n \text{ even}}} a_n^{1/n} \tag{4.1}$$

exists and lies between 2 and 3 (in fact, it equals[49] the connective constant for self-avoiding walks on $\mathbf{Z}^2$).

Now, for real β, define the *partition function*

$$Z(\beta) = \sum_{\gamma \in X} e^{-\beta|\gamma|} \tag{4.2}$$

Clearly, $Z(\beta)$ is finite for all β greater than the *critical point*

$$\beta_c \equiv \log \mu \tag{4.3}$$

(it is possibly finite also at β_c). Whenever $Z(\beta)$ is finite, define a probability measure π_β on X by

$$\pi_\beta(\gamma) = Z(\beta)^{-1} e^{-\beta|\gamma|} \tag{4.4}$$

From now on we fix $\beta > \beta_c$.

The next step is to define a reversible Markovian jump process with state space X, having π_β as its (unique) invariant measure. The jump rates $j(\gamma, \gamma')$ are assumed to satisfy the following two conditions:

(a) *Local motion condition.* $j(\gamma, \gamma') = 0$ unless $\gamma \triangle \gamma'$ is the perimeter of a unit lattice square.

(b) *Detailed balance condition.* For all $\gamma, \gamma' \in X$,

$$\pi_\beta(\gamma)\, j(\gamma, \gamma') = \pi_\beta(\gamma')\, j(\gamma', \gamma) \tag{4.5}$$

Provided that the jump rates do not grow too fast as $|\gamma| \to \infty$, this stochastic process is well-defined:

Proposition 4.1. Assume that there exist constants $C < \infty$ and $\varepsilon > 0$ such that

$$j(\gamma, \gamma') \leqslant C \exp[(\beta - \beta_c - \varepsilon)|\gamma|] \tag{4.6}$$

for all γ, γ'. Then there is no explosion.

Proof. By the local motion condition, the number of contours γ' that are accessible from a given contour γ grows at most linearly in $|\gamma|$. The claim is then an immediate consequence of the last sentence of Proposition 2.2. ∎

Under the hypothesis of Proposition 4.1 (which we assume henceforth), there is a well-defined reversible Markovian jump process, with invariant measure π_β, whose generator G on the space $L^2(\pi_\beta)$ is given by the quadratic form

$$(f, Gf) = \frac{1}{2} \sum_{\gamma, \gamma'} \pi_\beta(\gamma)\, j(\gamma, \gamma')\, |f(\gamma') - f(\gamma)|^2 \tag{4.7}$$

4.2. Absence of Mass Gap

Our main result is that if the total jump rates $r(\gamma) \equiv j(\gamma, \gamma')$ grow slower than linearly in the length of the contour as $|\gamma| \to \infty$, then G has zero mass gap. More precisely:

Theorem 4.2. Let $\beta > \beta_c$, and assume that the jump rates satisfy

$$\lim_{n \to \infty} \frac{1}{n} \sup_{|\gamma| = n} \sum_{\gamma'} j(\gamma, \gamma') = 0 \tag{4.8}$$

Then G has zero mass gap, i.e.,

$$m \equiv \inf \operatorname{spec}(G \upharpoonright \mathbf{1}^{\perp}) = 0 \tag{4.9}$$

We note that this theorem has an immediate application to certain discrete-time Markov chains analogous to the continuous-time jump process defined above. Suppose that the jump rates for G satisfy

$$r(\gamma) \equiv \sum_{\gamma'} j(\gamma, \gamma') \leqslant K_0 \tag{4.10}$$

for some constant K_0 independent of γ. Then we can define a reversible discrete-time Markov chain with transition matrix

$$P = I - \frac{1}{K} G \tag{4.11}$$

for any constant $K \geqslant K_0$. Since (4.10) implies (4.8) (it is roughly one power of n *stronger*), we conclude that:

Corollary 4.3. Let $\beta > \beta_c$, and assume that the jump rates satisfy (4.10). Then G has zero mass gap. Moreover, the discrete-time transition probability P defined by (4.11) also has zero mass gap, i.e.,

$$\sup \operatorname{spec}(P \upharpoonright \mathbf{1}^{\perp}) = 1 \tag{4.12}$$

To give the flavor of our methods, we first prove Corollary 4.3 (this proof is considerably simpler than that of Theorem 4.2). The main tool is the minimum hitting-time argument, Theorem 3.1. Let A_n be a one-element set consisting of a single square contour γ_n of side n (so that $|\gamma_n| = 4n$), and let B be the one-element set consisting of the null contour $\varnothing$. Because of the local motion condition, the area $\mathscr{A}(\gamma)$ bounded by the contour γ can change by at most one unit in each jump; so at least n^2 jumps are required to pass from γ_n to the null contour. Thus, by Theorem 3.1 we have

$$\begin{aligned} m &\leqslant \inf_n \frac{1}{8} K_0 \frac{1}{(n^2-1)^2} \log^2\left(\frac{4Z(\beta)^2}{\exp(-\beta|\gamma_n|)}\right) \\ &= \inf_n \frac{1}{8} K_0 \frac{1}{(n^2-1)^2} \log^2[4Z(\beta)^2 e^{4\beta n}] \\ &= 0 \end{aligned} \tag{4.13}$$

The point is that it takes a minimum of n^2 jumps in order to get from γ_n to $\varnothing$ (or vice versa), but only a time of order $\log \pi(\gamma_n) \sim n$ is "justified" by the rarity of γ_n in the invariant measure π. This simple geometric fact is at heart of the absence of mass gap for this class of stochastic contour models.

We now turn to the proof of Theorem 4.2, which is considerably deeper than Corollary 4.3. The main tool is the mean hitting-time argument (Proposition 3.10).

The pre-Liapunov function $\tilde{h}(\gamma)$ will be taken to be essentially the area $\mathscr{A}(\gamma)$, as suggested by our proof of Corollary 4.3. However, since the total rates $r(\gamma) \equiv \sum_{\gamma'} j(\gamma, \gamma')$ may be unbounded as $|\gamma| \to \infty$, there arises the possibility of a fast "back road" from γ_n to a short contour, by passing through a set of very long contours with not so large area. We will rule out this possibility by showing that such transitions are rare.

Let γ_n be a square contour of side n [so that $|\gamma_n| = 4n$ and $\mathscr{A}(\gamma_n) = n^2$]. Next, for fixed $\delta > 0$ (its value will be determined below), define

$$S_n = \{\gamma : \mathscr{A}(\gamma) \leqslant \tfrac{1}{4} n^2\} \tag{4.14a}$$

$$T_n = \{\gamma : |\gamma| \geqslant 4(1+\delta) n\} \tag{4.14b}$$

$$A_n = X \backslash (S_n \cup T_n) \tag{4.14c}$$

Clearly, $\gamma_n \in A_n$. Recall that G_{A_n} is the operator on $L^2(A_n, \pi)$ obtained from G by imposing Dirichlet boundary conditions on the complement of A_n, and that $\tau_{A_n^c}$ is the (random) time of first escape from A_n. We shall write $m_n \equiv m_{A_n} \equiv \inf \operatorname{spec} G_{A_n}$. Finally, let us define

$$r_n \equiv \sup_{|\gamma| = n} \sum_{\gamma'} j(\gamma, \gamma') \tag{4.15}$$

The crucial estimate is the following lower bound on the expected escape time for $\gamma(t)$ to leave A_n:

Lemma 4.4. Let $\beta > \beta_c$, and assume that the jump rates satisfy (4.8). Then, for δ and n sufficiently large (how large depends on β), the escape time $\tau_{A_n^c}$ for $\gamma(t)$ to leave A_n satisfies

$$E_{\gamma_n}(\tau_{A_n^c}) \geqslant \frac{1}{nc_n}\left(\frac{3}{4}n^2 - \frac{c}{m_n}C_1^n\right) \tag{4.16}$$

where $c_n \to 0$ as $n \to \infty$ (c_n depends on β and on the rates j), and $0 < C_1 < 1$ (c and C_1 depend only on β).

Proof. Let $\gamma \in A_n$; then the hypothesis (4.8), the definition of A_n, and the local motion condition together imply that

$$\begin{aligned} |(G\mathscr{A})(\gamma)| &\leqslant \sum_{\gamma'} j(\gamma, \gamma') |\mathscr{A}(\gamma') - \mathscr{A}(\gamma)| \\ &= \sum_{\gamma'} j(\gamma, \gamma') \\ &\leqslant |\gamma| \frac{r_{|\gamma|}}{|\gamma|} \\ &\leqslant (1+\delta)\, nc_n \end{aligned} \tag{4.17}$$

for some constant c_n that tends to zero as $n \to \infty$, where in the last step we have used that $2n < |\gamma| < 4(1+\delta)n$ for $\gamma \in A_n$. {This is a consequence of the isoperimetric inequality for lattice contours, $|\gamma| \geqslant 4\mathscr{A}(\gamma)^{1/2}$. Alternatively, one could use the usual isoperimetric inequality, $|\gamma| \geqslant [4\pi\mathscr{A}(\gamma)]^{1/2}$, with only a slight weakening of the constants.}

So we use the pre-Liapunov function $\tilde{h}(\gamma) = \mathscr{A}(\gamma) - \frac{1}{4}n^2$. Clearly, $\tilde{h} \geqslant 0$ on A_n. Note that $\tilde{h} \leqslant 0$ on S_n, but not on T_n. Therefore, in accordance with the discussion at the end of Section 3.2, we need to estimate the function

$$k_n(\gamma) \equiv E_\gamma[\tilde{h}_+(\gamma(\tau_{A^c}))] = E_\gamma[(\chi_{T_n}\tilde{h}_+)(\gamma(\tau_{A^c}))] \tag{4.18}$$

which solves the boundary-value problem

$$Gk_n = 0 \qquad \text{on } A_n \tag{4.19a}$$

$$k_n = \tilde{h}_+ = \chi_{T_n}\tilde{h}_+ \qquad \text{on } A_n^c \tag{4.19b}$$

Provided that $m_n > 0$, we can write this solution as

$$k_n = (G_{A_n})^{-1} I_{A_n} G I_{T_n} \tilde{h}_+ \tag{4.20}$$

where

$$(I_{A_n} G I_{T_n} \tilde{h}_+)(\gamma) = \sum_{\gamma' \in T_n} j(\gamma, \gamma')[\mathscr{A}(\gamma') - \tfrac{1}{4}n^2]_+$$
$$\leqslant \sum_{\gamma' \in T_n} j(\gamma, \gamma')(1+\delta)^2 n(n+1) \tag{4.21}$$

for $\gamma \in A_n$, where we have used the isoperimetric inequality $\mathscr{A}(\gamma') \leqslant |\gamma'|^2/16$ together with the fact that $|\gamma'| < 4(1+\delta)n + 2$ if $\gamma \in A_n$ and $j(\gamma, \gamma') \neq 0$. Clearly,

$$(I_{A_n} G I_{T_n} \tilde{h}_+)(\gamma) \leqslant (1+\delta)^2 n^2 (n+1) c_n \equiv (1+\delta)^2 n^3 c'_n$$

for all γ, and

$$(I_{A_n} G I_{T_n} \tilde{h}_+)(\gamma) = 0$$

unless $4(1+\delta)n - 2 \leqslant |\gamma| < 4(1+\delta)n$. Moreover, since $\beta > \beta_c$, we know that there exists a constant $0 < C_0 < 1$ (depending on β) such that

$$\pi_\beta(\{\gamma : |\gamma| = n\}) \leqslant C_0^n \tag{4.22}$$

for n sufficiently large. It follows that

$$\|I_{A_n} G I_{T_n} \tilde{h}_+\|^2_{L^2(A_n, \pi)}$$
$$\leqslant (1+\delta)^4 n^6 c_n'^2 \pi_\beta(\{\gamma \in A_n : 4(1+\delta)n - 2 \leqslant |\gamma| < 4(1+\delta)n\})$$
$$\leqslant (1+\delta)^4 n^6 c_n''^2 C_0^{4(1+\delta)n} \tag{4.23}$$

for n sufficiently large. Combining (4.23) with (4.20) and the Schwarz inequality, we get

$$k_n(\gamma_n) \equiv \pi_\beta(\gamma_n)^{-1} (\delta_{\gamma_n}, (G_{A_n})^{-1} I_{A_n} G I_{T_n} \tilde{h}_+)_{L^2(A_n, \pi)}$$
$$\leqslant (1+\delta)^2 \frac{n^3 c_n''}{m_n} \pi_\beta(\gamma_n)^{-1/2} C_0^{2(1+\delta)n}$$
$$\leqslant (1+\delta)^2 \frac{n^3 c_n''}{m_n} Z(\beta)^{1/2} C_0^{2(1+\delta)n} e^{2\beta n} \tag{4.24}$$

We now pick δ large enough so that

$$C_0^{1+\delta} e^\beta < 1 \tag{4.25}$$

Then, by enlarging δ slightly, we can summarize (4.24) as

$$k_n(\gamma_n) \leqslant \frac{c}{m_n} C_1^n \tag{4.26}$$

for n sufficiently large, where c and C_1 are independent of n (but of course dependent on β) and $0 < C_1 < 1$. It follows that the Liapunov function $h \equiv \tilde{h} - k_n$ satisfies

$$Gh \leqslant (1+\delta)\, nc_n \qquad \text{on } A_n \tag{4.27a}$$

$$h \leqslant 0 \qquad \text{on } A_n^c \tag{4.27b}$$

$$h(\gamma_n) \geqslant \frac{3}{4} n^2 - \frac{c}{m_n} C_1^n \tag{4.27c}$$

Then Proposition 3.7(b) yields (4.16). ∎

Proof of Theorem 4.2. We use the function $h(\gamma) = \bar{h}_0(\gamma) \equiv E_\gamma(\tau_{A_n^c})$ in Proposition 3.10. Clearly

$$\begin{aligned} \|\Phi_{0,\varepsilon'} \circ h\|_{L^2(\pi)} &\geqslant \pi(\gamma_n)^{1/2} \Phi_{0,\varepsilon'}(h(\gamma_n)) \\ &= Z(\beta)^{-1/2}[\exp(-2\beta n)] \frac{1}{\varepsilon'} \{\exp[\varepsilon' h(\gamma_n)] - 1\} \end{aligned} \tag{4.28}$$

so that Proposition 3.10 yields

$$\begin{aligned} m_n &\leqslant \inf_{\varepsilon' \geqslant 0} \varepsilon' \left(1 + Z(\beta)^{1/2}[\exp(2\beta n)]\{\exp[\varepsilon' h(\gamma_n)] - 1\}^{-1}\right) \\ &\leqslant \frac{2\beta n}{h(\gamma_n)} C(\beta) \end{aligned} \tag{4.29}$$

by taking $\varepsilon' = 2\beta n / h(\gamma_n)$. Thus, by Lemma 4.4,

$$m_n \leqslant \text{const}(\beta) \times c_n \left(\frac{3}{4} - \frac{c}{m_n} C_1^n n^{-2}\right)^{-1} \tag{4.30}$$

for n sufficiently large. Solving for m_n, we get

$$\begin{aligned} m_n &\leqslant \text{const}(\beta) \times (c_n + C_1^n n^{-2}) \\ &\to 0 \qquad \text{as } n \to \infty \end{aligned} \tag{4.31}$$

On the other hand, by Proposition 3.3,

$$m \leqslant \frac{m_n}{\pi(A_n^c)} \leqslant \frac{m_n}{\pi(\varnothing)} = Z(\beta) m_n \tag{4.32}$$

for all n. Hence $m = 0$. ∎

Remarks. 1. Consider the case in which the jump rates satisfy a bound $j(\gamma, \gamma') \leqslant C$ (uniformly in β), so that the total rates $r(\gamma)$ grow at most linearly in $|\gamma|$. Then, imitating the above analysis leads to a mass gap estimate $m \leqslant \text{const} \times \beta$. In fact, a much more careful analysis exploiting cancellations in (4.17) leads to an estimate $m \leqslant \text{const} \times \beta e^{-4\beta}$. *We conjecture that the mass gap is identically zero*, at least for β sufficiently large, but we are at present unable to prove this.

2. Numerous variants on our model can be treated by similar methods. For example:

(a) One can allow an arbitrary finite repertoire of local moves. This means that the local motion condition is replaced by: There exists a finite set $\mathscr{R}$ of closed contours such that $j(\gamma, \gamma') = 0$ unless $\gamma \triangle \gamma'$ is a translate of a member of $\mathscr{R}$.

(b) One can consider Peierls contours in $\mathbf{Z}^d$, $d > 2$. These contours would be closed surfaces in $d = 3$, closed hypersurfaces in $d = 4$, etc. The function $\mathscr{A}(\gamma)$ now denotes the *volume* in $\mathbf{Z}^d$ enclosed by γ.

(c) One can consider the BFACF[11–14,7] dynamics for self-avoiding walks with fixed endpoints x, y in $\mathbf{Z}^d$ for any dimension $d \geqslant 2$. Now $\mathscr{A}(\gamma)$ denotes the *minimal area* of a surface bounded by $\gamma \circ \gamma^*$, where γ^* is some fixed path from y to x. The primary application of the BFACF dynamics is as a discrete-time Monte Carlo algorithm: in this case the total rates $r(\gamma)$ are bounded by 1, and the absence of mass gap follows from Corollary 4.3.

(d) By similar methods one can consider the Sterling–Greensite[15–20] dynamics for self-avoiding surfaces with fixed boundary in any dimension $d \geqslant 3$.

4.3. Discussion

Huse and Fisher[10] argue that in the Glauber dynamics[8,9] for the Ising model at low temperature, the temporal autocorrelation function of a single spin

$$C_i(t) \equiv \langle \sigma_i(0)\, \sigma_i(t) \rangle - \langle \sigma_i \rangle^2 \tag{4.33}$$

(taken in a pure phase, e.g., the + phase) should exhibit the following asymptotic behavior as $t \to \infty$:

$$C_i(t) \sim \begin{cases} \exp[-(t/\tau)^{(d-1)/2}], & 1 < d < 3 \\ t^{-p} \exp(-t/\tau), & d = 3 \\ \exp(-t/\tau), & d > 3 \end{cases} \tag{4.34}$$

In particular, for $1<d<3$ they predict that the low-temperature Glauber dynamics has zero mass gap.

Their argument is based on a consideration of droplet fluctuations that contribute to $C_i(t)$. They first assume that the temperature is low enough so that droplets are dilute and hence noninteracting. This amounts to approximating the Glauber dynamics (which can be thought of as a dynamics for a gas of interacting Peierls contours) by a single-contour dynamics of essentially the form considered in this paper, with jump rates $j(\gamma, \gamma')=1$ for $\mathscr{A}(\gamma \triangle \gamma')=1$ [hence total jump rates $r(\gamma) \sim |\gamma|$]. The Huse–Fisher argument for the single-contour model then goes as follows:

Suppose that initially the contour is very large ($|\gamma| \gg$ the mean value in the equilibrium distribution π_β) and roughly spherical. Then its radius r is claimed to evolve roughly according to the Langevin equation

$$\frac{dr}{dt} \approx -\frac{\Gamma}{r} + \frac{\eta(t)}{r^{(d-1)/2}} \tag{4.35}$$

where Γ is a constant and $\eta(t)$ is white noise. The first term is the deterministic motion of the contour in response to its own surface tension, as discussed some years ago by Lifshitz[50]; the second term is the stochastic part of the evolution, averaged over the contour. Let us consider first only the deterministic evolution

$$dr/dt \approx -\Gamma/r \tag{4.36}$$

Under this evolution, a contour of initial radius r_0 will survive for a time of order r_0^2 before becoming "small" (or empty). Otherwise put, only contours of initial radius greater than about $t^{1/2}$ will survive for a time t. On the other hand, the equilibrium probability of a contour of radius r_0 is $\sim \exp(-cr_0^{d-1})$. Thus, the autocorrelation function

$$C_{FF}(t) \equiv \langle F(0)\, F(t)\rangle - \langle F\rangle^2 \tag{4.37}$$

for the observable $F(\gamma)=\chi(\gamma \neq \varnothing)$ (or any similar observable) should receive a contribution of order $\exp[-\text{const} \times t^{(d-1)/2}]$ from large contours.[12] If $d<3$, this contribution is larger (as $t \to \infty$) than the usual exponentially decaying fluctuations; assuming that this is the *largest* contribution, we obtain (4.34a). On the other hand, if $d>3$, the dominant contributions come from small contours and are pure exponentials. A more detailed analysis of (4.35), taking into account the stochastic term, shows that the results are unchanged except possibly in $d=3$, where the exponential decay may be modified by multiplicative power-law corrections.

[12] It would be interesting to prove general theorems about Markov chains and processes that make rigorous the type of reasoning underlying this last argument.

Our proof of Theorem 4.2, by contrast, is based on an argument of the form

$$\left| \frac{d}{dt} (\text{“volume” enclosed by } \gamma) \right| \leqslant \text{const} \times (\text{“surface area” of } \gamma) \quad (4.38)$$

Assuming spherical symmetry, this becomes

$$\left| \frac{d}{dt} (r^d) \right| \leqslant \text{const} \times r^{d-1} \quad (4.39)$$

or

$$|dr/dt| \leqslant \text{const} \quad (4.40)$$

which is one power of r weaker than (4.36). It follows from (4.40) that a contour of initial radius r_0 will survive for a time *at least* of order r_0 before becoming small. Since the equilibrium probability of such a contour is of order $\exp(-cr_0^{d-1})$, the mean hitting-time argument (or the Huse–Fisher heuristic argument on contributions to the autocorrelation function) implies the absence of mass gap for $d<2$. This should be compared with the absence of mass gap for $d<3$ obtained from (4.36).

Our bounds are too crude for two reasons:

1. In bounding $(G\mathscr{A})(\gamma)$ [first step of (4.17)], we replaced $\Delta\mathscr{A} \equiv \mathscr{A}(\gamma') - \mathscr{A}(\gamma)$ by its *absolute value*—this is, we made the worst-case pretense that all transitions are in the same direction. In fact, for most configurations the $\Delta\mathscr{A} = +1$ and $\Delta\mathscr{A} = -1$ transitions have almost equal probabilities; the difference is expected to be of order $1/r$, where r is a typical "radius of curvature" of the contour γ. So we need a more delicate estimation of $(G\mathscr{A})(\gamma)$.

2. Our present version of the mean hitting-time argument requires that $(Gh)(\gamma) \leqslant c$ for *all* $\gamma \in A_n$. On the other hand, for the conclusion to hold, it presumably suffices to have $Gh \leqslant c$ on A_n in some *average* sense. Moreover, this latter is probably the true behavior, i.e., there presumably do exist "rare" configurations $\gamma \in A_n$ for which $(Gh)(\gamma)$ is much larger than it is "on the average." So we need a more flexible version of the mean hitting-time argument.

We believe that with these two improvements, our methods could yield a rigorous proof of the Huse–Fisher conjectures for the single-contour model. A proof for the full Glauber dynamics (at very low temperature) would involve considerably more technical complication (to show that interactions between contours are irrelevant), but possibly not any additional physical ideas.

Finally, we mention the interesting Monte Carlo work of Ogielski,[51] which generally confirms the Huse–Fisher conjectures, but also indicates that the asymptotic behavior (4.34) is reached only at unbelievably long times, and would be unobservable in any conceivable experiment; the observable preasymptotic decay appears to be even *slower* than (4.34), i.e., $\exp[-(t/\tau)^p]$ with $p<(d-1)/2$.

ACKNOWLEDGMENTS

We thank David Brydges for bringing us together to begin this collaboration (December 1985). Also, we thank S. R. S. Varadhan for essential discussions regarding Section 2 (though he is not to blame for the errors that undoubtedly remain!). Finally, we thank Jean Bricmont and Joel Lebowitz for pointing out the work of Huse and Fisher (ref. 10).

This work was supported in part by NSF grants PHY-8413569 and DMS-8705599.

NOTE ADDED IN PROOF

Takano *et al.*[52] have also predicted a stretched-exponential decay for the spin-autocorrelation function in the low-temperature Ising model with Glauber dynamics, but their predicted exponent differs from that of Huse and Fisher.[10]

REFERENCES

1. A. D. Sokal and L. E. Thomas, Exponential convergence to equilibrium for a class of random-walk models, *J. Stat. Phys.*, submitted.
2. G. F. Lawler and A. D. Sokal, Bounds on the L^2 Spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality, *Trans. Am. Math. Soc.*, to appear.
3. A. Berretti and A. D. Sokal, *J. Stat. Phys.* **40**:483 (1985).
4. N. Madras and A. D. Sokal, *J. Stat. Phys.* **47**:573 (1987).
5. H. J. Hilhorst and J. M. Deutch, *J. Chem. Phys.* **63**:5153 (1975).
6. H. Boots and J. M. Deutch, *J. Chem. Phys.* **67**:4608 (1977).
7. S. Caracciolo and A. D. Sokal, *J. Phys. A* **19**:L797 (1986).
8. Yu. Ya. Gotlib, *Fiz. Tverd. Tela* **3**:2170 (1961) [*Sov. Phys. Solid State* **3**:1574 (1962)].
9. R. J. Glauber, *J. Math. Phys.* **4**:294 (1963).
10. D. A. Huse and D. S. Fisher, *Phys. Rev. B* **35**:6841 (1987).
11. J. Fröhlich, in *Recent Developments in Gauge Theories*, G. 'tHooft *et al.*, eds. (Plenum Press, New York, 1980), p. 65.
12. B. Berg and D. Foerster, *Phys. Lett.* **106B**:323 (1981).
13. C. Aragão de Carvalho and S. Caracciolo, *J. Phys. (Paris)* **44**:323 (1983).
14. C. Aragão de Carvalho, S. Caracciolo, and J. Fröhlich, *Nucl. Phys. B* **215**[FS7]:209 (1983).

15. T. Sterling and J. Greensite, *Phys. Lett.* **121B**:345 (1983).
16. B. Berg and A. Billoire, *Phys. Lett.* **139B**:297 (1984).
17. B. Berg, A. Billoire, and D. Foerster, *Nucl. Phys. B* **251**[FS13]:665 (1985).
18. B. Baumann and B. Berg, *Phys. Lett.* **164B**:131 (1985).
19. U. Glaus and T. L. Einstein, *J. Phys. A* **20**:L105 (1987).
20. U. Glaus, *J. Stat. Phys.* **50**:1141 (1988).
21. E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators* (Cambridge University Press, Cambridge, 1984).
22. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, 2nd ed. (Springer, New York, 1967).
23. L. Breiman, *Probability* (Addison-Wesley, Reading, Massachusetts, 1968), Chapter 15.
24. E. B. Dynkin, *Markov Processes*, 2 vols. (Springer, Berlin, 1965).
25. R. M. Blumenthal and R. K. Getoor, *Markov Processes and Potential Theory* (Academic Press, New York, 1968).
26. I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes* (Saunders, Philadelphia, 1969).
27. D. Freedman, *Markov Chains* (Holden-Day, San Francisco, 1971).
28. A. D. Sokal and L. E. Thomas, Lower bounds on the autocorrelation time of a reversible Markov chain, with applications to statistical mechanics, in preparation.
29. J. Neveu, *Bases Mathématiques du Calcul des Probabilités*, 2nd ed. (Masson, Paris, 1980) [English translation: *Mathematical Foundations of the Calculus of Probability* (Holden-Day, San Francisco, 1965)].
30. M. Fukushima, *Dirichlet Forms and Markov Processes* (North-Holland, Amsterdam, 1980).
31. T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. (Springer-Verlag, Berlin, 1976).
32. E. B. Dynkin, *Teor. Veroya Prim.* **3**:41 (1958) [*Theor. Prob. Appl.* **3**:38 (1958)].
33. I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes II* (Springer-Verlag, New York, 1975).
34. J. Neveu, *Discrete-Parameter Martingales* (North-Holland, Amsterdam, 1975).
35. S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence* (Wiley, New York, 1986).
36. D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes* (Springer-Verlag, Berlin, 1979).
37. C. Costantini, A. Gerardi, and G. Nappo, *Stat. Prob. Lett.* **1**:155 (1983).
38. J. L. Doob, *Trans. Am. Math. Soc.* **58**:455 (1945).
39. W. Feller, *Ann. Math.* **65**:527 (1957).
40. G. E. H. Reuter, *Acta Math.* **97**:1 (1957); *J. Lond. Math. Soc.* **34**:81 (1959).
41. D. Williams, *Z. Wahrsch. Verw. Gebiete* **3**:227 (1964).
42. P. Echeverría, *Z. Wahrsch. Verw. Gebiete* **61**:1 (1982).
43. M. Fukushima and D. Stroock, in *Probability, Statistical Mechanics, and Number Theory* (Academic Press, New York, 1986).
44. T. J. Rivlin, *The Chebyshev Polynomials* (Wiley, New York, 1974).
45. W. G. Sullivan, *Z. Wahrsch. Verw. Gebiete* **67**:387 (1984).
46. J.-F. Maitre and F. Musy, *SIAM J. Numer. Anal.* **21**:657 (1984).
47. R. Carmona and A. Klein, *Ann. Prob.* **11**:648 (1983).
48. L. E. Thomas and Zhong Yin, *J. Math. Phys.* **27**:2475 (1986).
49. J. M. Hammersley, *Proc. Camb. Phil. Soc.* **57**:516 (1961).
50. I. M. Lifshitz, *Zh. Eksp. Teor. Fiz.* **42**:1354 (1962) [*Sov. Phys. JETP* **15**:939 (1962)].
51. A. T. Ogielski, *Phys. Rev. B* **36**:7315 (1987).
52. H. Takano, H. Nakanishi and S. Miyashita, *Phys. Rev. B* **37**:3716 (1988).