On the Swapping Algorithm

Neal Madras, Zhongrong Zheng
Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada; e-mail: madras@mathstat.yorku.ca

Received 13 July 2001; accepted 18 May 2002

DOI 10.1002/rsa.10066

ABSTRACT: The Metropolis-coupled Markov chain method (or "Swapping Algorithm") is an empirically successful hybrid Monte Carlo algorithm. It alternates between standard transitions on parallel versions of the system at different parameter values, and swapping two versions. We prove rapid mixing for two bimodal examples, including the mean-field Ising model. © 2002 Wiley Periodicals, Inc. Random Struct. Alg., 22: 66–97, 2003

Keywords: Markov chain Monte Carlo; spectral gap; Metropolis-coupled Markov chains; mean-field Ising model; decomposition

1. INTRODUCTION

Markov chain Monte Carlo (MCMC) methods have become a standard tool for simulating observations from complicated distributions in a variety of fields of application, particularly physics (Binder and Heermann [1], Frenkel and Smit [12]) and statistics (Gilks, Richardson, and Spiegelhalter [15]). However, standard MCMC methods encounter serious difficulties if there are isolated modes in the target distribution from which we wish to sample. This is because the Markov chains get stuck in some local modes and rarely move between those isolated modes, and hence the rate at which the Markov chains converge to the target distribution is too slow to be practical. Isolated modes arise fairly often in statistical problems, and they are especially common in physics models that

Correspondence to: N. Madras

exhibit phase transitions. One of the classical examples of this is the Ising model at low temperature, which we shall describe below.

To speed up the convergence rates of MCMC methods when sampling from multimodal distributions, several new approaches have been proposed. Among these, the following three algorithms are based on similar intuition: the swapping algorithm of Geyer [13], also called Metropolis-coupled Markov chains or parallel tempering (Orlandini [24]); the simulated tempering algorithm (Marinari and Parisi [21], Geyer and Thompson [14], Madras [18]); and the tempered transition method (Neal [23]).

In practical use, these new methods seem to converge much faster than do standard methods, but there is little rigorous theory to back these empirical observations. In this paper, we shall focus on the swapping algorithm and rigorously show that it can be used to sample from two examples of bimodal distributions. The examples are the following:

Example I (Distribution with exponential valley on an interval). *Fix a real constant* $C > 1$, *and let* J *be a* (*large*) *positive integer. Consider the state space of all integers in the interval* $[-J, +J]$, *with the bimodal distribution*

$$\pi(x) = \frac{C^{|x|}}{Z} \qquad (x = -J, -J+1, \ldots, J),$$

where Z *is the normalizing constant. Observe that as* J *gets large* (*with* C *fixed*), $\pi(0)$ *is exponentially smaller than* $\pi(J)$. *The analysis turns out to be simpler if we can divide the state space into two exact halves. For this reason, we shall consider the more unusual state space* $\mathcal{A} = \mathcal{A}^{\mathrm{Int}}$, *which is defined to be the set of all odd integers in the interval* $[-2M-1, 2M+1]$, *with the distribution*

$$\pi(x) \equiv \pi^{\mathrm{Int}}(x|C) = \frac{C^{|x|}}{Z} \qquad \text{for } x \in \mathcal{A}.$$

(*See Remark* 3 *at the end of Section* 4 *for a brief discussion of how one can handle the state space* $\{-J, -J+1, \ldots, J\}$.)

Example II (Mean field Ising model). *Fix a real constant* $\beta > 0$ *and let* M *be a* (*large*) *positive integer. The state space is defined to be* $\mathcal{A} = \mathcal{A}^{\mathrm{Ising}} = \{-1, +1\}^M$. *The mean field Ising model* (*in zero field*) *is the probability distribution*

$$\pi(x) \equiv \pi^{\mathrm{Ising}}(x|\beta) = \frac{e^{\beta(\sum_{j=1}^{M} x[j])^2/2M}}{Z(\beta)} \qquad \text{for } x=(x[1], \ldots, x[M]) \in \mathcal{A},$$

where $Z(\beta)$ *is the normalizing constant. It can be shown that if* $\beta > 1$, *then the distribution of the "total spin"* $\sum_{j=1}^{M} x[j]$ *is bimodal for large* M, *and the probability at the modes is exponentially larger than it is at zero* (*e.g., Madras and Piccioni* [19]). *For simplicity, as in Example I, we assume throughout this paper that* M *is an odd integer.* (*See Remark* 3 *at the end of Section* 4 *for comments on the case of* M *even.*)

Both examples can be expressed in the form of a Gibbs distribution in statistical mechanics, that is,

$$\pi(x) = \frac{e^{-\beta H(x)}}{Z}, \tag{1}$$

where H is an explicit function (corresponding to the energy of configuration x) and β is a nonnegative parameter (inversely proportional to the temperature). In Example I, $H(x) = -|x|$ and $\beta = \log C$.

One of the most basic general-purpose Markov chain Monte Carlo methods for sampling from a distribution π is the Metropolis algorithm (for a recent theoretical review, see Diaconis and Saloff-Coste [8]). The method is the following. Suppose π is a probability distribution supported on the finite set $\mathcal{S}$. Let K be an irreducible, symmetric Markov chain on $\mathcal{S}$ (called the proposal chain or base chain). In many applications K is a random walk on a regular graph. The Metropolis chain is the new Markov chain T on $\mathcal{S}$ defined as follows:

$$T(x, y) = \begin{cases} K(x, y) & \text{if } \pi(y) \geq \pi(x) \text{ and } y \neq x, \\ K(x, y)\dfrac{\pi(y)}{\pi(x)} & \text{if } \pi(y) < \pi(x), \\ 1 - \displaystyle\sum_{z \neq x} T(x, z) & \text{if } y = x. \end{cases} \tag{2}$$

That is, if the Metropolis chain is currently at state x, then it chooses a proposed state y from the distribution $K(x, y)$, but it only accepts y as the next state with probability $\min\{1, \pi(y)/\pi(x)\}$ (if y is not accepted, then the next state is x again). The chain T is irreducible and reversible with respect to π; in particular, π is the equilibrium distribution of T (see Diaconis and Saloff-Coste [8, p. 20] for a proof of this). Therefore, we can sample from π by simulating the chain T for a sufficiently long time.

In our examples, we take two natural Metropolis chains. For Example I, the base chain is simple symmetric random walk on $\mathcal{A}^{\text{Int}}$:

$$K(i, j) = \begin{cases} \frac{1}{2} & \text{if } |i - j| = 2 \text{ or } i = j = \pm(2M + 1), \\ 0 & \text{otherwise.} \end{cases}$$

Thus the Metropolis chain for Example I looks like random walk on the interval $\mathcal{A}^{\text{Int}}$ with drift away from the center. For large M, this chain requires an exponentially long time to have much chance of reaching the negative half from initial state $2M + 1$. For Example II, the base chain chooses a site $i \in \{1, \ldots, M\}$ uniformly at random and then reverses the sign of $x[i]$. That is,

$$K(x, y) = \begin{cases} \dfrac{1}{M} & \text{if } x, y \in \mathcal{A}^{\text{Ising}} \text{ and } \|x - y\|_1 = 2, \\ 0 & \text{otherwise} \end{cases}$$

(where we write $\|z\|_1 = |z[1]| + \cdots + |z[M]|$ for $z \in \mathbf{R}^M$). The Metropolis chain of Example II reaches equilibrium exponentially slowly in M for any fixed $\beta > 1$ (the meaning of this assertion will be made precise in Section 2).

To discuss the intuition for the swapping and related methods, we will refer to a general distribution of the form (1) on a set $\mathcal{A}$. At the desired parameter value $\beta = \beta^*$, the natural Metropolis chain reaches equilibrium very slowly. But, at smaller values of β, in particular for $\beta = 0$, the equilibrium distribution is uniform on $\mathcal{A}$ and the corresponding Metropolis chain reaches equilibrium rapidly (after about M^2 or M iterations in our examples). Suppose that we could let the parameter β be a random variable, and let it perform a random walk on a set of values from 0 to β^*. We alternate β steps with ordinary Metropolis updates at the current value of β. If we could do all of this in a reversible way, then the chain would spend some of its time with β near 0, where it would change rapidly; and it would spend some of its time with β at β^*, where it would give us samples from π. If there are N intermediate values of β, then we would guess that it would only take about N^2 changes of β to get from β^* to 0 or vice versa. So we can hope that the overall time to equilibrium is a polynomial function of M and N instead of an exponential. This is the idea behind the simulated tempering method mentioned earlier, and the reasoning has been verified rigorously in some examples (Madras and Piccioni [19]). The method of tempered transitions has similar intuition, but we refer the interested reader to the papers mentioned earlier for more discussion of these two algorithms.

The swapping method also uses an interpolating set of β's,

$$0 = \beta_0 < \beta_1 < \cdots < \beta_N = \beta^*,$$

but it has one copy of the system for each β value. That is, the state space consists of $(N + 1)$-tuples $(x_0, x_1, \ldots, x_N) \in \mathcal{A}^{N+1}$. Roughly speaking, the swapping algorithm alternates between two basic steps: the usual Metropolis update of each x_i at β_i, and the "swapping" step in which the two configurations x_i and x_{i+1} try to trade places (for some randomly chosen i). The attempted swap is accepted with a probability dictated by the Metropolis algorithm designed to keep the product distribution on $\mathcal{A}^{N+1}$ (i.e., all copies independent) as the equilibrium distribution. A detailed description of the algorithm is given in Section 2. The intuition here is that if we follow a given copy as it changes β and undergoes Metropolis updates, it looks very much like the simulated tempering chain. Thus we expect the swapping chain to behave like $N + 1$ copies of the simulated tempering chain, subject to an "exclusion" condition that the β's of all the copies must be different. We remark that the simulated tempering method requires good guesses of ratios of the normalizing constants for different β's, but the swapping method does not.

Our main result is that the spectral gap of the Markov chain of the swapping algorithm for each of Examples I and II is bounded below by the reciprocal of a polynomial in the problem size M, provided that we take N proportional to M. This means that the swapping algorithm is a "fast" algorithm (in polynomial time) unlike the standard Metropolis Monte Carlo chain. Markov chains that reach equilibrium in polynomial time are often said to be "rapidly mixing."

The two examples are really "toy models," in that they are not inherently interesting (or, at least, they are tractable analytically and would not normally require simulation). However, they do capture some important qualitative features of more complex and more realistic models, and for this reason increase our optimism that the swapping method is efficient for wider classes of problems. Our intuition is also supported by Zheng [29, 30], who proves that in fairly general situations the spectral gap of the simulated tempering chain is bounded below by a multiple of the gap of the associated swapping chain.

As a point of technical interest, it is worth emphasizing that, for the mean field Ising model, our proof shows that the swapping method equilibrates in polynomial time for every value of β. It is not known whether this is true for the alternative Monte Carlo method known as the Swendsen-Wang algorithm (Swendsen and Wang [28]) when β is close to its critical value of 1. See Gore and Jerrum [16] and Cooper et al. [5] for results about Swendsen-Wang chains on mean field models.

The rest of this paper is organized as follows. The main result is stated in Section 2, together with a review of some convergence rate theory. Section 3 contains several general lemmas. An outline of the proof of the main result is given in Section 4. Details for steps of the proof are given in Sections 5, 6, and 7.

2. CONVERGENCE RATES AND THE MAIN RESULT

Markov chain Monte Carlo methods are based on the fact that a Markov chain with transition matrix P and equilibrium distribution π satisfies

$$P^t(x, y) \to \pi(y) \qquad \text{as } t \to \infty \text{ for all } x \text{ and } y$$

under suitable assumptions. In practice, this is not really enough, and we really need to know how large t should be to have $P^t(x, y)$ suitably close to $\pi(y)$. One standard way of quantifying the closeness to stationarity is to use the total variation distance:

$$\|P_x^t - \pi\| \stackrel{\text{def}}{=} \frac{1}{2} \sum_y |P^t(x, y) - \pi(y)|.$$

There are many techniques to study the convergence rates of Markov chains. For a review, see the expository paper by Rosenthal [26]. For geometrical and other techniques, see Diaconis and Stroock [9], Diaconis and Saloff-Coste [7, 8], and references therein. For our present purpose, we briefly review basic definitions and results relating to comparison techniques.

Let $P(x, y)$ be the transition matrix of an irreducible, aperiodic Markov chain on a finite state space $\mathcal{S}$. We also view P as an operator on functions on $\mathcal{S}$ in the standard way, whereby P maps the function f to the function $(Pf)(x) = \sum_y P(x, y) f(y)$. Suppose P is reversible with respect to the stationary distribution π, i.e., $P(x, y)\pi(x) = P(y, x)\pi(y)$ for all $x, y \in \mathcal{S}$. Define an inner product on real-valued functions on $\mathcal{S}$ by $(f, g)_\pi = \sum f(x) g(x) \pi(x)$. Then reversiblity of the chain P is equivalent to saying that the operator P is self-adjoint on the Hilbert space $\ell^2(\pi)$, i.e., $(Pf, g)_\pi = (f, Pg)_\pi$ for all $f, g \in \ell^2(\pi)$. This implies that P has real eigenvalues,

$$1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{|\mathcal{S}|-1} > -1,$$

and an orthonormal basis of real eigenvectors f_k.

One major concern is to bound the total variation distance between $P^t(x, \cdot)$ and $\pi(\cdot)$. We have

$$\Delta_x(t) \stackrel{\text{def}}{=} \|P_x^t - \pi\| \le \frac{1}{2\sqrt{\pi(x)}}\,\lambda_*^t, \tag{3}$$

where $\lambda_* = \max(\lambda_1, |\lambda_{|\mathcal{S}|-1}|)$. See Diaconis and Saloff-Coste [8, p. 24] for this bound.

Thus, in order to get bounds on rates of convergence, we may try to get bounds on eigenvalues. To this end, we may apply the minimax principle. If we define the *Dirichlet form* of P by

$$\mathcal{E}(f, f) = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 P(x, y)\,\pi(x),$$

for any real-valued function f on $\mathcal{S}$, and write

$$\begin{aligned}\mathrm{Var}(f) &= E_\pi(f^2) - (E_\pi f)^2 \\ &= \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 \pi(x)\,\pi(y),\end{aligned}$$

then the *spectral gap* of P (or say "of the chain") is

$$Gap(P) = 1 - \lambda_1 = \inf\left\{\frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)} : f \in \ell^2(\pi), \mathrm{Var}(f) \ne 0\right\}. \tag{4}$$

For instance, consider the Metropolis chain T for Example I described in Section 1. If f is the indicator function of the positive half line, then $\mathcal{E}(f, f) = C/Z$ and $\mathrm{Var}(f) = 1/4$, which implies that $Gap(T) \le 4C/Z$. Since $Z > C^{2M+1}$ and $C > 1$, this shows that the gap is exponentially small (in M). That is, the chain approaches equilibrium exponentially slowly. A similar argument for Example II is given in Madras and Piccioni [19, Section 3], also yielding exponentially small spectral gap for the Metropolis chain T for any fixed $\beta > 1$.

We now give a precise description of the steps of the swapping algorithm in the context of Examples I and II.

To use the swapping algorithm, we need to choose $N + 1$ distributions. For Example I, we shall make the following choices:

$$h_i(x) = \pi^{\mathrm{Int}}(x|C^{\beta_i}) = \frac{C^{\beta_i|x|}}{Z_i}, \qquad x \in \mathcal{A}^{\mathrm{Int}}, \tag{5}$$

where Z_i's are normalizing constants, and the β_i's are given by

$$\beta_i = \frac{i}{N}, \qquad i = 0, 1, \ldots, N. \tag{6}$$

For Example II, we proceed similarly, taking

$$h_i(x) = \pi^{\mathrm{Ising}}(x|\beta_i), \qquad x \in \mathcal{A}^{\mathrm{Ising}}, \tag{7}$$

where the β_i's are given by

$$\beta_i = \frac{i}{N}\beta^*, \qquad i = 0, 1, \ldots, N. \tag{8}$$

In each example, h_N is the desired distribution π, and h_0 is the uniform distribution on the state space.

Let T_i be the Metropolis chain with stationary distribution h_i based on the walk K defined in Section 1. By definition, the transition matrix of the chain T_i, also denoted by T_i, is as follows:

$$T_i(j, k) = \begin{cases} K(j, k)\min\left\{1, \dfrac{h_i(k)}{h_i(j)}\right\} & \text{if } j \neq k \\ 1 - \displaystyle\sum_{l \neq j} T_i(j, l) & \text{if } j = k \end{cases} \tag{9}$$

The state space Ω of a swapping chain with $N + 1$ components is given by the Cartesian product of $N + 1$ copies of $\mathcal{A}$:

$$\Omega = \mathcal{A}^{N+1}. \tag{10}$$

We shall write an element of Ω in the form $\vec{x} = (x_0, x_1, \ldots, x_N)$. For $i \in \{0, 1, \ldots, N - 1\}$ and $\vec{x} \in \Omega$, let $(i, i + 1)\vec{x}$ denote the element of Ω obtained from $\vec{x}$ by interchanging the ith component with the $(i + 1)$th component, i.e.,

$$(i, i + 1)\vec{x} = (x_0, x_1, \ldots, x_{i-1}, x_{i+1}, x_i, x_{i+2}, \ldots, x_N). \tag{11}$$

In this sense, the transposition $(i, i + 1)$ acts on the state $\vec{x}$.

Let ψ be the product measure distribution on Ω:

$$\psi(\vec{x}) = h_0(x_0)h_1(x_1) \cdots h_N(x_N). \tag{12}$$

That is, under ψ the components of $\vec{x}$ are independent with marginal distributions h_i. The swapping algorithm is designed to have ψ as its equilibrium distribution. It is made up of two chains, a "pure swapping" chain Q and an "updating" chain $\tilde{P}$, which we describe in turn.

The "pure swapping" part of the algorithm proceeds by choosing a random $i \in \{0, \ldots, N - 1\}$, proposing the new state $(i, i + 1)\vec{x}$, and accepting this as the next state with probability

$$\rho_{i,i+1}(\vec{x}) = \min\left\{1, \frac{\psi((i, i + 1)\vec{x})}{\psi(\vec{x})}\right\} = \min\left\{1, \frac{h_i(x_{i+1})h_{i+1}(x_i)}{h_i(x_i)h_{i+1}(x_{i+1})}\right\}. \tag{13}$$

For mathematical convenience, we add a "self-loop" (i.e., do nothing) with probability 1/2. More precisely, we define a pure swapping chain Q on Ω whose transition probabilities are the following. If $\vec{y} \neq \vec{x}$ but $\vec{y} = (i, i + 1)\vec{x}$ for some $i \in \{0, \ldots, N - 1\}$, then

$$Q(\vec{x}, \vec{y}) = \frac{1}{2N} \rho_{i,i+1}(\vec{x}). \tag{14}$$

If $\vec{y} \neq \vec{x}$ and $\vec{y} \neq (i, i+1)\vec{x}$ for each $i \in \{0, \ldots, N-1\}$, then $Q(\vec{x}, \vec{y}) = 0$. Finally,

$$Q(\vec{x}, \vec{x}) = 1 - \sum_{\vec{y} \neq \vec{x}} Q(\vec{x}, \vec{y}). \tag{15}$$

The Metropolis acceptance probability (13) ensures that Q is reversible with respect to ψ. Observe that $Q(\vec{x}, \vec{x}) \geq 1/2$ for every $\vec{x} \in \Omega$ as a result of the self-loops. Hence by Lemma 3 of Section 3 below, we see that Q is a positive operator.

Next, we try to update some component of the current state. The updating step is given by a *product chain* $\tilde{P}$ which corresponds to choosing one of the $N+1$ components of the current state $\vec{x} = (x_0, x_1, \ldots, x_N)$ uniformly at random and updating that component. (For mathematical convenience, we also make sure that the chain does nothing half of the time.) In terms of transition matrices, $\tilde{P}$ is given by

$$\begin{aligned}\tilde{P}(\vec{x}, \vec{y}) = \frac{1}{2}\,\delta(\vec{x}, \vec{y}) + \frac{1}{2(N+1)} \sum_{i=0}^{N} \delta(x_0, y_0) \cdots \delta(x_{i-1}, y_{i-1}) \\ \times\, T_i(x_i, y_i)\delta(x_{i+1}, y_{i+1}) \cdots \delta(x_N, y_N),\end{aligned}$$

where $\vec{x}$ and $\vec{y}$ are in Ω, $\delta(u, v) = 1$ if $u = v$ and 0 otherwise, and T_i is the Metropolis chain as defined before. In terms of operators, this means that

$$\tilde{P} = \frac{1}{N+1} \sum_{i=0}^{N} \underbrace{I \otimes \cdots \otimes I}_{i} \otimes \left(\frac{I + T_i}{2}\right) \otimes \underbrace{I \otimes \cdots \otimes I}_{N-i}. \tag{16}$$

As with Q, we observe that $\tilde{P}$ is reversible with respect to ψ and is a positive operator.

The swapping algorithm consists of alternating the above two steps. There are different Markov chains that are possible by this strategy (e.g., $\tilde{P}Q$ or $Q\tilde{P}$ or $(\tilde{P} + Q)/2$ or $\tilde{P}^7 Q$). We choose to define the swapping chain to be given by $Q\tilde{P}Q$. One advantage of this choice is that it is reversible, unlike some of the others (such as $\tilde{P}Q$). It is also convenient for the decomposition methods that are key to the proofs.

Remarks.

1. The pure swapping chain Q is reducible. Indeed, if the components of $\vec{y}$ are not a permutation of those of $\vec{x}$, then $Q^t(\vec{x}, \vec{y}) = 0$ for every t.
2. Product chains have been studied by Diaconis and Saloff-Coste [7, pp. 712–717]. One useful basic result is that the spectral gap of the product chain is equal to the minimum over spectral gaps of its component chains divided by the number of components (see Lemma 4 below). This tells us that if T_N is not rapidly mixing, then neither is the product chain $\tilde{P}$. The point of the swapping method is that $Q\tilde{P}Q$ can be rapidly mixing even if $\tilde{P}$ is not.

The main result of this paper is a lower bound for the spectral gap of the swapping chain $Q\tilde{P}Q$. It is expressed in terms of a model-dependent constant $\Theta \in (0, 1)$ defined by

$$\Theta = \begin{cases} C^{-2} & \text{in Example I,} \\ e^{-\beta^*/2} & \text{in Example II.} \end{cases}$$

Intuitively, the constant Θ measures the amount of "overlap" between adjacent distributions h_i and h_{i+1}, as is reflected in the following lemma.

Lemma 1. *Define the constant Θ as above. Then*

$$\rho_{i,i+1}(\vec{x}) \geq \Theta^{M/N}$$

for every $i \in \{0, \ldots, N-1\}$ and $\vec{x} \in \Omega$.

Proof. We will prove the result for Example II. The proof for Example I is similar. Recall $\beta_i = i\beta^*/N$. For $x \in \mathcal{A}$, let

$$G(x) = \frac{\left(\sum_{j=1}^{M} x[j]\right)^2}{2M}.$$

Then $0 \leq G(x) \leq M/2$. Thus we have

$$\begin{aligned} \rho_{i,i+1}(\vec{x}) &= \min\left\{1, \frac{h_{i+1}(x_i)h_i(x_{i+1})}{h_i(x_i)h_{i+1}(x_{i+1})}\right\} \\ &= \min\left\{1, \frac{\exp((i+1)\beta^*G(x_i)/N)\exp(i\beta^*G(x_{i+1})/N)}{\exp(i\beta^*G(x_i)/N)\exp((i+1)\beta^*G(x_{i+1})/N)}\right\} \\ &= \min\{1, \exp(\beta^*G(x_i)/N - \beta^*G(x_{i+1})/N)\} \\ &\geq \exp(-\beta^*M/2N). \end{aligned}$$

The result follows. ∎

Now we can state our main result.

Theorem 2. *The spectral gap of the swapping chain $Q\tilde{P}Q$ has the following lower bound*:

$$\mathrm{Gap}(Q\tilde{P}Q) \geq \begin{cases} \dfrac{\Theta^{2M/N}}{384M^2(N+1)^6} & \textit{in Example I} \\ \dfrac{\Theta^{2M/N+8}}{768M^3(N+1)^6} & \textit{in Example II.} \end{cases} \tag{17}$$

In particular, if M/N approaches a constant $\gamma \in (0, \infty)$, then the lower bound is asymptotically $\tilde{c}/M^8$ in Example I and $\tilde{c}/M^9$ in Example II, where $\tilde{c}$ is a constant depending on Θ and γ.

3. SOME USEFUL LEMMAS

Before we get into a detailed outline of the proof of the main theorem, this section presents several useful lemmas. The reader can survey them now, or skip over them until they are needed. We state them in the context of finite state spaces, but this is for simplicity rather than necessity.

The following well-known lemma was invoked in the preceding section, and is a convenient way to show that an operator is positive.

Lemma 3. *Let P be a Markov chain that is reversible with respect to π on the finite state space $\mathcal{S}$. Also assume that $P(x, x) \geq 1/2$ for every $x \in \mathcal{S}$. Then P is a positive operator.*

Proof. The proof is obvious once we observe that we can write $P = (I + P^{\#})/2$ where $P^{\#}$ is a reversible transition probability matrix. ∎

The next lemma deals with product chains, which we have already encountered in the definition of $\tilde{P}$ in the preceding section. A proof appears in Diaconis and Saloff-Coste [7, Lemma 3.2].

Lemma 4. *For each $i = 1, \ldots, m$, let K_i be a reversible Markov chain on a finite state space $\mathcal{T}_i$. Consider the product Markov chain K on the product space $\prod_1^m \mathcal{T}_i$, defined by*

$$K = \frac{1}{m} \sum_{i=1}^{m} \underbrace{I \otimes \cdots \otimes I}_{i-1} \otimes K_i \otimes \underbrace{I \otimes \cdots \otimes I}_{m-i}.$$

Then

$$Gap(K) = \frac{1}{m} \min\{Gap(K_i), i = 1, \ldots, m\}.$$

The next lemma says that if we have comparisons between the Dirichlet forms and between the stationary distributions of two chains on the same state space, then we can deduce a comparison between their spectral gaps. See Diaconis and Saloff-Coste [7, Lemma 3.3] for a proof.

Lemma 5. *Let (K, π) and (K', π') be two Markov chains on the same finite state space $\mathcal{S}$, with respective Dirichlet forms $\mathcal{E}$ and $\mathcal{E}'$. Assume that there exist constants A, $a > 0$ such that*

$$\mathcal{E}' \leq A\mathcal{E} \qquad \text{and} \qquad a\pi \leq \pi'.$$

Then

$$Gap(K') \le \frac{A}{a}\, Gap(K).$$

Remark. A sufficient condition for $\mathscr{E}' \le A\mathscr{E}$ is that

$$\pi'(x)K'(x, y) \le A\,\pi(x)K(x, y) \qquad \text{for all } x, y \in \mathcal{S} \quad \text{such that } x \ne y.$$

The next result relates the spectral gap of a transition matrix P to that of P^m, where m is a positive integer.

Lemma 6. *For any reversible finite Markov chain P,*

$$Gap(P) \ge \frac{1}{m}\, Gap(P^m), \quad \textit{for all } m \in \mathbb{N}.$$

Proof. Let λ_1 be the second-largest eigenvalue of the transition matrix P, with corresponding eigenvector v_1, i.e., $Pv_1 = \lambda_1 v_1$. Then $P^m v_1 = \lambda_1^m v_1$, so the second-largest eigenvalue of P^m cannot be smaller than λ_1^m. Therefore,

$$Gap(P^m) \le 1 - \lambda_1^m \le m(1 - \lambda_1) = m\ Gap(P).$$

The result follows. ∎

The next result concerns the comparison of spectral gaps for two operators, which are not necessarily Markov transition kernels. Let π be a probability measure on a state space $\mathcal{S}$. Recall the definition of the inner product $(\cdot, \cdot)_\pi$ and the Hilbert space $\ell^2(\pi)$ from Section 2. Let $\mathbf{1}^\perp$ be the orthogonal complement of the constant functions in $\ell^2(\pi)$:

$$\mathbf{1}^\perp := \{f \in \ell^2(\pi) : (f, 1)_\pi = 0\}.$$

Let A and B be self-adjoint operators on $\ell^2(\pi)$. Further, we assume that the following conditions are satisfied:

(i) Both A and B have the largest eigenvalue 1;
(ii) A is invariant on $\mathbf{1}^\perp$ (i.e., $Af \in \mathbf{1}^\perp$ if $f \in \mathbf{1}^\perp$);
(iii) $\|A\| \le 1$;
(iv) B is positive [i.e., $(Bf, f)_\pi \ge 0$ for all $f \in \ell^2(\pi)$].

Lemma 7. *Under the above assumptions, the following holds*:

$$Gap(ABA) \ge Gap(B). \tag{18}$$

Proof. In the following, for simplicity, we omit the subscript π in the inner product notation. If $f \in \mathbf{1}^\perp$, then, by assumption (ii), $Af \in \mathbf{1}^\perp$. By definition of spectral gaps, we have

$$
\begin{aligned}
Gap(ABA) &= \inf_{f \in \mathbf{1}^{\perp}} \frac{(f, (I - ABA)f)}{(f,f)} \\
&= \inf_{f \in \mathbf{1}^{\perp}} \frac{(f,f) - (f, ABAf)}{(f,f)} \\
&= \inf_{f \in \mathbf{1}^{\perp}} \left\{ 1 - \frac{(Af, BAf)}{(f,f)} \right\} \\
&\geq \inf_{f \in \mathbf{1}^{\perp}} \left\{ 1 - \frac{(Af, BAf)}{(Af, Af)} \right\} \qquad \text{[by (iv)]} \\
&= \inf_{g \in \mathbf{1}^{\perp}, g = Af} \left\{ 1 - \frac{(g, Bg)}{(g,g)} \right\} \\
&\geq \inf_{g \in \mathbf{1}^{\perp}} \left\{ 1 - \frac{(g, Bg)}{(g,g)} \right\} \\
&= \mathrm{Gap}(B). \qquad \blacksquare
\end{aligned}
$$

Remarks.

1. If $A1 = 1$, then A is invariant on $\mathbf{1}^{\perp}$.
2. If P is the transition kernel of a Markov chain reversible with respect to π, then the corresponding operator satisfies all assumptions for A in Lemma 7.
3. If A is a nonnegative self-adjoint operator that satisfies all assumptions in Lemma 7, then the same is true for its square root $A^{1/2}$. In particular, this applies to the operator $Q^{1/2}$ since Q is nonnegative and reversible.

4. STRATEGY OF THE PROOF

The proof of the main theorem relies on the decomposition of the Markov chain into simpler chains. The key result for this is a theorem due to Caracciolo, Pelissetto, and Sokal [2]. For more detailed discussion and a proof of this result, see Madras and Randall [20]. For convenience here, we will assume that the state space is finite.

The idea of the decomposition is the following. When trying to show that a complicated Markov chain is rapidly mixing, we may notice that equilibration is rapid within certain parts of the chain. Indeed, suppose that we can partition the state space into several pieces such that the restrictions of the chain to each of these pieces are rapidly mixing. Also suppose that the original chain moves easily among these different pieces. That is, consider an "aggregate" Markov chain in which each state corresponds to one of the pieces of the original chain, and whose transitions model the movements of the original chain between the pieces. Then if the aggregate chain is rapidly mixing, and if the restriction to each piece is rapidly mixing, then the original chain should be rapidly mixing. Indeed, the spectral gap ($\approx$ inverse time to equilibrium) should be bounded below by the product of the gap of the aggregate chain times the smallest gap of the pieces. This is the spirit of the results of Caracciolo, Pelissetto, and Sokal [2] and Madras and Randall

[20], and it is the basic strategy of our proof. We shall now describe the decomposition method more formally.

Let θ be a probability distribution on a finite state space $\mathcal{S}$, and let $\mathcal{P}$ be a transition matrix that is reversible with respect to θ. Suppose further that the state space is partitioned into m disjoint pieces:

$$\mathcal{S} = \bigcup_{i=1}^{m} \mathcal{S}_i, \qquad \text{where } \mathcal{S}_i \cap \mathcal{S}_j = \varnothing \qquad \text{if } i \neq j. \tag{19}$$

For each $i = 1, \ldots, m$, define $\mathcal{P}_i$, the restriction of $\mathcal{P}$ to $\mathcal{S}_i$, by rejecting jumps that leave $\mathcal{S}_i$:

$$\mathcal{P}_i(x, B) = \mathcal{P}(x, B) + 1_{\{x \in B\}} \mathcal{P}(x, \mathcal{S} \backslash \mathcal{S}_i), \quad \text{for } x \in \mathcal{S}_i, \quad B \subset \mathcal{S}_i. \tag{20}$$

It is easy to see that $\mathcal{P}_i$ is reversible with respect to the restriction of θ to $\mathcal{S}_i$. Let $b_i = \theta(\mathcal{S}_i) = \sum_{x \in \mathcal{S}_i} \theta(x)$. Then $(b_1, \ldots, b_m)$ can be viewed as a probability distribution on $\{1, \ldots, m\}$.

Let $\mathcal{Q}$ be another transition matrix that is reversible with respect to θ. Define an aggregated chain $\bar{\mathcal{Q}}$ as follows:

$$\bar{\mathcal{Q}}(i, j) = \frac{1}{b_i} \sum_{y \in \mathcal{S}_j} \sum_{x \in \mathcal{S}_i} \theta(x) \mathcal{Q}(x, y) \qquad (i, j = 1, \ldots, m). \tag{21}$$

Observe that $\bar{\mathcal{Q}}$ is reversible with respect to $b = (b_1, \ldots, b_m)$. The chain $\bar{\mathcal{Q}}$ describes a kind of overall rate at which the chain jumps from one piece of the partition to another. Indeed, if the chain reached equilibrium within each piece as soon as it jumped to that piece, then this would be exactly what $\mathcal{Q}$ is describing.

The key result which quantifies the intuition described above is the following.

Theorem 8 (Caracciolo–Pelissetto–Sokal [2]). *Assume that $\mathcal{Q}$ is a positive operator. Let $\mathcal{Q}^{1/2}$ be the positive square root of $\mathcal{Q}$. Then*

$$Gap(\mathcal{Q}^{1/2} \mathcal{P} \mathcal{Q}^{1/2}) \geq Gap(\bar{\mathcal{Q}}) \min_{1 \leq i \leq m} Gap(\mathcal{P}_i). \tag{22}$$

The proof of our main theorem requires us to choose $\mathcal{P}$ and $\mathcal{Q}$ appropriately, so that $\mathcal{Q}^{1/2} \mathcal{P} \mathcal{Q}^{1/2}$ can be related to the original chain $(Q \tilde{P} Q)$, and so that the aggregate chain $\bar{\mathcal{Q}}$ and the pieces $\mathcal{P}_i$ have good spectral gaps.

Now we outline the strategy of proof of the main result. First we introduce a symmetric partition of the state space $\mathcal{A}$ into "positive" and "negative" pieces $\mathcal{A}_+$ and $\mathcal{A}_-$, as follows. In Example I ($\mathcal{A} = \mathcal{A}^{\text{Int}}$), let

$$\mathcal{A}_+ = \mathcal{A}_+^{\text{Int}} = \{x \in \mathcal{A}^{\text{Int}} : x > 0\},$$

and in Example II ($\mathcal{A} = \mathcal{A}^{\text{Ising}}$), let

$$\mathcal{A}_+ = \mathcal{A}_+^{\text{Ising}} = \left\{ x \in \mathcal{A}^{\text{Ising}} : \sum_{i=1}^{M} x[i] > 0 \right\}.$$

In both examples, let $\mathcal{A}_- = \mathcal{A} \backslash \mathcal{A}_+$. By symmetry, we have $\pi(\mathcal{A}_+) = 1/2 = \pi(\mathcal{A}_-)$.

Step One. This is a technical step to prepare for the application of Theorem 8. Observe that $Q\tilde{P}Q$ is an irreducible, aperiodic and reversible finite Markov chain. We see that

$$\begin{aligned} Gap(Q\tilde{P}Q) &\geq \tfrac{1}{3} Gap(Q\tilde{P}QQ\tilde{P}QQ\tilde{P}Q) \qquad \text{(by Lemma 6)} \\ &= \tfrac{1}{3} Gap(Q\tilde{P}Q^{1/2}(Q^{1/2}Q\tilde{P}QQ^{1/2})Q^{1/2}\tilde{P}Q) \\ &\geq \tfrac{1}{3} Gap(Q^{1/2}(Q\tilde{P}Q)Q^{1/2}) \qquad \text{(by Lemma 7, three times).} \end{aligned}$$

As we shall see in the following steps, we need these extra $Q^{1/2}$ terms because we will want to apply decomposition (Theorem 8) twice. First we apply it with $\mathcal{Q} = Q$ and $\mathcal{P} = Q\tilde{P}Q$ (Step 2). It will turn out that we will need to use decomposition again to get good bounds on the gaps of the $\mathcal{P}_i$ chains (Step 4). Step 1 ensures that we still have a $Q\tilde{P}Q$ term after the first decomposition [see right-hand side of Eq. (24) below].

Step Two. In this step we decompose Ω according to the number of positive components among $x_1, \ldots, x_N$, and then apply Theorem 8. Let $\tilde{\Omega} = \{+, -\}^N$, and consider $\vec{x} = (x_0, x_1, \ldots, x_N) \in \Omega = \mathcal{A}^{N+1}$. We define the signature $\text{sgn}(\vec{x})$ of $\vec{x}$ to be the vector $(v_1, \ldots, v_N) \in \tilde{\Omega}$, where

$$v_i = \begin{cases} + & \text{if } x_i \in \mathcal{A}_+ \\ - & \text{if } x_i \in \mathcal{A}_- \end{cases} \qquad (1 \leq i \leq N). \tag{23}$$

In other words, the signature of $\vec{x}$ is the vector consisting of signs of its last N components. In particular, we don't count the sign of its first component. It will become apparent later why we ignore x_0 (see Remark 2 at the end of this section).

If $k \in \{0, 1, \ldots, N\}$, define

$$\tilde{\Omega}_k = \{v \in \tilde{\Omega} : v \text{ has exactly } k + \text{'s}\}.$$

Then $\{\tilde{\Omega}_k, 0 \leq k \leq N\}$ is a partition of $\tilde{\Omega}$, i.e.,

$$\tilde{\Omega} = \bigcup_{k=0}^{N} \tilde{\Omega}_k \qquad \text{and} \qquad \tilde{\Omega}_i \cap \tilde{\Omega}_j = \varnothing \qquad \text{whenever } i \neq j.$$

Analogously, define Ω_k to be the subset of all elements in Ω which contain exactly k nonnegative components, i.e.,

$$\Omega_k = \{\vec{x} \in \Omega : sgn(\vec{x}) \in \tilde{\Omega}_k\}.$$

Then $\{\Omega_k,\ 0 \le k \le N\}$ is a partition of Ω.

Consider the decomposition

$$\Omega = \bigcup_{k=0}^{N} \Omega_k.$$

Define the aggregated transition matrix $\bar{Q}$ as prescribed by Eq. (21) (with $\mathcal{S}_k$ corresponding to Ω_k; see Section 5 for details). Since Q is a positive operator, we can apply Theorem 8 to get the following:

$$Gap(Q^{1/2}(Q\tilde{P}Q)Q^{1/2}) \ge Gap(\bar{Q}) \cdot \min_{0\le k\le N} Gap((Q\tilde{P}Q)|_{\Omega_k}). \tag{24}$$

Observe that $\bar{Q}$ is essentially a one-dimensional nearest-neighbor chain on $\{0, 1, \ldots, N\}$, and as such it is relatively easy to bound its spectral gap. This is done in Section 5. The next two steps are devoted to getting lower bounds for $Gap((Q\tilde{P}Q)|_{\Omega_k})$.

Step Three. We need a lower bound for $Gap((Q\tilde{P}Q)|_{\Omega_k})$. This chain is somewhat awkward to work with, since it is the restriction to Ω_k of the product of three kernels. It is easier to work with a product of three kernels, each of which is itself a restriction to Ω_k. The purpose of Step 3 is to get a lower bound for the gap of this restricted product of kernels in terms of the gap of a product of restricted kernels.

Let $\psi(\vec{x}) = \Pi_{i=0}^{N} h_i(x_i)$, and let ψ_k be the normalized restriction of $\psi = \Pi_0^N h_i$ to Ω_k:

$$\psi_k(A) = \frac{\psi(A \cap \Omega_k)}{b_k}, \qquad A \subset \Omega, \tag{25}$$

where $b_k = \psi(\Omega_k)$. Let P_k be the restriction of the updating chain $\tilde{P}$ to Ω_k:

$$P_k(\vec{x}, A) = \tilde{P}(\vec{x}, A \cap \Omega_k) + 1_{\{\vec{x}\in A\}}\tilde{P}(\vec{x}, \Omega\backslash\Omega_k), \qquad \text{for } \vec{x} \in \Omega_k, \quad A \subset \Omega.$$

Let Q_k be the restriction of the pure swapping chain Q to Ω_k:

$$Q_k(\vec{x}, A) = Q(\vec{x}, A \cap \Omega_k) + 1_{\{\vec{x}\in A\}}Q(\vec{x}, \Omega\backslash\Omega_k), \qquad \text{for } \vec{x} \in \Omega_k, \quad A \subset \Omega.$$

Since $\tilde{P}$ and Q are both reversible with respect to ψ, so are their restrictions P_k and Q_k.

With these definitions in place, we can state the main inequality of this step:

$$Gap((Q\tilde{P}Q)|_{\Omega_k}) \ge \tfrac{1}{8}\, Gap(Q_kP_kQ_k). \tag{26}$$

The proof is straightforward, as follows. First we claim that

$$Q(\vec{x}, \vec{y}) \ge \tfrac{1}{2}\, Q_k(\vec{x}, \vec{y}) \qquad \text{and} \qquad \tilde{P}(\vec{x}, \vec{y}) \ge \tfrac{1}{2}\, P_k(\vec{x}, \vec{y}), \qquad \text{for all } \vec{x}, \vec{y} \in \Omega_k. \tag{27}$$

To see this, first consider Q. The inequality (27) is obvious if $\vec{x} \neq \vec{y}$ because then $Q(\vec{x}, \vec{y}) = Q_k(\vec{x}, \vec{y})$, and it is obvious if $\vec{x} = \vec{y}$ because then $Q(\vec{x}, \vec{x}) \geq 1/2$. This proves (27) for Q, and the argument for $\tilde{P}$ is identical. Next, for $\vec{x}, \vec{y} \in \Omega_k$, we have

$$\begin{aligned}
(Q\tilde{P}Q)|_{\Omega_k}(\vec{x}, \vec{y}) &= (Q\tilde{P}Q)(\vec{x}, \vec{y}) \\
&= \sum_{\vec{z},\vec{w}\in\Omega} Q(\vec{x}, \vec{z})\tilde{P}(\vec{z}, \vec{w})Q(\vec{w}, \vec{y}) \\
&\geq \sum_{\vec{z},\vec{w}\in\Omega_k} Q(\vec{x}, \vec{z})\tilde{P}(\vec{z}, \vec{w})Q(\vec{w}, \vec{y}) \\
&\geq \frac{1}{8} \sum_{\vec{z},\vec{w}\in\Omega_k} Q_k(\vec{x}, \vec{z})P_k(\vec{z}, \vec{w})Q_k(\vec{w}, \vec{y}) \\
&= \frac{1}{8} (Q_kP_kQ_k)(\vec{x}, \vec{y}).
\end{aligned}$$

Since both $(Q\tilde{P}Q)|_{\Omega_k}$ and $Q_kP_kQ_k$ are reversible with respect to ψ_k, the above yields a corresponding inequality for their Dirichlet forms. Then the desired inequality (26) for their spectral gaps follows from the comparison lemma, Lemma 5.

Step Four. Next we perform a further decomposition, this time of each Ω_k. Each state $\vec{x}$ in Ω_k has exactly k positive components among $x_1, \ldots, x_N$. This step decomposes Ω_k according to which components are positive. For $\sigma \in \tilde{\Omega}_k$, define

$$\Omega_\sigma = \{\vec{x} \in \Omega : sgn(\vec{x}) = \sigma\}. \tag{28}$$

Let P_σ be the restriction of P_k to Ω_σ by rejecting jumps that leave Ω_σ:

$$P_\sigma(\vec{x}, A) = P_k(\vec{x}, A \cap \Omega_\sigma) + 1_{\{\vec{x}\in A\}}P_k(\vec{x}, \Omega\backslash\Omega_\sigma), \tag{29}$$

for $\vec{x} \in \Omega_\sigma$ and $A \subset \Omega_k$. Note that $P_\sigma(\vec{x}, \vec{y}) = \tilde{P}(\vec{x}, \vec{y})$ as long as $\vec{x}, \vec{y} \in \Omega_\sigma$ and $\vec{x} \neq \vec{y}$.

For each $k \in \{0, \ldots, N\}$, consider the partition

$$\Omega_k = \bigcup_{\sigma\in\tilde{\Omega}_k} \Omega_\sigma.$$

As in Step Two, we define another aggregated chain $\bar{Q}_k$. The details are in Section 6, but here is the intuition. The chain $\bar{Q}_k$ models the movements of positive coordinates as they swap with negative coordinates. Thus we have k positive coordinates, each performing a "random walk" on $\{1, \ldots, N\}$. As we explain in Section 6, this chain is similar to the well known exclusion process, whose spectral gap is well understood. This similarity will be used to get a good bound on $Gap(\bar{Q}_k)$.

Since Q_k is also positive, we can first apply Lemma 7 to obtain that

$$Gap(Q_kP_kQ_k) \geq Gap[(Q_k)^{1/2}P_k(Q_k)^{1/2}]. \tag{30}$$

Then, applying Theorem 8 again, we can conclude that

$$Gap[(Q_k)^{1/2}P_k(Q_k)^{1/2}] \geq Gap(\bar{Q}_k) \cdot \min_{\sigma \in \Omega_k} Gap(P_\sigma) \tag{31}$$

for each $k \in \{0, \ldots, N\}$.

Now the above inequalities can be combined to give us the following:

$$\begin{aligned}
\mathrm{Gap}(Q\tilde{P}Q) &\geq \tfrac{1}{3} Gap(Q^{1/2}(Q\tilde{P}Q)Q^{1/2}) \qquad \text{(by Step One)} \\
&\geq \tfrac{1}{3} Gap(\bar{Q}) \cdot \min_{0 \leq k \leq N} Gap((Q\tilde{P}Q)|_{\Omega_k}) \qquad \text{(by Step Two)} \\
&\geq \tfrac{1}{24} Gap(\bar{Q}) \cdot \min_{0 \leq k \leq N} Gap(Q_k P_k Q_k) \qquad \text{(by Step Three)} \\
&\geq \tfrac{1}{24} Gap(\bar{Q}) \cdot \min_{0 \leq k \leq N} Gap(\bar{Q}_k) \cdot \min_{\sigma \in \tilde{\Omega}_k, 0 \leq k \leq N} Gap(P_\sigma) \qquad \text{(by Step Four).}
\end{aligned} \tag{32}$$

The only thing left is to get a good lower bound for each quantity that appears in the right side of the last inequality. This is accomplished in later sections. Specifically, we shall prove

$$Gap(\bar{Q}) \geq \frac{1}{4N^2}\, \Theta^{M/N} \tag{33}$$

in Proposition 10 of Section 5,

$$Gap(\bar{Q}_k) \geq \frac{4\Theta^{M/N}}{N^3} \tag{34}$$

in Proposition 11 of Section 6, and

$$Gap(P_\sigma) \geq \begin{cases} \dfrac{1}{16M^2(N+1)} & \text{for Example I,} \\ \dfrac{e^{-4\beta^*}}{32M^3(N+1)} & \text{for Example II.} \end{cases} \tag{35}$$

in Proposition 12 of Section 7. Inserting these bounds into Eq. (32) proves Theorem 2.

Remarks.

1. A similar overall strategy was devised independently by Martin and Randall [22] for a very different problem.
2. The component x_0 plays a special role in our proof. The equilibrium distribution h_0 is uniform, and the T_0 chain is rapidly mixing. In contrast, for large i, the T_i chains are slowly mixing. Our proof essentially pretends that all T_i transitions between $\mathcal{A}_+$

and $\mathcal{A}_-$ are suppressed for $i \geq 1$. This is a negligible change for large i, and not a serious one for small non-zero i. By omitting x_0 from the definition of $\mathrm{sgn}(\vec{x})$, we allow $\bar{Q}$ to change state by swapping x_1 with x_0. The x_0 coordinate is free to change sign between swaps. It is the rapid mixing of T_0 together with the swapping that permits the number of nonnegative components of $\vec{x}$ to change frequently in the long run.

3. It is not difficult to extend the analysis to the more natural cases where there is no restriction on the odd parity of integers. For example I, one can split the middle state 0 into two copies, each with half of the original weight $1/Z$. For example II, "middle states" (i.e., those states whose total spins are zero) only exist when M is even, and there are $\binom{M}{M/2}$ such states in this case. We may fix any half of these states to be in the positive piece $\mathcal{A}_+$ with each state carrying a weight of $1/Z(\beta)$. Then the analysis can be done in the same manner. Another approach is simply to count the middle states as "positive": More precisely, in the definition of $\mathcal{A}_+$, replace $>$ by $\geq$. Though this change entails replacing a few equalities by corresponding inequalities, the analysis can still be carried over, and the results on estimates of spectral gaps remain true because only inequalities are needed to get these results. It is significant that the probability of $\mathcal{A}_+$ under h_i is exponentially close to 1/2 in this modification. The details of this approach to Example I appeared in Zheng [29].

5. A LOWER BOUND FOR Gap($\bar{Q}$)

Define the aggregated chain $\bar{Q}$ on the state space $\{0, \ldots, N\}$ as follows:

$$\bar{Q}(k, j) = \frac{1}{b_k} \sum_{\vec{y} \in \Omega_j} \sum_{\vec{x} \in \Omega_k} \psi(\vec{x}) Q(\vec{x}, \vec{y}), \tag{36}$$

where $b_k = \psi(\Omega_k)$. Since Q is reversible with respect to ψ, we see that $\bar{Q}$ is reversible with respect to the probability distribution $(b_0, \ldots, b_N)$.

By symmetry, we have

$$h_k(\mathcal{A}_-) = h_k(\mathcal{A}_+) = \tfrac{1}{2}. \tag{37}$$

Since ψ is a product measure, we see from (37) that $\psi(\Omega_\sigma) = 2^{-N}$. The cardinality of the set $\tilde{\Omega}_k$ is $\binom{N}{k}$, so we have

$$b_k = \psi(\Omega_k) = \sum_{\sigma \in \tilde{\Omega}_k} \psi(\Omega_\sigma) = \binom{N}{k} \frac{1}{2^N}. \tag{38}$$

Note that for the pure swapping chain Q, if $|i - j| > 1$, then it is impossible to change a state in Ω_i to a state in Ω_j in one swap, i.e.,

$$Q(\vec{x}, \vec{y}) = 0 \qquad \text{if} \quad \vec{x} \in \Omega_i, \quad \vec{y} \in \Omega_j, \quad \text{and} \quad |i - j| > 1.$$

Hence,

$$\bar{Q}(i, j) = 0 \qquad \text{if } |i - j| > 1.$$

In the following, we use the notation of Eq. (11) to write $Q(\vec{x}, (0, 1)\vec{x})$ to denote $Q((x_0, x_1, x_2, \ldots, x_N), (x_1, x_0, x_2, \ldots, x_N))$. Observe that the only way that the chain can leave Ω_i is by interchanging x_0 and x_1 when $x_0 \in \mathcal{A}_+$ and $x_1 \in \mathcal{A}_-$ or vice versa. Recalling Lemma 1, for $i \in \{0, \ldots, N-1\}$ we have

$$\begin{aligned}
b_i \bar{Q}(i, i+1) &= \sum_{\vec{y} \in \Omega_{i+1}} \sum_{\vec{x} \in \Omega_i} \psi(\vec{x}) Q(\vec{x}, \vec{y}) \\
&= \sum_{x_0 \in \mathcal{A}_+} \sum_{x_1 \in \mathcal{A}_-} \sum_{\substack{\sigma \in \tilde{\Omega}_i \\ x_j \in \mathcal{A}_{\sigma_j}, 2 \le j \le N}} \psi(\vec{x}) Q(\vec{x}, (0, 1)\vec{x}) \\
&= \sum_{x_0 \in \mathcal{A}_+} \sum_{x_1 \in \mathcal{A}_-} \sum_{\substack{\sigma \in \tilde{\Omega}_i \\ x_j \in \mathcal{A}_{\sigma_j}, 2 \le j \le N}} \psi(\vec{x}) \frac{1}{2N} \rho_{0,1}(\vec{x}) \\
&\ge \frac{1}{2N} \sum_{x_0 \in \mathcal{A}_+} \sum_{x_1 \in \mathcal{A}_-} h_0(x_0) h_1(x_1) \Theta^{M/N} \sum_{\substack{\sigma \in \tilde{\Omega}_i \\ x_j \in \mathcal{A}_{\sigma_j}, 2 \le j \le N}} \prod_{2}^{N} h_j(x_j) \\
&= \frac{1}{2N} \sum_{x_0 \in \mathcal{A}_+} \sum_{x_1 \in \mathcal{A}_-} h_0(x_0) h_1(x_1) \binom{N-1}{i} \frac{1}{2^{N-1}} \Theta^{M/N} \\
&= \frac{1}{2^{N+2} N} \binom{N-1}{i} \Theta^{M/N}. \qquad (39)
\end{aligned}$$

In order to study the chain $\bar{Q}$, we introduce another chain and use the comparison argument. Consider the symmetric simple random walk $\bar{K}$ on $\{0, 1, \ldots, N\}$:

$$\begin{aligned}
\bar{K}(0, 1) &= \bar{K}(0, 0) = \bar{K}(N, N-1) = \bar{K}(N, N) \\
&= \bar{K}(i, i-1) = \bar{K}(i, i+1) = \tfrac{1}{2}, \qquad 0 < i < N.
\end{aligned}$$

Define R to be the Metropolis chain with base chain $\bar{K}$ for the binomial distribution $(b_0, \ldots, b_N)$ as given by (38). Then R is irreducible and reversible with respect to $(b_0, \ldots, b_N)$. Moreover, its spectral gap is well understood.

Lemma 9. *The spectral gap of the Metropolis chain R satisfies the following inequality*:

$$\frac{1}{N} \le Gap(R) \le \frac{2}{N}. \qquad (40)$$

For more details about this chain and a proof of this lemma, see Diaconis and Saloff-Coste [7, pp. 698, 719].

Now we shall compare the chain $\bar{Q}$ with the chain R. First, if $0 \le i < N$, then we have

$$R(i, i+1) = \bar{K}(i, i+1)\min\left\{1, \frac{b_{i+1}}{b_i}\right\}$$
$$= \frac{1}{2}\min\left\{1, \frac{\binom{N}{i+1}}{\binom{N}{i}}\right\}$$
$$= \frac{1}{2}\min\left\{1, \frac{N-i}{i+1}\right\}. \tag{41}$$

By direct calculation, one can show

$$\frac{\binom{N-1}{i}}{\binom{N}{i}} = \frac{N-i}{N} \geq \frac{1}{2}\min\left\{1, \frac{N-i}{i+1}\right\}. \tag{42}$$

So, for $i = 0, \ldots, N-1$, we use Eqs. (39), (42), and (41) to obtain

$$\bar{Q}(i, i+1) \geq \frac{1}{2^{N+2}N}\binom{N-1}{i}\Theta^{M/N}\left[\binom{N}{i}\frac{1}{2^N}\right]^{-1}$$
$$\geq \frac{1}{4N}\frac{1}{2}\min\left\{1, \frac{N-i}{i+1}\right\}\Theta^{M/N}$$
$$= \frac{1}{4N}\Theta^{M/N}R(i, i+1).$$

Using this inequality and reversibility, along with the fact that $R(i, j) = \bar{Q}(i, j) = 0$ whenever $|i - j| > 1$, we conclude that

$$\bar{Q}(i, j) \geq \frac{1}{4N}\Theta^{M/N}R(i, j) \qquad \text{for all } i \neq j. \tag{43}$$

Let $\mathscr{E}_R$ and $\mathscr{E}_{\bar{Q}}$ denote the Dirichlet forms for R and $\bar{Q}$, respectively. Since R and $\bar{Q}$ have the same equilibrium distribution, the bound (43) and Lemma 5 imply that

$$\mathscr{E}_{\bar{Q}} \geq \frac{1}{4N}\Theta^{M/N}\mathscr{E}_R,$$

and hence

$$Gap(\bar{Q}) \geq \frac{1}{4N}\Theta^{M/N}Gap(R).$$

The above inequality and Lemma 9 enable us to get the following.

Proposition 10. *The spectral gap of the aggregated chain $\bar{Q}$ satisfies*

$$Gap(\bar{Q}) \geq \frac{1}{4N^2}\,\Theta^{M/N}. \tag{44}$$

Remark. Analogous bounds can be used to show that $Gap(\bar{Q}) \leq 1/N^2$ (Zheng [29]).

6. A LOWER BOUND FOR $Gap(\bar{Q}_k)$

First we define $\bar{Q}_k$, the aggregated chain on $\tilde{\Omega}_k$. Fix $k \in \{0, \ldots, N\}$. For $\sigma \in \tilde{\Omega}_k$, let $\bar{b}_\sigma = \psi_k(\Omega_\sigma) = \psi(\Omega_\sigma)/b_k$. If $\sigma, \tau \in \tilde{\Omega}_k$ and $\sigma \neq \tau$, then let

$$\bar{Q}_k(\sigma, \tau) = \frac{1}{\bar{b}_\sigma} \sum_{\vec{y} \in \Omega_\tau} \sum_{\vec{x} \in \Omega_\sigma} \psi_k(\vec{x}) Q_k(\vec{x}, \vec{y}).$$

Also let

$$\bar{Q}_k(\sigma, \sigma) = 1 - \sum_{\tau \neq \sigma} \bar{Q}_k(\sigma, \tau).$$

Note that if $\sigma, \tau \in \tilde{\Omega}_k$, $\tau \neq \sigma$, and $\tau \neq (i, i+1)\sigma$ for every $i \in [1, N)$, then

$$Q_k(\vec{x}, \vec{y}) = Q(\vec{x}, \vec{y}) = 0 \qquad \text{for all } \vec{x} \in \Omega_\sigma, \vec{y} \in \Omega_\tau,$$

and hence

$$\bar{b}_\sigma \bar{Q}_k(\sigma, \tau) = \sum_{\vec{y} \in \Omega_\tau} \sum_{\vec{x} \in \Omega_\sigma} \psi_k(\vec{x}) Q_k(\vec{x}, \vec{y}) = 0.$$

Since Q_k is reversible with respect to ψ_k, it follows that

$$\bar{b}_\sigma \bar{Q}_k(\sigma, \tau) = \bar{b}_\tau \bar{Q}_k(\tau, \sigma), \tag{45}$$

i.e., $\bar{Q}_k$ is reversible with respect to the probability distribution $\{\bar{b}_\sigma : \sigma \in \tilde{\Omega}_k\}$.

In this section, we obtain a lower bound for $Gap(\bar{Q}_k)$ for each $k \in \{0, \ldots, N\}$. To do this, we will compare the chain $(\bar{Q}_k, \bar{b}_\sigma)$ with another chain, which turns out to be a version of the exclusion process studied by Quastel [25], Diaconis and Saloff-Coste [6], and others.

If $\sigma, \tau \in \tilde{\Omega}_k$, $\tau \neq \sigma$, and $\tau = (i, i+1)\sigma$ for some $i \in [1, N)$, then

$$\begin{aligned}
\bar{b}_\sigma \bar{Q}_k(\sigma, \tau) &= \sum_{\vec{y} \in \Omega_\tau} \sum_{\vec{x} \in \Omega_\sigma} \psi_k(\vec{x}) Q_k(\vec{x}, \vec{y}) \\
&= \sum_{\vec{x} \in \Omega_\sigma} \frac{\psi(\vec{x})}{b_k} Q(\vec{x}, (i, i+1)\vec{x}) \\
&= \frac{1}{b_k} \sum_{\vec{x} \in \Omega_\sigma} \psi(\vec{x}) \frac{1}{2N} \rho_{i,i+1}(\vec{x}) \\
&\geq \frac{1}{2Nb_k} \sum_{\vec{x} \in \Omega_\sigma} \psi(\vec{x}) \Theta^{M/N} \qquad \text{(by Lemma 1)} \\
&= \frac{\Theta^{M/N}}{2Nb_k} \psi(\Omega_\sigma) \\
&= \frac{\Theta^{M/N}}{2N} \binom{N}{k}^{-1}.
\end{aligned}$$

In the last equality we have used the fact that

$$\psi(\Omega_\sigma) = \frac{1}{2^N} \qquad \text{and} \qquad b_k = \binom{N}{k} \frac{1}{2^N},$$

as we saw in Section 5.

We will compare the chain $\bar{Q}_k$ with another chain U on $\tilde{\Omega}_k$, which we now define. If $\sigma, \tau \in \tilde{\Omega}_k$, $\sigma \neq \tau$, and $\tau = (i, i+1)\sigma$ for some $i \in [1, N)$, then let $U(\sigma, \tau) = 1/(2k)$. For other distinct σ and τ, let $U(\sigma, \tau) = 0$. Finally, let

$$U(\sigma, \sigma) = 1 - \sum_{\tau \in \tilde{\Omega}_k, \tau \neq \sigma} U(\sigma, \tau).$$

The chain U is reversible with respect to the uniform distribution on $\tilde{\Omega}_k$ (that is, $\bar{b}_\sigma$). We can interpret the chain U as an exclusion process as follows. Think of k particles at k distinct sites on the line segment $\{1, \ldots, N\}$. Each configuration of k particles corresponds to a unique σ in $\tilde{\Omega}_k$ if we interpret

$$\sigma_i = \begin{cases} + & \text{if there is a particle at site } i, \\ - & \text{if site } i \text{ is vacant.} \end{cases}$$

At each time, one of the k particles is chosen at random, and it tries to jump one site to the left or right with equal probability. The attempt fails if the destination site is occupied (this is the exclusion property) or if the jump would leave the segment (i.e., a jump to the left from site 1 fails and a jump to the right from site N fails). The jump succeeds if the destination site is a vacant site of the segment. Then the Markov chain U exactly describes this exclusion process. Diaconis and Saloff-Coste [6, p. 708] showed that the second largest eigenvalue of U is at most $1 - 4/kN^2$, which is equivalent to

$$Gap(U) \geq \frac{4}{kN^2}. \tag{46}$$

From the above, we see that if $\sigma, \tau \in \tilde{\Omega}_k$ and $\sigma \neq \tau$, then

$$\bar{Q}_k(\sigma, \tau) \geq \frac{k}{N} \Theta^{M/N} U(\sigma, \tau). \tag{47}$$

From this, we obtain a corresponding inequality for the Dirichlet forms:

$$\mathscr{E}_{\bar{Q}_k} \geq \frac{k}{N} \Theta^{M/N} \mathscr{E}_U. \tag{48}$$

From this and Eq. (46) and the Comparison Lemma 5, we obtain

Proposition 11. *For each $k = 0, 1, \ldots, N$, the spectral gaps of the chains $\bar{Q}_k$ and U satisfy*

$$Gap(\bar{Q}_k) \geq \frac{k}{N} \Theta^{M/N} Gap(U) \geq \frac{4\Theta^{M/N}}{N^3}.$$

7. A LOWER BOUND FOR $Gap(P_\sigma)$

In this section, we study the mixing rate of the chain P_σ, which is the restriction to Ω_σ of the pure updating chain $\tilde{P}$ defined on Ω. This part is very dependent upon the details of the particular chain under study, and consequently our argument for Example I will be quite separate from that for Example II. In both cases, however, the basic strategy is the same: We show that each component of the product chain P_σ is rapidly mixing. We shall prove the following.

Proposition 12. *P_σ is (uniformly in σ) rapidly mixing on Ω_σ. More precisely,*

$$Gap(P_\sigma) \geq \begin{cases} \dfrac{1}{4(2M-1)^2(N+1)} & \text{for Example I} \\ \dfrac{\exp(-4\beta^*)}{32M^3(N+1)} & \text{for Example II.} \end{cases}$$

Throughout this section, we fix a $\sigma = (\sigma_1, \ldots, \sigma_N)$ in $\tilde{\Omega}_k$. In the notation of Section 4, we have

$$\Omega_\sigma = \mathscr{A} \times \prod_{i=1}^{N} \mathscr{A}_{\sigma_i}.$$

Recall from Section 2 that T_i is the Metropolis chain with respect to h_i on $\mathcal{A}$. For $i \geq 1$, let $T_i^{\sigma_i}$ be the restriction of T_i to $\mathcal{A}_{\sigma_i}$. Then in terms of operators, we have

$$P_\sigma = \frac{1}{N+1} \sum_{i=0}^{N} \underbrace{I \otimes \cdots \otimes I}_{i} \otimes \left(\frac{I + T_i^{\sigma_i}}{2} \right) \otimes \underbrace{I \otimes \cdots \otimes I}_{N-i}. \tag{49}$$

By convention, $T_0^{\sigma_0} = T_0$. By Lemma 4 about product chains, we see that our task is to obtain lower bounds for the spectral gap of every $T_i^{\sigma_i}$. We need to do this separately for our two Examples.

First we shall prove the result for Example I. From the definition of T_i and the relation

$$\frac{h_i(k)}{h_i(j)} = \frac{C^{|k|i/N}/Z_i}{C^{|j|i/N}/Z_i} = C^{(|k|-|j|)i/N},$$

it follows that, for $i \geq 1$,

$$T_i^{\sigma_i}(2j-1, 2j+1) = \begin{cases} \frac{1}{2} & \text{if } \sigma_i = + \text{ and } 1 \leq j \leq M \\ \frac{1}{2} C^{-2i/N} & \text{if } \sigma_i = - \text{ and } -M \leq j \leq -1, \end{cases}$$

$$T_i^{\sigma_i}(2j+1, 2j-1) = \begin{cases} \frac{1}{2} & \text{if } \sigma_i = - \text{ and } -M \leq j \leq -1 \\ \frac{1}{2} C^{-2i/N} & \text{if } \sigma_i = + \text{ and } 1 \leq j \leq M, \end{cases}$$

$$T_i^{\sigma_i}(2j+1, 2j+1) = \tfrac{1}{2} - \tfrac{1}{2} C^{-2i/N} \qquad \text{if } 1 \leq |j| < M,$$

$$T_i^{-}(-1, -1) = T_i^{+}(1, 1) = \tfrac{1}{2},$$

$$T_i^{-}(-2M-1, -2M-1) = T_i^{+}(2M+1, 2M+1) = 1 - \tfrac{1}{2} C^{-2i/N}.$$

Note that $T_i^{\sigma_i}$ differs from T_i only in transitions involving boundary states (i.e., state -1 for $\sigma_i =$ "$-$" and state 1 for $\sigma_i =$ "$+$").

It is easy to check that $T_i^{\sigma_i}$ is the Metropolis chain based on nearest neighbor symmetric random walk on Ω_{σ_i} and with stationary distribution $\pi_i^{\sigma_i}$ given by

$$\pi_i^{\sigma_i}(x) = \frac{2C^{i|x|/N}}{Z_i}, \qquad x \in \mathcal{A}_{\sigma_i}, \tag{50}$$

with

$$Z_i = \sum_{x=1}^{M} 2C^{i(2x+1)/N}.$$

For each $i \neq 0$, $\pi_i^{\sigma_i}$ has a unique local maximum at j such that $|j| = 2M+1$. So by Proposition 6.3 in Diaconis and Saloff-Coste [8], we have

$$Gap(T_i^{\sigma_i}) = 1 - \lambda_1(T_i^{\sigma_i}) \geq \frac{1}{2M^2}. \tag{51}$$

The following result is well known.

Lemma 13. *Let K be the symmetric simple random walk on $\{0, 1, \ldots, L\}$ with $K(0, 0) = K(L, L) = 1/2$. Then*

$$Gap(K) = \cos\frac{\pi}{L+1} \geq \frac{1}{2L^2}. \tag{52}$$

(The equality in this lemma is proven in Feller [11, p. 437] and the inequality is a simple calculation.) As for the spectral gap of T_0, this lemma with $K = T_0$ and $L = 2M - 1$ implies that

$$Gap(T_0) \geq \frac{1}{2(2M-1)^2}. \tag{53}$$

Now we know spectral gaps for each component of the pure updating chain $T_i^{\sigma_i}$; hence we can conclude from Lemma 4 that the spectral gap of the product chain P_σ has the following lower bound:

$$\begin{aligned} Gap(P_\sigma) &= \frac{1}{N+1} \min_{0 \leq i \leq N} Gap([I + T_i^{\sigma_i}]/2) \\ &\geq \frac{1}{N+1}\frac{1}{2}\min\left\{\frac{1}{2(2M-1)^2}, \frac{1}{2M^2}\right\} \\ &= \frac{1}{4(2M-1)^2(N+1)}. \end{aligned} \tag{54}$$

We now turn to the proof for Example II, the mean field Ising model. For now, fix any odd integer M and any $\beta \geq 0$. (Later, we will consider $\beta = \beta_i = i\beta^*/N$.)

For $x \in \mathcal{A} = \{-1, +1\}^M$, let

$$k(x) = \tfrac{1}{2}\left(M + \sum_{j=1}^{M} x[j]\right),$$

that is, $k(x)$ is the number of j's such that $x[j] = +1$. Then the Ising distribution can be written as

$$\pi(x) = \frac{e^{\beta(2k(x)-M)^2/2M}}{Z(\beta)}.$$

We can also write

$$\mathcal{A}_+ = \left\{ x \in \mathcal{A} : k(x) > \frac{M}{2} \right\}.$$

Since M is odd, we have

$$\sum_{x \in \mathcal{A}_+} \pi(x) = \frac{1}{2}.$$

For $x, y \in \mathcal{A}_+$, observe that $\pi(x) < \pi(y)$ if and only if $k(x) < k(y)$.

Write T^+ for the restriction of the Metropolis algorithm to $\mathcal{A}_+$, and consider the chain P defined by $P = (I + T^+)/2$. This is the Markov chain on $\mathcal{A}_+$ determined by

$$P(x, y) = \begin{cases} \dfrac{1}{2M} & \text{if } \|x - y\|_1 = 2, \quad k(y) = k(x) + 1 \\ \dfrac{1}{2M} e^{2\beta(M+1-2k(x))/M} & \text{if } \|x - y\|_1 = 2, \quad k(y) = k(x) - 1 \\ 0 & \text{if } \|x - y\|_1 > 2 \\ 1 - \displaystyle\sum_{z \in \mathcal{A}_+ \setminus \{x\}} P(x, z) & \text{if } x = y, \end{cases}$$

for $x, y \in \mathcal{A}_+$. Observe that if $k(y) = k(x) - 1$, then

$$\begin{aligned} \frac{\pi(y)}{\pi(x)} &= \frac{e^{\beta(2(k(x)-1)-M)^2/2M}}{e^{\beta(2k(x)-M)^2/2M}} \\ &= e^{2\beta(M+1-2k(x))/M} \qquad (55) \\ &\geq e^{2\beta(M+1-2M)/M} \\ &\geq e^{-2\beta}. \qquad (56) \end{aligned}$$

We need a good lower bound for the spectral gap of the above chain. To this end, we consider the following decomposition of its state space:

$$\mathcal{A}_+ = \bigcup_{i=M_0}^{M} \mathcal{A}_{[i]},$$

where $M_0 = (M + 1)/2$, and $\mathcal{A}_{[i]} = \{ x \in \mathcal{A} : k(x) = i \}$. We shall apply Theorem 8 with $\mathcal{S}_i = \mathcal{A}_{[i]}$, $\mathcal{P} = P^2$, and $\mathcal{Q} = P^2$. Since P is reversible, $\mathcal{Q}$ is positive. Then, with the help of Lemma 6, we have

$$Gap(P) \geq \tfrac{1}{4} Gap(P^4) = \tfrac{1}{4} Gap(\mathcal{Q}^{1/2}\mathcal{P}\mathcal{Q}^{1/2})$$

$$\geq \tfrac{1}{4} Gap(\bar{\mathcal{Q}}) \cdot \min_{M_0 \leq i \leq M} Gap(\mathcal{P}_i).$$

We shall prove the following two claims:

$$Gap(\bar{\mathcal{Q}}) \geq \frac{e^{-2\beta}}{2M^2}, \tag{57}$$

$$Gap(\mathcal{P}_i) \geq \frac{e^{-2\beta}}{4M}. \tag{58}$$

From these, it follows that

$$Gap(P) \geq \frac{1}{4} \cdot \frac{e^{-2\beta}}{2M^2} \cdot \frac{e^{-2\beta}}{4M}$$

$$\geq \frac{e^{-4\beta}}{32M^3}. \tag{59}$$

Proof of Eq. (57). In our terminology from the beginning of Section 4, $\theta = 2\pi_\beta$ and

$$b_i = \sum_{x \in \mathcal{A}_{[i]}} \theta(x) = \frac{2 \binom{M}{i} e^{\beta(2i-M)^2/2M}}{Z(\beta)}, \qquad M_0 \leq i \leq M.$$

Notice that $P(x, x) \geq \frac{1}{2}$ for every $x \in \mathcal{A}_+$. Therefore,

$$\mathcal{Q}(x, y) = P^2(x, y) \geq \tfrac{1}{2} P(x, y) \qquad \text{for every } x, y \in \mathcal{A}_+.$$

Observe that for every x in Ω_i there are exactly $M - i$ y's in Ω_{i+1} and i y's in Ω_{i-1} such that $\|x - y\|_1 = 2$. Using this observation and the preceding inequality, we obtain

$$\bar{\mathcal{Q}}(i, i+1) = \frac{1}{b_i} \sum_{x \in \mathcal{A}_{[i]}} \sum_{y \in \mathcal{A}_{[i+1]}} \theta(x) \mathcal{Q}(x, y)$$

$$\geq \frac{1}{b_i} \sum_{x \in \mathcal{A}_{[i]}} \theta(x)(M - i) \times \frac{1}{2} \times \frac{1}{2M}$$

$$= \frac{M - i}{4M} \tag{60}$$

and

$$\begin{aligned}\bar{\mathcal{Q}}(i, i-1) &\geq \frac{1}{b_i} \sum_{x \in \mathcal{A}_{[i]}} \theta(x) i \times \frac{1}{2} \times \frac{1}{2M} e^{-2\beta} \\ &= \frac{i e^{-2\beta}}{4M} \\ &\geq \frac{e^{-2\beta}}{8}. \end{aligned} \tag{61}$$

At this point, we need the following basic result.

Lemma 14. *The distribution $b_{M_0}, \ldots, b_M$ is unimodal. That is, there exists an integer $i_0 \in [M_0, M]$ such that $b_i \leq b_{i+1}$ whenever $M_0 \leq i < i_0$, and $b_i \geq b_{i+1}$ whenever $i_0 \leq i < M$.*

Proof of Lemma 14. Notice that

$$\frac{b_{n-1}}{b_n} = \frac{n}{M-n+1} e^{2\beta(M+1-2n)/M}.$$

Let

$$g(t) = \log t - \log(M - t + 1) + \frac{2\beta(M+1-2t)}{M} \qquad \text{for real } t \in [M_0, M].$$

Then

$$\frac{b_{n-1}}{b_n} = e^{g(n)}.$$

Thus, $b_{n-1} < b_n$ if and only if $g(n) < 0$. We must show that either g does not change sign on $[M_0, M]$, or else g changes sign exactly once (say at t_0) with $g(t) < 0$ for $M_0 \leq t < t_0$ and $g(t) > 0$ for $t_0 < t \leq M$. This will follow if we prove that (a) g is convex on $[M_0, M]$ and (b) $g(M_0) = 0$. Part (b) is immediate. Part (a) holds since

$$g''(t) = -\frac{1}{t^2} + \frac{1}{(M-t+1)^2},$$

which is nonnegative for $t \in [M_0, M]$. ∎

The rest of the proof of Eq. (57) is similar to the proof of Proposition 6.3 in Diaconis and Saloff-Coste [8]. The only difference is in their calculation of Q (here we use their notation, not ours). For the edge $e = (i, i+1)$ $(i \in \{M_0, \ldots, M\})$, define

$$Q(e) = b_i \bar{\mathfrak{Q}}(i, i+1) = b_{i+1} \bar{\mathfrak{Q}}(i+1, i).$$

By Lemma 14, there exists $i_0 \in \{M_0, \ldots, M\}$ such that

$$b_{M_0} \le b_{M_0+1} \le \cdots \le b_{i_0-1} \le b_{i_0} \qquad \text{and} \qquad b_{i_0} \ge b_{i_0+1} \ge \cdots \ge b_M.$$

Equation (61) gives the bound

$$Q(e) = b_{i+1} \bar{\mathfrak{Q}}(i+1, i) \ge b_{i+1} \tfrac{1}{8} e^{-2\beta}.$$

Then the Poincaré inequality tells us that

$$Gap(\bar{\mathfrak{Q}}) \ge \frac{1}{A},$$

where

$$A = \max_i \frac{M - M_0}{b_{i+1} \frac{1}{8} e^{-2\beta}} \left(\sum_{j=M_0}^{i} b_j \right) \left(\sum_{j=i+1}^{M} b_j \right).$$

If $i + 1 \le i_0$, then we bound A using

$$\sum_{j=M_0}^{i} b_j \le (i + 1 - M_0) b_{i+1} \le (M - M_0) b_{i+1} \qquad \text{and} \qquad \sum_{j=i+1}^{M} b_j \le 1.$$

If $i \ge i_0$, then we use

$$\sum_{j=M_0}^{i} b_j \le 1 \qquad \text{and} \qquad \sum_{j=i+1}^{M} b_j \le (M - M_0) b_{i+1}.$$

Thus we obtain

$$A \le 8 e^{2\beta} (M - M_0)^2 < 2 e^{2\beta} M^2,$$

which proves Eq. (57).

Proof of Eq. (58). $\mathcal{P}_i$ is the restriction of P^2 to $\mathcal{A}_{[i]}$. This is trivial if $i = M$, so assume $M_0 \le i \le M - 1$. Consider $x, y \in \mathcal{A}_{[i]}$ such that $\|x - y\|_1 = 2$. Then there exists $z \in \mathcal{A}_{[i+1]}$ such that $\|x - z\|_1 = 1 = \|z - y\|_1$. Hence for such x and y,

$$\begin{aligned}\mathcal{P}_i(x, y) = P^2(x, y) &\geq P(x, z)P(z, y) \\ &= \frac{1}{2M} \times \frac{1}{2M}\, e^{2\beta(M+1-2(i+1))/M} \qquad \text{[by Eq. (55)]} \\ &\geq \frac{e^{-2\beta}}{4M^2}.\end{aligned}$$

We shall compare the chain $\mathcal{P}_i$ with another chain $V_{[i]}$ on $\mathcal{A}_{[i]}$. This new chain is known as the exclusion process with i particles on the complete graph with M sites, interpreting a particle to be at the jth site if $x[j] = +1$. At each time, a randomly chosen particle jumps to a randomly chosen vacant site. (This process is also called "Bernoulli Laplace model of diffusion.") More precisely, the transition probabilities are given by

$$V_{[i]}(x, y) = \begin{cases} \dfrac{1}{i(M-i)} & \text{if } x, y \in \mathcal{A}_{[i]} \quad \text{and} \quad \|x - y\|_1 = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$P^2(x, y) \geq e^{-2\beta} \cdot \frac{i(M-i)}{4M^2} \cdot V_{[i]}(x, y), \qquad \forall\, x \neq y \in \mathcal{A}_{[i]}.$$

Diaconis and Saloff-Coste [6, p. 709] prove that

$$Gap(V_{[i]}) = \frac{M}{i(M-i)}.$$

Notice that $\mathcal{P}_i$ and $V_{[i]}$ both have the same equilibrium distribution on $\mathcal{A}_{[i]}$, namely, the uniform distribution. Therefore,

$$\begin{aligned} Gap(\mathcal{P}_i) &\geq e^{-2\beta} \cdot \frac{i(M-i)}{4M^2} \cdot Gap(V_{[i]}) \\ &= e^{-2\beta} \cdot \frac{i(M-i)}{4M^2} \cdot \frac{M}{i(M-i)} \\ &= \frac{e^{-2\beta}}{4M}. \end{aligned} \tag{62}$$

This proves Eq. (57).

Now Eq. (59) tells us that

$$Gap\left(\frac{I + T_i^{\sigma_i}}{2}\right) \geq \frac{\exp(-4\beta_i)}{32M^3} \geq \frac{\exp(-4\beta^*)}{32M^3} \tag{63}$$

for $i = 1, \ldots, N$. Next, observe that the chain T_0 is the same as the base chain K on $\mathcal{A}^{\text{Ising}}$ defined in Section 1. But this is exactly the random walk on the hypercube $\{-1,$

$+1\}^M$ from Example 3.2 of Diaconis and Saloff-Coste [7], where its gap is shown to be $2/M$. Hence

$$Gap\left(\frac{I+T_0}{2}\right) = \frac{1}{M}. \tag{64}$$

Finally, the bound of Proposition 12 follows from Lemma 4, using Eqs. (63) and (64) in the definition of P_σ in Eq. (49).

ACKNOWLEDGMENTS

The research of the first author was partially supported by a Research Grant from the Natural Sciences and Engineering Research Council of Canada, and as well as by part of a group grant from Communications and Information Technology Ontario. The second author is grateful to Professor N. Madras for his guidance and support during his graduate study and postdoc research and to Professors G. L. O'Brien and T. S. Salisbury for their encouragement, support, and comments, which have greatly improved the presentation of this work. Both authors want to thank Professor Dana Randall for discussions and comments and for raising a challenging question for further research. We also thank an anonymous referee for numerous detailed comments that helped us to improve our exposition.

REFERENCES

[1] K. Binder and D. W. Heermann, Monte Carlo simulation in statistical physics, Springer-Verlag, Berlin, 1992.

[2] S. Caracciolo, A. Pelissetto, and A. D. Sokal, Two remarks on simulated tempering, unpublished manuscript, 1992.

[3] M.-F. Chen, Estimation of spectral gap for Markov chains, Acta Math Sin New Ser 12(4) (1996), 337–360.

[4] M.-F. Chen and F.-Y. Wang, Estimation of spectral gap for elliptic operators, Trans AMS 349 (1997), 1239–1267.

[5] C. Cooper, M. E. Dyer, A. M. Frieze, and R. Rue, Mixing properties of the Swendsen-Wang process on the complete graph and narrow grids, J Math Phys 41 (2000), 1499–1527.

[6] P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains, Ann Appl Probab 3 (1993), 696–730.

[7] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, Ann Appl Probab 6 (1996), 695–750.

[8] P. Diaconis and L. Saloff-Coste, What do we know about the Metropolis algorithm? J Comput Syst Sci 57 (1998), 20–36.

[9] P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains, Ann Appl Probab 1 (1991), 36–61.

[10] R. Durrett, Probability, theory and examples, 2nd edition, Wadsworth, Belmont, CA, 1996.

[11] W. Feller, An introduction to probability theory and its applications, Vol. I, 3rd edition, Wiley, Toronto, 1968.

[12] D. Frenkel and B. Smit, Understanding molecular simulation, Academic Press, San Diego, 1996.

[13] C. J. Geyer, "Markov chain Monte Carlo maximum likelihood," Computing science and statistics: Proceedings of the 23rd Symposium on the Interface, Ed. E. M. Keramidas, 156–163. Interface Foundation, Fairfax Station, VA, 1991.

[14] C. J. Geyer and E. A. Thompson, Annealing Markov chain Monte Carlo with applications to ancestral inference, J Am Stat Assoc 90 (1995), 909–920.

[15] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds., Markov chain Monte Carlo in practice, Chapman and Hall, London, 1996.

[16] V. K. Gore and M. R. Jerrum, The Swensen-Wang process does not always mix rapidly, J Stat Phys 97 (1999), 67–86.

[17] M. Luby, D. Randall, and A. Sinclair, Markov chain algorithms for planar lattice structures, Proc 36th Annu IEEE Symp Foundations of Computer Science, IEEE Computer Society Press, Los Alamitos, CA, 1995, pp. 150–159.

[18] N. Madras, "Umbrella sampling and simulated tempering," Numerical methods for polymeric systems, Ed. S. G. Whittington, IMA Volume in Mathematics and Its Applications 102, 19–32. Springer-Verlag, New York, 1998.

[19] N. Madras and M. Piccioni, Importance sampling for families of distributions, Ann Appl Probab 9 (1999), 1202–1225.

[20] N. Madras and D. Randall, Markov chain decomposition for convergence rate analysis, Ann Appl Probab 12 (2002), 581–606.

[21] E. Marinari and G. Parisi, Simulated tempering: A new Monte Carlo scheme, Europhys Lett 19 (1992), 451–458.

[22] R. A. Martin and D. Randall, Sampling adsorbing staircase walks using a new Markov chain decomposition method, Proc 41st Annu IEEE Symp Foundations of Computer Science, IEEE Computer Society Press, Los Alamitos, CA, 2000, pp. 492–502.

[23] R. Neal, Sampling from multimodal distributions using tempered transitions, Stat Comput 6 (1996), 353–366.

[24] E. Orlandini, "Monte Carlo study of polymer systems by multiple Markov chain method," Numerical methods for polymeric systems, Ed., S. G. Whittington, IMA Volume in Mathematics and Its Applications 102, 33–57. Springer-Verlag, New York, 1998.

[25] J. Quastel, Diffusion of colour in the simple exclusion process, Commun Pure Appl Math 42 (1992), 115–137.

[26] J. S. Rosenthal, Convergence rates for Markov chains, SIAM Rev 37 (1995), 387–405.

[27] A. Sinclair, Algorithms for random generation and counting: A Markov chain approach, Birkhäuser, Boston, 1993.

[28] R. H. Swendsen and J.-S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, Phys Rev Lett 58 (1987), 86–88.

[29] Z. Zheng, Analysis of swapping and tempering Monte Carlo algorithms, Dissertation, York University, 1999.

[30] Z. Zheng, On swapping and simulated tempering algorithms, Stoch Proc Appl (2002), to appear.