METROPOLIS ALGORITHM AND EQUIENERGY SAMPLING FOR TWO MEAN FIELD SPIN SYSTEMS

FEDERICO BASSETTI AND FABRIZIO LEISEN

ABSTRACT. In this paper we study the Metropolis algorithm in connection with two mean–field spin systems, the so called mean–field Ising model and the Blume–Emery–Griffiths model. In both this examples the naive choice of proposal chain gives rise, for some parameters, to a slowly mixing Metropolis chain, that is a chain whose spectral gap decreases exponentially fast (in the dimension N of the problem). Here we show how a slight variant in the proposal chain can avoid this problem, keeping the mean computational cost similar to the cost of the usual Metropolis. More precisely we prove that, with a suitable variant in the proposal, the Metropolis chain has a spectral gap which decreases polynomially in $1/N$. Using some symmetry structure of the energy, the method rests on allowing appropriate jumps within the energy level of the starting state, and it is strictly connected to both the *small world Markov chains* of [15, 16] and to the *equi-energy sampling* of [22] and [26].

1. INTRODUCTION.

The Metropolis algorithm, introduced in [29] and later generalized in [18], is currently (together with other Monte Carlo Markov Chain methods) one of the most used simulation techniques both in statistics and in physics. See, among others, [33, 32, 39, 17, 35, 34, 25, 6].

In a finite setting the Metropolis algorithm can be described as follows. Suppose that, given a probability $\pi(x)$ on a finite set $\mathcal{X}$, want to approximate

$$\mu = \sum_x f(x)\pi(x), \tag{1.1}$$

for $f : \mathcal{X} \to \mathbb{R}$. As a first step, take a reversible Markov chain $K(x, y)$ (the proposal chain) on $\mathcal{X}$ and change its output in order to have a new chain with stationary distribution π. This can be achieved by constructing a new (π–reversible) chain

$$M(x,y) = \begin{cases} K(x,y)A(x,y) & x \neq y \\ K(x,x) + \sum_{z\neq x} K(x,z)(1 - A(x,z)) & x = y \end{cases} \tag{1.2}$$

where $A(x, y) := \min(\frac{\pi(y)K(y,x)}{\pi(x)K(x,y)}, 1)$. Then, the *metropolis estimate* of μ is given by

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(Y_i), \tag{1.3}$$

where Y_0 is generated from some initial distribution π_0 and $Y_1, \ldots, Y_n$ from $M(x, y)$.

It is clear that, from a computational point of view, the speed of convergence to the stationary distribution and the (asymptotic) variance of the estimate are two very important features of the Markov chain M.

It is well-known that in some situation a Markov chain can converge very slowly to its stationary distribution and, moreover, that the asymptotic variance of the estimate (1.3) can be much bigger than the variance of f, i.e. $Var_\pi(f) := \sum_x (f(x) -$

Key words and phrases. asymptotic variance, Chain decomposition theorem, fast/slowly mixing chain, mean-field Ising model, Metropolis, spectral gap analysis.

$\mu)^2\pi(x)$, which is equal to the asymptotic variance of the crude Montecarlo estimator. In these cases (1.3) turns out to be a very inefficient estimate of μ.

For the Metropolis chain a classical situation in which the convergence is slow (and the variance big) is when the target distribution π has many peaks and K is somehow too "local".

This is well known in statistical physics, where, typically, a distribution of a system with energy function h and in thermal equilibrium at temperature T is described by the Gibbs distribution

$$\pi_{h,T}(x) = \exp\{-h(x)/T\} Z_T^{-1}$$

with $Z_T = \sum_x \exp\{-h(x)/T\}$. In point of fact, the Metropolis algorithm has been proposed in [29] to compute average with respect to such distributions. Indeed, if h is nice, the Metropolis algorithm is very efficient, but it can perform very poorly if the energy has many local minima separated by high barriers that cannot be crossed by the proposal moves K. This problem can be bypassed, for specific energy, designing appropriate moves that have higher chance to cut across the energy barrier (see, e.g, [4, 5]), or constructing clever alternative approaches to the problem, for instance using a reparametrization of the problem (see, e.g., [12, 13]) or using auxiliary variables (see, e.g., [40, 9, 1, 30]). A different kind of solution has been proposed in [14] and in [28] by introducing the so called *simulated tempering*, which essentially means that T is changed (stochastically or not) to flatten h. A remarkable variant of these methods is the *parallel tempering*, see, for instance, [19]. More recently new algorithms based on the so called *equi–energy levels sampling* have been proposed (see [26] and [22]). In particular, the algorithm proposed in [22] relies on the so–called equi-energy jump, which enables the chain to reach regions of the sample space with energy close to the one of the starting state, but that may be separated by steep energy barriers. In point of fact, even if, according to some simulations, the method seems to be efficient nothing has been formally proved. Finally, let us mention a recent algorithm, called *small world Markov chains* (see [15, 16]), that combine a local chain with long jumps. In these papers, it has been shown that a simple modification of the proposal mechanism results in faster convergence of the chain. That mechanism, which is based on an idea from the field of small-world networks, amounts to adding occasional wild proposals to any local proposal scheme.

In the present paper we study two simple examples: the so called mean field Ising model and the mean field Blume–Emery–Griffiths model. As for the former, it is well-known that the usual choice of K gives rise, for low temperature, to a slowly mixing Metropolis chain (see, e.g., [26]). Here we show that a slight variant in the proposal chain can completely solve this problem, keeping the mean computational cost similar to the cost of the usual Metropolis. The idea again rests on allowing appropriate jumps in the same energy level of the starting state. As for the Blume–Emery–Griffyths mean–field model, we first show that there is a critical region of the parameters space for which the naive Metropolis chain is slowly mixing. Then we show how one can modify the proposal chain in order to obtain a better mixing for the Metropolis chain. The present paper should be intended as a further step in the direction of a better mathematical understanding of both small world Markov chains and equi-energy sampling.

The rest of the paper is organized as follows. In Section 2 some general considerations are given. In Section 3 some basic tools concerning Markov chain, which will be used in the paper, are reviewed. Section 4 contains a warming up example. In Section 5 the mean field Ising model is treated, while Section 6 deals with the more complex case of the mean field Blume-Emery-Griffiths model. All the proofs are deferred to the Appendix.

2. A general strategy

In an abstract setting, what we shall do in the next examples can be summarized as follows. Let $\mathcal{G}$ be a group acting on $\mathcal{X}$ for which

$$\pi(x) = \pi(g(x)) \qquad \forall\ x\ \in \mathcal{X},\ \forall\ g\ \in \mathcal{G}. \tag{2.1}$$

For every x in $\mathcal{X}$ let $O_x := \{y = g(x) : g \in \mathcal{G}\}$ be the orbit of x (of course if y belongs to O_x then $O_x = O_y$).

Assume now that we have a reversible Markov chain $K_E(x, y)$ (the proposal) on $\mathcal{X}$ and suppose that the Metropolis chain M_E with proposal K_E is slowly mixing (see next section for more details). To speed up the mixing one can try to exploit (2.1) by taking a proposal of the following form:

$$K_\epsilon(x, y) = \epsilon K_E(x, y) + (1 - \epsilon)K_{\mathcal{G}}(x, y) \tag{2.2}$$

where

$$K_{\mathcal{G}}(x, y) = \sum_{z \in O_x} q_x(z)\mathbb{I}_z(y),$$

$0 < q_x(z) < 1$ and $\sum_{z \in O_x} q_x(z) = 1$.

In point of fact, usually K_E is "local"; for instance frequently

$$K_E(x, y) = 0$$

whenever $y \neq x$ belongs to O_x, hence with $K_{\mathcal{G}}$ we are adding "long" jumps to the chain. Moreover, note that if K_E is such that $K_E(x, g(x)) = K_E(g(x), x)$, for every x in $\mathcal{X}$ and g in $\mathcal{G}$, then the Metropolis always accepts the move $x \to g(x)$ and

$$M(x, g(x)) = \epsilon K_E(x, g(x)) + (1 - \epsilon)q_x(g(x)).$$

In particular this holds when K_E is symmetric.

The heuristics under (2.2) is to combining small world Markov chains and equienergy sampling.

Before presenting some examples in which one can actually improve the performances of the Metropolis chain using this idea, we collect in the next section some useful facts concerning Markov chains.

3. Preliminaries

Let $P(x, y)$ be a reversible and ergodic Markov chain on the finite set $\mathcal{X}$ with (unique) stationary distribution $p(x)$. Thus, $p(x)P(x, y) = p(y)P(y, x)$. Let $L^2(p) = \{f : \mathcal{X} \to \mathbb{R}\}$ with $< f, g >_p = E_p(fg) = \sum_x f(x)g(x)p(x)$. Reversibility is equivalent to $P : L^2 \to L^2$ being self–adjoint. Here $Pf(x) = \sum_y f(y)P(x, y)$. The spectral theorem implies that P has real eigenvalues $1 = \lambda_0(P) > \lambda_1(P) \geq \lambda_2(P) \geq \cdots \geq \lambda_{|\mathcal{X}|-1}(P) > -1$ with orthonormal basis of eigen–functions $\psi_i : \mathcal{X} \to \mathbb{R}$ ($P\psi_i(x) = \lambda_i\psi_i(x)$, $< \psi_i, \psi_j >_p = \delta_{ij}$).

3.1. Spectral gap, variance and speed of convergence.

A very important quantity related to the eigenvalues is the spectral gap, defined by

$$Gap(P) = 1 - \max\{\lambda_1, |\lambda_{|\mathcal{X}|-1}|\}.$$

It turns out that the spectral gap is a good index to measure the mixing of a chain. To better understand this point, assume that f belongs to $L^2(p)$ and write $f(x) = \sum_{i \geq 0} a_i\psi_i(x)$ (with $a_i =< f, \psi_i >_p$). Now let Y_0 be chosen form some distribution p_0 and $Y_1, \dots, Y_n$ be a realization of the $P(x, y)$ chain, then

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(Y_i)$$

has asymptotic variance given by

$$AVar(f,p,P) := \lim_{n\to+\infty} n \cdot Var(\hat{\mu}_n) = \sum_{k\geq 1} |a_k|^2 \frac{1+\lambda_k}{1-\lambda_k}.$$

See, for instance, Theorem 6.5 in Chapter 6 of [3]. From the last expression, the classical inequality

$$AVar(f,p,P) \leq \frac{2}{1-\lambda_1} Var_p(f), \tag{3.1}$$

follows easily. The last inequality is the usual way of relating spectral gap to asymptotic variance and, hence, to the efficency of a chain.

The spectral gap is very important also to give bounds on the speed of convergence to the stationary distribution. For example, if $\|\cdot\|_{TV}$ denotes the total variation norm, one has

$$\|\delta_x P^k - p\|_{TV}^2 = \left(\sup_{A\subset\mathcal{X}} |P^k(x,A) - p(x)| \right)^2 \leq \frac{1-p(x)}{4p(x)} (\max\{\lambda_1, |\lambda_{|\mathcal{X}|-1}|\})^{2k}$$

See, e.g., Proposition 3 in [7]. Another classical bound is

$$\|p_0 P^k/p - 1\|_{2,p} \leq Gap(P^k) \|p_0/p - 1\|_{2,p}$$

valid for every probability p_0. See, for instance, [39].

Roughly speaking one can say that a sequence of Markov chains defined on a sequence of state space $\mathcal{X}_N$ is slowly mixing (in the dimension of the problem N) if the spectral gap decreases exponentially fast in N.

3.2. **Cheeger's inequality.** As already recalled, problems of slowly mixing typically occur when π has two or more peaks and the chain K can only move in a neighborhood of the starting peak. Usually this phenomenon is called bottleneck. A powerful tool to detect the presence of a bottleneck is the conductance and the related Cheeger's inequality. Recall that the conductance of a chain P with stationary distribution p is defined by

$$h = h(p,P) := \inf_{A\ :p(A)\leq\frac{1}{2}} \frac{1}{p(A)} \sum_{x\in A, y\in A^c} p(x)P(x,y),$$

and the well-known Cheeger's inequality is

$$1 - 2h \leq \lambda_1(P) \leq 1 - \frac{h^2}{2}. \tag{3.2}$$

See, for instance, [3, 37, 7]. Note that, since P is reversible,

$$h \leq \frac{1}{p(A)} \sum_{x\in A} \sum_{y\in A^c} p(x)P(x,y) = \frac{1}{p(A)} \sum_{x\in A} \sum_{y\in A^c} p(y)P(y,x) \tag{3.3}$$

for every A such that $p(A) \leq 1/2$.

3.3. **Chain decomposition theorem.** In this subsection we briefly describe a useful technique to obtain bounds on the spectral gap: the so called chain decomposition technique. Following [16] assume that $A_1, \dots, A_m$ is a partition of $\mathcal{X}$. Moreover, for each $i = 1, \dots, m$, define a new Markov chain on A_i by setting

$$P_{A_i}(x,y) := P(x,y) + \mathbb{I}_x(y) \left(\sum_{z\in A_i^c} P(x,z) \right) \qquad (x, y \in A_i).$$

P_{A_i} is a reversible chain on the state space A_i with respect to the probability measure

$$p_i(x) := p(x)/p(A_i).$$

The movement of the original chain among the "pieces" $A_1, \ldots, A_m$ can be described by a Markov chain with state space $\{1, \ldots, m\}$ and transition probabilities

$$P_H(i,j) := \frac{1}{2p(A_i)} \sum_{x \in A_i, y \in A_j} P(x,y)p(x)$$

for $i \neq j$ and

$$P_H(i,i) := 1 - \sum_{j \neq i} P_H(i,j),$$

which is reversible with stationary distribution

$$\bar{p}(i) := p(A_i).$$

A variant of a result of Caracciolo, Pelisetto and Sokal (published in [27]), states that

$$Gap(P) \geq \frac{1}{2} Gap(P_H) \left(\min_{i=1,\ldots,m} Gap(P_{A_i}) \right) \tag{3.4}$$

holds true, see Theorem 2.2 in [16]. Other results about chain decompositions can be found, for instance, in [20].

In the next very simple example we shall show how this technique can be used, starting from a slowly mixing chain, to suggest how to modify the proposal chain in order to obtain a fast mixing chain.

4. Warming up example

Set $\mathcal{X} = \{-N, -N+1, \ldots, 0, 1, \ldots, N\}$ and define a probability measure on $\mathcal{X}$ by

$$\pi(x) = \frac{(\theta - 1)\theta^{|x|}}{2\theta^{N+1} + 1 - \theta},$$

θ being a given parameter bigger than 1. Here we can consider $\mathcal{G} = \{+1, -1\}$ (with group operation given by the usual product) acting on $\mathcal{X}$ by $g(x) = gx$, hence $O_x = \{x, -x\}$.

Now let K_E be a chain defined by

$$\begin{aligned} &K_E(x, x+1) = 1/2 \qquad x \neq N \\ &K_E(x, x-1) = 1/2 \qquad x \neq -N \\ &K_E(N,N) = K_E(-N,-N) = 1/2 \\ &K_E(x,y) = 0 \quad \text{otherwise} \end{aligned}$$

and denote by M_E the Metropolis chain with stationary distribution π derived by K_E. It is clear that in this case $K_E(x,y) = 0$ whenever y belongs to O_x. In this example it is very easy to bound the conductance on M_E, indeed, taking $A = \{-N, \ldots, -1\}$, by (3.3), it follows that

$$h(\pi, M_E) \leq \frac{\pi(0)}{1 - \pi(0)}.$$

Hence,

$$h(\pi, M_E) \leq C\theta^{-N},$$

and then (3.2) yields

$$1 - \lambda_1 \leq 2C\theta^{-N}.$$

This means that, if f is such that $a_1 \neq 0$ and $\theta > 1$, then the asymptotic variance of f blows up exponentially fast, indeed

$$AVar(f, \pi, M_E) \geq 2Ce^{\log(\theta)N}.$$

Now, instead of K_E consider

$$K_\epsilon(x,y) = (1-\epsilon)K_E(x,y) + \epsilon \mathbb{I}_{\{-x\}}(y)$$

and let $M^{(\epsilon)}$ be the Metropolis chain derived by K_ϵ. Decompose $\mathcal{X}$ as follows

$$\mathcal{X} = A_1 \cup A_2 \cdots \cup A_N$$

with $A_1 = \{-1, 0, 1\}$ and $A_i = \{x \in \mathcal{X} : |x| = i\}$, for $i > 1$. Moreover let

$$\bar{\pi}(i) = \pi(A_i) = \begin{cases} (2\theta+1)/Z & \text{for } i = 1 \\ 2\theta^i/Z & \text{for } i > 1 \end{cases}$$

where

$$Z = \frac{2\theta^{N+1} + 1 - \theta}{(\theta - 1)}$$

and set

$$M_H^{(\epsilon)}(i,j) = \frac{1}{2\pi(A_i)} \sum_{l \in A_i, m \in A_j} M^{(\epsilon)}(l,m)\pi(l), \qquad M_H^{(\epsilon)}(i,i) = 1 - \sum_{j \neq i} M_H^{(\epsilon)}(i,j).$$

For $i \neq 1, N$, one has

$$M_H^{(\epsilon)}(i, i+1) = \frac{1}{2\pi(A_i)}[M^{(\epsilon)}(i,i+1)\pi(i) + M^{(\epsilon)}(-i,-i-1)\pi(-i)]$$

and, since $\pi(i) = \pi(-i)$ and $\pi(i+1) \geq \pi(i)$

$$M_H^{(\epsilon)}(i,i+1) = \frac{1-\epsilon}{4}.$$

In the same way it is easy to see that

$$\begin{aligned}
&M_H^{(\epsilon)}(i,i-1) = \frac{1-\epsilon}{4\theta}, \qquad i \neq 1, N \\
&M_H^{(\epsilon)}(i,i) = 1 - \frac{1-\epsilon}{4}(1+\theta^{-1}) \qquad i \neq 1, N \\
&M_H^{(\epsilon)}(N,N-1) = \frac{1-\epsilon}{4\theta} \qquad M_H^{(\epsilon)}(N,N) = 1 - \frac{1-\epsilon}{4\theta} \\
&M_H^{(\epsilon)}(1,2) = \frac{1-\epsilon}{4(1+1/(2\theta))} \qquad M_H^{(\epsilon)}(1,1) = 1 - \frac{1-\epsilon}{4(1+1/(2\theta))}.
\end{aligned}$$

Moreover, for every $i \neq 1$, $M_{A_i}^{(\epsilon)}$ in matrix form is given by

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix},$$

and hence

$$Gap(M_{A_i}^{(\epsilon)}) = 1 - |1 - 2\epsilon|.$$

While $M_{A_1}^{(\epsilon)}$ is given by

$$\begin{pmatrix} (2\theta-1)(1-\epsilon)/(2\theta) & (1-\epsilon)/(2\theta) & \epsilon \\ (1-\epsilon)/2 & \epsilon & (1-\epsilon)/2 \\ \epsilon & (1-\epsilon)/(2\theta) & (2\theta-1)(1-\epsilon)/(2\theta) \end{pmatrix}$$

and hence

$$Gap(M_{A_i}^{(\epsilon)}) = k(\theta, \epsilon) > -1.$$

Moreover, since

$$\begin{aligned}
&\min\left[\min_{i \neq 1,N}(M_H^{(\epsilon)}(i, i\pm 1)), M_H^{(\epsilon)}(1,2), M_H^{(\epsilon)}(N,N-1)\right] \\
&\qquad \geq \min\left[(1-\epsilon)/(4\theta), \frac{1-\epsilon}{4(1+1/(2\theta))}\right] =: m(\epsilon,\theta) > 0
\end{aligned}$$

and $\bar{\pi}(i) \leq 3\bar{\pi}(j)$ for every $i < j$, Lemma A.1 in the appendix yields that

$$1 - \lambda_1(M_H^{(\epsilon)}) \geq \frac{m(\epsilon, \theta)}{3N^2}.$$

In the same way, since $M_H^{(\epsilon)}(i, i+1) + M_H^{(\epsilon)}(i, i-1) \leq (1-\epsilon)M(\theta)/4$, with $M(\theta) = \max(1 + \theta^{-1}, 2\theta/(2\theta+1)) \leq 2$, inequality (A.1) in the Appendix yields that

$$\lambda_{N-1}(M_H^{(\epsilon)}) \geq 1 - \frac{1-\epsilon}{2} \geq \frac{1+\epsilon}{2}.$$

Hence

$$Gap(M_H^{(\epsilon)}) \geq \frac{m(\epsilon, \theta)}{3N^2}$$

and (3.4) yield

$$Gap(M^{(\epsilon)}) \geq \frac{h(\theta, \epsilon)}{N^2}$$

for a suitable h. This shows that $M^{(\epsilon)}$ is fast mixing for every $\epsilon > 0$ and for every $\theta > 1$ while M_E is slowly mixing for every $\theta > 1$.

5. The mean field Ising model

Let $\mathcal{X} = \{-1, 1\}^N$, N being an even integer. For every $\beta > 0$ let $\pi = \pi_{\beta,N}$ be a probability on $\mathcal{X}$ defined by

$$\pi(x) = \pi_{\beta,N}(x) := \exp\left\{\beta \frac{S_N^2(x)}{2N}\right\} Z_N^{-1}(\beta) \qquad (x \in \mathcal{X})$$

where

$$Z_N(\beta) = Z_N := \sum_{x \in \mathcal{X}} \exp\left\{\beta \frac{S_N^2(x)}{2N}\right\}$$

is the normalization constant ("partition function") and

$$S_N(x) := \sum_{i=1}^{N} x_i \qquad x = (x_1, \ldots, x_N).$$

This is the so called *mean field Ising model*, or *Curie-Weiss model*, in which every particle i, with spin x_i, interacts equally with every other particle. It is probably the most simple but also the most studied example of spin system on a complete graph. The usual Metropolis algorithm uses as proposal chain

$$K_E(x, y) = \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}_{\{x^{(j)}\}}(y)$$

where $x^{(j)}$ denotes the vector $(x_1, \ldots, -x_j, \ldots, x_N)$. It has been proved in [26] that, whenever $\beta > 1$,

$$1 - \lambda_1 \leq Ce^{-D^2 N}$$

where λ_1 is the first eigenvalues smaller than 1 of the Metropolis chain M_E derived K_E. This yields that the variance of an estimator obtained from this Metropolis algorithm can blow up exponentially fast in N.

The aim of this section is to show how one can construct a different Metropolis chain avoiding this problem. In the notation of Section 2, we consider

$$\mathcal{G} = \mathcal{S}_N \times \{+1, -1\}$$

($\mathcal{S}_N$ being the symmetric group of order N) and we define the action of $\mathcal{G}$ on $\mathcal{X} = \{-1, 1\}^N$ by

$$g(x) = (e \cdot x_{\sigma(1)}, \ldots, e \cdot x_{\sigma(N)}) \qquad g = (\sigma, e).$$

In order to introduce a new proposal, it is useful to write $\mathcal{X}$ as the union of its "energy sets", that is

$$\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_2 \cup \mathcal{X}_4 \cup \cdots \cup \mathcal{X}_N$$

where

$$\mathcal{X}_i := \{x \in \mathcal{X} : |S_N(x)| = i\} \qquad (i = 0, 2, \dots, N).$$

Note that energy takes only even values and that $O_x = \mathcal{X}_{|S_N(x)|}$. Moreover, for $i \neq 0$, set

$$\mathcal{X}_i^+ := \{x \in \mathcal{X} : S_N(x) = i\} \text{ and } \mathcal{X}_i^- := \{x \in \mathcal{X} : S_N(x) = -i\}.$$

The new proposal chain will be

$$\begin{aligned} &K(x,y) = p_1 K_E(x,y) + (1-p_1)K_0(x,y) \qquad \text{if } x \in \mathcal{X}_0 \\ &K(x,y) = p_1 K_E(x,y) + p_2 \mathbb{I}_{\{-x\}}(y) + (1-p_1-p_2)K_i(x,y) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{if } x \in \mathcal{X}_i,\ i \neq 0 \end{aligned} \tag{5.1}$$

where p_1, p_2 belong to $(0,1)$, $p_1 + p_2 < 1$, and

$$K_i(x,y) = \mathbb{I}_{\mathcal{X}_i^+}\{x\} K_i^+(x,y) + \mathbb{I}_{\mathcal{X}_i^-}\{x\} K_i^-(x,y) \qquad (i \neq 0).$$

We shall assume that $K_i^\pm$ (K_0, respectively) are irreducible, symmetric and aperiodic chains on $\mathcal{X}_i^\pm$ ($\mathcal{X}_0$, respectively).

As a leading example we shall take

$$\begin{aligned} K_0(x,y) &= \frac{1}{\binom{N}{N/2}} \qquad y \in \mathcal{X}_0 \\ K_i^\pm(x,y) &= \frac{1}{\binom{N}{(N-i)/2}} \qquad y \in \mathcal{X}_i^\pm, \end{aligned} \tag{5.2}$$

that is: a realization of a chain $K_i^\pm$ (K_0, respectively) is simply a sequence of independent uniform random sampling from $\mathcal{X}_i^\pm$ ($\mathcal{X}_0$, respectively).

Remark 1. *Note that (5.2) is the (n,k)-Bose-Einstein distribution with $n = (N+i)/2$ and $k = (N-i)/2 + 1$ and recall that there is a very easy way to directly generate Bose-Einstein configurations. One may place n balls sequentially into k boxes, each time choosing a box with probability proportional to its current content plus one. Starting from the empty configuration this results in a Bose-Einstein distribution for every stage.*

Now let M be the Metropolis chain defined by the transition kernel (1.2) with K as in (5.1), i.e. for every x in $\mathcal{X}_i^\pm$ $(i \neq 0)$

$$M(x,y) = \begin{cases} \frac{p_1}{N} \min\left(1, \frac{\pi(y)}{\pi(x)}\right) & \text{if } y = x^{(j)}, \quad j = 1...N \\ p_2 & \text{if } y = -x \\ (1-p_1-p_2)K_i^\pm(x,y) & \text{if } y \in \mathcal{X}_i^\pm, y \neq x \\ 1 - \sum_{z \neq x} M(x,z) & \text{if } y = x \end{cases}$$

while for x in $\mathcal{X}_0$

$$M(x,y) = \begin{cases} \frac{p_1}{N} \min\left(1, \frac{\pi(y)}{\pi(x)}\right) & \text{if } y = x^{(j)}, \quad j = 1...N \\ (1-p_1)K_0(x,y) & \text{if } y \in \mathcal{X}_0, y \neq x \\ 1 - \sum_{z \neq x} M(x,z) & \text{if } y = x. \end{cases}$$

By construction M is an aperiodic, irreducible and reversible chain with stationary distribution π. Then, when (5.2) holds true,

$$M(x,y)=\begin{cases} \frac{p_1}{N}\min\left(1,\frac{\pi(y)}{\pi(x)}\right) & \text{if } y=x^{(j)},\quad j=1...N \\ p_2 & \text{if } y=-x \\ (1-p_1-p_2)\frac{1}{\binom{N}{(N-i)/2}} & \text{if } y\in\mathcal{X}_i^{\pm}, y\neq x \\ 1-\sum_{z\neq x}M(x,z) & \text{if } y=x \end{cases}$$

for x in $\mathcal{X}_i^{\pm}$ $(i\neq 0)$, while if x belongs to $\mathcal{X}_0$

$$M(x,y)=\begin{cases} \frac{p_1}{N}\min\left(1,\frac{\pi(y)}{\pi(x)}\right) & \text{if } y=x^{(j)},\quad j=1...N \\ (1-p_1)\frac{1}{\binom{N}{N/2}} & \text{if } y\in\mathcal{X}_0, y\neq x \\ 1-\sum_{z\neq x}M(x,z) & \text{if } y=x. \end{cases}$$

In order to bound the spectral gap of M we shall use the decomposition theorem described in Subsection 3.3. To this end, for every $i=0,2,\dots,N$ and every $j\neq i$ set

$$\bar{P}(i,j):=\frac{1}{2\pi(\mathcal{X}_i)}\sum_{x\in\mathcal{X}_i}\sum_{y\in\mathcal{X}_j}M(x,y)\pi(x)$$

and

$$\bar{P}(i,i):=1-\sum_{j\neq i}\bar{P}(i,j).$$

As already noted, $\bar{P}$ is a reversible chain on $\{0,2,\dots,N\}$ with stationary distribution

$$\bar{\pi}(i):=\pi(\mathcal{X}_i).$$

Moreover define for every $i=0,2,\dots,N$ a chain on $\mathcal{X}_i$ setting

$$P_{\mathcal{X}_i}(x,y):=M(x,y)+\mathbb{I}_x(y)\left(\sum_{z\in\mathcal{X}_i^c}M(x,z)\right)$$

where both x and y belong to $\mathcal{X}_i$. In the same way, define chains on $\mathcal{X}_i^+$ and $\mathcal{X}_i^-$ for $i=2,\dots,N$ setting

$$P_{\mathcal{X}_i^{\pm}}(x,y):=P_{\mathcal{X}_i}(x,y)\quad (y\neq x,\ x,y\in X_i^{\pm})$$

and

$$P_{\mathcal{X}_i^{\pm}}(x,x):=1-\sum_{y\in\mathcal{X}_i^{\pm}y\neq x}P_{\mathcal{X}_i}(x,y).$$

These chains are reversible on $\mathcal{X}_i$ ($\mathcal{X}_i^{\pm}$, respectively) and have as stationary distributions

$$\pi_{\mathcal{X}_i}(x):=\frac{\pi(x)}{\pi(\mathcal{X}_i)}=\frac{1}{|\mathcal{X}_i|}\quad\text{and}\quad \pi_{\mathcal{X}_i^{\pm}}(x):=\frac{\pi_{\mathcal{X}_i}(x)}{\pi_{\mathcal{X}_i}(\mathcal{X}_i^{\pm})}=\frac{1}{|\mathcal{X}_i^{\pm}|},$$

respectively. Finally, for every $i = 2, 4, \dots, N$, define a chain on $\{+, -\}$ setting

$$P_i(+,-) := \frac{1}{2\pi_{\mathcal{X}_i}(\mathcal{X}_i^+)} \sum_{x \in \mathcal{X}_i^+} \sum_{y \in \mathcal{X}_i^-} P_{\mathcal{X}_i}(x,y)\pi_{\mathcal{X}_i}(x)$$

$$P_i(-,+) := \frac{1}{2\pi_{\mathcal{X}_i}(\mathcal{X}_i^-)} \sum_{x \in \mathcal{X}_i^-} \sum_{y \in \mathcal{X}_i^+} P_{\mathcal{X}_i}(x,y)\pi_{\mathcal{X}_i}(x).$$

Now the lower bound (3.4), applied two times yields

$$\begin{aligned} Gap(M) &\geq \frac{1}{2} Gap(\bar{P}) \min_{i=0,2,\dots,N} \{Gap(P_{\mathcal{X}_i})\} \\ &\geq \frac{1}{2} Gap(\bar{P}) \min \Big[Gap(P_{\mathcal{X}_0}), \\ &\quad \min_{i=2,\dots,N} \Big\{ \frac{1}{2} Gap(P_i) \min\{Gap(P_{\mathcal{X}_i^+}), Gap(P_{\mathcal{X}_i^-})\} \Big\} \Big]. \end{aligned} \tag{5.3}$$

Hence, to get a lower bound on $Gap(M)$ it is enough to obtain bounds on the gaps of the chains $\bar{P}$, $P_{\mathcal{X}_0}$, P_i, $P_{\mathcal{X}_i^\pm}$.

The most important of these bounds is given by the following

Proposition 5.1. *$\bar{P}$ is a birth and death chain on $\{0, 2, \dots, N\}$, more precisely*

$$\begin{aligned} &\bar{P}(0,2) = \tfrac{p_1}{2} \\ &\bar{P}(i,i+2) = \tfrac{p_1}{4}\tfrac{N-i}{N} && i \neq N, 0 \\ &\bar{P}(i,i-2) = \tfrac{p_1}{4}\tfrac{N+i}{N}\exp\{2\beta(1-i)/N\} && i \neq 0. \end{aligned} \tag{5.4}$$

Moreover

$$\lambda_1(\bar{P}) \leq 1 - \frac{p_1}{16}\frac{1}{(N/2+1)^3}.$$

and

$$\lambda_{N/2}(\bar{P}) \geq 1 - p_1.$$

The proof of the previous proposition is based on a bound for a birth and death chain, given in the Appendix, which can be of its own interest.

As for the others chains, we have the following

Lemma 5.2. *For every $i = 2, 4, \dots, N$*

$$Gap(P_{\mathcal{X}_i^\pm}) \geq (1 - p_1 - p_2) Gap(K_i^\pm)$$

$$Gap(P_i) = p_2,$$

moreover

$$Gap(P_{\mathcal{X}_0}) \geq (1 - p_1) Gap(K_0).$$

In this way, using (5.3), we can prove the main result of this section.

Proposition 5.3. *Let M be the Metropolis chain derived by the chain K defined as in (5.1) then*

$$\begin{aligned} Gap(M) \geq \frac{p_1 p_2}{32}\frac{1}{(N/2+1)^3} \min \Big[&\frac{(1-p_1)}{p_2} Gap(K_0), \\ &\frac{(1-p_1-p_2)}{2} \min_{i \neq 0} \min\{Gap(K_i^+), Gap(K_{i,}^-)\} \Big]. \end{aligned}$$

If $K_i^\pm$ and K_0 are defined as in (5.2) then

$$Gap(M) \geq \frac{p_1 p_2}{32}\frac{1}{(N/2+1)^3} \min \Big[\frac{(1-p_1-p_2)}{2}, \frac{(1-p_1)}{p_2} \Big]$$

for every $\beta > 0$ and $N \geq N_0$.

Proposition 5.3 shows that the gap is polynomial in $1/N$ independently of β. Hence, even when $\beta > 1$, the variance of the metropolis estimate obtained with this proposal can not grow up faster than a polynomial in N.

Note that if in Proposition 5.3 we choose

$$p_1 = 1 - a/(2N), \quad p_2 = a/N \tag{5.5}$$

we get

$$Gap(M) \geq \frac{C}{N^5}.$$

Hence, even with this choice, the Metropolis algorithm is still fast mixing for every β. It is worth noticing that the mean computational cost of this Metropolis does not change with respect to the Metropolis which uses the proposal K_E. Indeed, in the case of the usual Metropolis, the computational cost needed to go from X_n to X_{n+1} is $O(N)$, since it is essentially due to a sample of one number among N numbers (we need to decide which coordinate to flip). In the case of the "modified" proposal, things are slight more complex. In this case, at the beginning, we have an extra "toss". If with this fist toss we decide to flip at random a coordinate the cost is still $O(N)$ but if we need to sample from $K_i^{\pm}$ the cost is $O(N^2)$ (in this last case we need to pick a sample from a Bose-Einstein distribution). Hence, although our algorithm is "sometime" more expensive, if we take p_1 and p_2 as in (5.5), we get that the mean cost of our algorithm is still $O(N)$.

6. The mean–field Blume-Emery-Griffiths model

The Blume-Emery-Griffiths (BEG) model (see [2]) is an important lattice–spin model in statistical mechanics, it has been studied extensively as a model of many diverse systems, including $He^3 - He^4$ mixtures as well solid–liquid–gas systems, microemulsions, semiconductor alloys and electronic conduction models. See, for instance, [2, 38, 23, 24, 31, 36, 21]. We will focus our attention on a simplified mean–field version of the BEG model. For a mathematical treatment of this mean–field model see [10]. In what follows let $\mathcal{X} := \{-1, 0, 1\}^N$, N being an even integer, and for every $\beta > 0$ and $K > 0$ let $\pi_{\beta,K,N}$ be the probability defined by

$$\pi(x) = \pi_{\beta,K,N}(x) = \exp\{-\beta R_N(x) + \frac{K\beta}{N} S_N^2(x)\} Z_N^{-1}(\beta, K) \qquad (x \in \mathcal{X})$$

where

$$Z_N(\beta, K) = Z_N := \sum_{x \in \mathcal{X}} \exp\left\{-\beta R_N(x) + \frac{K\beta}{N} S_N^2(x)\right\}$$

is the normalization constant,

$$S_N(x) := \sum_{i=1}^N x_i \qquad \text{and} \qquad R_N(x) := \sum_{i=1}^N x_i^2 \qquad x = (x_1, x_2, ..., x_N).$$

A natural Metropolis algorithm can be derived by using the proposal chain

$$K_E(x, y) = \frac{1}{2N} \sum_{j=1}^N [\mathbb{I}_{\{x^{(+j)}\}}(y) + \mathbb{I}_{\{x^{(-j)}\}}(y)] \tag{6.1}$$

where $x^{(\pm j)}$ denotes the vector $(x_1, \ldots, x_j \pm 1, \ldots, x_N)$, with the convention that $2 = -1$ and $-2 = 1$.

The next proposition shows that there exists a critical region of the parameters space in which the Metropolis chain is slowly mixing. More precisely, using some results of [10] it is quite straightforward to proove the following

Proposition 6.1. *Ler M_E be the Metropolis chain (with stationary distribution π) with proposal chain K_E defined in* (6.1)*.Then, there exists a non decreasing function $\Gamma : (0, +\infty) \to (0, +\infty)$ with $\lim_{x\to 0} \Gamma(x) = +\infty$ and $\lim_{x\to\infty} \Gamma(x) = \gamma_c \simeq 1.082$ such that for every couple of positive parametrs (β, K) with $K > \Gamma(\beta)$*

$$Gap(M_E) \leq Ce^{-\Delta N}$$

for suitable constants $C = C(\gamma, K) > 0$ and $\Delta = \Delta(\gamma, K) > 0$.

As in the case of the mean–field Ising model, we intend to by pass the slowly mixing problem of this Metropolis chain by choosing a different proposal. To understand which kind of proposal is reasonable, here we choose

$$\mathcal{G} = \mathcal{S}_N \times \{+1, -1\}$$

with $\mathcal{G}$ acting on $\mathcal{X} = \{-1, 0, 1\}^N$ by

$$g(x) = (e \cdot x_{\sigma(1)}, \dots, e \cdot x_{\sigma(N)}) \qquad g = (\sigma, e).$$

At this stage, decompose $\mathcal{X}$ as the union of its "energy sets", that is

$$\mathcal{X} = \mathcal{X}_{0,0} \cup \mathcal{X}_{1,1} \cup \mathcal{X}_{0,2} \cup \mathcal{X}_{1,3} \cup X_{3,3} \cup ... \cup \mathcal{X}_{0,N} \cup \mathcal{X}_{2,N} \cup ...\mathcal{X}_{N,N}$$

where

$$\mathcal{X}_{s,r} := \{x \in X : |S_N| = s \text{ and } R_N(x) = r\}$$

$r = 0, 1, 2, ..., N$ and $s = 1, 3, ..., r$ if r is odd and $s = 0, 2, ..., N$ if r is even. Moreover, for $s = 1, 2, ..., N$, set

$$\mathcal{X}^+_{s,r} := \{x \in X : S_N = s \text{ and } R_N(x) = r\}$$

and

$$\mathcal{X}^-_{s,r} := \{x \in X : S_N = -s \text{ and } R_N(x) = r\}.$$

Note again that $O_x = \mathcal{X}_{s,r}$ with $s = S_N(x)$ and $r = R_N(x)$. The new proposal chain will be

$$\begin{aligned} &K(x, y) = p_1 K_E(x, y) + (1 - p_1)K_{0,r}(x, y) \qquad \text{if } x \in \mathcal{X}_{0,r}, \quad r = 0, 2, ..., N \\ &K(x, y) = p_1 K_E(x, y) + p_2 \mathbb{I}_{\{-x\}}(y) + (1 - p_1 - p_2)K_{s,r}(x, y) \\ &\qquad\qquad \text{if } x \in \mathcal{X}_{s,r}, s \neq 0 \end{aligned} \tag{6.2}$$

where p_1, p_2 belong to $(0, 1)$, $p_1 + p_2 < 1$, and

$$K_{s,r}(x, y) = \mathbb{I}_{\mathcal{X}^+_{s,r}}\{x\}K^+_{s,r}(x, y) + \mathbb{I}_{\mathcal{X}^-_{s,r}}\{x\}K^-_{s,r}(x, y) \qquad (s \neq 0)$$

with

$$\begin{aligned} K_{0,r}(x, y) &= \frac{1}{\binom{N}{r}\binom{r}{r/2}} \qquad y \in \mathcal{X}_{0,r} \\ K^\pm_{s,r}(x, y) &= \frac{1}{\binom{N}{r}\binom{r}{(r-s)/2}} \qquad y \in \mathcal{X}^\pm_{s,r}. \end{aligned} \tag{6.3}$$

Now let M be the Metropolis chain defined by the transition kernel (1.2) with K as in (6.2), i.e. for every x in $\mathcal{X}^\pm_{s,r}$ $(s \neq 0)$

$$M(x, y) = \begin{cases} \frac{p_1}{2N} \min\left(1, \frac{\pi(y)}{\pi(x)}\right) & \text{if } y = x^{(\pm j)}, \quad j = 1...N \\ p_2 & \text{if } y = -x \\ (1 - p_1 - p_2)\frac{1}{\binom{N}{r}\binom{r}{(r-s)/2}} & \text{if } y \in \mathcal{X}^\pm_{s,r}, y \neq x \\ 1 - \sum_{z \neq x} M(x, z) & \text{if } y = x, \end{cases}$$

while if x belongs to $\mathcal{X}_{0,r}$

$$M(x,y) = \begin{cases} \frac{p_1}{2N} \min\left(1, \frac{\pi(y)}{\pi(x)}\right) & \text{if } y = x^{(\pm j)}, \quad j = 1...N \\ (1-p_1)\frac{1}{\binom{N}{r}\binom{r}{r/2}} & \text{if } y \in \mathcal{X}_{0,r}, y \neq x \\ 1 - \sum_{z \neq x} M(x,z) & \text{if } y = x. \end{cases}$$

By construction M is an aperiodic, irreducible and reversible chain with stationary distribution π.

Also in this case, to bound the spectral gap of M, we shall use the chain decomposition tools. Let

$$\mathbb{D}_N = \{(0,0),(1,1),(0,2),(2,2),(1,3),(3,3),(0,4),(2,4),(4,4),...,(0,N),(2,N),...,(N,N)\}$$

and, for every couple $(s,r),(\tilde{s},\tilde{r})$ in $\mathbb{D}_N$, with $(s,r) \neq (\tilde{s},\tilde{r})$, let

$$\bar{P}((s,r),(\tilde{s},\tilde{r})) := \frac{1}{2\pi(\mathcal{X}_{s,r})} \sum_{x \in \mathcal{X}_{s,r}} \sum_{y \in \mathcal{X}_{\tilde{s},\tilde{r}}} M(x,y)\pi(x)$$

and

$$\bar{P}((s,r),(s,r)) := 1 - \sum_{(\tilde{s},\tilde{r}) \neq (s,r)} \bar{P}((s,r),(\tilde{s},\tilde{r})).$$

Once again, note that $\bar{P}$ is a reversible chain on $\mathbb{D}_N$ with stationary distribution

$$\bar{\pi}(s,r) := \pi(\mathcal{X}_{s,r}).$$

Moreover, for every (s,r) in $\mathbb{D}_N$, define a chain on $\mathcal{X}_{s,r}$ setting

$$P_{\mathcal{X}_{s,r}}(x,y) := M(x,y) + \mathbb{I}_x(y)\left(\sum_{z \in \mathcal{X}_{s,r}^c} M(x,z)\right)$$

where both x and y belong to $\mathcal{X}_{s,r}$. In the same way, define chains on $\mathcal{X}_{s,r}^+$ and $\mathcal{X}_{s,r}^-$ for (s,r) in $\mathbb{D}_N$, $s \neq 0$, setting

$$P_{\mathcal{X}_{s,r}^\pm}(x,y) := P_{\mathcal{X}_{s,r}}(x,y) \quad (y \neq x,\ x,y \in X_{s,r}^\pm)$$

and

$$P_{\mathcal{X}_{s,r}^\pm}(x,x) := 1 - \sum_{y \in \mathcal{X}_{s,r}^\pm y \neq x} P_{\mathcal{X}_{s,r}}(x,y).$$

These chains are reversible on $\mathcal{X}_{s,r}$ ($\mathcal{X}_{s,r}^\pm$, respectively) and have as stationary distributions

$$\pi_{\mathcal{X}_{s,r}}(x) := \frac{\pi(x)}{\pi(\mathcal{X}_{s,r})} = \frac{1}{|\mathcal{X}_{s,r}|} \quad \text{and} \quad \pi_{\mathcal{X}_{s,r}^\pm}(x) := \frac{\pi_{\mathcal{X}_{s,r}}(x)}{\pi_{\mathcal{X}_{s,r}}(\mathcal{X}_{s,r}^\pm)} = \frac{1}{|\mathcal{X}_{s,r}^\pm|},$$

respectively. Finally, for every (s,r) in $\mathbb{D}_N$, $s \neq 0$, define a chain on $\{+,-\}$ setting

$$P_{s,r}(+,-) := \frac{1}{2\pi_{\mathcal{X}_{s,r}}(\mathcal{X}_{s,r}^+)} \sum_{x \in \mathcal{X}_{s,r}^+} \sum_{y \in \mathcal{X}_{s,r}^-} P_{\mathcal{X}_{s,r}}(x,y)\pi_{\mathcal{X}_{s,r}}(x)$$

$$P_{s,r}(-,+) := \frac{1}{2\pi_{\mathcal{X}_{s,r}}(\mathcal{X}_{s,r}^-)} \sum_{x \in \mathcal{X}_{s,r}^-} \sum_{y \in \mathcal{X}_{s,r}^+} P_{\mathcal{X}_{s,r}}(x,y)\pi_{\mathcal{X}_{s,r}}(x).$$

At this stage, the lower bound (3.4), applied two times, yields

$$
\begin{aligned}
Gap(M) &\geq \frac{1}{2} Gap(\bar{P}) \min_{(s,r)\in\mathbb{D}_N} \left\{ Gap(P_{\mathcal{X}_{s,r}}) \right\} \\
&\geq \frac{1}{2} Gap(\bar{P}) \min \Big[\min_{r=0,2,\dots,N} \left\{ Gap(P_{\mathcal{X}_{0,r}}) \right\}, \\
&\min_{(s,r)\in\mathbb{D}_N, s\neq 0} \left\{ \frac{1}{2} Gap(P_{s,r}) \min\{Gap(P_{\mathcal{X}^+_{s,r}}), Gap(P_{\mathcal{X}^-_{s,r}})\} \right\} \Big].
\end{aligned} \tag{6.4}
$$

To derive from the last bound a more explicit bound we need some preliminary work. The first result we need is exactly the analogous of Lemma 5.2.

Lemma 6.2. *Fore every $r = 1, \dots, N$*

$$Gap(P_{\mathcal{X}_{0,r}}) \geq (1 - p_1) Gap(K_{0,r}) = (1 - p_1),$$

moreover, for every (s, r) in $\mathbb{D}_N$ with $s \neq 0$,

$$Gap(P^{\pm}_{\mathcal{X}_{s,r}}) \geq (1 - p_1 - p_2) Gap(K^{\pm}_{s,r}) = (1 - p_1 - p_2).$$

Finally, for every (s, r) in $\mathbb{D}_N$,

$$Gap(P_{s,r}) = p_2.$$

Hence, (6.4) can be rewritten as

$$Gap(M) \geq Gap(\bar{P}) \frac{p_2}{2} \min\{(1 - p_1)/2, (1 - p_1 - p_2)/2\}. \tag{6.5}$$

It remains to bound $Gap(\bar{P})$. Unfortunately the the analogous of Proposition 5.1 is not so simple, hence we shall require an additional hypothesis. In what follows let

$$
\begin{aligned}
q_{[\![N]\!]}(r) &:= \binom{N}{r} e^{-\beta r} \left[\binom{r}{\frac{r}{2}} + 2 \sum_{i=0}^{\frac{r}{2}-1} \binom{r}{i} e^{\frac{k\beta}{N}(r-2i)^2} \right] && \text{if } r \text{ is even} \\
q_{[\![N]\!]}(r) :&= \binom{N}{r} e^{-\beta r} \left[2 \sum_{i=0}^{\frac{r-1}{2}} \binom{r}{i} e^{\frac{k\beta}{N}(r-2i)^2} \right] && \text{if } r \text{ is odd}
\end{aligned}
$$

$r = 0, 1, \dots, N$ and set

$$\mathcal{A} = \{\beta > 0, K > 0 : \exists N_0 \text{ such that } \forall N \geq N_0, \quad q_{[\![N]\!]} \text{ is unimodal}\}.$$

Lemma 6.3. *For every (β, K) in $\mathcal{A}$*

$$Gap(\bar{P}) \geq \frac{C p_1^2}{N^6}$$

for a suitable constant $C = C(\beta, K)$.

Under the same assumptions of the previous Lemma we can state the main results of this section.

Proposition 6.4. *For every (β, K) in $\mathcal{A}$*

$$Gap(M) \geq \frac{\tilde{C} p_1^2}{N^6}$$

for a suitable constant $\tilde{C} = \tilde{C}(\beta, K)$.

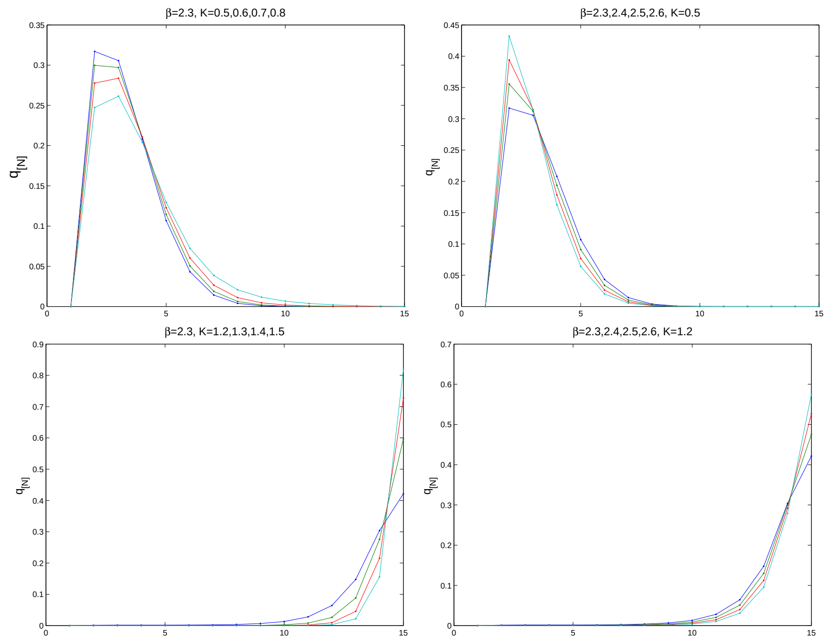

FIGURE 1. The function $q_{[\![N]\!]}$ for $N = 15$ and few values of β and K.

We conjecture that $Gap(\bar{P})$ is polynomial in N for every (β, K) such that $\beta \neq \Gamma(K)$ (where Γ is the function of Proposition 6.1), but we are not able to prove this conjecture. In point of fact we conjecture that $\mathbb{R}^+ \times \mathbb{R}^+ \setminus \{(\beta, K) : \Gamma(K) = \beta\} \subset \mathcal{A}$. We plotted $q_{[\![N]\!]}$ for different N, β and K, and these plotts seem, at least, to confirm that $\mathbb{R}^+ \times \mathbb{R}^+ \setminus \{(\beta, K) : |\Gamma(K) - \beta| \leq \epsilon\} \subset \mathcal{A}$ for a suitable small ϵ. In Figure 1 we show the graph of $q_{[\![N]\!]}$ for few different N, β and K.

APPENDIX A. THE SPECTRAL GAP OF A BIRTH AND DEATH CHAIN

We derive here some bounds on the eigenvalues of a birth and death chain that we shall use later. These bounds are obtained using the so called geometric techniques, see [7]. Let P_n be a birth and death chain on $\Omega_n = \{1, \dots, n\}$. Assume that P_n is reversible with respect to a probability p_n, that is $p_n(i)P_n(i,j) = p_n(j)P_n(j,i)$. Moreover let

$$1 > \lambda_1 \geq \lambda_2 \geq \dots \lambda_{n-1} \geq -1$$

the eigenvalues of P_n.

We can now prove the following variant of Proposition 6.3 in [6].

Lemma A.1. *If there exist positive constants A, q, B and an integer k such that*

$$\begin{aligned} &P_n(i, i \pm 1) \geq A n^{-q} \qquad (i \neq 1, n) \\ &P_n(1,2) \geq A n^{-q} \\ &P_n(n, n-1) \geq A n^{-q} \end{aligned}$$

and

$$\begin{aligned} p_n(i) \leq B p_n(j) \quad i \leq j \leq k \\ p_n(j) \leq B p_n(i) \quad k \leq i \leq j \end{aligned}$$

then

$$\lambda_1 \leq 1 - \frac{A}{B} \frac{1}{n^{q+2}}.$$

Proof. We use the notation and the techniques of [7], see also [3] and [6]. Choose the set of paths

$$\Gamma = \{\gamma_{ij} = (i, i+1, ..., j); i \leq j; i, j \in \Omega_n\}$$

and for $e = (i, i+1)$ $(i < n)$ let

$$\psi(e) = \frac{1}{p_n(i, i+1)} \sum_{\substack{\gamma_{l,m} \in \Gamma \\ \gamma_{l,m} \ni e}} |\gamma_{l,m}| \frac{p_n(l) p_n(m)}{p_n(i)}$$

where $|\gamma|$ is the length of the path γ. Setting $K := \sup_e \psi(e)$ one has

$$\lambda_1 \leq 1 - \frac{1}{K}$$

(see Proposition 1' in [7], or Exercise 6.4 page 248 in [3]). So, for our purposes, it suffices to give an upper bound on K. Assume first that $e = (i, i+1)$ with $i < k \leq n$, since $|\gamma_{l,m}| \leq n$, it follows that

$$\begin{aligned}
\psi(e) &\leq \frac{n^q}{A} n \left(\sum_{\substack{s \geq i+1 \\ r \leq i}} \frac{p_n(r) p_n(s)}{p_n(i)} \right) \\
&\leq \frac{n^{q+1}}{A} \left(\sum_{r \leq i} \frac{p_n(r)}{p_n(i)} \right) \left(\sum_{s \geq i+1} p_n(s) \right) \\
&\leq \frac{n^{q+1}}{A} \left(\sum_{r \leq i} B \right) \left(\sum_{s=1}^{n} p_n(s) \right) \\
&\leq n^{q+2} \frac{B}{A}.
\end{aligned}$$

All the other cases can be treated in the same way. Hence,

$$\sup_e \psi(e) \leq \frac{B}{A} n^{q+2}$$

and then

$$\lambda_1 \leq 1 - \frac{A}{B} \frac{1}{n^{q+2}}.$$

□

As for the smaller eigenvalues, Gershgorin theorem yields that

$$\lambda_{n-1} \geq -1 + 2 \min_i P(i, i).$$

See, for instance, Corollary 2.1 in the Appendix of [3]. Hence, if there exists a positive constant D such that

$$P_n(i, i+1) + P_n(i, i-1) \leq D/2$$

for every i, then

$$\lambda_{n-1} \geq 1 - D. \tag{A.1}$$

Appendix B. Proofs

To prove Proposition 5.1 we need first to show that $\bar{\pi}$ is essentially unimodal.

Lemma B.1. *Let*

$$q_N(i) = \binom{N}{\frac{N-i}{2}} \exp\left\{\frac{\beta}{2N} i^2\right\} \qquad i = 0, 2, 4, \dots, N.$$

For every $\beta < 1$ there exists an integer N_0 such that for every $N \geq N_0$

$$q_N(i) \leq q_N(j)$$

whenever $j \leq i$. For every $\beta \geq 1$ there exists an integer N_0 such that for every $N \geq N_0$

$$q_N(i) \leq q_N(j)$$

whenever $i \leq j \leq k_N$ and

$$q_N(i) \geq q_N(j)$$

whenever $k_N \leq i \leq j$, k_N being a suitable integer.

Proof. Let $\Delta_N(i)$ be the ratio

$$\Delta_N(i) = \frac{q_N(i+2)}{q_N(i)} \qquad i = 0, 2, 4, ..., N-2,$$

so that

$$\begin{aligned}\Delta_N(i) &= \frac{\binom{N}{\frac{N-i}{2}-1}}{\binom{N}{\frac{N-i}{2}}} \exp\left\{\frac{2\beta}{N}(1+i)\right\} \\ &= \frac{N-i}{N+2+i} \exp\left\{\frac{2\beta}{N}(1+i)\right\}.\end{aligned}$$

Setting $\Delta_N(x) = \frac{N-x}{N+2+x} \exp\left\{\frac{2\beta}{N}(1+x)\right\}$, x in $[0, N-2]$, it is enough to prove that $x \mapsto \Delta_N(x)$ takes the value 1 at most once in $[0, N-2]$, for sufficiently large N. To prove this last claim first note that

$$\begin{aligned}\Delta_N(0) &= \frac{N}{N+2} \exp\left\{\frac{2\beta}{N}\right\} = \frac{1}{1+\frac{2}{N}} \exp\left\{\frac{2\beta}{N}\right\} \\ &= \left[1 - \frac{2}{N} + 2\left(\frac{2}{N}\right)^2 + o\left(\frac{1}{N^2}\right)\right]\left[1 + \frac{2\beta}{N} + \frac{1}{2}\left(\frac{2\beta}{N}\right)^2 + o\left(\frac{1}{N^2}\right)\right] \\ &= 1 - \frac{2}{N}(1-\beta) + \left(\frac{2}{N}\right)^2\left(\frac{\beta^2}{2} - \beta + 2\right) + o\left(\frac{1}{N^2}\right).\end{aligned}$$

Hence, there exists N_0 in $\mathbb{N}$ such that for $N \geq N_0$:

$$\beta \geq 1 \quad \Rightarrow \quad \Delta_N(0) > 1$$

$$\beta < 1 \quad \Rightarrow \quad \Delta_N(0) < 1.$$

As for the first derivative note that

$$\Delta'_N(x) = \frac{-2(N+1) + 2\beta(N+2) - \frac{2\beta}{N}(x^2+2x)}{(N+x+2)^2} \exp\left\{\frac{2\beta}{N}(1+x)\right\},$$

hence $\Delta'_N(x) = 0$ if and only if

$$-2(N+1) + 2\beta(N+2) - \frac{2\beta}{N}(x^2+2x) = 0.$$

Rearranging the last equation as

$$-\frac{2\beta}{N}x^2 - \frac{4\beta}{N} + 2[(\beta-1)N + 2\beta - 1] = 0$$

one sees that the roots are

$$x_{1,2} = 1 \pm \sqrt{1 + \frac{2\beta-1}{\beta}N + \frac{\beta-1}{\beta}N^2}.$$

Hence, after setting

$$r := 1 + \sqrt{1 + \frac{2\beta-1}{\beta}N + \frac{\beta-1}{\beta}N^2} \qquad \text{and} \qquad \overline{r} := 1 + \sqrt{1+N}$$

one has

$$\begin{array}{lll} \beta < 1 & \Rightarrow \quad \Delta'_N(x) < 0 & \forall x \in [0, N-2] \\ \\ \beta > 1 & \Rightarrow \quad \Delta'_N(x) > 0 & \text{for } x \in [0, r) \\ & \quad\;\;\; \Delta'_N(x) < 0 & \text{for } x \in (r, N-2] \\ \\ \beta = 1 & \Rightarrow \quad \Delta'_N(x) < 0 & \text{for } x \in [0, \overline{r}) \\ & \quad\;\;\; \Delta'_N(x) < 0 & \text{for } x \in (\overline{r}, N-2] \end{array}$$

and this concludes the proof. □

Proof of Proposition 5.1. By direct computations it is easy to prove (5.4). Hence

$$P(i, i \pm 2) \geq \frac{p_1}{4N} \geq \frac{p_1}{4(N+2)} \geq \frac{p_1}{8(\frac{N}{2}+1)},$$

and

$$P(i, i+2) + P(i, i-2) \leq \frac{p_1}{2}.$$

Now observe that

$$\bar{\pi}(0) = \frac{1}{Z_N(\beta)} q_N(0)$$

and

$$\bar{\pi}(i) = \frac{2}{Z_N(\beta)} q_N(i) \qquad i \neq 0.$$

Hence, by Lemma B.1, if $\beta < 1$

$$\bar{\pi}(i) \leq 2\bar{\pi}(j)$$

whenever $j \leq i$ and N is large enough. While for $\beta > 1$

$$\bar{\pi}(i) \leq \bar{\pi}(j)$$

whenever $i \leq j \leq k_N$ and

$$\bar{\pi}(i) \geq \bar{\pi}(j)$$

whenever $k_N \leq i \leq j$. The thesis follows now by Lemma A.1 and by (A.1). □

In order to prove Lemma 5.2 we recall that by Rayleigh's theorem

$$1 - \lambda_1(P) = \inf\left\{\frac{\mathcal{E}_p(f,f)}{Var_p(f)} : f \text{ nonconstant}\right\} \tag{B.1}$$

where

$$\mathcal{E}_p(f,f) := < (I-P)f, f >_p = \frac{1}{2}\sum_{x,y}(f(x) - f(y))^2 P(x,y)p(x),$$

P being a reversible chain w.r.t. p, moreover

$$1 - |\lambda_{N-1}| = \inf \left\{ \frac{\frac{1}{2}\sum_{x,y}(f(x)+f(y))^2 P(x,y)p(x)}{Var_p(f)} : f \text{ nonconstant} \right\} \tag{B.2}$$

(see, for instance, Theorem 2.3 in Chapter 6 of [3] and Section 2.1 of [8]). At this stage set

$$P_\epsilon(x,y) := (1-\epsilon)P(x,y) + \epsilon \mathbb{I}_x(y).$$

Hence, (B.1) yields

$$\begin{aligned} 1 - \lambda_1(P_\epsilon) &= \inf_{f \in L^2_p f \neq const} \frac{\frac{1}{2}\sum_{x,y}(f(x)-f(y))^2 P_\epsilon(x,y)p(x)}{Var_p(f)} \\ &= \inf_{f \in L^2_p f \neq const} (1-\epsilon) \frac{\frac{1}{2}\sum_{x \neq y}(f(x)-f(y))^2 P(x,y)p(x)}{Var_p(f)} \\ &= (1-\epsilon)(1-\lambda_1(P)). \end{aligned}$$

Arguing in the same way and using (B.2) we get

$$1 - |\lambda_{|\mathcal{X}|-1}(P_\epsilon)| \geq (1-\epsilon)(1 - |\lambda_{|\mathcal{X}|-1}(P)|).$$

Hence,

$$Gap(P_\epsilon) \geq (1-\epsilon) Gap(P). \tag{B.3}$$

Proof of Lemma 5.2. Note that

$$P_{\mathcal{X}_i^\pm}(x,y) = (1-p_1-p_2)K_i^\pm(x,y) + (p_1+p_2)\mathbb{I}_x(y)$$

and, analogously,

$$P_{\mathcal{X}_0}(x,y) = (1-p_1)K_0(x,y) + p_1 \mathbb{I}_x(y).$$

Hence, by (B.3),

$$Gap(P_{\mathcal{X}_i^\pm}) \geq (1-p_1-p_2) Gap(K_i^\pm)$$

as well

$$Gap(P_{\mathcal{X}_0}) \geq (1-p_1) Gap(K_0).$$

Finally note that P_i is given by

$$\begin{pmatrix} 1-\frac{p_2}{2} & \frac{p_2}{2} \\ \frac{p_2}{2} & 1-\frac{p_2}{2} \end{pmatrix}$$

for every i, hence $Gap(P_i) = p_2$. □

Proof of Proposition 5.3. To prove the first part of the proposition it is enough to combine Lemma 5.2, Proposition 5.1 and (5.3). To complete the proof observe that $Gap(K_i^\pm) = Gap(K_0) = 1$, when $K_i^\pm$ and K_0 are given by (5.2). □

In order to prove Proposition 6.1 we need some results obtained in [10].

Theorem B.2 (Ellis-Otto-Touchette)**.** *Let ρ_N be the distribution of $S_N(x)/N$ under $\pi_{\beta,K,N}$, then ρ_N satisfies a large deviation principle on $[-1,1]$ with rate function*

$$\tilde{I}_{\beta,K}(z) = J_\beta(z) - \beta K z^2 - \inf_{t \in \mathbb{R}} \{J_\beta(t) - \beta K t^2\}$$

with

$$J_\beta(z) = \sup_{t \in \mathbb{R}} \left\{ tz - \log\left[\frac{1+e^{-\beta}(e^t+e^{-t}}{1+2e^{-\beta}}\right]\right\}.$$

Moreover, if $\tilde{\mathcal{E}}_{\beta,K} := argmin \tilde{I}_{\beta,K}$, then there exists a non decreasing function $\Gamma : (0,+\infty) \to (0,+\infty)$ with $\lim_{x\to 0}\Gamma(x) = +\infty$ and $\lim_{x\to\infty}\Gamma(x) = \gamma_c \simeq 1.082$ such that for every (β, K) with $K > \Gamma(\beta)$ then

$$\tilde{\mathcal{E}}_{\beta,K} = \{\pm z(\beta,K) \neq 0\}.$$

In particular, for such (β, K) *and for every* $0 < \epsilon < |z(\beta, K)|$ *there exists a constant* $C_1 = C_1(\epsilon, \beta, K)$ *such that*

$$\rho([0,\epsilon]) \leq C_1 \exp\{-\frac{N}{2}\gamma_{\epsilon,\beta,K}\} \tag{B.4}$$

with

$$\gamma_{\epsilon,\beta,K} = \inf_{z\in[0,\epsilon]} \tilde{I}_{\beta,K}(z) > 0. \tag{B.5}$$

Proof. For the first part see Theorems 3.3, 3.6 and 3.8 in [10]. As for (B.4)-(B.5), they are standard consequences of the theory of the large deviations and of the first part of the proposition, see, e.g., Proposition 6.4 of [11]. □

Proof of Proposition 6.1. We intend to use the Chegeer's inequality. To do this, let $A := \{x : S_N(x) < 0\}$, $B := \{x : S_N(x) > 0\}$, $C := \{S_N(x) = 0\}$. First of all note that, by symmetry, $\pi(A) = \pi(B) = (1-\pi(C))/2 \leq 1/2$. The main task is to bound

$$\phi(A) = \sum_{x\in A}\sum_{y\in A^c} \pi(x)M_E(x,y) = \sum_{y\in A^c}\sum_{x\in A} \pi(y)M_E(y,x).$$

Now, observe that if $S_N(y) > 1$ then $M_E(y,x) = 0$ for every x in A, hence

$$\begin{aligned}\phi(A) &= \sum_{y:S_N(y)=0} \pi(y)\sum_{x\in A} M_E(y,x) + \sum_{y:S_N(y)=1} \pi(y)\sum_{x\in A} M_E(y,x)\\ &\leq \pi\left\{y : S_N(y) \in \{0,1\}\right\}.\end{aligned}$$

This yields a bound on the conductance

$$h = h(\pi, M_E) \leq \phi(A)/\pi(A) \leq \frac{2\pi\left\{y : S_N(y) \in \{0,1\}\right\}}{1-\pi\{y : S_N(y) = 0\}}.$$

Now by Proposition B.2 we get

$$h(\pi, M_E) \leq C_2 e^{-\Delta N}$$

for suitable constants C_2 and $\Delta > 0$. The thesis follows by Cheeger inequality (3.2). □

Proof of Lemma 6.2. The proof is exactly the same as the proof of Lemma 5.2. □

In order to prove Lemma 6.3 it is convenient to fix some simple properties of the chain $\bar{P}$.

Lemma B.3. $\bar{P}$ *is a random walk on* $\mathbb{D}_N$*. If* $\bar{P}((s,r),(\tilde{s},\tilde{r})) \neq 0$*,*

$$\bar{P}((s,r),(\tilde{s},\tilde{r})) \geq \frac{p_1 C_3}{N}$$

for a suitable constant $C_3 = C_3(\beta, K)$*, moreover*

$$\bar{P}((s,r),(\tilde{s},\tilde{r})) \leq \frac{p_1}{4}$$

for every $(s,r),(\tilde{s},\tilde{r})) \neq ((0,0),(1,1))$*.*

Proof of Lemma B.3. Easy but tedious computations show that

$$\begin{aligned}&\bar{P}((0,0),(1,1)) = \frac{p_1}{2}\min\left(1, \exp\{\frac{K\beta}{N} - \beta\}\right)\\ &\bar{P}((0,N),(1,N-1)) = \frac{p_1}{4}\\ &\bar{P}((0,N),(2,N)) = \frac{p_1}{4}\end{aligned}$$

$$\bar{P}((0,r),(2,r)) = \frac{p_1}{4N} \qquad r = 0,2,4,...,N-2$$
$$\bar{P}((0,r),(1,r-1)) = \frac{p_1}{4N} \qquad r = 0,2,4,...,N-2$$
$$\bar{P}((0,r),(1,r+1)) = \frac{p_1}{2N} \min\left(1, \exp\{\frac{K\beta}{N} - \beta\}\right) \qquad r = 0,2,4,...,N-2$$
$$\bar{P}((s,r),(s+2,r)) = \frac{p_1}{8N}(r-s)$$
$$(s,r) \in \mathbb{D}_N, 0 < s \le N-2, r \le N$$
$$\bar{P}((s,r),(s-2,r)) = \frac{p_1}{8N}(r+s)\exp\{4\frac{K\beta}{N}(1-s)\}$$
$$(s,r) \in \mathbb{D}_N, 0 < s \le N, r \le N$$
$$\bar{P}((s,r),(s+1,r+1)) = \frac{p_1}{4N}(N-r)\min\left(1, \exp\{\frac{K\beta}{N}(2s+1) - \beta\}\right)$$
$$(s,r) \in \mathbb{D}_N, 0 < s, r \le N-1,$$
$$\bar{P}((s,r),(s-1,r+1)) = \frac{p_1}{4N}(N-r)\exp\{\frac{K\beta}{N}(-2s+1) - \beta\}$$
$$(s,r) \in \mathbb{D}_N, 0 < s, r \le N-1,$$
$$\bar{P}((s,r),(s+1,r-1)) = \frac{p_1}{8N}(r-s)$$
$$(s,r) \in \mathbb{D}_N, 0 < r \le N, 0 < s \le N-2$$
$$\bar{P}((s,r),(s-1,r-1)) = \frac{p_1}{8N}(r+s)\min\left(1, \exp\{\frac{K\beta}{N}(2s+1) - \beta\}\right)$$
$$(s,r) \in \mathbb{D}_N, 0 < r \le N, 0 < s \le r.$$

At this stage the statement follows easily. □

Proof of Lemma 6.3. In order to obtain a bound on the gap of $\bar{P}$ we shall apply another time the decomposition technique. Write

$$\mathbb{D}_N = \bar{\mathcal{X}}_1 \cup \bar{\mathcal{X}}_2 \cup \bar{\mathcal{X}}_3 \cup ... \cup \bar{\mathcal{X}}_N,$$

where

$$\bar{\mathcal{X}}_1 = \{(0,0),(1,1)\} \qquad \bar{\mathcal{X}}_r = \{(u_1,u_2) \in \mathbb{D}_n : u_2 = r\}.$$

On $[\![N]\!] := \{1,...,N\}$ define a chain $P_{[\![N]\!]}$ setting

$$P_{[\![N]\!]}(i,j) := \frac{1}{2\bar{\pi}(\bar{\mathcal{X}}_i)} \sum_{a \in \bar{\mathcal{X}}_i} \sum_{b \in \bar{\mathcal{X}}_j} \bar{P}(a,b)\bar{\pi}(a)$$

and

$$P_{[\![N]\!]}(i,i) := 1 - \sum_{j \ne i} P_{[\![N]\!]}(i,j).$$

Again $P_{[\![N]\!]}$ is a reversible chain on $[\![N]\!]$ with stationary distribution

$$\bar{\pi}_{[\![N]\!]}(i) := \bar{\pi}(\bar{\mathcal{X}}_i).$$

Finally for every $r = 1, 2, \dots, N$ we define a chain on $\bar{\mathcal{X}}_r$ by setting

$$P_{\bar{\mathcal{X}}_r}(a,b) := \bar{P}(a,b) + \mathbb{I}_a(b)\left(\sum_{z \in \bar{\mathcal{X}}_r^c} \bar{P}(a,z)\right)$$

where both a and b belong to $\bar{\mathcal{X}}_r$. Now note that for every $r = 2, 3, \dots, N$ $P_{\bar{\mathcal{X}}_r}$ is a birth and death chain on the state space $\{(1,r),(3,r),\dots,(r,r)\}$ for r odd and

$\{(0,r),(2,r),\dots,(r,r)\}$ for r even. Let

$$q_r(s) := \binom{r}{(r-s)/2} e^{\frac{\beta K}{N}s^2}$$

and, for r even,

$$q_r(0) := 2\binom{r}{r/2}.$$

Now observe that $P_{\bar{\mathcal{X}}_r}$ has stationary distribution

$$\pi_r(s) \propto q_r(s)$$

with $s = 0,2,\dots,r$ if r is even and $s = 1,3,\dots,r$ if r is odd. First of all let $r \neq 1$, by Lemma B.3 and Lemma B.1, it is easy to check that $(P_{\bar{\mathcal{X}}_r}, \pi_r)$ meets the condition of Lemma A.1 with

$$B = 2, \quad n = [(r+2)/2], \quad A = C_3 p_1 [(r+2)/2] N^{-1}$$

($[x]$ being the integer part of x) and then

$$1 - \lambda_1(P_{\bar{\mathcal{X}}_r}) \geq \frac{C_3 p_1 [(r+2)/2]}{2N[(r+2)/2]^3} \geq \frac{C_3 p_1}{2N^3}.$$

Finally, Lemma B.3 with (A.1) yields

$$\lambda_{|\bar{\mathcal{X}}_r|-1}(P_{\bar{\mathcal{X}}_r}) \geq 1 - p_1.$$

Hence, for every $r \neq 1$, we have proved that

$$Gap(P_{\bar{\mathcal{X}}_r}) \geq C_3/2 p_1 N^{-3}. \tag{B.6}$$

For $r = 1$

$$P_{\bar{\mathcal{X}}_1} = \begin{pmatrix} 1-\alpha_1/2 & \alpha_1/2 \\ \alpha_2/2 & 1-\alpha_2/2 \end{pmatrix}$$

where

$$\alpha_1 := \frac{p_1}{2N}\min\left(1, \exp\{\frac{3K\beta}{N} - \beta\}\right)$$

$$\alpha_2 := p_1 \min\left(1, \exp\{\frac{K\beta}{N} - \beta\}\right)$$

So

$$Gap(P_{\bar{\mathcal{X}}_1}) \geq 1 - |\frac{2-\alpha_1-\alpha_2}{2}| = \frac{\alpha_1+\alpha_2}{2}$$

where the last equality follows from the fact that $\frac{\alpha_1}{2} \leq \frac{1}{2}$ and $\frac{\alpha_2}{2} \leq \frac{1}{2}$. Hence, for sufficiently large N, it's easy to see that

$$Gap(P_{\bar{\mathcal{X}}_1}) \geq C_4 p_1 N^{-3} \tag{B.7}$$

with $C_4 = C_4(\beta, K)$. At this stage (B.6) with (B.7) gives

$$Gap(P_{\bar{\mathcal{X}}_r}) \geq C_5 p_1 N^{-3} \tag{B.8}$$

for all $r \in [\![N]\!]$. As for the gap of $P_{[\![N]\!]}$, first of all note that $P_{[\![N]\!]}$ is a birth and death chain on $[\![N]\!]$. From Lemma B.3

$$P_{[\![N]\!]}(i, i+1) := \frac{1}{2\bar{\pi}(\bar{\mathcal{X}}_i)} \sum_{a\in\bar{\mathcal{X}}_i} \sum_{b\in\bar{\mathcal{X}}_{i+1}} \bar{P}(a,b)\bar{\pi}(a) \geq \frac{p_1 C_3}{N} \frac{1}{2\bar{\pi}(\bar{\mathcal{X}}_i)} \sum_{a\in\bar{\mathcal{X}}_i} \sum_{b\in\bar{\mathcal{X}}_{i+1}} \bar{\pi}(a) \geq \frac{p_1 C_3}{2N}$$

and analogously,

$$P_{[\![N]\!]}(i, i-1) \geq \frac{p_1 C_3}{2N}.$$

Now, for $r \neq 1$

$$\bar{\pi}_{[\![N]\!]}(r) = q_{[\![N]\!]}(r) / (\sum_{i=0}^{N} q_{[\![N]\!]}(i))$$

while

$$\bar{\pi}_{[\![N]\!]}(1) = (q_{[\![N]\!]}(1) + q_{[\![N]\!]}(0))/(\sum_{i=0}^{N} q_{[\![N]\!]}(i)).$$

So, using the unimodality of $q_{[\![N]\!]}$, we can apply Lemma B.3 with

$$A = \frac{p_1 C_3}{2} \qquad B = \frac{e^{-2\beta}}{2}$$

which gives

$$\lambda_1(P_{[\![N]\!]}) \le 1 - \frac{p_1 C_3}{e^{-2\beta}} \frac{1}{N^3} \le 1 - \frac{p_1 C_3}{N^3}.$$

Using another time Lemma B.3, by (A.1), we get

$$\lambda_N(P_{[\![N]\!]}) \ge 1 - p_1.$$

Combining this two bounds we have

$$Gap(P_{[\![N]\!]}) \ge \frac{C_3 p_1}{N^3} \tag{B.9}$$

and so from (3.4)

$$Gap(\bar{P}) \ge \frac{C p_1^2}{2N^6},$$

C being a suitable constant that depends by β, K, C_3, C_4, C_5. □

Proof of Proposition 6.4. Combine Lemma 6.3 with 6.5. □

Acknowledgments

We should like to thank Persi Diaconis for useful discussions and for having encouraged us during this work, Antonietta Mira for suggesting some interesting references and Claudio Giberti for helping to improve an earlier version of the paper.

References

[1] J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B*, 55(1):25–37, 1993.

[2] M. Blume, V. J. Emery, and R. B. Griffiths. Ising model for the λ transition and phase separation in he3-he4 mixtures. *Phys. Rev. A*, 4:1071–1077, 1971.

[3] P. Bremaud. *Markov Chains*. Springer-Verlag, New York, 1998.

[4] S. Caracciolo, A. Pelissetto, and A. D. Sokal. Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Statist. Phys.*, 60(1-2):1–53, 1990.

[5] S. Caracciolo, A. Pelissetto, and A. D. Sokal. Dynamic critical exponent of the BFACF algorithm for self-avoiding walks. *J. Statist. Phys.*, 63(5-6):857–865, 1991.

[6] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, 57(1):20–36, 1998. 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV).

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

[8] M. Dyer, L. A. Goldberg, M. Jerrum, and R. Martin. Markov chain comparison. *Probab. Surv.*, 3:89–111 (electronic), 2006.

[9] R. G. Edwards and A. D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D (3)*, 38(6):2009–2012, 1988.

[10] R. S. Ellis, P. T. Otto, and H. Touchette. Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model. *Ann. Appl. Probab.*, 15(3):2203–2254, 2005.

[11] R.S. Ellis. *The Theory of Large Deviation and Applications to Statistical Mechanics*. 2006. Lectures for the international seminar on Extreme Events in Complex Dynamics. http://www.math.umass.edu/ ˜ rsellis/pdf-files/Dresden-lectures.pdf Max-Planck-Institut.

[12] A. E. Gelfand, S. K. Sahu, and B. P. Carlin. Efficient parameterisations for normal linear mixed models. *Biometrika*, 82(3):479–488, 1995.

[13] A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, New York, 1996.

[14] C. J. Gey er and E. A. Thompson. Annealing markov chain monte carlo with applications to ancestral inference. *J. Amer. Statist. Assoc.*, 90:909–920, 1995.
[15] Y. Guan, R. Fleißner, P. Joyce, and S. M. Krone. Markov chain Monte Carlo in small worlds. *Stat. Comput.*, 16(2):193–202, 2006.
[16] Y. Guan and S. M. Krone. Small-world mcmc and convergence to multi-modal distributions: From slow mixing to fast mixing. *Ann. Appl. Probab.*, 17(1):284–304, 2007.
[17] J. M. Hammersley and D. C. Handscomb. *Monte Carlo methods*. Methuen & Co. Ltd., London, 1965.
[18] W.K. Hastings. Monte carlo sampling methods using markov chains and their application. *Biometrika*, 57:97–109, 1970.
[19] K. Hukushima and K. Nemoto. Exchange monte carlo method and application to spin glass simulations. *J.Phys.Soc.Jpn.*, 65:1604–1608, 1996.
[20] M. Jerrum, J. Son, P. Tetali, and E. Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.*, 14(4):1741–1765, 2004.
[21] S. A. Kivelson, V. J. Emery, and H. Q. Lin. Doped antiferromagnets in the weak-hopping limit. *Phys. Rev. B*, 42:6523–6530, 1990.
[22] S. C. Kou, Qing Zhou, and Wing Hung Wong. Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.*, 34(4):1581–1619, 2006.
[23] J. Lajzerowicz and J. Sivardiére. Spin 1 lattice gas model. ii. condensation and phase separation in a binary fluid. *Phys. Rev. A*, 11:2090–2100, 1975.
[24] J. Lajzerowicz and J. Sivardiére. Spin 1 lattice gas model. iii. tricritical points in binary and ternary fluids. *Phys. Rev. A*, 11:2101–2110, 1975.
[25] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer-Verlag, New York, 2001.
[26] N. Madras and M. Piccioni. Importance sampling for families of distributions. *Ann. Appl. Probab.*, 9(4):1202–1225, 1999.
[27] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Ann. Appl. Probab.*, 12(2):581–606, 2002.
[28] E. Marinari and G. Parisi. Simulated tempering: a new monte carlo scheme. *Europhy s. Lett*, 19:451–458, 1992.
[29] N. Metrpolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state caluclations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.
[30] A. Mira and L. Tierney. Efficiency and convergence properties of slice samplers. *Scand. J. Statist.*, 29(1):1–12, 2002.
[31] K.E. Newman and J.D. Dow. Zinc blende diamond order disorder transition in metastable crystalline (gaas)1-xge2x alloys. *Phys. Rev. B*, 27:7495–7508, 1983.
[32] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973.
[33] P. H. Peskun. Guidelines for choosing the transition matrix in Monte Carlo methods using Markov chains. *J. Comput. Phys.*, 40(2):327–344, 1981.
[34] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.
[35] R. Y. Rubinstein. *Simulation and the Monte Carlo method*. John Wiley & Sons Inc., New York, 1981. Wiley Series in Probability and Mathematical Statistics.
[36] M. Schick and Wei-Heng Shih. Spin 1 model of a microemulsion. *Phys. Rev. B*, 34:1797–1801, 1986.
[37] A. Sinclair. *Algorithms for random generation and counting*. Progress in Theoretical Computer Science. Birkhäuser Boston Inc., Boston, MA, 1993. A Markov chain approach.
[38] J. Sivardiére and J. Lajzerowicz. Spin 1 lattice gas model. i. condensation and solidification of a simple fluid. *Phys. Rev. A*, 11:2090–2100, 1975.
[39] A. Sokal. Monte Carlo methods in statistical mechanics: foundations and new algorithms. In *Functional integration (Cargèse, 1996)*, volume 361 of *NATO Adv. Sci. Inst. Ser. B Phys.*, pages 131–192. Plenum, New York, 1997.
[40] R.H. Swendsen and J. S. Wang. Non–universal critical dynamics in monte carlo simulations. *Phy. Rev. Lett.*, pages 86–88, 1987.

Università degli Studi di Pavia, Dipartimento di Matematica, via Ferrata 1, 27100 Pavia, Italy

Università dell'Insubria, Dipartimento di economia, via monte generoso 71, 21100 Varese , Italy

E-mail address: federico.bassetti@unipv.it
E-mail address: leisen.fabrizio@unimore.it