The Annals of Applied Probability
2017, Vol. 27, No. 2, 846–882
DOI: 10.1214/16-AAP1217
© Institute of Mathematical Statistics, 2017

A PIECEWISE DETERMINISTIC SCALING LIMIT OF LIFTED METROPOLIS–HASTINGS IN THE CURIE–WEISS MODEL[1]

BY JORIS BIERKENS AND GARETH ROBERTS

Delft University of Technology and University of Warwick

In Turitsyn, Chertkov and Vucelja [*Phys. D* **240** (2011) 410–414] a nonreversible Markov Chain Monte Carlo (MCMC) method on an augmented state space was introduced, here referred to as Lifted Metropolis–Hastings (LMH). A scaling limit of the magnetization process in the Curie–Weiss model is derived for LMH, as well as for Metropolis–Hastings (MH). The required jump rate in the high (supercritical) temperature regime equals $n^{1/2}$ for LMH, which should be compared to n for MH. At the critical temperature, the required jump rate equals $n^{3/4}$ for LMH and $n^{3/2}$ for MH, in agreement with experimental results of Turitsyn, Chertkov and Vucelja (2011). The scaling limit of LMH turns out to be a nonreversible piecewise deterministic exponentially ergodic "zig-zag" Markov process.

1. Introduction. Markov chain Monte Carlo [MCMC, Metropolis et al. (1953)] has been extremely successful in providing a generic simulation framework with wide-ranging applications. It works by composing collections of move types, each which leave the target distribution of interest invariant. Invariance is assured through detailed balance making the building blocks of MCMC reversible, giving advantages in terms of accessibility to mathematical investigation and practicality of implementation. Yet there is a growing interest in the phenomenon that, where comparative mathematical studies are possible, nonreversible Markov chains often outperform their reversible counterparts.

A fundamental approach for obtaining nonreversible Markov processes is by "lifting" or "augmenting" the state space. In this case, the states are augmented by one or more additional variables, which can often be interpreted as, for example, momentum or direction. Let us provide a (nonexhaustive) overview of the literature concerning this approach. In Chen, Lovász and Pak (1999), it is shown that "lifting" a Markov chain may result in a reduced mixing time that is at best of order square root of the original mixing time. In order to achieve this improvement, a nonreversible lifting is required. In Diaconis, Holmes and Neal (2000), a simple reversible Markov chain on a finite state space of size n, is lifted to a nonreversible Markov chain on a space of size $2n$. It is shown that this construction reduces the

Received July 2015; revised March 2016.

[1]Supported by EPSRC Grants EP/D002060/1 (CRiSM) and EP/K014463/1 (iLike).

MSC2010 subject classifications. Primary 60F05; secondary 65C05.

Key words and phrases. Weak convergence, Markov chain Monte Carlo, piecewise deterministic Markov process, phase transition, exponential ergodicity.

mixing time of the chain from $O(n^2)$ to $O(n)$. In Turitsyn, Chertkov and Vucelja (2011) a nonreversible lifting of Metropolis–Hastings is introduced, which we will refer to as *Lifted Metropolis–Hastings (LMH)*, and applied to the Ising model on a fully connected graph (i.e., the *Curie–Weiss model*). In a numerical experiment, it appears that at the critical temperature, the "relaxation time" is reduced from $O(n^{1.43})$ to $O(n^{0.85})$, where n denotes the number of spins. The "lifting approach" is not the only way of obtaining nonreversible Markov processes. For interesting approaches to constructing and analyzing the benefits of nonreversible Markov processes see, for example, Hwang, Hwang-Ma and Sheu (1993), Sun, Gomez and Schmidhuber (2010), Chen and Hwang (2013), Rey-Bellet and Spiliopoulos (2015), Bierkens (2016), Lelièvre, Nier and Pavliotis (2013), Duncan, Lelièvre and Pavliotis (2016).

It is the goal of this paper to shed light on the general theory of lifted nonreversible Markov chains, and in particular on the recent experimental result of Turitsyn, Chertkov and Vucelja (2011) mentioned above. This is achieved by obtaining a scaling limit of Lifted Metropolis–Hastings, in its application to the Curie–Weiss model. This scaling limit may be compared to a similar scaling limit for (classical) Metropolis–Hastings.

Initiated by Roberts, Gelman and Gilks (1997), a large amount of understanding of particular Markov Chain Monte Carlo (MCMC) algorithms has been obtained by identifying a suitable diffusion limit: Given a sequence of Markov chains of increasing size or dimensionality n, a suitable scaling of the state space and of the amount of steps per unit time interval is determined. As n tends towards infinity, the scaled Markov process converges (in the sense of weak convergence on Skorohod path space) to a diffusion process, which is often of an elementary nature. In particular the required number of Markov chain transitions per unit time interval as a function of n provides a fundamental measure of the speed of the Markov chain.

The Curie–Weiss model is an exchangeable probability distribution on $\{-1, 1\}^n$ which depends on two parameters, the "external field" h, and the "inverse temperature" β (which describes interactions between components). At inverse temperature $\beta = 1$, the model undergoes a phase transition. This results in differences in behaviour for $\beta < 1$, $\beta = 1$ and $\beta > 1$, and we shall analyse the behaviour of both standard Metropolis–Hastings and Lifted Metropolis–Hastings in the first two of these cases. We will determine a scaling limit of Metropolis–Hastings [Hastings (1970), Metropolis et al. (1953)] as well as Lifted Metropolis–Hastings [Turitsyn, Chertkov and Vucelja (2011)] for the magnetization in the Curie–Weiss model, for the supercritical temperature regime ($\beta < 1$) with external field $h \in \mathbb{R}$ and at the critical temperature ($\beta = 1$), without external field ($h = 0$). To obtain these results, we depend on nonasymptotic concentration results of Chatterjee (2007), Chatterjee and Dey (2010). The case of subcritical temperature ($\beta > 1$) is typically more difficult to analyse. In this paper, we do not obtain results for this case because, as far as we know, no nonasymptotic concentration results are available.

TABLE 1
Time to convergence for Metropolis-Hastings and Lifted Metropolis-Hastings

	$\beta < 1$	$\beta = 1$
Metropolis–Hastings	$O(n)$ (Theorem 1)	$O(n^{3/2})$ (Theorem 2)
Lifted Metropolis–Hastings	$O(n^{1/2})$ (Theorem 3)	$O(n^{3/4})$ (Theorem 4)

In both the supercritical and critical cases, our results demonstrate that the lifted chain convergence is an order of magnitude faster (as a function of dimension n) than the regular Metropolis–Hastings algorithm as is summarised in Table 1.

The results for Lifted Metropolis–Hastings are surprising since it would require at least $O(n)$ iterations to update each component. Therefore, magnetization is converging significantly more rapidly than should be expected. This is explained by the strong concentration of the magnetization around its mean, so that only relatively few spin updates suffice to update the magnetization at the appropriate scale.

As is common for weak limits of light-tailed Metropolis–Hasting algorithms, the limits of Metropolis–Hastings for Curie–Weiss are simple univariate diffusion processes. Interestingly, in determining the scaling limit of Lifted Metropolis–Hastings we obtain an elementary Markov process which has so far received only very limited attention in the literature. The limit process is a one-dimensional piecewise deterministic Markov process which we will refer to as a *zig-zag process*: the process moves at a deterministic and constant speed, until it switches direction and moves at the same speed but in the opposite direction. The switching occurs at a time-inhomogeneous rate which is directly related to the derivative of the density function of its stationary distribution. We analyse this zig-zag process in some detail, establishing in particular exponential ergodicity under mild conditions.

Piecewise deterministic Markov processes were first introduced in Davis (1984) and discussed extensively in Davis (1993). A zig-zag process with a constant switching rate appears in Goldstein (1951) and is discussed further in Kac (1974). A similar process on the torus is discussed in Miclo and Monmarché (2013). In Peters and De With (2012), a multi-dimensional version of the zig-zag process with space inhomogeneous switching rates is introduced and used for MCMC. This MCMC method is analysed in detail in Bouchard-Côté, Vollmer and Doucet (2015). Monmarché (2016) discusses the use of the one-dimensional zig-zag process for simulated annealing, and in Monmarché (2014) the exponential ergodicity of the zig-zag process is discussed in case of bounded switching rates. In Fontbona, Guérin and Malrieu (2012, 2016) the exponential ergodicity of the one-dimensional zig-zag process is discussed under seemingly stronger conditions than in the current paper.

1.1. *Outline*. This article is structured as follows. In Section 2, we briefly provide the necessary background on Metropolis–Hastings (MH), Lifted Metropolis–Hastings [LMH, based on Turitsyn, Chertkov and Vucelja (2011)], and the Curie–Weiss model, including the relatively recent nonasymptotic concentration results of Chatterjee (2007), Chatterjee and Dey (2010). Also we briefly describe the basic random walk Markov chain, used as proposal chain in MH and LMH, in terms of magnetization.

In Section 3, we consider the time evolution of the magnetization as $n \to \infty$ for MH applied to the Curie–Weiss model. By a suitable rescaling of both space (i.e., the magnetization variable) and time (i.e., the jump rate within a unit time interval) we obtain a diffusion limit of this stochastic process, at supercritical temperature $\beta < 1$ (Theorem 1) as well as at critical temperature, $\beta = 1$, $h = 0$ (Theorem 2). It is perhaps not very surprising that this diffusion limit corresponds to the Langevin diffusion of the known limiting distribution, that is, a Gaussian distribution in case $\beta < 1$ and non-Gaussian in case $\beta = 1$. Also not surprisingly, the required jump rate to obtain this diffusion limit is in line with known results on mixing time for Curie–Weiss [Levin, Luczak and Peres (2009), Ding, Lubetzky and Peres (2009)]: if $\beta < 1$, the required speed up is equal to a factor n, while for $\beta = 1$ and $h = 0$ the required speed up is equal to a factor $n^{3/2}$.

The main result of this paper may be found in Section 4. In this section, we obtain the scaling limit of the magnetization for LMH applied to Curie–Weiss, again for the cases $\beta < 1$ (Theorem 3) and $\beta = 1$, $h = 0$ (Theorem 4). The limiting process is a piecewise deterministic Markov process which has received only a small amount of attention in the mathematics and physics literature. Naturally, it has the same limiting invariant distribution as for Metropolis–Hastings. Interestingly, the required time scaling for LMH corresponds exactly to the square root of the time scaling for MH: this time scaling is $n^{1/2}$ for $\beta < 1$ and $n^{3/4}$ for $\beta = 1$, $h = 0$. This "square root" improvement is in agreement with the theory developed in Chen, Lovász and Pak (1999) and in line with the numerical result of Turitsyn, Chertkov and Vucelja (2011).

In Section 5, the limiting "zig-zag" process is analysed in detail. First, the process is generalized to allow for general one-dimensional invariant distributions satisfying mild conditions on the derivative of the density function. In particular, it is established that this process is a nonexplosive process satisfying the strong Markov property (Proposition 3) which is weak Feller (Proposition 4) but not strong Feller (Observation 1). A straightforward relation between the switching rate of the process and its invariant distribution is obtained in Proposition 5. Furthermore, under a reasonable strengthening of the assumptions exponential ergodicity is obtained (Theorem 5).

Section 6 is devoted to the proofs of the mentioned results, including necessary technical lemmas. In particular, let us mention the following intermediate results: The Feller property is obtained by a coupling argument (Proof of Proposition 4), it is shown that all compact sets are "petite sets" for the zig-zag process (Lemma 15),

and a Foster–Lyapunov function is constructed to establish exponential ergodicity (Lemma 16).

2. Preliminaries.

2.1. *Metropolis–Hastings* (*MH*). For a given proposal transition probability matrix Q and target distribution π on a discrete state space S, the Metropolis–Hastings transition probabilities are given for $x \neq y$ by

$$P(x,y) = \begin{cases} Q(x,y)\left(1 \wedge \dfrac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right), & \text{if } Q(x,y) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

As is well established, the transition probabilities P are reversible with respect to π, that is, $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x, y \in S$. This implies that π is invariant for P.

2.2. *Lifted Metropolis–Hastings* (*LMH*). In Turitsyn, Chertkov and Vucelja (2011), a nonreversible chain T is constructed with invariant distribution $\frac{1}{2}(\pi, \pi)$ on an augmented state space $S^\sharp := S \times \{-1, +1\}$. The set $S \times \{+1\}$ is called the *forward replica* and $S \times \{-1\}$ is called the *backward replica*. The construction is as follows. Let $T^+(x,y)$ and $T^-(x,y)$, $x \neq y$, denote probabilities satisfying the following conditions:

(i) $T^\pm(x,y) \geq 0$ for all $x, y \in S$, $x \neq y$;
(ii) $\sum_{y \in S, y \neq x} T^\pm(x,y) \leq 1$ for all $x \in S$;
(iii) skew detailed balance:

$$\pi(x)T^+(x,y) = \pi(y)T^-(y,x) \qquad \text{for all } x \neq y. \tag{2}$$

The values T^+ and T^- will represent transition probabilities within the respective replicas. Define transition probabilities between the forward and backward replicas by

$$\begin{aligned} T^{-+}(x) &= \max\Bigg(0, \sum_{\substack{y \in S \\ y \neq x}} \big(T^+(x,y) - T^-(x,y)\big)\Bigg), \\ T^{+-}(x) &= \max\Bigg(0, \sum_{\substack{y \in S \\ y \neq x}} \big(T^-(x,y) - T^+(x,y)\big)\Bigg). \end{aligned} \tag{3}$$

Finally, for $x \in S$, define $T^+(x,x)$ and $T^-(x,x)$ by

$$T^+(x,x) = 1 - T^{+-}(x) - \sum_{\substack{y \in S \\ y \neq x}} T^+(x,y) \quad \text{and}$$

$$T^-(x,x) = 1 - T^{-+}(x) - \sum_{\substack{y\in S\\ y\neq x}} T^-(x,y),$$

so that the rows sums equal 1. Define the full matrix of transition probabilities T with state space $S \times \{-1,+1\}$ by

$$\begin{aligned}
&T\big((x,-1),(y,-1)\big) = T^-(x,y), \qquad T\big((x,+1),(y,+1)\big) = T^+(x,y),\\
&T\big((x,+1),(x,-1)\big) = T^{+-}(x), \qquad T\big((x,-1),(x,+1)\big) = T^{-+}(x),\\
&T\big((x,-1),(y,+1)\big) = 0 \qquad \text{for } x \neq y,\\
&T\big((x,+1),(y,-1)\big) = 0 \qquad \text{for } x \neq y,
\end{aligned}$$

or in matrix notation,

$$T = \begin{pmatrix} T^+ & \operatorname{diag}(T^{+-}) \\ \operatorname{diag}(T^{-+}) & T^- \end{pmatrix}.$$

A few important properties of T are stated in the following proposition. Most importantly, the marginal invariant distribution of T over S is equal to π.

PROPOSITION 1. *Let T be as constructed above. Then*:

(i) *T is a Markov transition matrix*,

(ii) *T has invariant probability distribution on $S \times \{-1,+1\}$ equal to $\frac{1}{2}(\pi,\pi)$, and*

(iii) *if, for some $x, y \in S$, $T^+(x,y) \neq T^-(x,y)$, then T is not reversible with respect to its invariant distribution.*

PROOF. The proofs of these results are immediate. □

REMARK 1. Once T^+ and T^- are picked, T is fixed according to the definitions above. However, there is still freedom in choosing T^+ and T^- satisfying (2). In Turitsyn, Chertkov and Vucelja (2011) and here, T is fixed as follows. Suppose that P is a transition matrix on S that is reversible with respect to π, and let $\eta : S \to \mathbb{R}$. Now define the off-diagonal components of $T^\pm$ by

$$T^+(x,y) = \begin{cases} P(x,y), & \text{if } \eta(y) \geq \eta(x),\\ 0, & \text{if } \eta(y) < \eta(x), \end{cases} \quad \text{and}$$

$$T^-(x,y) = \begin{cases} 0, & \text{if } \eta(y) > \eta(x),\\ P(x,y), & \text{if } \eta(y) \leq \eta(x). \end{cases}$$

Then $T^\pm$ satisfies the skew detailed balance condition (2). This way, Lifted Metropolis–Hastings creates a nonreversible lifted chain T out of a given reversible chain P, which has (marginally) the same invariant distribution as P. In particular, this construction may be applied to the Metroplis–Hastings transition probabilities P given by (1).

REMARK 2. There is some freedom in the choice of transition probabilities between replicas, that is, T^{+-}, T^{-+}. In general, transition probabilities between replicas need to satisfy the conditions

$$T^{+-}(x) - T^{-+}(x) = \sum_{\substack{y \in S \\ y \neq x}} \big[T^{-}(x, y) - T^{+}(x, y)\big], \tag{4}$$

in order for $\frac{1}{2}(\pi, \pi)$ to be invariant. Here, as in Turitsyn, Chertkov and Vucelja (2011), we choose (3). See Sakai and Hukushima (2013) for other variants.

2.3. *The Curie–Weiss model.* Let $S^n := \{-1, 1\}^n$ and let target invariant distributions π^n on S^n be given by

$$\pi^n(x) = Z_n \exp\big(-\beta H^n(x)\big), \tag{5}$$

with

$$H^n(x) = -\frac{1}{2n} \sum_{i,j=1}^{n} x_i x_j - h \sum_{i=1}^{n} x_i, \tag{6}$$

where (Z_n) are normalization constants, β is a parameter usually referred to as inverse temperature, and $h \in \mathbb{R}$ a parameter known as the external magnetization. As remarked in the Introduction, we will later specialize to the case $0 \leq \beta \leq 1$, but for now we allow general $\beta \geq 0$. Define the *magnetization* $m^n : \{-1, 1\}^n \to \mathbb{R}$ by $m^n(x) = \frac{1}{n} \sum_{i=1}^{n} x_i$. The crucial observation for the Curie–Weiss model is that the Hamiltonian may be expressed in terms of m, as

$$H^n(x) = -n\left(\frac{1}{2}\big(m^n(x)\big)^2 + h m^n(x)\right). \tag{7}$$

We may consider m^n and other mappings from S^n into $\mathbb{R}$ as random variables on the probability space (S^n, π^n); in particular, we will suppress the dependence on $x \in S^n$ where this does not cause confusion.

For $0 \leq \beta \leq 1$, as well as for $\beta > 1$ and $h \neq 0$, there exists a unique $m_0 = m_0(h, \beta)$ around which the magnetization will concentrate. The value of m_0 can be obtained as the unique minimizer of

$$i(m) = -\left(\frac{1}{2}\beta m^2 + \beta h m\right) + \frac{1-m}{2}\log(1-m) + \frac{1+m}{2}\log(1+m), \qquad m \in (-1, 1). \tag{8}$$

This value m_0 satisfies

$$\beta m_0 + \beta h = \frac{1}{2}\log\frac{1+m_0}{1-m_0}, \tag{9}$$

or equivalently $m_0 = \tanh(\beta(m_0 + h))$. In case $\beta > 1$ and for h sufficiently small, there exist two other solutions to (9) but these are not global minima of (8). As

$h \to 0$, $m_0(h, \beta) \to 0$. For $h = 0$ or $\beta = 0$, $m_0 = 0$. For $h \neq 0$, the sign of m_0 is equal to the sign of h. These results are well known; see, for example, Ellis (2006), Section IV.4.

As $n \to \infty$, the random variables m^n will be increasingly concentrated around m_0.

PROPOSITION 2 (Concentration for Curie–Weiss). (i) *For all* $\beta \geq 0$, $h \in \mathbb{R}$ *and* $t \geq 0$,

$$\pi^n\left(|m^n - \tanh(\beta(m^n + h))| \geq \frac{\beta}{n} + \frac{t}{\sqrt{n}}\right) \leq 2\exp\left(-\frac{t^2}{4(1+\beta)}\right).$$

(ii) *If* $h = 0$ *and* $\beta = 1$, *then there exists a constant* $c > 0$ *such that for any* $n \in \mathbb{N}$ *and* $t \geq 0$,

$$\pi^n\left(|m^n| \geq t^{1/4}\right) \leq 2e^{-cnt}.$$

PROOF. Claim (i) is Chatterjee (2007), Proposition 1.3. Claim (ii) is a simple consequence of Chatterjee and Dey (2010), Proposition 5. $\square$

REMARK 3. In case $\beta > 1$ and $h \neq 0$, there is a unique global minimum of (8). However, to develop scaling limits for Metropolis–Hastings and Lifted Metropolis–Hastings we require nonasymptotic concentration results as given in Lemmas 5 and 7, which are based upon Proposition 2. Even though Proposition 2 includes the case $\beta > 1$, the proof of Lemma 5 seems to depend crucially on the assumption that $\beta < 1$. Therefore, we have to restrict our attention to $0 \leq \beta < 1$ (along with the critical case $h = 0$, $\beta = 1$).

As quantity of interest (which is a necessary ingredient in the formulation of the lifted Markov chain; see Remark 1), we will consider suitably shifted and renormalized magnetization,

$$\eta^n(x) := n^\gamma\left(m^n(x) - m_0\right), \qquad x \in S^n.$$

In view of Proposition 2, for η^n to be of $O(1)$ as $n \to \infty$, we will need to choose $\gamma = 1/2$ for $0 \leq \beta < 1$ and $\gamma = 1/4$ for $\beta = 1$. For smaller choices of γ, any limiting random variable would be trivially concentrated at a single point, whereas for larger choices of γ a suitable limiting random variable would not exist. The precise concentration statements we will use are given in Lemmas 5 (for $0 \leq \beta < 1$) and 7 (for $\beta = 1$). For now, we will only assume that $\gamma \in (0, 1)$.

Rather than using x as state space variable, it will be useful to express all quantities and probabilities in terms of $\eta^n(x)$. For example, the Hamiltonian H^n can be re-expressed in terms of $\eta^n(x)$ by $H^n(x) = c_n + \Phi^n(\eta^n(x))$, where the constants c_n do not depend on η^n, and

$$\Phi^n(\eta) := -\frac{1}{2}n^{1-2\gamma}\eta^2 - n^{1-\gamma}(m_0 + h)\eta, \qquad \eta \in \mathbb{R}. \tag{10}$$

2.4. *Random walk on the discrete hypercube.* Consider the Markov transition probabilities on $S^n = \{-1, 1\}^n$ given by

$$\mathrm{Prob}(x \to y) = \begin{cases} \dfrac{1}{n}, & \text{when } y = F_k(x) \text{ for some } k = 1, \dots, n, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $F_k : S^n \to S^n$ denotes the operation of flipping the sign of $x(k)$, that is,

$$\big[F_k(x)\big]_i := \begin{cases} x_i, & \text{for } i \neq k, \\ -x_i, & \text{for } i = k. \end{cases}$$

In words, a transition consists of flipping the sign of x_i, where i is selected uniformly among $\{1, \dots, n\}$. This Markov chain corresponds to a random walk on the discrete hypercube S^n. We will express the above transition probabilities in terms of $\eta = \eta^n(x)$ rather than x. For $\eta = \eta^n(x)$, a fraction $\frac{1}{2}(1 - m^n(x)) = \frac{1}{2}(1 - m_0 - n^{-\gamma}\eta)$ of entries of x has value -1, and similarly a fraction $\frac{1}{2}(1 + m_0 + n^{-\gamma}\eta)$ has value $+1$. If one entry of x flips, there is a change in η^n by $2n^{\gamma-1}$. Therefore, for $\eta \in X^n := \eta^n(S^n)$, we define

$$Q^n\big(\eta, \eta \pm 2n^{\gamma-1}\big) := \frac{1}{2}\big(1 \mp \big(m_0 + n^{-\gamma}\eta\big)\big), \tag{11}$$

and $Q^n(\eta, \zeta) := 0$ for all $\eta, \zeta \in X^n$ for which $|\zeta - \eta| \neq 2n^{\gamma-1}$. Defined this way, Q^n is a matrix of transition probabilities on X^n.

3. Diffusion limit of Metropolis–Hastings applied to Curie–Weiss. In this section, we consider the limit of Metropolis–Hastings for the Curie–Weiss model as $n \to \infty$ in terms of the scaled magnetization $\eta^n(x) = n^\gamma(m^n(x) - m_0)$. In terms of η^n, the invariant distribution is given by

$$\mu^n(\eta) := \big(\pi^n \circ \big(\eta^n\big)^{-1}\big)(\eta) \propto \exp\big(-\beta\Phi^n(\eta)\big), \qquad \eta \in X^n. \tag{12}$$

Using the random walk transition probabilities Q^n and the target distribution μ^n for the Curie–Weiss model, we obtain for the MH transition probabilities

$$\begin{aligned} &P^n\big(\eta, \eta \pm 2n^{\gamma-1}\big) \\ &\quad = Q^n\big(\eta, \eta \pm 2n^{\gamma-1}\big)\big(1 \wedge \exp\big(\beta\{\Phi^n(\eta) - \Phi^n\big(\eta \pm 2n^{\gamma-1}\big)\}\big)\big), \end{aligned} \tag{13}$$

for $\eta \in X^n$, with Φ^n given by (10).

Let Y^n denote the stationary continuous time Markov chain that jumps at rate n^α according to P^n with stationary distribution $\mu^n \propto \exp(-\beta\Phi^n(\eta))$. Let $D([0, \infty), \mathbb{R})$ denote the space of cadlag paths in $\mathbb{R}$, equipped with the Skorohod topology. We are now in a position to state our first two results concerning the high-dimensional limit of Y^n in the supercritical and critical cases, respectively.

THEOREM 1 (Diffusion limit of Metropolis–Hastings in the supercritical temperature regime). *Suppose $0 \le \beta < 1$ and $h \in \mathbb{R}$. Suppose Y^n jumps at rate n, that is, we let $\alpha = 1$ in the above definition of Y^n. Let the spatial scaling in the transition probabilities P^n be determined by $\gamma = \frac{1}{2}$. Then Y^n converges weakly in $D([0,\infty),\mathbb{R})$ to Y, where Y is the stationary Ornstein–Uhlenbeck process satisfying the stochastic differential equation*

$$dY(t) = -2l(h,\beta)Y(t)\,dt + \sigma(h,\beta)\,dB(t), \qquad Y(0) \sim \mu,$$

and with stationary distribution μ, where $(B(t))$ is a standard Brownian motion, μ is the centred normal distribution with variance

$$v(h,\beta) := \frac{1 - m_0^2(h,\beta)}{1 - \beta(1 - m_0^2(h,\beta))} \tag{14}$$

and with

$$\sigma(h,\beta) := 2\sqrt{1 - |m_0(h,\beta)|} \quad \text{and}$$
$$l(h,\beta) := \frac{1}{1 + |m_0(h,\beta)|} - \beta\big(1 - |m_0(h,\beta)|\big).$$

The proof depends on the convergence of the infinitesimal generator of the Markov chain semigroup as, for example, Roberts, Gelman and Gilks (1997), Theorem 1.1, and is provided in Section 6.

THEOREM 2 (Diffusion limit of Metropolis–Hastings at the critical temperature). *Suppose $\beta = 1$ and $h = 0$. Suppose Y^n jumps at rate $n^{3/2}$, that is, we let $\alpha = 3/2$ in the definition of Y^n. Let the spatial scaling in the transition probabilities P^n be determined by $\gamma = \frac{1}{4}$. Then Y^n converges weakly in $D([0,\infty),\mathbb{R})$ to Y, where Y is the stationary Langevin process satisfying the stochastic differential equation*

$$dY(t) = -(2/3)\big(Y(t)\big)^3\,dt + 2\,dB(t), \qquad Y(0) \sim \mu$$

with $(B(t))$ a standard Brownian motion, where μ is the probability distribution on $\mathbb{R}$ with Lebesgue density

$$\frac{d\mu}{dy} = \left(\frac{4}{3}\right)^{1/4} \frac{\exp(-y^4/12)}{\Gamma(1/4)}.$$

The expression for the limiting non-Gaussian distribution for the Curie–Weiss model is well known; see, for example, Chatterjee and Dey (2010), page 4.

4. Scaling limit for lifted Metropolis–Hastings applied to Curie–Weiss. Carrying out the construction of Section 2.2, the Lifted Metropolis(–Hastings) scheme with random walk proposal leads to transition probabilities T^n in the space $X^n \times \{-1, +1\}$ given by

$$\begin{aligned} &T^n\big((\eta, +1), (\eta + 2n^{\gamma-1}, +1)\big) = p_+^n(\eta), \\ &T^n\big((\eta, -1), (\eta - 2n^{\gamma-1}, -1)\big) = p_-^n(\eta), \\ &\qquad T^n\big((\eta, +1), (\eta, -1)\big) = \max\big(0, p_-^n(\eta) - p_+^n(\eta)\big), \\ &\qquad T^n\big((\eta, -1), (\eta, +1)\big) = \max\big(0, p_+^n(\eta) - p_-^n(\eta)\big), \end{aligned} \tag{15}$$

and all other transition probabilities from $(\eta, \pm 1)$ to a different state are equal to zero. Here, $p_\pm^n = P^n(\eta, \eta \pm 2n^{\gamma-1})$, with P^n the transition probabilities of MH for the Curie–Weiss model, as given by (13). Recall that $p_\pm^n$, and hence T^n, depends on the choice of the spatial scaling parameter γ.

Let (Y^n, J^n) denote the stationary continuous time Markov chain which jumps at rate n^α according to T^n. Let

$$a(h, \beta) := 1 - |m_0| \tag{16}$$

and let $l(h, \beta)$ be as given in Theorem 1. In the supercritical temperature regime, with $0 \le \beta < 1$ and $h \in \mathbb{R}$, the limiting Markov process will be shown to have generator

$$L\varphi(\eta, j) = a(h, \beta) j \frac{\partial \varphi}{\partial \eta} + \max\big(0, jl(h, \beta)\eta\big)\big(\varphi(\eta, -j) - \varphi(\eta, j)\big), \tag{17}$$

with domain

$$D(L) = \left\{\varphi : \mathbb{R} \times \{-1, 1\} \to \mathbb{R}, \eta \mapsto \frac{\partial \varphi}{\partial \eta}(\eta, j) \in C_0(\mathbb{R}) \text{ for } j = \pm 1\right\},$$

where $C_0(\mathbb{R})$ is the Banach space of continuous functions on $\mathbb{R}$, vanishing at infinity. This scaling limit is obtained provided we choose the right speed factor: we have to jump at rate $n^{1/2}$. This is formulated in the following theorem.

THEOREM 3. *Suppose $0 \le \beta < 1$ and $h \in \mathbb{R}$. Suppose (Y^n, J^n) jumps at rate $n^{1/2}$, that is, we let $\alpha = 1/2$ in the definition of (Y^n, J^n). Let the spatial scaling in the transition probabilities T^n be determined by $\gamma = \frac{1}{2}$. Then (Y^n, J^n) converges weakly in $D([0, \infty), \mathbb{R} \times \{-1, 1\})$ to (Y, J), where (Y, J) is the stationary Markov process with generator L and stationary distribution $\frac{1}{2}\mu \otimes (\delta_{-1} + \delta_{+1})$, with $\mu = N(0, v(h, \beta))$ and $v(h, \beta)$ given by (14).*

It will be established in Section 5 that L is the generator of a Markov–Feller process. Let (Y, J) denote the continuous time Markov process with generator L. The interpretation of (Y, J) is straightforward: Y moves with constant drift $a(h, \beta)$

in the direction J, until it changes its direction to $-J$. The changes in direction occur at events generated by a time inhomogeneous Poisson process with switching rate given by $\max(0, J(t)l(h,\beta)Y(t))$. See Section 5 for a detailed discussion of this process. At the critical temperature, we have to jump at a faster rate $n^{3/4}$ to obtain a nontrivial limiting Markov process. The limiting process is slightly different (compared to the supercritical temperature regime) in the sense that it switches replicas at a modified (cubic) rate.

THEOREM 4. *Suppose $\beta = 1$ and $h = 0$. Suppose (Y^n, J^n) jumps at rate $n^{3/4}$, that is, we let $\alpha = 3/4$ in the definition of (Y^n, J^n). Let the spatial scaling in the transition probabilities T^n be determined by $\gamma = \frac{1}{4}$. Then (Y^n, J^n) converges weakly in $D([0,\infty), \mathbb{R}\times\{-1,1\})$ to (Y, J), where (Y, J) is the stationary Markov process with generator L given*

$$L\varphi(\eta, j) = j\frac{d\varphi}{d\eta}(\eta, j) + \max\bigl(0, 1/3j\eta^3\bigr)\bigl(\varphi(\eta, -j) - \varphi(\eta, j)\bigr), \tag{18}$$

with stationary distribution $\frac{1}{2}\mu \otimes (\delta_{-1} + \delta_{+1})$, where μ is as in Theorem 2.

REMARK 4. Analogous results can be obtained for the closely related Glauber dynamics and its lifted version. The only difference is that the resulting Langevin diffusion (for Glauber dynamics) and zig-zag process (for lifted Glauber dynamics) are a factor $2/(1 + |m_0|) \in (1, 2]$ slower than for MH and LMH.

5. The limiting zig-zag process. In this section, we will investigate a generalization of the Markov process with generator (17). Let $E = \mathbb{R}\times\{-1,+1\}$. For $\varphi : E \to \mathbb{R}$, we often write $\varphi^+(y) := \varphi(y, +1)$ and $\varphi^-(y) := \varphi(y, -1)$. If we write $\varphi^\pm$, we mean both φ^+ and φ^-. Equip E with the product topology and let $C(E)$ denote the space of continuous functions $\varphi : E \to \mathbb{R}$. Note that $\varphi \in C(E)$ if and only if $\varphi^\pm \in C(\mathbb{R})$. Let $C_0(E)$ denote the linear subspace of $\varphi \in C(E)$ which vanish at infinity, that is, $\varphi^\pm \in C_0(\mathbb{R})$ [where $C_0(\mathbb{R})$ denotes the Banach space of continuous functions on $\mathbb{R}$, vanishing at infinity]. Let $C^1(\mathbb{R})$ denote the space of continuously differentiable functions on $\mathbb{R}$.

Throughout this section, let $\lambda : E \to [0,\infty)$ be continuous, and $a > 0$. Introduce a densely defined linear operator on $C_0(E)$

$$L\varphi(y, j) = aj\frac{\partial\varphi}{\partial y}(y, j) + \lambda(y, j)\bigl(\varphi(y, -j) - \varphi(y, j)\bigr), \qquad y \in \mathbb{R},\ j = \pm 1, \tag{19}$$

with domain $D(L) = \{\varphi : E \to \mathbb{R}, \varphi^\pm \in C^1(\mathbb{R}), (L\varphi)^\pm \in C_0(\mathbb{R})\}$. It is easy to verify that L is closable.

5.1. *Construction of the zig-zag process.*

ASSUMPTION 1. There exist constants $y_0 \geq 0$ and $\lambda_{\min} > 0$ such that $\lambda(y, j) \geq \lambda_{\min}$ for $jy \geq y_0$.

We will call a switch from the (j)-replica to the $(-j)$-replica a "good switch" when $jy \geq y_0$, and a "bad switch" when $jy \leq -y_0$. For example, a switch from $+1$ to -1 is good for $y \geq y_0$, but bad for $y \leq -y_0$. Good switches make the process direct itself towards the origin, whereas bad switches do the opposite. If "too few" good switches occur, the process might wander off to infinity. Assumption 1 states that for $|y| \geq y_0$ there is a lower bound for the rate at which good switches occur.

For $(y, j) \in E$, define the survival function

$$F(t; y, j) := \exp\left(-\int_0^t \lambda(y + ajs, j)\, ds\right), \qquad t \geq 0. \tag{20}$$

Since λ is continuous, and hence bounded on compact sets, for every $(y, j) \in E$ and $t \geq 0$, $F(t; y, j) > 0$. It is established in Lemma 13 that for every $(y, j) \in E$, $1 - F(\cdot, y, j)$ is the distribution function of a strictly positive random variable that is almost surely finite. In fact, $1 - F(\cdot, y, j)$ will serve as the distribution of the random time at which the value of j will be switched, starting from (y, j).

Given $(y, j) \in E$, define the process $(Y(t), J(t))$ along with random variables $(Z_i)_{i \in \{1,2,\dots\}}$ and $(T_i)_{i \in \{0,1,2,\dots\}}$ as follows:

- Let $T_0 = 0$, $J(0) = j$, $Y(0) = y$.
- For $i = 1, 2, \dots$:
  - Let Z_i be distributed according to $\mathbb{P}_{y,j}(Z_i > t | Z_1, \dots, Z_{i-1}) = F(t; Y(T_{i-1}), J(T_{i-1}))$.
  - Let $T_i := T_{i-1} + Z_i$.
  - Define $J(t) = J(T_{i-1})$ for $T_{i-1} < t < T_i$ and $J(T_i) = -J(T_{i-1})$.
  - Define $Y(t) = Y(T_{i-1}) + J(T_{i-1})a(t - T_{i-1})$ for $T_{i-1} < t \leq T_i$.

Then $T_k := \sum_{i=1}^k Z_i$. The process $(Y(t))$ is continuous and piecewise linear, and $(J(t))$ is piecewise constant and right-continuous. It follows that $(Y(t), J(t))_{t \geq 0}$ is cadlag. For $t \geq 0$, let

$$N(t) := \sup\{k \in \mathbb{N} : T_k \leq t\},$$

the number of switches that have occurred up to time t. We have defined $(Y(t), J(t))$ up to $t < T_\infty := \lim_{k\to\infty} T_k \leq \infty$. By Lemma 14, we can exclude the possibility that $\lim_{k\to\infty} T_k < \infty$.

Let $\mathbb{P}_{y,j}$ denote the probability distribution conditional over these random variables given that $Y(0) = y$, $J(0) = j$. Let $\mathcal{F}_t := \sigma(\{(Y(s), J(s)) : s \leq t\})$.

PROPOSITION 3. *Suppose Assumption 1 holds. Then under $\mathbb{P}_{y,j}$, the process (Y, J) is a nonexplosive strong Markov process with respect to $(\mathcal{F}_t)$, with generator equal to the closure of* (19).

PROOF. This follows directly from general theory for piecewise deterministic Markov processes; see Davis (1984). □

5.2. *Regularity.* Let $P = (P(t))_{t\geq 0}$ denote the Markov semigroup corresponding to the zig-zag process (Y, J). By a coupling argument, we can establish the Feller property for P. The Feller property of piecewise deterministic Markov processes is established in Davis (1993) under the assumption of bounded switching rates, which is not satisfactory in our setting. The proofs of this proposition and subsequent results are located in Section 6.3.

PROPOSITION 4. *Suppose Assumption 1 holds. The Markov transition semigroup P with infinitesimal generator L is Feller, that is, for every $\varphi \in C_0(E)$ and $t \geq 0$, we have $P(t)\varphi \in C_0(E)$.*

Let $B_b(E)$ and $C_b(E)$ denote the sets of bounded Borel measurable functions and bounded continuous functions on E, respectively. Recall that $(P(t))_{t\geq 0}$ is strong Feller if $P(t)\varphi \in C_b(E)$ for any $t > 0$ and any $\varphi \in B_b(E)$. The transition semigroup corresponding to the zig-zag process does not satisfy this property.

OBSERVATION 1. *Suppose Assumption 1 holds. Then $(P(t))_{t\geq 0}$ is not strong Feller.*

PROOF. Let $j = +1$ and $y \in \mathbb{R}$. Let $t > 0$ and let $A = [y + at, \infty)$. Let $\varphi(y, j) = \mathbb{1}_A(y)$. Because $t < T_1$ implies $Y(t) \in A$, it follows that $P(t)\varphi(y, j) = \mathbb{P}_{y,j}(Y(t) \in A) \geq \mathbb{P}_{y,j}(T_1 > t) > 0$. However, $P(t)\varphi(z, j) = \mathbb{P}_{z,j}(Y(t) \in A) = 0$ for every $z < y$, so that $P(t)\varphi$ is not continuous. □

5.3. *Invariant measure.* Let us strengthen Assumption 1 into the following assumption.

ASSUMPTION 2. There exist constants $y_0 \geq 0$ and $\lambda_{\min} > 0$ such that:

(i) $\lambda(y, j) \geq \lambda_{\min}$ for $jy \geq y_0$, and
(ii) $\lambda(y, -j) \leq \lambda(y, j)$ for $jy \geq y_0$.

We strengthened Assumption 1 by requiring that in the tails the rate at which "good switches" (i.e., mean reverting switches) occur is higher than the rate of "bad switches".

PROPOSITION 5. *Suppose Assumption 2 holds. Let $\Psi : \mathbb{R} \to \mathbb{R}$ be defined by*

$$\Psi(y) = \frac{1}{a}\int_0^y \{\lambda^+(\eta) - \lambda^-(\eta)\}\, d\eta, \qquad y \in \mathbb{R}.$$

Then Ψ is bounded from below, $\Psi(y) < \infty$ for all $y \in \mathbb{R}$, and the Markov process (Y, J) has invariant measure μ with density $(y, j) \mapsto \exp(-\Psi(y))$ with respect to $\mathrm{Leb} \otimes (\delta_{-1} + \delta_{+1})$ *on E.*

Under the stated assumption, we cannot yet make any claims as to whether μ is a finite measure. As an example, consider the case in which $\lambda(y, \pm j) = \lambda_0 > 0$ for all (y, j), which satisfies Assumption 2. By Proposition 5, this corresponds to a uniform invariant density.

The proof of Proposition 5 is a simple computation that we will include here.

PROOF OF PROPOSITION 5. Note that Ψ and λ are related by

$$a\frac{d\Psi(y)}{dy} + \lambda^-(y) - \lambda^+(y) = 0, \qquad y \in \mathbb{R}. \tag{21}$$

It follows from Assumption 2 that Ψ is bounded from below and $\Psi(y) < \infty$ for all $y \in \mathbb{R}$. Suppose $\varphi \in D(L)$ and suppose μ is as specified. Then, using that Ψ is bounded from below and $\varphi \in C_0(E)$ in the partial integration below,

$$\begin{aligned}
&\sum_{j=-1,+1} \int_{-\infty}^{\infty} L\varphi(y, j)\, d\mu(y, j) \\
&\quad = \sum_{j=-1,+1} \int_{-\infty}^{\infty} \Big\{ aj \frac{\partial \varphi(y, j)}{\partial y} \\
&\qquad + \lambda(y, j)\big(\varphi(y, -j) - \varphi(y, j)\big) \Big\} \exp\big(-\Psi(y)\big)\, dy \\
&\quad = \sum_{j=-1,+1} \int_{-\infty}^{\infty} \Big\{ aj \frac{d\Psi(y)}{dy} - \lambda(y, j) \Big\} \varphi(y, j) \exp\big(-\Psi(y)\big)\, dy \\
&\qquad + \sum_{k=-1,+1} \int_{-\infty}^{\infty} \lambda(y, -k)\varphi(y, k) \exp\big(-\Psi(y)\big)\, dy \\
&\quad = \sum_{j=-1,+1} \int_{-\infty}^{\infty} \Big\{ aj \frac{d\Psi(y)}{dy} - \lambda(y, j) + \lambda(y, -j) \Big\} \varphi(y, j) \exp\big(-\Psi(y)\big)\, dy \\
&\quad = 0.
\end{aligned}$$

Note that we first let $k = -j$ and in the next step replaced k by j. It follows that

$$\sum_{j=-1,+1} \int_{-\infty}^{\infty} P(t)\varphi(y, j)\, d\mu(y, j) = \mu(\varphi), \qquad \varphi \in D(L), t \geq 0.$$

By a standard approximation argument, this holds for any $\varphi \in B_b(E)$, and it follows that μ is invariant for P. $\square$

5.4. *Exponential ergodicity.* We will further strengthen Assumption 2 into the following assumption, which therefore also implies Assumption 1.

ASSUMPTION 3. There is a $y_0 > 0$ such that:

(i) $\inf_{y \geq y_0} \lambda^+(y) > \sup_{y \geq y_0} \lambda^-(y)$, and
(ii) $\inf_{y \leq -y_0} \lambda^-(y) > \sup_{y \leq -y_0} \lambda^+(y)$.

LEMMA 1 (Invariant measure is finite). *Suppose Assumption 3 holds and Ψ satisfies (21). Then μ defined in Proposition 5 is finite, that is, $\mu(E) < \infty$.*

PROOF. Using Assumption 3, we have $\lambda^+(y) - \lambda^-(y) \geq c$ on $[y_0, \infty)$ for some $c > 0$. Therefore,

$$\int_{y_0}^{\infty} \exp\bigl(-\Psi(y)\bigr)\, dy \leq \int_{y_0}^{\infty} \exp\bigl(-\Psi(y_0) - c(y - y_0)\bigr)\, dy < \infty,$$

and similarly for the integral over $(-\infty, -y_0]$. $\square$

Without loss of generality, we will assume below that μ is a probability measure, that is, $\mu(E) = 1$. For $f : E \to [1, \infty)$ define the f-norm by

$$\|\mu\|_f = \sup_{|g| \leq f} \bigl|\mu(g)\bigr| \qquad \mu \text{ signed measure on } \mathcal{B}(E),$$

which is a stronger norm than the total variation norm. By characterizing the "petite sets" and using a Foster–Lyapunov function (Lemmas 15 and 16, respectively, located in Section 6), we can establish exponential ergodicity. We acknowledge the recommendation of a referee to use the Lyapunov function of Fontbona, Guérin and Malrieu (2016) instead of our earlier construction, which allowed us to further weaken the conditions under which we obtain exponential ergodicity. A function $V \in C(E)$ is norm-like if $\lim_{|x| \to \infty} V(x) = \infty$.

THEOREM 5. *Suppose Assumption 3 holds. Then $(Y(t), J(t))_{t \geq 0}$ is exponentially ergodic, that is, there exist constants $0 < \rho < 1$ and $0 < \kappa < \infty$ and a norm-like function V such that*

$$\bigl\|\mathbb{P}_{y,j}\bigl((Y(t), J(t)) \in \cdot\bigr) - \mu\bigr\|_f \leq \kappa f(y, j) \rho^t, \qquad t \geq 0,$$

where $f(y, j) = 1 + V(y, j)$.

PROOF. By Lemma 15 and Lemma 16, all conditions of Meyn and Tweedie (1993), Theorem 6.1, are satisfied, so that the stated result follows. $\square$

5.5. *Application to Curie–Weiss.* In the Curie–Weiss model, the generator obtained in Theorems 3 and 4 is given by (19) with $a = a(h, \beta)$ given by (16) and $\lambda(y, j) = \max(0, j \frac{d\Psi(y)}{dy})$, with

$$\Psi(y) = \begin{cases} y^4/12, & \beta = 1, h = 0, \\ \dfrac{1}{2} l(h, \beta) y^2, & 0 \leq \beta < 1, h \in \mathbb{R}, \end{cases}$$

with $l(h, \beta)$ given by (28). In particular $\lambda(y, j) > 0$ for $jy > 0$ and $\lambda(y, j) = 0$ for $jy \leq 0$. It follows that Assumption 3 is satisfied, taking any $y_0 > 0$. Assumptions 1 and 2 are weaker than Assumption 3. To summarize, we have the following corollary.

COROLLARY 1. *L given by (19), with a and $\lambda(y, j)$ as above, is the generator of a Markov–Feller transition semigroup on $C_0(E)$. The associated Markov process (Y, J) has finite invariant measure μ on E as in Proposition 5 and is exponentially ergodic.*

PROOF. This is a combination of Propositions 3, 4, 5 and Theorem 5. □

6. Proofs.

6.1. *Estimates on Metropolis–Hastings applied to Curie–Weiss.* We can easily compute the difference in interaction energy for increments in η,

$$\Phi^n(\eta) - \Phi^n\big(\eta \pm 2n^{\gamma-1}\big) = \pm 2n^{-\gamma}\eta \pm 2(m_0 + h) + 2n^{-1}. \tag{22}$$

Combined with (11) and (13), it follows that

$$\begin{aligned} p^n_{\pm}(\eta) &:= P^n\big(\eta, \eta \pm 2n^{\gamma-1}\big) \\ &= \frac{1}{2}\big(1 \mp \big(m_0 + n^{-\gamma}\eta\big)\big) \\ &\quad \times \big(1 \wedge \exp\big(\beta\{\pm 2n^{-\gamma}\eta \pm 2(m_0 + h) + 2n^{-1}\}\big)\big). \end{aligned} \tag{23}$$

Due to the possibility of rejection, there will be positive mass on transition probabilities $P^n(\eta, \eta)$. These values are fully determined by the off-diagonal transition probabilities and will not appear in the analysis below.

To rephrase slightly, for $\eta \in X^n$, define probability distributions $\mathbb{P}^n_\eta$ on X^n, and let Y denote X^n-valued random variables with distribution $P^n(\eta, \cdot)$. In other words, under $\mathbb{P}^n_\eta$, Y is distributed according to $P^n(\eta, \cdot)$. Expectation with respect to $\mathbb{P}^n_\eta$ will be denoted by $\mathbb{E}^n_\eta$, so that $\mathbb{E}^n_\eta[\varphi(Y)] = P^n\varphi(\eta)$ for $\varphi : X^n \to \mathbb{R}$.

We will be particularly interested in values of η that are concentrated on the following sets

$$F^{n,\delta} := \{\eta \in X^n : |\eta| \leq n^\delta\}, \tag{24}$$

where $\delta < \gamma$. In the computations that follow, we will frequently need to approximate the exponent in the Metropolis–Hastings acceptance probability by its Taylor approximation. The following lemma helps in determining the required order of approximation. Let $p_k(x)$ denote the kth order Taylor approximation of $\exp(x)$,

that is, $p_k(x) = \sum_{i=0}^{k} \frac{x^i}{i!}$. Define approximate transition probabilities

$$p_{\pm}^{n,k}(\eta) := \frac{1}{2}\big(1 \mp (m_0 + n^{-\gamma}\eta)\big) \times \begin{cases} 1, & \text{if } h + m_0 = 0 \text{ and } \pm\eta \geq 0, \\ p_k(\pm 2\beta n^{-\gamma}\eta), & \text{if } h + m_0 = 0 \text{ and } \pm\eta < 0, \\ \exp\big(\pm 2\beta(m_0 + h)\big) p_k(\pm 2\beta n^{-\gamma}\eta), & \text{if } \pm(h + m_0) < 0, \\ 1, & \text{if } \pm(h + m_0) > 0. \end{cases} \tag{25}$$

For example, if $\eta < 0$, and $h + m_0 = 0$, then

$$p_{+}^{n,k}(\eta) = \frac{1}{2}\big(1 - (m_0 + n^{-\gamma}\eta)\big) p_k(2\beta n^{-\gamma}\eta).$$

LEMMA 2. *Let $0 < \delta < \gamma < 1$. Suppose $h \neq 0$, $\beta > 0$ or $h = 0$, $0 \leq \beta \leq 1$. Then, for $r < \min(1, (k+1)(\gamma - \delta))$,*

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} n^r \big|p_{\pm}^{n}(\eta) - p_{\pm}^{n,k}(\eta)\big| \to 0.$$

PROOF. The result is trivial in case $\beta = 0$. In the remainder, therefore, assume $\beta > 0$. Define

$$\widetilde{p}_{\pm}^{n}(\eta) := \frac{1}{2}\big(1 \mp (m_0 + n^{-\gamma}\eta)\big)\big(1 \wedge \exp\big(\pm 2\beta\{n^{-\gamma}\eta + (m_0 + h)\}\big)\big).$$

[This is just $p_{\pm}^{n}(\eta)$ without the $O(n^{-1})$ term in the exponent.] We estimate, using 1-Lipschitz continuity of $x \mapsto 1 \wedge e^{-x}$ (for $x \geq 0$), $|m_0 + n^{-\gamma}\eta| \leq 1$, and $r < 1$,

$$\sup_{\eta \in F^{n,\delta}} n^r \big|p_{\pm}^{n}(\eta) - \widetilde{p}_{\pm}^{n}(\eta)\big| \leq \sup_{\eta \in F^{n,\delta}} \frac{1}{2}\big|1 \mp (m_0 + n^{-\gamma}\eta)\big| 2\beta n^{r-1} \leq 2\beta n^{r-1} \to 0$$

as $n \to \infty$. For $k \in \mathbb{N}$, define further approximate transition probabilities

$$\widetilde{p}_{\pm}^{n,k}(\eta) = \frac{1}{2}\big(1 \mp (m_0 + n^{-\gamma}\eta)\big) \min\big\{1, \exp\big(\pm 2\beta(m_0 + h)\big) p_k\big(\pm 2\beta n^{-\gamma}\eta\big)\big\}.$$

Then, using 1-Lipschitz continuity of $x \mapsto 1 \wedge x$, $|n^{-\gamma}\eta| \leq n^{\delta-\gamma} \leq 1$ on $F^{n,\delta}$, and $|p_k(x) - \exp(x)| \leq e^{\xi} \frac{x^{k+1}}{(k+1)!}$ for some $\xi \in (\min(0, x), \max(0, x))$, it follows that

$$\begin{aligned} &\sup_{\eta \in F^{n,\delta}} n^r \big|\widetilde{p}_{\pm}^{n}(\eta) - \widetilde{p}_{\pm}^{n,k}(\eta)\big| \\ &\quad \leq n^r \exp\big(\pm 2\beta(m_0 + h)\big)\big|\exp\big(\pm 2\beta n^{-\gamma}\eta\big) - p_k\big(\pm 2\beta n^{-\gamma}\eta\big)\big| \\ &\quad \leq n^r \exp\big(\pm 2\beta(m_0 + h)\big) \exp(2\beta) \frac{(2\beta n^{-\gamma}\eta)^{k+1}}{(k+1)!} \\ &\quad = c n^{r-(k+1)(\gamma-\delta)} \to 0 \end{aligned}$$

as $n \to \infty$. In the limit as $n \to \infty$, the minimization in the expression for $\widetilde{p}_{\pm}^{n,k}(\eta)$ will only depend on the lowest order terms. Since the convergence of $n^{-\gamma}\eta$ is uniform on $F_{n,\delta}$, the stated result follows after distinguishing cases for $h + m_0 = 0$ and $h + m_0 \neq 0$. $\square$

As a first example of the use of Lemma 2, we have the following result for the second moment of Metropolis–Hastings updates. We introduce a multiplicative factor n^α which will represent speeding up the Markov chain: within a time interval of length $t \in \mathbb{R}$ we will make $N(t)$ switches according to P^n, where $N(t) \sim \text{Poisson}(n^\alpha t)$. One of the results of our analysis is the correct value of α for which a suitable scaling limit is obtained, which turns out to be related to γ by $\alpha = 2(1-\gamma)$.

LEMMA 3 (Metropolis–Hastings second moment for Curie–Weiss). *Let* $0 < \delta < \gamma < 1$. *Let* $\alpha = 2(1-\gamma)$. *Suppose* $h \neq 0$, $\beta > 0$ *or* $h = 0$, $0 \leq \beta \leq 1$. *Define*

$$\sigma(h,\beta) := 2\sqrt{1 - |m_0(h,\beta)|}. \tag{26}$$

Then

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} \left| n^\alpha \mathbb{E}_\eta^n\big[(Y-\eta)^2\big] - \sigma(h,\beta)^2 \right| = 0. \tag{27}$$

PROOF. We have

$$n^\alpha \mathbb{E}_\eta^n\big[(Y-\eta)^2\big] = n^\alpha \big(p_+^n(\eta) + p_-^n(\eta)\big)\big(2n^{\gamma-1}\big)^2 = 4\big(p_+^n(\eta) + p_-^n(\eta)\big).$$

We may apply Lemma 2 with $r = 0$ and $k = 0$, to deduce that

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} \left| n^\alpha \mathbb{E}\big[(Y-\eta)^2\big] - 4\big(p_+^{n,0}(\eta) + p_-^{n,0}(\eta)\big) \right| = 0,$$

where

$$\begin{aligned} &p_+^{n,0}(\eta) + p_-^{n,0}(\eta) \\ &\quad = \frac{1}{2}\big(1 - m_0 - n^{-\gamma}\eta\big)\big(1 \wedge \exp\big(2\beta(m_0+h)\big)\big) \\ &\qquad + \frac{1}{2}\big(1 + m_0 + n^{-\gamma}\eta\big)\big(1 \wedge \exp\big(-2\beta(m_0+h)\big)\big). \end{aligned}$$

On $F^{n,\delta}$, we have $|n^{-\gamma}\eta| \leq n^{\delta-\eta} \to 0$, so the remaining dependence on n in the above expression vanishes asymptotically, and we conclude that (27) holds for

$$\begin{aligned} \sigma(h,\beta) = \big\{&2(1-m_0)\big(1 \wedge \exp\big(2\beta(m_0+h)\big)\big) \\ &+ 2(1+m_0)\big(1 \wedge \exp\big(-2\beta(m_0+h)\big)\big)\big\}^{1/2}. \end{aligned}$$

Distinguishing cases and using (9), this is equal to the stated expression for σ. $\square$

Another useful observation is that higher-order moments of $Y - \eta$ vanish.

LEMMA 4. *Let* $0 < \delta < \gamma < 1$ *and let* $\alpha = 2(1-\gamma)$. *Suppose* $h \neq 0$, $\beta > 0$ *or* $h = 0$, $0 \leq \beta \leq 1$. *Then*

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} n^{\alpha} \mathbb{E}_{\eta}^{n}\big[|Y-\eta|^{p}\big] = 0$$

for any $p > 2$.

PROOF. We have

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} n^{\alpha} \mathbb{E}_{\eta}^{n}\big[|Y-\eta|^{p}\big] = \lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} 2^{p}\big(p_{+}^{n}(\eta) + p_{-}^{n}(\eta)\big) n^{(2-p)(1-\gamma)} = 0,$$

using that the sum of the probabilities is bounded by 1. □

6.1.1. *Supercritical temperature regime.* We already mentioned that in the supercritical temperature case ($0 \leq \beta < 1$), the correct scaling of the magnetization would be $\gamma = \frac{1}{2}$.

LEMMA 5. *Suppose* $\gamma = \frac{1}{2}$ *and* $\delta \in (0, \frac{1}{2})$. *If* $0 \leq \beta < 1$, *then for any* $\alpha > 0$,

$$\lim_{n\to\infty} n^{\alpha} \pi^{n}\big(\eta^{n}(x) \notin F^{n,\delta}\big) = 0.$$

PROOF. Note

$$\big\{x : \eta^{n}(x) \notin F^{n,\delta}\big\} = \big\{x : |m^{n}(x) - m_{0}| > n^{\delta-\gamma}\big\}.$$

By the mean value theorem, $|m - \tanh(\beta(m+h))| \geq (1-\beta)|m - m_0|$ for $m \in \mathbb{R}$. Therefore, using Proposition 2, with $t_n := ((1-\beta)n^{\delta-\gamma} - \beta n^{-1})n^{1/2}$, we find that

$$\begin{aligned}
\pi^{n}\big(|m^{n}(x) - m_{0}| \geq n^{\delta-\gamma}\big) &\leq \pi^{n}\big(|m - \tanh\beta(m+h)| \geq (1-\beta)n^{\delta-\gamma}\big) \\
&= \pi^{n}\left(|m - \tanh\beta(m+h)| \geq \frac{\beta}{n} + \frac{t_n}{\sqrt{n}}\right) \\
&\leq 2\exp\left(-\frac{t_n^2}{4(1+\beta)}\right)
\end{aligned}$$

from which the result follows, using that $t_n \sim n^{1/2+\delta-\gamma} \to \infty$. □

LEMMA 6. *Suppose* $0 \leq \beta < 1$, $\gamma = \frac{1}{2}$, $\alpha = 2(1-\gamma) = 1$ *and* $\delta \in (0, 1/4)$. *Let*

$$l(h,\beta) := \frac{1}{1+|m_0|} - \beta\big(1 - |m_0|\big). \tag{28}$$

Then $\lim_{n\to\infty} \sup_{\eta \in F^{n}} |n^{\alpha}\mathbb{E}_{\eta}^{n}[Y-\eta] + 2l(h,\beta)\eta| = 0$.

In other words, the "drift" function of the Metropolis–Hastings transitions is given by $-2l(h,\beta)\eta$.

PROOF OF LEMMA 6. We have, using $\alpha = 1$,

$$n^{\alpha}\mathbb{E}_{\eta}^{n}[Y-\eta] = n\big(p_{+}^{n}(\eta) - p_{-}^{n}(\eta)\big)\big(2n^{-1/2}\big).$$

Therefore, applying Lemma 2 with $r = \frac{1}{2}$ and $k = 1$, we may approximate $n^{\alpha}\mathbb{E}_{\eta}^{n}[Y-\eta]$ to sufficient precision by

$$2n^{1/2}\big(p_{+}^{n,1}(\eta) - p_{-}^{n,1}(\eta)\big).$$

Now distinguish the following cases:

- Suppose $m_0 = 0$ and (therefore) $h + m_0 = 0$. We will show the result for $\eta \geq 0$, the case $\eta < 0$ is analogous. If $\eta \geq 0$, then

$$\begin{aligned}
&2n^{1/2}\big(p_{+}^{n,1}(\eta) - p_{-}^{n,1}(\eta)\big)\\
&\quad = n^{1/2}\big(1 - n^{-1/2}\eta\big) - n^{1/2}\big(1 + n^{-1/2}\eta\big)\big(1 - 2\beta n^{-1/2}\eta\big)\\
&\quad = -2(1-\beta)\eta + 2\beta n^{-1/2}\eta^2.
\end{aligned}$$

  Now using that $n^{-1/2}\eta^2 \leq n^{2\delta - 1/2} \to 0$ on $F^{n,\delta}$, the result follows.
- Suppose $h + m_0 > 0$ (the case $h + m_0 < 0$ is analogous). Then, using (9),

$$\begin{aligned}
&2n^{1/2}\big(p_{+}^{n,1}(\eta) - p_{-}^{n,1}(\eta)\big)\\
&\quad = n^{1/2}\big(1 - (m_0 + n^{-1/2}\eta)\big)\\
&\qquad - n^{1/2}\big(1 + m_0 + n^{-1/2}\eta\big)\exp\big(-2\beta(m_0 + h)\big)\big(1 - 2\beta n^{-1/2}\eta\big)\\
&\quad = n^{1/2}\big(1 - (m_0 + n^{-1/2}\eta)\big)\\
&\qquad - n^{1/2}\big(1 + m_0 + n^{-1/2}\eta\big)\left(\frac{1-m_0}{1+m_0}\right)\big(1 - 2\beta n^{-1/2}\eta\big)\\
&\quad = -\eta - \eta\left(\frac{1-m_0}{1+m_0}\right) + 2\beta(1-m_0)\eta + 2\beta\left(\frac{1-m_0}{1+m_0}\right)n^{-1/2}\eta^2\\
&\quad = -\frac{2\eta}{1+m_0} + 2\beta(1-m_0)\eta + 2\beta\left(\frac{1-m_0}{1+m_0}\right)n^{-1/2}\eta^2,
\end{aligned}$$

  where again the $O(n^{-1/2}\eta^2)$-term vanishes. $\square$

PROOF OF THEOREM 1. The generator of Y^n is given by

$$G^{n,\alpha}\varphi(\eta) := n^{\alpha}\big(P^n\varphi(\eta) - \varphi(\eta)\big), \qquad \varphi : X^n \to \mathbb{R}. \tag{29}$$

Let G denote the unbounded operator $G : D(G) \subset C_0(\mathbb{R}) \to C_0(\mathbb{R})$, where

$$D(G) = \big\{\varphi \in C_0^2(\mathbb{R}) : \eta \mapsto \eta\varphi'(\eta) \in C_0(\eta)\big\}$$

and

$$G\varphi(\eta) = -2l(h,\beta)\eta\frac{d\varphi}{d\eta} + \frac{1}{2}\sigma^2(h,\beta)\frac{d^2\varphi}{d\eta^2}, \qquad \varphi \in D(G).$$

The space of infinitely differentiable functions with compact support $C_c^\infty(\mathbb{R})$ is strongly separating [in the sense of Ethier and Kurtz (2005), Section 3.4]. Let $\varphi \in C_c^\infty(\mathbb{R})$. For $\eta, \zeta \in \mathbb{R}$, we have

$$\left|\varphi(\zeta) - \left\{\varphi(\eta) + \varphi'(\eta)(\zeta-\eta) + \frac{1}{2}\varphi''(\eta)(\zeta-\eta)^2\right\}\right| \le (1/6)\|\varphi^{(3)}\|_\infty|\zeta-\eta|^3.$$

Since $\varphi^{(3)}$ is bounded, we may approximate, for $\alpha = 1$ and $\delta = 1/8$, using Lemmas 3, 4 and 6,

$$\begin{aligned}
&\sup_{\eta\in F^{n,\delta}}\left|G^{n,\alpha}\varphi(\eta) - G\varphi(\eta)\right| \\
&\quad\le \sup_{\eta\in F^{n,\delta}}\left|n^\alpha\mathbb{E}_\eta^n[\varphi_n(Y)-\varphi_n(\eta)] - n^\alpha\mathbb{E}_\eta^n\left[\varphi'(\eta)(Y-\eta) + \frac{1}{2}\varphi''(\eta)(Y-\eta)^2\right]\right| \\
&\quad\le (1/6)n^\alpha\|\varphi^{(3)}\|_\infty\mathbb{E}_\eta^n[|Y-\eta|^3] \to 0.
\end{aligned}$$

Let $\mathbb{P}^n$ denote the distribution of the stationary Markov process Y^n with invariant distribution μ^n. Then, for $T > 0$, by Lemma 5,

$$\mathbb{P}^n\big(Y^n(t) \notin F^{n,\delta} \text{ for some } 0 \le t \le T\big) \le n^\alpha\pi^n\big(\eta^n(x) \notin F^{n,\delta}\big) \to 0.$$

We may now apply Ethier and Kurtz (2005), Corollary 4.8.7, to arrive at the stated result. □

6.1.2. *At critical temperature*. In this section, we assume the "critical" case $h = 0$ and $\beta = 1$. The correct scaling of the magnetization will be $\eta^n = n^{\gamma-1}m^n$ with $\gamma = 1/4$.

LEMMA 7. *Suppose $h = 0$ and $\beta = 1$. Let $\gamma = 1/4$ and $\delta > 0$. Then, for any $\alpha > 0$,*

$$\lim_{n\to\infty} n^\alpha\pi^n\big(\eta^n(x) \notin F^{n,\delta}\big) = 0.$$

PROOF. This follows since, by Proposition 2(ii),

$$n^\alpha\pi^n\big(\eta^n \notin F^n\big) = n^\alpha\pi^n\big(|m^n| > n^{\delta-\gamma}\big) \le 2n^\alpha\exp\big(-cn^{1+4(\delta-\gamma)}\big) \to 0. \qquad \square$$

It turns out that in this case, the correct speed-up factor is n^α with $\alpha = 3/2$. In order to obtain the generator in the critical regime, we will require higher-order Taylor expansions, resulting in a nonlinear drift in the diffusion limit, and accordingly, a non-Gaussian invariant distribution.

LEMMA 8. *Suppose $\beta = 1$ and $h = 0$. Let $\delta \in (0, 1/16)$. Then*

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} \left| n^{3/2}\mathbb{E}\left[Y^n - \eta\right] + 2/3\eta^3 \right| = 0.$$

PROOF. We have $n^{3/2}\mathbb{E}^n_\eta[Y - \eta] = n^{3/2}(p^n_+(\eta) - p^n_-(\eta))(2n^{-3/4}) = 2(p^n_+(\eta) - p^n_-(\eta))n^{3/4}$. Applying Lemma 2 with $r = 3/4$ and $\gamma = 1/4$, we find that we may approximate $p^n_\pm$ by the third-order approximation $p^{n,3}_\pm$. Assuming $\eta > 0$ (the other case is analogous),

$$\begin{aligned}
&2\left(p^{n,3}_+(\eta) - p^{n,3}_-(\eta)\right)n^{3/4} \\
&\quad = \left(\left(1 - n^{-1/4}\eta\right) - \left(1 + n^{-1/4}\eta\right)\right. \\
&\qquad \times \left.\left(1 - 2n^{-1/4}\eta + 2n^{-1/2}\eta^2 - (4/3)n^{-3/4}\eta^3\right)\right)n^{3/4} \\
&\quad = \left(-(2/3)n^{-3/4}\eta^3 + (4/3)n^{-1}\eta^4\right)n^{3/4}.
\end{aligned}$$

On $F^{n,\delta}$ with $\delta < 1/16$, we have $\eta^4 \le n^{4\delta} < n^{1/4}$. It follows that the fourth-order term in η vanishes asymptotically, and the stated result follows. □

PROOF OF THEOREM 2. The proof is completely analogous to that of Theorem 1, taking $\alpha = 3/2$, $\gamma = 1/4$, $\delta = 1/32$, and applying Lemma 8 instead of Lemma 6. □

6.2. *Estimates for lifted Metropolis–Hastings applied to Curie–Weiss.* Let $(Y, J) \in \mathbb{R} \times \{-1, 1\}$ denote the random variable indicating the new state after a single jump. Under $\mathbb{P}^n_{\eta,j}$, let (Y, J) have distribution $T^n((\eta, j), \cdot)$, so that

$$\mathbb{P}^n_{\eta,j}\left[\varphi(Y, J)\right] = \sum_{y,k} T^n\left((\eta, j), (y, k)\right)\varphi(y, k).$$

We will see that the correct speed-up factor for the LMH chain is $\alpha = (1 - \gamma)$ [as opposed to $\alpha = 2(1 - \gamma)$ for Metropolis–Hastings]. At this scaling, the second moment of the increments vanishes for the LMH Markov chain.

LEMMA 9 (LMH second moment for Curie–Weiss). *Let $0 < \gamma < 1$. Let $\alpha = 1 - \gamma$. Suppose $h \neq 0$, $\beta \ge 0$ or $h = 0$, $0 \le \beta \le 1$. Then for any $p > 1$ and $j \in \{-1, +1\}$,*

$$\lim_{n\to\infty} \sup_{\eta \in X^n} n^\alpha \mathbb{E}^n_{\eta,j}\left[|Y - \eta|^p\right] = 0.$$

PROOF. We compute

$$n^\alpha \mathbb{E}^n_{y,j=\pm}\left[|Y - \eta|^p\right] = n^\alpha p^n_\pm(\eta)\left(2n^{\gamma-1}\right)^p = 2^p p^n_\pm(\eta) n^{(p-1)(\gamma-1)}.$$

Since $|p^n_\pm(\eta)| \le 1$, the supremum over $\eta \in X^n$ converges to zero. $\square$

Asymptotically, the first moment of the increments does not depend on η. Let $a(h, \beta)$ be given by (16).

LEMMA 10 (LMH drift for Curie–Weiss). *Let $0 < \delta < \gamma < 1$. Let $\alpha = 1 - \gamma$. Suppose $h \neq 0$, $\beta \ge 0$ or $h = 0$, $0 \le \beta \le 1$. Then for $j \in \{-1, +1\}$,*

$$\lim_{n\to\infty} \sup_{\eta \in X^n} \left| n^\alpha \mathbb{E}^n_{\eta,j}[Y - \eta] - a(h, \beta) j \right| = 0.$$

PROOF. We compute $n^\alpha \mathbb{E}^n_{\eta,\pm 1}[Y - \eta] = \pm p^n_\pm(\eta) 2 n^{\gamma-1} n^\alpha = \pm 2 p^n_\pm(\eta)$. We may apply Lemma 2 with $r = 0$ and $k = 0$, to replace $p_\pm(\eta)$ by $p^{n,0}_\pm(\eta)$, given by (25). Since $n^{-\gamma}|\eta| \le n^{\delta-\gamma} \to 0$ in the supremum over $F^{n,\delta}$, we find that as $n \to \infty$, using (9),

$$\begin{aligned} \pm 2 p^{n,0}_\pm(\eta) &\to \pm(1 \mp m_0) \min\bigl(1, \exp\bigl(\pm 2\beta(m_0 + h)\bigr)\bigr) \\ &= \pm(1 \mp m_0) \min\left(1, \left(\frac{1 + m_0}{1 - m_0}\right)^{\pm 1}\right). \end{aligned}$$

By distinguishing cases, this can be seen to equal (16). $\square$

It only remains to determine the switching rates between the replicas. This will depend on whether $0 \le \beta < 1$ or $\beta = 1$.

6.2.1. *Supercritical temperature regime.* As we have seen, for $0 \le \beta < 1$ the correct scaling is given by $\gamma = \frac{1}{2}$. In this case, we have the following asymptotic result for the switching rate between replicas.

LEMMA 11. *Let $0 \le \beta < 1$ and $h \in \mathbb{R}$. Suppose $\gamma = \frac{1}{2}$, $\alpha = 1 - \gamma = \frac{1}{2}$ and $\delta \in (0, 1/4)$. Then for $j = \pm 1$,*

$$\lim_{n\to\infty} \sup_{\eta \in F^{n,\delta}} \left| n^\alpha \mathbb{P}^n_{\eta,j}(J = -j) - \max\bigl(0, j l(h, \beta)\eta\bigr) \right| = 0,$$

with $l(h, \beta)$ given by (28).

PROOF. We have

$$n^\alpha \mathbb{P}^n_{\eta, j=\pm 1}(J = -j) = n^{1/2} \max\bigl(0, p^n_\mp(\eta) - p^n_\pm(\eta)\bigr).$$

Applying Lemma 2 with $r = \frac{1}{2}$ and $k = 1$, we find that $p^n_\pm(\eta)$ may be approximated to sufficient accuracy by $p^{n,1}_\pm(\eta)$, given by (25). We distinguish cases.

- Suppose $h = 0$ (and hence $m_0 = 0$) and $j = +1$. Then

$$\begin{aligned} p_-^{n,1}(\eta) - p_+^{n,1}(\eta) &= \frac{1}{2}\{(1 + n^{-1/2}\eta)\min(1, 1 - 2\beta n^{-1/2}\eta) \\ &\quad - (1 - n^{-1/2}\eta)\min(1, 1 + 2\beta n^{-1/2}\eta)\} \\ &= (1-\beta)n^{-1/2}\eta - \operatorname{sign}(\eta)\beta n^{-1}\eta^2. \end{aligned}$$

  Using Lipschitz continuity of $x \mapsto \max(0, x)$ and $\delta < 1/4$,

$$\begin{aligned} &\left|n^{1/2}\max(0, p_-^{n,1}(\eta) - p_+^{n,1}(\eta)) - \max(0, (1-\beta)\eta)\right| \\ &\quad \le \beta n^{-1/2}\eta^2 \le \beta n^{-1/2+2\delta} \to 0 \end{aligned}$$

  in the supremum over η, as $n \to \infty$. The case $j = -1$ is analogous.
- Suppose $h \neq 0$. Let us say without loss of generality $h > 0$, and hence $m_0 > 0$. Taking $j = +1$, we compute using (9),

$$\begin{aligned} &p_-^{n,1}(\eta) - p_+^{n,1}(\eta) \\ &\quad = \frac{1}{2}(1 + m_0 + n^{-1/2}\eta)\exp(-2\beta(m_0 + h))(1 - 2\beta n^{-1/2}\eta) \\ &\qquad - \frac{1}{2}(1 - m_0 - n^{-1/2}\eta) \\ &\quad = \frac{1}{2}(1 + m_0 + n^{-1/2}\eta)\left(\frac{1-m_0}{1+m_0}\right)(1 - 2\beta n^{-1/2}\eta) - \frac{1}{2}(1 - m_0 - n^{-1/2}\eta) \\ &\quad = \frac{1}{2}\left\{n^{-1/2}\eta\left(\frac{1-m_0}{1+m_0}\right) - 2\beta n^{-1/2}\eta(1-m_0) + n^{-1/2}\eta\right\} + O(\eta^2 n^{-1}) \\ &\quad = \left(\frac{1}{1+m_0} - \beta(1-m_0)\right)\eta. \end{aligned}$$

  The other cases follow by analogous computations, or by exploiting the symmetry transformations $(\eta, j) \leftrightarrow (-\eta, -j)$ and $(h, m_0) \leftrightarrow (-h, -m_0)$. $\square$

PROOF OF THEOREM 3. The generator of (Y^n, J^n) is given by

$$L^{n,\alpha}\varphi(\eta, j) = n^\alpha\big(T^n\varphi(\eta, j) - \varphi(\eta, j)\big), \qquad \varphi : X^n \times \{-1, +1\} \to \mathbb{R}. \tag{30}$$

It is established in Proposition 3 that L given by (30) generates a Markov process in $\mathbb{R} \times \{-1, 1\}$. By Proposition 4, the Markov process corresponds to a Feller semigroup $(P(t))$ on $C_0(E)$. Note

$$\begin{aligned} L^{n,\alpha}\varphi(\eta, j) &= n^\alpha \mathbb{E}^n_{\eta,j}[\varphi(Y, J) - \varphi(y, j)] \\ &= n^\alpha \mathbb{E}^n_{\eta,j}[(\varphi(Y, j) - \varphi(y, j))\mathbb{1}_{\{J=j\}}] \\ &\quad + (\varphi(y, -j) - \varphi(y, j))\mathbb{P}^n_{\eta,j}(J = -j). \end{aligned}$$

Consider the set of functions $M = \{\varphi : \mathbb{R} \times \{-1, 1\} \to \mathbb{R}, \varphi(\cdot, j) \in C_c^\infty(\mathbb{R})$ for $j = \pm 1\}$. Then M is strongly separating. Using an analogous Taylor approximation argument as in the proof of Theorem 1, for $j = \pm 1$,

$$\begin{aligned}
&\sup_{\eta \in F^{n,\delta}} \left| L^{n,\alpha}\varphi(\eta, j) - L\varphi(\eta, j) \right| \\
&\quad \le \sup_{\eta \in F^{n,\delta}} \Bigg| L^{n,\alpha}\varphi(\eta, j) - \frac{\partial}{\partial \eta}\varphi(\eta, j)\mathbb{E}^n_{\eta,j}[Y - \eta] \\
&\qquad - n^\alpha \mathbb{P}^n_{\eta,j}(J = -j)\big(\varphi(\eta, -j) - \varphi(\eta, j)\big)\Bigg| \\
&\qquad + \sup_{\eta \in F^{n,\delta}} \Bigg| \frac{\partial}{\partial \eta}\varphi(\eta, j)\mathbb{E}^n_{\eta,j}[Y - \eta] \\
&\qquad + n^\alpha \mathbb{P}^n_{\eta,j}(J = -j)\big(\varphi(\eta, -j) - \varphi(\eta, j)\big) - L\varphi(\eta, j)\Bigg| \\
&\quad \le \frac{1}{2} n^\alpha \sup_{\eta \in F^{n,\delta}} \left\| \frac{\partial^2 \varphi}{\partial \eta^2} \right\|_\infty \mathbb{E}^n_{\eta,j}\big[(Y - \eta)^2\big] \\
&\qquad + 2 \sup_{\eta \in F^{n,\delta}} \left| n^\alpha \mathbb{P}^n_{\eta,j}(J = -j) - \max\big(0, jl(h, \beta)\eta\big) \right| \|\varphi\|_\infty \\
&\qquad + \sup_{\eta \in F^{n,\delta}} \left\| \frac{\partial \varphi}{\partial \eta} \right\|_\infty \left| n^\alpha \mathbb{E}^n_{\eta,j}[Y - \eta] - a(h, \beta) j \right|,
\end{aligned}$$

which converges to zero by applying Lemmas 9, 10 and 11, taking $\alpha = 1/2$, $\gamma = 1/2$ and $\delta = 1/8$. As in the proof of Theorem 1, using Lemma 5 (Y^n, J^n) are increasingly concentrated on $F^{n,\delta}$ for $\delta = 1/8$. We may now apply Ethier and Kurtz (2005), Corollary 4.8.7, to deduce the stated weak convergence. It is established in Proposition 5 that (Y, J) has the stated stationary distribution. □

6.2.2. *At critical temperature.* As above for $h = 0$ and $\beta = 1$, we consider the scaled magnetization $\eta^n = n^\gamma m^n$ with $\gamma = 1/4$.

LEMMA 12. *Let* $\beta = 1$, $h = 0$, $\gamma = 1/4$, $\alpha = 1 - \gamma = 3/4$, *and* $\delta \in (0, 1/16)$. *Then*

$$\lim_{n \to \infty} \sup_{\eta \in F^{n,\delta}} \left| n^\alpha \mathbb{P}^n_{\eta,j}(J = -j) - \max\big(0, 1/3 j \eta^3\big) \right| = 0.$$

PROOF. As before

$$n^\alpha \mathbb{P}^n_{\eta,j}(J = -j) = n^{3/4} \max\big(0, p^n_\mp(\eta) - p^n_\pm(\eta)\big).$$

Applying Lemma 2 with $r = 3/4$, $\gamma = 1/4$ and $k = 3$, we find that a sufficiently precise approximation is $p^n_\pm(\eta) \approx p^{n,3}_\pm(\eta)$. The computation of $p^n_\mp(\eta) - p^n_\pm(\eta)$ has already been performed in Lemma 8, resulting in the stated expression. □

PROOF OF THEOREM 4. The proof is fully analogous to the proof of Theorem 3, now taking $\delta = 1/32$, $\gamma = 1/4$ and $\alpha = 3/4$, and applying Lemmas 7 and 12 instead of Lemmas 5 and 11. $\square$

6.3. *The limiting zig-zag process.* By rescaling the time variable and $\lambda^{\pm}$ if necessary, we may assume $a = 1$ without loss of generality throughout the proofs below.

6.3.1. *Construction.*

LEMMA 13. *Suppose Assumption 1 holds. Then for every $(y, j) \in E$, $\lim_{t\to\infty} F(t; y, j) = 0$. In particular, for every $(y, j) \in E$, $1 - F(\cdot; y, j)$ is the distribution function of a positive random variable that is almost surely finite.*

PROOF. We fix $(y, j) \in E$. Suppose T is distributed according to $1 - F$. Since F is continuous at 0, $\mathbb{P}(T = 0) = 0$. By Assumption 1, there exist t_0 and $\lambda_{\min}$ such that $\lambda(y + js, j) \geq \lambda_{\min}$ for $s \geq t_0$. Then, for $t \geq t_0$,

$$\begin{aligned} F(t; y, j) &= F(t_0; y, j) \exp\biggl(-\int_{t_0}^{t} \lambda(y + js, j)\, ds\biggr) \\ &\leq F(t_0; y, j) \exp\bigl(-(t - t_0)\lambda_{\min}\bigr), \end{aligned}$$

and the stated result follows. $\square$

LEMMA 14. *Suppose Assumption 1 holds. Then for every $t \geq 0$ and $(y, j) \in E$, $\mathbb{P}_{y,j}(N(t) < \infty) = 1$.*

PROOF. We assume y, j are fixed and suppress the (y, j)-subscript in $\mathbb{P}_{y,j}$, etc. Introduce the notation $Y_k = Y(T_k)$, $J_k = J(T_k)$. Observe that on $\{T_{k-1} \leq t \leq T_k\}$,

$$\begin{aligned} \bigl|Y(t) - y\bigr| &= \bigl|J_{k-1}(t - T_{k-1}) + Y_{k-1} - y\bigr| \leq t - T_{k-1} + \biggl|\sum_{i=1}^{k-1}(Y_i - Y_{i-1})\biggr| \\ &\leq t - T_{k-1} + T_{k-1} = t. \end{aligned}$$

It follows that on $\{0 \leq t \leq T_k\}$, for every $s \leq t$, $Y(s) \in [y - s, y + s] \subset [y - t, y + t]$. By Assumption 1, λ is bounded on $[y - t, y + t]$, say by a constant $\lambda_{\max} > 0$.

We will show by induction that, for $0 \leq s \leq t$ and $k \in \mathbb{N} \cup \{0\}$,

$$\mathbb{P}(T_k \leq s) \leq 1 - \exp(-\lambda_{\max} s) \sum_{j=0}^{k-1} \frac{(\lambda_{\max} s)^j}{j!}.$$

For $k = 0$, this is trivial. We have for any $k \in \mathbb{N}$ and $0 \le s \le t$,

$$\begin{aligned}
&\mathbb{P}(T_k \le s | T_{k-1}) \\
&\quad = \mathbb{E}\big[\mathbb{P}(T_k \le s | Y_{k-1}, J_{k-1}) \mathbb{1}_{\{T_{k-1} \le s\}} | T_{k-1}\big] \\
&\quad = \mathbb{E}_{y,j}\left[\left(1 - \exp\left(-\int_0^{s-T_{k-1}} \lambda\big(Y(r), J(r)\big)\, dr\right)\right) \mathbb{1}_{\{T_{k-1} \le s\}} | T_{k-1}\right] \\
&\quad \le \big(1 - \exp\big(-\lambda_{\max}(s - T_{k-1})\big)\big) \mathbb{1}_{\{T_{k-1} \le s\}}.
\end{aligned}$$

For $s \le t$, it follows that

$$\begin{aligned}
\mathbb{P}_{y,j}(T_k \le s) &= \mathbb{E}_{y,j}\big[\mathbb{P}_{y,j}(T_k \le s | T_{k-1})\big] \\
&\le \mathbb{E}\big[\big(1 - \exp\big(-\lambda_{\max}(s - T_{k-1})\big)\big) \mathbb{1}_{\{T_{k-1} \le s\}}\big] \\
&= \mathbb{P}(T_{k-1} \le s) - \exp(-\lambda_{\max} s) \mathbb{E}\big[\exp(\lambda_{\max} T_{k-1}) \mathbb{1}_{\{T_{k-1} \le s\}}\big].
\end{aligned} \tag{31}$$

Let G denote the distribution function of T_{k-1} and note by the induction hypothesis for $k - 1$,

$$G(r) = \mathbb{P}(T_{k-1} \le r) \le 1 - \exp(-\lambda_{\max} r) \sum_{i=0}^{k-2} \frac{(\lambda_{\max} r)^i}{i!}, \qquad 0 \le r \le t.$$

Then

$$\begin{aligned}
&\mathbb{E}\big[\exp(\lambda_{\max} T_{k-1}) \mathbb{1}_{\{T_{k-1} \le s\}}\big] \\
&\quad = \int_0^s \exp(\lambda_{\max} r)\, dG(r) \\
&\quad = \big[\exp(\lambda_{\max} r) G(r)\big]_0^s - \lambda_{\max} \int_0^s \exp(\lambda_{\max} r) G(r)\, dr \\
&\quad \ge \exp(\lambda_{\max} s) \mathbb{P}(T_{k-1} \le s) \\
&\qquad - \lambda_{\max} \int_0^s \exp(\lambda_{\max} r) \left(1 - \exp(-\lambda_{\max} r) \sum_{i=0}^{k-2} \frac{(\lambda_{\max} r)^i}{i!}\right) dr \\
&\quad = \exp(\lambda_{\max} s) \mathbb{P}(T_{k-1} \le s) + 1 - \exp(\lambda_{\max} s) + \sum_{i=0}^{k-2} \frac{(\lambda_{\max} s)^{i+1}}{(i+1)!}.
\end{aligned}$$

Inserting this expression into (31), the induction hypothesis follows for k. It now follows by the Fatou lemma that

$$P\big(N(t) = \infty\big) = \mathbb{P}\Big(\liminf_{k \to \infty} \{T_k \le t\}\Big) \le \liminf_{k \to \infty} \mathbb{P}(T_k \le t) = 0. \qquad \square$$

6.3.2. *Regularity.* The total variation distance between measures on a Polish space is defined as usual by

$$\|\nu - \mu\|_{\mathrm{TV}} := \sup_A |\nu(A) - \mu(A)|,$$

where the supremum is over all Borel sets.

PROOF OF PROPOSITION 4. Let $\varphi \in C_0(E)$. The value of $P(t)\varphi(y, j)$ only depends on values of φ within the bounded set $([y - t, y + t], \pm 1) \subset E$. Since φ vanishes at infinity $P(t)\varphi$ vanishes at infinity as well. It remains to establish continuity of $P(t)\varphi$.

We construct a coupling of (Y, J) starting from two different initial conditions, (y, j) and (z, j), as follows. Let $(y, j), (z, j) \in E$ and suppose $z \in \mathbb{R}$. Without loss of generality, assume $j = +1$ and $z \geq y$. Let ν_1 denote the distribution of $T_1 + (z - y)/2$, with initial condition (z, j), that is, ν_1 has distribution function

$$\begin{aligned} H_1(t; y, z) &= \mathbb{P}_{z,j}\big(T_1 + (z - y)/2 \leq t\big) = \mathbb{P}_{z,j}\big(T_1 \leq t - (z - y)/2\big) \\ &= 1 - F\big(t - (z - y)/2; z, j\big), \end{aligned}$$

and let ν_2 denote the distribution of T_1 with initial condition (y, j), that is, ν_2 has distribution function

$$H_2(t; y, z) = \mathbb{P}_{y,j}(T_1 \leq t) = 1 - F(t; y, j).$$

Let $c_1(y, z) := \|\nu_1 - \nu_2\|_{\mathrm{TV}}$. There exists a "maximal" coupling (R_1, R_2) under a probability measure $\mathbb{P}$ of ν_1 and ν_2 such that $\mathbb{P}(R_1 \neq R_2) = c_1(y, z)$; see, for example, Lindvall (2002), Theorem I.5.2. Use (Y^y, J^y) to denote the process starting from initial condition (y, j) and (Y^z, J^z) for the process starting from initial condition (z, j). We introduce a dependence between the two processes through the distribution of the first replica switch time, T_1. Using the same superscript notation here, we let $T_1^z = R_1 - (z - y)/2$ and $T_1^y := R_2$. Let all other switch times T_i^y and T_i^z be defined as usual, that is, $T_{i+1}^y = T_i^y + Z_{i+1}^y$ where $\mathbb{P}(Z_{i+1} \geq \zeta) = F(\zeta; Y_{T_i^y}^y, J_{T_i^y}^y)$, etc. and construct the continuous time processes (Y^y, J^y) and (Y^z, J^z) as in Section 5.1. Define an event

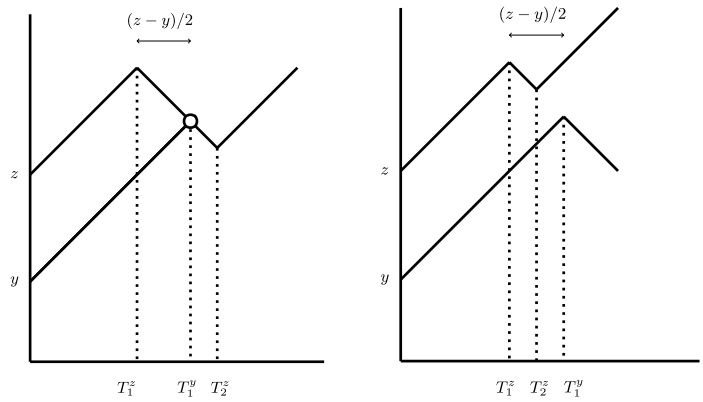
$$\Omega_{\mathrm{coupling}} := \big\{R_1 = R_2 \text{ and } T_2^z > T_1^y\big\},$$

that is, on $\Omega_{\mathrm{coupling}}$ a coupling occurs between R_1 and R_2, and (Y^z, J^z) does not switch a second time before T_1^y; see also Figure 1. On $\Omega_{\mathrm{coupling}}$,

$$T_1^y = R_2 = R_1 = T_1^z + (z - y)/2$$

and hence

$$Y^z\big(T_1^y\big) = z + T_1^z - \big(T_1^y - T_1^z\big) = z + T_1^z - (z - y)/2 = (y + z)/2 + T_1^z,$$



(a) Coupling occurs: $T_1^y = T_1^z + (z-y)/2$ and $T_2^z > T_1^y$. (b) No coupling occurs: $T_1^y = T_z^1 + (z-y)/2$, but $T_2^z < T_1^y$.

FIG. 1. *Illustration of the coupling used in the proof of Proposition* 4.

and

$$Y^y(T_1^y) = y + T_1^y = y + T_1^z + (z-y)/2 = (y+z)/2 + T_1^z,$$

that is, $Y^z(T_1^y) = Y^y(T_1^y)$. By the Strong Markov property, the process

$$(\widetilde{Y}^y, \widetilde{J}^y)(t,\omega) := \begin{cases} (Y^z(t,\omega), J^z(t,\omega)), & \omega \in \Omega_{\text{coupling}}, t \geq T_1^y, \\ (Y^y(t,\omega), J^y(t,\omega)), & \text{otherwise}, \end{cases}$$

is a Markov process with generator L. Since H_1 and H_2 have densities, we may evaluate

$$\begin{aligned} c_1(y,z) &= \|\nu_1 - \nu_2\|_{\text{TV}} \\ &= \frac{1}{2}\int_0^\infty |H_1'(t) - H_2'(t)|\,dt = \frac{1}{2}\int_0^{\frac{z-y}{2}} |H_2'(t)|\,dt \\ &\quad + \frac{1}{2}\int_{\frac{z-y}{2}}^\infty |H_1'(t) - H_2'(t)|\,dt \\ &= \frac{1}{2}\int_0^{\frac{z-y}{2}} \lambda(y+t)F(t;y,j)\,dt \\ &\quad + \frac{1}{2}\int_{\frac{z-y}{2}} \big|\lambda\big(z+t-(z-y)/2\big)F\big(t-(z-y)/2;z,j\big) \\ &\quad - \lambda(y+t)F(t;y,j)\big|\,dt. \end{aligned}$$

The second integrand is trivially dominated by

$$\lambda\big(z+t-(z-y)/2\big)F\big(t-(z-y)/2;z,j\big) + \lambda(y+t)F(t;y,j),$$

which is integrable (since it is the sum of two density functions). Since λ and F depend continuously on y, z, we may apply the dominated convergence theorem to conclude that $c_1(y, z)$ is continuous in y, z. Also note that $c_1(y, y) = c_1(z, z) = 0$. Hence, $\lim_{y\to z} c_1(y, z) = \lim_{z\to y} c_1(y, z) = 0$. Also let

$$c_2(y, z) = \mathbb{P}_{z,j}\big(T_2^z \le T_1^z + (z-y)/2\big) = 1 - \mathbb{E}_{z,j}\big[F\big((z-y)/2; Y(T_1), J(T_1)\big)\big],$$

and note that c_2 is continuous in (y, z) and $\lim_{y\to z} c_2(y, z) = 0$. We estimate

$$\begin{aligned}\mathbb{P}(\Omega \setminus \Omega_{\text{coupling}}) &= \mathbb{P}\big(R_1 \ne R_2 \text{ or } T_2^z \le T_1^y\big)\\ &= \mathbb{P}\big(R_1 \ne R_2 \text{ or } T_2^z \le T_1^z + (z-y)/2\big)\\ &\le \mathbb{P}(R_1 \ne R_2) + \mathbb{P}\big(T_2^z \le T_1^z + (z-y)/2\big)\\ &= c_1(y, z) + c_2(y, z).\end{aligned}$$

Fix $t \ge 0$. Let $\varepsilon > 0$ and let $y \in \mathbb{R}$. Pick $\delta > 0$ such that $c_1(y, z) + c_2(y, z) < \varepsilon/(2\|\varphi\|_\infty)$ for all z for which $|y - z| < \delta$ and (using uniform continuity) $|\varphi(\zeta_1) - \varphi(\zeta_2)| < \varepsilon/2$ for all $\zeta_1, \zeta_2 : |\zeta_1 - \zeta_2| < \delta$ with $\zeta_i \in [y - t, z + t]$. Then, for $|y - z| < \delta$, using that on Ω_{coupling}, the processes $Y^y(t)$ and $Y^z(t)$ remain within distance $|y - z|$ of each other and within $[y - t, z + t]$, we estimate

$$\begin{aligned}&\big|\mathbb{E}_{y,j}\varphi\big(Y(t), J(t)\big) - \mathbb{E}_{z,j}\varphi\big(Y(t), J(t)\big)\big|\\ &\quad\le \mathbb{E}\big|\varphi\big(Y^y(t), J^y(t)\big) - \varphi\big(Y^z(t), J^z(t)\big)\big|\\ &\quad\le \mathbb{P}(\Omega_{\text{coupling}})\varepsilon/2 + \big(c_1(y, z) + c_2(y, z)\big)\|\varphi\|_\infty\\ &\quad< \varepsilon,\end{aligned}$$

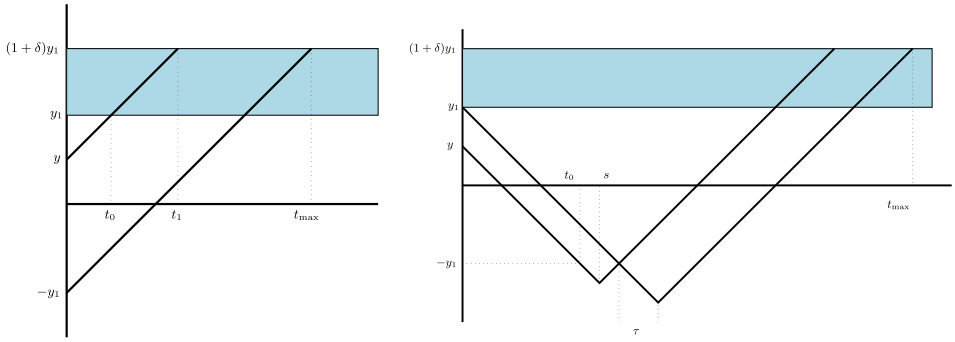
which establishes continuity of $P(t)\varphi(y, j)$ in y for $j = +1$. The case $j = -1$ is analogous. $\square$

6.3.3. *Petite sets.* Let K denote the resolvent Markov kernel given by

$$K\big((y, j), A\big) = \int_0^\infty \exp(-t)\mathbb{P}_{y,j}\big(\big(Y(t), J(t)\big) \in A\big)\,dt, \qquad (y, j) \in E, A \in \mathcal{B}(E).$$

The notion of a petite set plays an important role in establishing exponential ergodicity for a continuous time Markov process; see, for example, Meyn and Tweedie (1993). A set $C \subset E$ is *petite* for K if there exists a nontrivial reference measure ν on E such that $K((y, j), A) \ge \nu(A)$, for any $(y, j) \in C$ and $A \in \mathcal{B}(E)$. The following lemma is instrumental in establishing exponential ergodicity (Theorem 5).

LEMMA 15. *Suppose Assumption 1 holds. Then every compact set $C \subset E$ is petite for K.*



(a) Claim (i) (b) Claim (ii)

FIG. 2. *Illustration of the proof of Lemma* 15.

PROOF. Let $y_0 \geq 0$ and $\lambda_{\min} > 0$ be as defined in Assumption 1(ii). Without loss of generality, it is sufficient to show that any set C of the form $C := [-y_1, y_1] \times \{-1, 1\}$, with $y_1 \geq y_0$, is petite. Indeed, given a compact set $\widetilde{C}$ choose $y_1 \geq y_0$ sufficiently large such that $\widetilde{C} \subset C$. If C is petite, then clearly $\widetilde{C}$ is petite.

Let $C = [-1, 1] \times \{-1, 1\}$ and $\delta > 0$. We will show that for any $\delta > 0$ there exists a constant $c > 0$ such that for every $(y, j) \in C$ and $\varphi \in B_b(E)$, $\varphi \geq 0$,

$$\begin{aligned} &\int_0^\infty \exp(-t)\mathbb{E}_{y,j}\big[\varphi\big(Y(t), J(t)\big)\big]\,dt \\ &\qquad \geq c\int_{y_1}^{(1+\delta)y_1} \varphi(z, +1)\,dz. \end{aligned} \tag{32}$$

This then establishes that C is ν-petite with ν proportional to Lebesgue measure on $[y_1, (1+\delta)y_1] \times \{+1\}$.

Claim (i): *There exists a constant $c > 0$ such that for $\varphi \in B_b(E)$, $\varphi \geq 0$, and $y \in [-y_1, y_1]$,* (32) *holds for $j = +1$; see Figure* 2(a).

Proof of Claim (i): Let $j = +1$. Let $\lambda_{\max} := \max_{y\in[-y_1,(1+\delta)y_1]} \lambda(y, +1)$, which is finite by Assumption 1(i). The time of reaching $(1+\delta)y_1$ from $-y_1$ is $t_{\max} := (2+\delta)y_1$. Let $c := \exp(-(\lambda_{\max}+1)t_{\max})$.

Let $\varphi \in B_b(E)$, $\varphi \geq 0$, $y \in [-y_1, y_1]$ and $0 \leq t \leq t_{\max}$. Then

$$\begin{aligned} &\exp(-t)\mathbb{E}_{y,j}\big[\varphi\big(Y(t), J(t)\big)\big] \\ &\qquad \geq \exp(-t)\mathbb{E}_{y,+1}\big[\varphi\big(Y(t), J(t)\big)\mathbb{1}_{\{T_1\geq t\}}\big] \\ &\qquad = \exp(-t)\varphi(y+t, +1)F(t; y, +1) \\ &\qquad \geq \exp\big(-(\lambda_{\max}+1)t_{\max}\big)\varphi(y+t, +1) \\ &\qquad = c\varphi(y+t, +1). \end{aligned}$$

Define $t_0(y) := \inf\{t \geq 0 : y + t = y_1\} = y_1 - y$ and $t_1(y) := \inf\{t \geq 0 : y + t = (1+\delta)y_1\} = (1+\delta)y_1 - y \leq t_{\max}$. Then

$$\begin{aligned}\int_0^\infty \exp(-t)\mathbb{E}_{y,j}\big[\varphi\big(Y(t), J(t)\big)\big]\,dt &\geq c\int_{t_0}^{t_1} \varphi(y+t,+1)\,dt \\ &= c\int_{y_1}^{(1+\delta)y_1} \varphi(z,+1)\,dz.\end{aligned}$$

Claim (ii): *There exists a constant $c > 0$ such that for $\varphi \in B_b(E)$, $\varphi \geq 0$, and $y \in [-y_1, y_1]$,* (32) *holds for $j = -1$; see Figure* 2(b).

Proof of Claim (ii): Let $j = -1$. We know by Assumption 1(ii), that $\lambda(z, -1)$ is bounded from below for $z \leq -y_1$ by $\lambda_{\min}$. Heuristically, in order to obtain a uniformly positive probability of switching to the $+$-replica, we need to spend at least a certain amount of time, $\tau > 0$ say, in the region $(-\infty, -y_1]$. For definiteness, let $\tau \in (0, 2y_1)$. Hence, from a given $y \in [-y_1, y_1]$ we will travel for a certain amount of time $t_0(y) := y_1 + y$ until we reach $-y_1$, and then continue moving in the negative direction up to time $t_1(y) := t_0(y) + \tau$. We will then have to move back in the positive direction from $-y_1 - \tau$ until reaching $y_1(1+\delta)$. The maximum amount of time required is obtained if we start from $y = +y_1$, which results in a value $t_{\max} := (4+\delta)y_1 + 2\tau$.

Let

$$\lambda_{\max} := \sup_{0 \leq t \leq t_{\max}} \lambda(y-t,-1) \vee \lambda(y+t,+1) \vee \lambda(y-t,+1) \vee \lambda(y+t,+1),$$

a crude but effective upper bound for the switching rate in either replica. Define

$$c := \lambda_{\min} \exp\big(-(\lambda_{\max}+1)t_{\max}\big)\tau.$$

Let $\varphi \in B_b(E)$, $\varphi \geq 0$, $y \in [-y_1, y_1]$ and $0 \leq t \leq t_{\max}$. Then

$$\begin{aligned}&\mathbb{E}_{y,j}\big[\exp(-t)\varphi\big(Y(t), J(t)\big)\mathbb{1}_{\{T_1 \leq t, T_2 > t\}}\big] \\ &\quad= \mathbb{E}_{y,j}\Big[\exp(-t)\int_0^t \varphi\big(y - s + (t-s), +1\big)\mathbb{1}_{\{T_1 \in ds, T_2 > t\}}\Big] \\ &\quad= \exp(-t)\int_0^t \varphi(y - 2s + t, +1)\lambda(y-s,-1) \\ &\qquad\times F(s; y, -1)F(t-s; y-s, +1)\,ds \\ &\quad\geq \int_{t_0(y)}^t \varphi(y-2s+t,+1)\lambda_{\min}\exp\big(-(\lambda_{\max}+1)t\big)\,ds \\ &\quad= \frac{c}{\tau}\int_{t_0(y)}^t \varphi(y-2s+t,+1)\,ds.\end{aligned} \tag{33}$$

We have

$$\begin{aligned}
&\int_0^\infty \mathbb{E}_{y,j}\big[\exp(-t)\varphi\big(Y(t), J(t)\big)\big]\,dt \\
&\quad\ge \int_0^{t_{\max}} \mathbb{E}_{y,j}\big[\exp(-t)\varphi\big(Y(t), J(t)\big)\mathbb{1}_{\{T_1\le t, T_2>t\}}\big]\,dt \\
&\quad\ge c/\tau \int_0^{t_{\max}} \int_{t_0(y)}^{t} \varphi(y-2s+t,+1)\,ds\,dt \qquad [\text{by } (33)] \\
&\quad\ge c/\tau \int_{t_0(y)}^{t_1(y)} \int_{t_1(y)}^{t_{\max}} \varphi(y-2s+t,+1)\,ds\,dt \\
&\qquad\qquad\qquad

$$m^- := \sup_{y \le -y_0} \lambda^+(y), \qquad M^- := \inf_{y \le -y_0} \lambda^-(y).$$

By Assumption 3, $M^+ > m^+$ and $M^- > m^-$. In particular, there exist constants $\beta^+ > 0$, $\beta^- > 0$ such that $m^+ \exp(2\beta^+) < M^+$ and $m^- \exp(2\beta^-) < M^-$. It follows that

$$m^\pm(\exp(2\beta^\pm) - 1) = m^\pm \exp(2\beta^\pm)(1 - \exp(-2\beta^\pm)) < M^\pm(1 - \exp(-2\beta^\pm)).$$

Therefore, we can pick positive constants $\alpha^\pm \in (m^\pm(\exp(2\beta^\pm) - 1), M^\pm(1 - \exp(-2\beta^\pm)))$. Let $y_1 \ge y_0$ be undefined for now. As a Lyapunov function, we take a function V such that, outside of $(-y_1, +y_1)$, and for $j \in \{-1, +1\}$,

$$V(y, j) = \begin{cases} \exp(\alpha^+ y + \beta^+ \operatorname{sign}(j)), & \text{if } y \ge y_1, \\ \exp(-\alpha^- y - \beta^- \operatorname{sign}(j)), & \text{if } y \le -y_1, \end{cases} \tag{34}$$

and such that V is positive and continuously differentiable on $(-y_1, +y_1)$. As long as $y_1 \ge y_0$ is taken sufficiently large, then V thus defined is positive and continuously differentiable on E. Now on $y \ge y_1$, we have

$$LV(y, +1) = (\alpha^+ - \lambda^+(y)(1 - \exp(-2\beta^+)))V(y, +1),$$
$$LV(y, -1) = (-\alpha^+ + \lambda^-(y)(\exp(2\beta^+) - 1))V(y, -1).$$

By the choice of $\alpha^\pm$, we have

$$\alpha^+ - \lambda^+(y)(1 - \exp(-2\beta^+)) \le \alpha^+ - M^+(1 - \exp(-2\beta^+)) < 0 \quad \text{and}$$
$$-\alpha^+ + \lambda^-(y)(\exp(2\beta^+) - 1) \le -\alpha^+ + m^+(\exp(2\beta^+) - 1) < 0.$$

It follows that there exists a constant $c^+ > 0$ such that $LV(y, j) \le -c^+ V(y, j)$ for $y \ge y_1$ and $j \in \{-1, +1\}$. Analogously, there exists a constant $c^- > 0$ such that $LV(y, j) \le -c^- V(y, j)$ for $y \le -y_1$ and $j \in \{-1, +1\}$. The proof is completed by taking $c := c^- \wedge c^+$. $\square$

Acknowledgements. We acknowledge the suggestions for improvement of two anonymous referees and the Associate Editor which have significantly contributed towards the accuracy and clarity of this work.

REFERENCES

BIERKENS, J. (2016). Non-reversible Metropolis–Hastings. *Stat. Comput.* **16** 1213–1228.

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2015). The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method. Preprint. Available at arXiv:1510.02451.

CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

CHATTERJEE, S. and DEY, P. S. (2010). Applications of Stein's method for concentration inequalities. *Ann. Probab.* **38** 2443–2485. MR2683635

CHEN, T.-L. and HWANG, C.-R. (2013). Accelerating reversible Markov chains. *Statist. Probab. Lett.* **83** 1956–1962. MR3079029

CHEN, F., LOVÁSZ, L. and PAK, I. (1999). Lifting Markov chains to speed up mixing. In *Annual ACM Symposium on Theory of Computing* (*Atlanta*, *GA*, 1999) 275–281 (electronic). ACM, New York. MR1798046

DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **46** 353–388. MR0790622

DAVIS, M. H. A. (1993). *Markov Models and Optimization*. *Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978

DING, J., LUBETZKY, E. and PERES, Y. (2009). The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Comm. Math. Phys.* **289** 725–764. MR2506768

DUNCAN, A. B., LELIÈVRE, T. and PAVLIOTIS, G. A. (2016). Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.* **163** 457–491.

ELLIS, R. S. (2006). *Entropy*, *Large Deviations*, *and Statistical Mechanics*. Springer, Berlin. MR2189669

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2012). Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.* **44** 977–994. MR3052846

FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.* **126** 3077–3101. MR3542627

GOLDSTEIN, S. (1951). On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.* **4** 129–156. MR0047963

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437

HWANG, C.-R., HWANG-MA, S.-Y. and SHEU, S. J. (1993). Accelerating Gaussian diffusions. *Ann. Appl. Probab.* **3** 897–913. MR1233633

KAC, M. (1974). A stochastic model related to the telegrapher's equation. *Rocky Mountain J. Math.* **4** 497–509. MR0510166

LELIÈVRE, T., NIER, F. and PAVLIOTIS, G. A. (2013). Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.* **152** 237–274. MR3082649

LEVIN, D. A., LUCZAK, M. J. and PERES, Y. (2010). Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability. *Probab. Theory Related Fields* **146** 223–265. MR2550363

LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover, Mineola, NY. MR1924231

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087.

MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295

MICLO, L. and MONMARCHÉ, P. (2013). Séminaire de Probabilités XLV (C. Donati-Martin, A. Lejay and A. Rouault, eds.). *Lecture Notes in Math.* **2078**. Springer, Cham. MR3202581

MONMARCHÉ, P. (2014). Hypocoercive relaxation to equilibrium for some kinetic models. *Kinet. Relat. Models* **7** 341–360. MR3195078

MONMARCHÉ, P. (2016). Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 357–398. MR3487077

PETERS, E. A. J. F. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* (3) **85** 1–5.

REY-BELLET, L. and SPILIOPOULOS, K. (2015). Irreversible Langevin samplers and variance reduction: A large deviations approach. *Nonlinearity* **28** 2081–2103. MR3366637

ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751

SAKAI, Y. and HUKUSHIMA, K. (2013). Dynamics of one-dimensional ising model without detailed balance condition. *J. Phys. Soc. Jpn.* **82** 064003.

SUN, Y., GOMEZ, F. and SCHMIDHUBER, J. (2010). Improving the asymptotic performance of Markov chain Monte-Carlo by inserting vortices. In *Advances in Neural Information Processing Systems* 23 (J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) 2235–2243.

TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D* **240** 410–414.

DELFT UNIVERSITY OF TECHNOLOGY
DIAM - EECMS
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS
E-MAIL: joris.bierkens@tudelft.nl

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM
E-MAIL: gareth.o.roberts@warwick.ac.uk