FROM METROPOLIS TO DIFFUSIONS: GIBBS STATES AND OPTIMAL SCALING

L.A. BREYER AND G.O. ROBERTS

ABSTRACT. This paper investigates the behaviour of the random walk Metropolis algorithm in high dimensional problems. Here we concentrate on the case where the components in the target density is a spatially homogeneous Gibbs distribution with finite range. The performance of the algorithm is strongly linked to the presence or absence of phase transition for the Gibbs distribution; the convergence time being approximately linear in dimension for problems where phase transition is not present. Related to this, there is an optimal way to scale the variance of the proposal distribution in order to maximise the speed of convergence of the algorithm. This turns out to involve scaling the variance of the proposal as the reciprocal of dimension (at least in the phase transition free case). Moreover the actual optimal scaling can be characterised in terms of the overall acceptance rate of the algorithm, the maximising value being 0.234, the value as predicted by studies on simpler classes of target density. The results are proved in the framework of a weak convergence result, which shows that the algorithm actually behaves like an infinite dimensional diffusion process in high dimensions.

1. INTRODUCTION AND DISCUSSION OF RESULTS

For Markov chain Monte Carlo algorithms, a crucial question of interest is how times needed to ensure convergence scale with the dimensionality of the problem. This question is complicated by the fact that its answer is fundamentally affected by the dependence structure between the one-dimensional components of the target distribution.

For the Gibbs sampler on the Ising model, Frigessi *et al.* (1986) demonstrate that for sub-critical temperatures, convergence times scale exponentially with dimension,

---

Date: March, 1998.

1991 *Mathematics Subject Classification.* Primary 60J, Secondary 60F.

Key words and phrases. Markov chain Monte Carlo, Hamiltonians, Hybrid algorithms.

whilst in the super-critical case, convergence is polynomial in dimension. This result accords with the heuristic thought to hold in much greater generality; that convergence times for algorithms tend to be polynomial or exponential in dimension, according to the presence or absence of phase transition.

In this paper we shall concentrate on the Random Walk Metropolis algorithm. Suppose π_n is an n-dimensional density with respect to Lebesgue measure, and let q denote the increment density of a symmetric Random Walk. The algorithm proceeds iteratively as follows. Given X_t, we propose a new value Y_{t+1} from the density $q(\cdot - X_t)$. Now we set $X_{t+1} = Y_{t+1}$ with probability

$$\alpha(X_t, Y_{t+1}) = \min\{1, \pi(Y_{t+1})/\pi(X_t)\} \ ,$$

Otherwise we set $X_{t+1} = X_t$. Therefore in the algorithms we are considering, the proposals are fully n-dimensional, as opposed to other schemes such as single site updating.

For the Random Walk Metropolis algorithm, a closely related implementational problem to the above *scaling* problem, is the following. For an n-dimensional problem, and given (for instance) an n-dimensional Gaussian proposal with variance σ_n^2, how should σ_n scale as a function of n? Furthermore, is it possible to characterise optimality of σ_n in a way that can be practically utilised.

A partial answer to these two questions is given in Roberts *et al.* (1997) where the problem is considered for the case where the proposal distribution consists of n independent identically distributed components from an arbitrary density f say. In this case, it turns out to be optimal to scale the proposal variance as $O(n^{-1})$, and the optimality criterion is most usefully expressed as scaling the variance so that the overall acceptance probability for the algorithm (that is $\int_{\mathbb{R}^n \times \mathbb{R}^n} \pi_n(x) q(x, y) \alpha(x, y) dx dy$) is approximately 0.234.

With independent components, phase transition is necessarily absent, so this result conforms with the phase transition heuristic mentioned above. Although in special cases (such as Gaussian target densities) it can be seen that the optimal scaling and acceptance rate criterion are rather robust to changes in dependence structure, no general results of this type appear to be available.

In this paper we generalise Roberts *et al.* (1997), giving a weak convergence result showing that for suitably behaved sequences of target densities with partial correlations of finite range, the algorithms behaves like an infinite dimensional Langevin diffusion. The result holds under the scaling of the proposal by $O(n^{-1})$, as in the independent component case.

In the case where no phase transition occurs, it follows that the optimal variance can be characterised as that which induces an overall acceptance rate of 0.234 as in the independent component case.

On the other hand, in the phase transition case, the limiting diffusion is in fact reducible, being unable to move between phases. Movement between phases for the n-dimensional algorithm therefore happens at a rate slower than $O(n^{-1})$. This is consistent with, and provides evidence to support the phase transition heuristic mentioned above.

In the phase transition case, the sequence of probability measures admits more than one limiting Gibbs measure or phase. It can be seen that the limiting diffusion then acts locally in a way which is independent of its phase, apart from its speed measure. An interesting consequence of this is the possibility of empirically *diagnosing* phase transition behaviour in high dimensional problems by monitoring overall acceptance rates of the algorithm.

2. Overview of results

We now describe in greater detail the results of this paper. Consider a collection V_n consisting of n sites arranged on the lattice $\mathbb{Z}^d$. In other words, V_n is a finite subset of $\mathbb{Z}^d$ with $|V_n| = n$. Each site $k \in V_n$ is given a real valued "colour" $x_k \in \mathbb{R}$, and we call the collection $(x_k : k \in V_n) \in \mathbb{R}^{V_n}$ a configuration. Viewed in this way, a configuration is a function $x : V_n \to \mathbb{R}$.

Most probability distributions π_n on $\mathbb{R}^{V_n}$ can be approximated by the Random Walk Metropolis algorithm $X^n = (X^n_t : t \geq 0)$, in the sense that $X^n_t \Rightarrow \pi_n$ as $t \to \infty$. In this paper, we discuss a scaling problem as the number of sites n tends to infinity.

More precisely, suppose at first the existence of an idealised system consisting of all sites in $\mathbb{Z}^d$ and a corresponding distribution π on

$$S = \mathbb{R}^{\mathbb{Z}^d} = \{\text{configurations } x : \mathbb{Z}^d \to \mathbb{R}\}.$$

The measures π_n are viewed as the conditional distribution of π, given the configuration of sites $x_{V_n^c} = (x_k : k \notin V_n)$. The n-th Markov chain algorithm X^n depends on a parameter σ_n^2 representing the variance of the Random Walk step. We shall show how to choose σ_n as a function of n, so as to optimise the speed of convergence of the algorithm in the limit $n \to \infty$.

Such a problem was worked on previously by Roberts *et al.* (1997), who considered the case when $\pi(dx) = \prod_{k\in\mathbb{Z}^d} f(x_k)dx_k$ is a product measure. This corresponds to assuming that the n sites in V_n take their colours $x(k)$ independently of each other.

In our generalisation, we take π as a "perturbed" product measure, i.e. the Gibbs measure heuristically written as

$$\pi(dx) = e^{-\sum_{k\in\mathbb{Z}^d} U_k(x)} \cdot \prod_{k\in\mathbb{Z}^d} \mu(dx_k). \tag{1}$$

Here, μ denotes a probability measure on $\mathbb{R}$, and each of the functions U_k, $k \in \mathbb{Z}^d$ will be assumed to depend only on a finite number of neighbouring sites (x_j) (finite range interactions). Moreover, we will assume also that the set of functions $(U_k : k \in \mathbb{Z}^d)$ is preserved under spatial translations. Both these assumptions are often satisfied in the statistical analysis of certain spatial models.

We now introduce some notation. Given a subset $W \subset \mathbb{Z}^d$, we define the σ-algebra $\mathcal{F}_W = \sigma(x_k : k \in W)$ and $\mathcal{F}_{W^c} = \sigma(x_k : k \notin W)$. It is useful to generalise the notation for the components x_k of a configuration $x \in S$. Given a set $W \subset \mathbb{Z}^d$, we let $x_W = (x_k : k \in W)$. Then we obviously have $x = (x_{W^c}, x_W)$.

Supposing now that $z \in S$ is some fixed "boundary" configuration, and $V_n \uparrow \mathbb{Z}^d$, we say that the family (π_n) of distributions on $\mathbb{R}^{V_n}$ is a *scaling family* if

$$\pi_n(dx) = \mathbb{P}[X_{V_n} \in dx \,|\, X_{V_n^c} = z_{V_n^c}], \quad \text{on } \mathcal{F}_{V_n} = \sigma(\mathbb{R}^{V_n}), \quad X \sim \pi. \tag{2}$$

Thus π_n is a regular conditional distribution of π with respect to $\mathcal{F}_{V_n^c}$, and z is a choice of *boundary condition*, that is a fixed configuration in S.

We have so far assumed the existence of π, but this is an idealisation. Suppose we specify a family Π of probability kernels on S as follows:

$$\Pi = (\pi_W(a, dx) : W \subset \mathbb{Z}^d \text{ finite}, \, a \in S), \tag{3}$$

where the heuristic interpretation is that $\pi_W(a, dx) = \mathbb{P}[X_W \in dx \,|\, X_{W^c} = a_{W^c}]$. For fixed $z \in S$, the scaling family will again be assumed of the form (2), but π no longer appears as part of the definition. While this shift in perspective allows a more realistic model (after all, MCMC is often done on $\mathbb{R}^n$, $n < \infty$), we are now faced with the added difficulty of identifying π, if this exists, so that (2) makes sense in full. Furthermore, (2) may be compatible with several distinct probability distributions π on S. This fundamental problem is addressed by the theory of random fields.

From this point onwards, we shall assume given a fixed family Π as in (3), satisfying the following consistency conditions:

$$\int \pi_W(z, dy)\pi_U(y, dx) = \pi_W(z, dx), \quad U \subset W \subset \mathbb{Z}^d, \tag{4}$$

which using (2) are simply

$$\mathbb{E}[\mathbb{P}[X_U \in dx \,|\, X_{U^c}] \,|\, X_{W^c} = z] = \mathbb{P}[X_U \in dx \,|\, X_{W^c} = z].$$

A probability ξ on S is called a *Gibbs distribution* if

$$\xi(dx \,|\, \mathcal{F}_{W^c}) = \pi_W(\cdot, dx), \quad W \subset \mathbb{Z}^d \text{ (finite)}, \, z \in S. \tag{5}$$

Thus Gibbs distributions are precisely the probability measures for which $X \sim \xi$ gives rise to the family of conditional distributions Π. The set $G(\Pi)$ of Gibbs distributions may consist of more than one measure. In this case, we say that there is a *phase transition.*

We shall be interested only in those Gibbs distributions which are translation invariant, i.e. $\xi \circ \oplus_k = \xi$ for all $k \in \mathbb{Z}^d$, where $\oplus_k$ denotes the *shift transformation* $\oplus_k x_j = x_{j+k}$.

The form taken by the specification Π will be important in the sequel, and we now describe the notation we shall use. Let $V = \{0, v^1, \dots, v^m\} \subset \mathbb{Z}^d$ be a finite neighbourhood of site 0. We assume given a collection of functions

$$h_k(x) = h_k(x_k, x_{k+v^1}, \dots, x_{k+v^m}), \quad x \in S, \quad k \in \mathbb{Z}^d,$$

which satisfy the conditions $h_k \circ \oplus_l = h_{k+l}$. The formal sum

$$H(x) = -\sum_{k\in\mathbb{Z}^d} h_k(x), \quad x \in S$$

is called the Hamiltonian. It is not a well defined function on S; however, note that its *partial derivatives* $D_{x_k}H(x)$ are always proper finite functions on S, by virtue of the imposed finite range condition. The Hamiltonian describes the energy of a configuration $x \in S$. If we restrict ourselves to a finite collection of sites $W \subset \mathbb{Z}^d$ only, the corresponding natural quantity is the finite volume Hamiltonian

$$H_W(z_{W^c}, x_W) = -\sum_{k\in W} h_k(z_{W^c}, x_W), \quad x, z \in S, \quad W \subset \mathbb{Z}^d. \tag{6}$$

Note that this is always a well defined function on S.

We can now specify a consistent family Π by setting

$$\pi_W(z, dx) = C_{W,z}^{-1} \exp[-H_W(z_{W^c}, x_W)]dx_W, \tag{7}$$

where $dx_W = \prod_{k\in W} dx_k$ is Lebesgue measure on $\mathbb{R}^W$, and $C_{W,z}$ is a normalising constant. Under Hypothesis (H1) in Section 2, the measures $\pi_n(\cdot) = \pi_{W_n}(z, \cdot)$ converge weakly, in a suitable topology, to some Gibbs distribution $\xi(\cdot)$, for "most" boundary conditions z. The limit will generally depend on z – If $z \sim \xi$, we get that limit.

Consider now, for a fixed set of sites V_n and boundary condition z, a Random Walk Metropolis chain $X_t^{V_n,z}, t = \{0, 1, 2, \dots\}$ for π_n, defined in Section 4. As shown in Roberts and Smith (1994), the law of $X_t^{V_n,z}$ converges as $t \to \infty$ to π_n. We shall investigate a diffusion approximation as $n \to \infty$.

It is shown in Section 3 that the discrete time generator of $X_t^{V_n,z}$ can be written, for any bounded differentiable test function $f : \mathbb{R}^{V_n} \to \mathbb{R}$, $x \in \mathbb{R}^{V_n}$,

$$A^{V_n,z}f(x) = \sigma_n^2\left(\frac{1}{2}\sum_{i\in V_n} a_{V_n,z}^{ii}(x)D_{x_i}^2 f(x) - \sum_{i\in V_n} b_{V_n,z}^i(x)D_{x_i}f(x)\right) + o(n\sigma_n^2)$$

where σ_n^2 is the proposal step variance for the algorithm $X_t^{V_n,z}$.

In Section 4, we show that, if $\sigma_n^2 = \ell^2/n$ with ℓ a constant, then

$$\lim_{n\to\infty} a_{V_n,z}^{ii}(x) = v(x), \quad \xi \text{ } a.e. \text{ } x, \text{ and} \tag{8}$$

$$\lim_{n\to\infty} b_{V_n,z}^i(x) = -\frac{1}{2}D_{x_i}H(x)v(x), \quad \xi \text{ } a.e. \text{ } x \tag{9}$$

Here the function v is given by the formula

$$v(x) = 2\ell^2 \Phi\left(-\frac{\ell}{2}\sqrt{\xi(D^2_{x_0}H \mid \mathcal{I})(x)}\right), \tag{10}$$

where $\mathcal{I}$ denotes the $(\oplus_k : k \in \mathbb{Z}^d)$ invariant σ-algebra and $D_{x_i}H$, $D^2_{x_0}H$ are partial derivatives of the Hamiltonian (these are well defined due to the finite range condition).

It will also be shown that

$$a^{ii}_{V_n,z}(x) = \mathbb{E}\left[1 \wedge \frac{d\pi_n}{dx_{V_n}}(X^{V_n,z}_1) \Big/ \frac{d\pi_n}{dx_{V_n}}(x)\right] \tag{11}$$

is the overall or expected acceptance probability for the next proposed move from x. Thus (8) states that the acceptance probability converges to a nontrivial quantity.

Combining (8) and (9) with the expression for the generator of $X^{V_n,z}_t$, we get for ξ a.e. x,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} A^{V_n,z} f(x) = \frac{1}{2} v(x)\Big(\Delta f(x) - \nabla H(x)\cdot\nabla f(x)\Big) \tag{12}$$

provided that the test function $f : S \to \mathbb{R}$ depends on at most a finite number of coordinates. This is proved as Theorem 8.

Probabilistically, this result is interpreted as follows (see Sections 5 and 6). Suppose that we run the Random Walk Metropolis algorithm $X^{V_n,z}_t$ from stationarity, that is with $X^{V_n,z}_0 \sim \pi_n$. If Z_t is the infinite dimensional Langevin diffusion solving the SDE

$$dZ_t = v(Z_t)^{1/2} dB_t - \frac{1}{2} v(Z_t)\nabla H(Z_t)dt, \quad Z_0 \sim \xi, \tag{13}$$

then we have the weak convergence result $X^{V_n,z}_{[tn/\ell^2]} \Rightarrow Z_t$ (Theorem 13), and the function v defined by (10) appears as a speed measure for Z.

We were only able to prove the probabilistic interpretation above under one further hypothesis, (H6). Unfortunately, this assumption precludes the existence of phase transitions. However, we believe the result to be true independently of the existence or not of phase transitions. The convergence of the generators (Theorem 8) is certainly true under phase transitions.

Note that the existence of Z_t is nontrivial and requires certain assumptions. In particular, the space S is too big as a state space to be useful, and we restrict attention to the set of those configurations $x : \mathbb{Z}^d \to \mathbb{R}$ satisfying a growth condition (see Section 5). Under this condition, the set of admissible configurations becomes a separable Hilbert space. We shall see that every Gibbs distribution is, under appropriate conditions a stationary distribution for Z. A study of the diffusion Z_t gives much insight into the scaling behaviour of the Metropolis algorithm.

Most interestingly, suppose that there are no phase transitions for the family Π in (3). In that case, there is only one possible Gibbs distribution ξ, independently of the chosen boundary condition for π_n, and the invariant σ-algebra $\mathcal{I}$ is trivial. We can thus maximise the speed (and in particular the speed of convergence) of Z_t by choosing

$$\hat{\ell} \approx 2.38 \Big/ \sqrt{\int D^2_{x_0}(x)\xi(dx)}.$$

In that case, the value of v becomes approximately 0.234. Since v is also the limit of the acceptance probabilities in (8), we get the following easy rule:

Optimization Rule: *In the absence of phase transitions, choose the proposal variance $\sigma_n^2 = \hat{\ell}/n$; equivalently, tune σ_n so that the average acceptance rate is approximately* 0.234*, and this will maximise the speed of convergence of the algorithm for large n.*

Suppose now on the contrary that there are phase transitions. Every Gibbs distribution ξ is now a mixture of extreme, ergodic Gibbs distributions λ, in the sense that there exists a probability γ_ξ on $G_\oplus(\Pi)$ such that (see Section 2, Standard Fact (iii))

$$\xi(\cdot) = \int \lambda(\cdot)\gamma_\xi(d\lambda).$$

The measures λ are mutually singular, and this gives the following behaviour for the process Z_t. Every realisation of Z belongs to the support of some unique λ (according to the probability γ_ξ) for all time, with excursions from one measure λ to another $\lambda' \neq \lambda$ being impossible.

Stated differently, the state space of Z_t is no longer irreducible. Accordingly, the Metropolis chain $X_t^{V_n,z}$ must, when n is large, take much longer to move about *its* state space $\mathbb{R}^{V_n}$ (which is still irreducible). Since the measure π_n approximates ξ, it is multimodal, with "valleys" of very low probability. Consequently, the speed of convergence to π_n, when n is large, reduces dramatically, as the process is trapped in each mode for a long time.

In the presence of phase transitions, optimal scaling means that the acceptance rate tends to zero with dimension.

3. Hypotheses and Gibbs distributions

In this section, we list three hypotheses which we shall make on the specification Π and the scaling family given by (2), (3), and (7). We illustrate these by various examples.

We begin with the hypothesis which underlies all subsequent developments. Below, we shall give examples of specifications which satisfy it

Hypothesis (H1): Let V be a finite subset of $\mathbb{Z}^d$ such that $0 \in V$ and $v \in V$ implies also that $-v \in V$ For each $k \in \mathbb{Z}^d$, let $h_k : \mathbb{R}^{V+k} \to \mathbb{R}$ be C^3, and such that $h_k \circ \oplus_l = h_{k+l}$. We assume that the family of probability measures Π defined by (2), (3) is tight in the local topology, and that the set $G_\oplus(\Pi)$ of *translation invariant* Gibbs distributions is nonempty.

Much is known about the applicability of Hypothesis (H1); a standard reference is (Georgii, 1988). The *local topology* referred to above is that generated by all those functions $S \to \mathbb{R}$ which each depend on at most a finite number of coordinates. We proceed to give some examples.

Example 1. Let μ be an absolutely continuous probability measure on $\mathbb{R}$ and set $h_k(x) = (d\mu/dx)(x_k)$. The specification Π reduces to that of the product measure $\prod_{k\in\mathbb{Z}^d} \mu(dx_k)$, which is the only Gibbs distribution. Thus there is no phase transition.

Example 2. Suppose that $h_k(x) = U \circ \oplus_k(x) - \log(d\mu/dx)(x_k)$, where μ is an absolutely continuous probability measure on $\mathbb{R}$ and $U : S_V \to \mathbb{R}$ is bounded. It is a well known result (Georgii, 1988, Theorem 4.23) that Hypothesis (H1) then holds. Phase transitions may occur. If we set $U_k(x) = U \circ \oplus_k(x)$, we recover the heuristic description (1).

Example 3. Let $(q_l : l \in V)$ be a collection of real numbers with $q_0 \neq 0$ and such that the matrix $q_{ij} = q_{|i-j|}$ is positive definite. We obtain a homogeneous Gaussian specification by setting

$$h_k(x) = \sum_{l\neq 0} q_l x_k x_{k+l} + q_0 x_k^2 \tag{14}$$

For $y \in (-1,1]^d$, let $\hat{J}(y) = \sum_{v \in V} q_v \cos(\pi \sum_{i=1}^d v_i y_i)$ be the discrete Fourier transform of (q_l). The following three cases are possible (Georgii, 1988, p.277).

- If $\int \hat{J}(y)^{-1} dy = \infty$, then $G_{\oplus}(\Pi) = \emptyset$.
- If $q_l = 0$ for all $l \neq 0$, then $G_{\oplus}(\Pi)$ contains a unique Gibbs measure (no phase transition).
- If $q_l \neq 0$ for some $l \neq 0$ and $\int \hat{J}(y)^{-1} dy < \infty$, there is phase transition.

Example 4. Let $p(x,y)dy$ be the transition probability function of some time homogeneous, Harris recurrent Markov chain on $\mathbb{R}$, with stationary distribution m. Let $(M_t : -\infty < t < \infty)$ be a Markov chain with this transition function, and such that $M_t \sim m$ for all $t \in \mathbb{Z}$. We shall view the path of M as a configuration on $\mathbb{Z}$, and define a family Π by setting

$$\pi_W(a, dx) = \mathbb{P}\Big(M_{m+1} = x_{m+1}, \ldots, M_{n-1} = x_{n-1} \,|\, M_m = a_m, M_n = a_n\Big),$$

for $W = \{j : m < j < n\} \subset \mathbb{Z}$. For this case, the Hamiltonian is built up of interactions of the form

$$h_k(x) = -\Big(\log p(x_{k-1}, x_k) + \log p(x_k, x_{k+1})\Big).$$

The following standard facts follow from Assumption (H1), and we shall make use of these throughout the paper.

Standard Facts: *(Georgii, 1988)*

(i) *When the set $G_{\oplus}(\Pi)$ is nonempty, it is convex and its extreme points consist of Gibbs distributions λ, any two of which are mutually singular on S.*

(ii) *A measure λ is extreme if and only if it is ergodic with respect to the group of translations $(\oplus_k : k \in \mathbb{Z}^d)$.*

(iii) *Any Gibbs measure $\xi \in G_{\oplus}(\Pi)$ can be written as a mixture of extremes: there exists a probability γ_ξ on $G_{\oplus}(\Pi)$ such that*

$$\xi(\cdot) = \int \lambda(\cdot) \gamma_\xi(d\lambda)$$

This measure satisfies

$$\gamma_\xi(A) = \xi(z : \lim_{V_n \uparrow \mathbb{Z}^d} \pi_{V_n}(z, \cdot) \in A), \quad A \subseteq G_{\oplus}(\Pi)$$

(iv) *Whenever* $\lambda \in G_{\oplus}(\Pi)$ *is extreme, the following Ergodic Theorem holds (Nguyen and Zessin, 1979): For any* $f \in L^p(d\lambda)$, $1 \le p < \infty$, *let* (V_n) *be an increasing sequence of finite subsets of* $\mathbb{Z}^d$ *such that*

$$\sup_n \frac{|V_n'|}{|V_n|} < \infty, \quad V_n' = \textit{convex hull of } V_n. \tag{15}$$

If the interior diameter of V_n,

$$d(V_n) = \sup\{ \textit{ radius of a sphere entirely contained in } V_n\}, \tag{16}$$

tends to infinity with n, *then*

$$\lim_{n\to\infty} \frac{1}{|V_n|} \sum_{k\in V_n} f \circ \oplus_k = \langle \lambda, f\rangle \quad \lambda \textit{ a.s. and in } L^p(d\lambda).$$

With a view towards applying the above ergodic theorem, we now make the assumption that

Hypothesis (H2): The scaling family (2), that is $\pi_n(dx) = \pi_{V_n}(z, dx)$, is constructed from a sequence (V_n) which is increasing, such that $|V_n| = n$, and satisfies both (15) and $d(V_n) \to \infty$ (with d defined in (16)). Moreover, let V be the neighbourhood in (H1); we set

$$\partial V_n = \{k \in \mathbb{Z}^d : k + V \not\subseteq V_n\},$$

and we shall assume that $|V + \partial V_n| \,/\, |V_n| < Cn^{-\alpha}$ for some $\alpha > 0$.

The condition involving α above restricts the growth of the boundary of V_n. It is clearly satisfied if the sets V_n are approximately cubes, for example.

The third hypothesis we make will be useful in Section 4.

Hypothesis (H3): For every $m \in \mathbb{Z}^d$, the second and third order derivatives of h_m are bounded:

$$\left\|D_{x_i x_j} h_k\right\|_\infty + \left\|D_{x_l x_m x_n} h_p\right\|_\infty < \infty, \quad i,j,k,l,m,n,p \in \mathbb{Z}^d.$$

Note that (H3) is satisfied by Example 3 and may often hold for Example 2. We believe that this condition can be relaxed considerably while keeping the results of this paper intact, but for simplicity, we do not pursue the matter here.

Hypothesis (H4): Every Gibbs measure $\xi \in G_{\oplus}(\Pi)$ satisfies, for some $\delta > 1$

$$\int |x_k|^{2\delta}\, \xi(dx) < \infty, \quad k \in \mathbb{Z}^d.$$

This hypothesis implies that, for any probability measure $\rho = (\rho_k : k \in \mathbb{Z}^d)$, the Gibbs distributions ξ satisfy

$$\xi\left(x : \sum_{k\in\mathbb{Z}^d} \rho_k \left|x_k\right|^2 < \infty\right) = 1.$$

As a result, we can restrict attention to a much smaller class of admissible configurations $x : S \to \mathbb{R}$, namely those which belong to $E = L^2(\rho)$. This will become the state space for the diffusion approximation of Section 5. The higher order moments will be used in conjunction with (H6) below, when we prove Lemma 12.

Example 2 revisited. Suppose that $\int x^{2\delta}\mu(dx) < \infty$ holds, then (H4) holds also. Indeed, we have

$$\begin{aligned}\int \left|x_k\right|^{2\delta} \xi(dx) &= \int \xi(dz) \int \left|x_k\right|^{2\delta} \pi_{\{k\}}(z, dx_k) \\ &\le \int \xi(dz) \int \left|x_k\right|^{2\delta} e^{\|U_k\|_\infty} \mu(dx_k) \\ &\le e^{\|U_k\|_\infty} \int \left|x_k\right|^{2\delta} \mu(dx_k) < \infty.\end{aligned}$$

Example 3 revisited. Here, (H4) holds always since

$$\sup_z \int \left|x_k\right|^{2\delta} \pi_{\{k\}}(z, dx_k) = \sup_z \int \left|x_k\right|^{2\delta} e^{-\sum_{v\in V} a_v x_k x_{k+v} - a_0 (x_k)^2} dx_k = C < \infty$$

so that

$$\int \left|x_k\right|^{2\delta} \xi(dx) \le \int \xi(dz) \sup_z \int \left|x_k\right|^{2\delta} \pi_{\{k\}}(z, dx_k) = C < \infty.$$

The following hypothesis is to be used in Section 5 for the existence of the infinite dimensional diffusion Z in (13). Note that this is satisfied by Example 3, and by Example 2 when (H3) holds.

Hypothesis (H5): For each $k \in \mathbb{Z}^d$, the function $h_k(x)$ given in (H1) satisfies the Lipschitz and growth conditions (with Euclidean norm)

$$\max_{v\in V} \|D_{x_v} h_k(x) - D_{x_v} h_k(y)\|_{\mathbb{R}^V} \le C \cdot \|x - y\|_{\mathbb{R}^V} \quad x, y \in \mathbb{R}^V$$

$$\max_{v\in V} \|D_{x_v} h_k(x)\|_{\mathbb{R}^V} \le C \cdot (1 + \|x\|_{\mathbb{R}^V}), \quad x \in \mathbb{R}^V$$

In Section 6, we shall prove the weak convergence (Theorem 13) referred to in the introduction. We shall use an assumption that the limiting Gibbs distribution

ξ is *strongly mixing.* There are various definitions of mixing in the literature. Here we use the following (for standard results on this topic, see Doukhan, 1994).

Let U, W be two subsets of $\mathbb{Z}^d$, and consider the corresponding σ-algebras $\mathcal{F}_U$, $\mathcal{F}_W$. The strong mixing coefficient of ξ is the number

$$\begin{aligned}\alpha_\xi(\mathcal{F}_U, \mathcal{F}_W) &= \sup\{|\xi(A \cap B) - \xi(A)\xi(B)| : A \subset \mathcal{F}_U, B \subset \mathcal{F}_W\} \\ &= \sup\{|\mathrm{Cov}(f(X_U), g(X_W))| : |f|, |g| \leq 1\}.\end{aligned} \tag{17}$$

Note that the coefficients can be calculated solely from the specification Π. We shall use the following assumption

Hypothesis (H6): There exists $\epsilon > 0$ such that

$$\sum_{r=1}^{\infty} (r+1)^{3d-1} |\alpha_\xi(r)|^{\epsilon/(4+\epsilon)} < \infty.$$

Here $\alpha_\xi(r) = \sup\{\alpha_\xi(\mathcal{F}_{A+V}, \mathcal{F}_{B+V}) : \mathrm{dist}(A, B) \geq r, |A| = |B| = 2\}$, and the distance between sets is defined as

$$\mathrm{dist}(A, B) = \min\{\max_{i \leq d} |a_i - b_i| : a \in A, b \in B\}.$$

Assumption (H6) is only used once, in the proof of Lemma 12. It stops the existence of phase transitions.

4. Random Walk Metropolis algorithm

In this section, we assume given the sequence $\pi_n(dx)$ of probability distributions on $\mathbb{R}^{V_n}$, where $|V_n| = n$. The dependence of π_n on the boundary condition z shall be temporarily ignored, and we shall identify $\mathbb{R}^{V_n}$ with $\mathbb{R}^n$.

For $n = 1, 2, 3, \ldots$, consider the Random Walk Metropolis algorithm for π_n, with proposal step

$$X_t \mapsto X_t + \sigma_n(R_1, \ldots, R_n),$$

where $R = (R_i)_{i=1}^{\infty}$ is a sequence of independent, identically distributed random variables with symmetric distribution and unit variance. The real number σ_n is used to control the variance of the proposal step generated from R. We shall assume that the variables R_i have at least four finite moments.

Lemma 1. *For any suitably bounded and differentiable test function $f : \mathbb{R}^n \to \mathbb{R}$, the discrete time generator of the Random Walk Metropolis chain for a distribution π_n with Lebesgue density*

$$\pi_n(x_1, \dots, x_n) = e^{-H_n(x_1,\dots,x_n)}, \quad n \geq 1$$

is given, as $\sigma_n \to 0$, by Taylor's theorem

$$A_n f(x) = -\sigma_n^2 \sum_{i=1}^n b_n^i(x) D_i f(x) + \frac{1}{2}\sigma_n^2 \sum_{i=1}^n a_n^{ii}(x) D_i^2 f(x) + o(n\sigma_n^2)$$

where

$$a_n^{ii}(x) = \mathbb{E}[1 \wedge e^{-K_n^i(x)}] \tag{18}$$

$$b_n^i(x) = \mathbb{E}\Big[D_i H_n(x + \sigma_n R \,|\, R_i = 0) e^{-K_n^i(x)}; K_n^i(x) > 0\Big] \tag{19}$$

and

$$\begin{aligned} K_n^i(x) &= H_n(x + \sigma_n R \,|\, R_i = 0) - H_n(x) \\ &= \sigma_n \sum_{r \neq i}^n D_r H_n(x) R_r + \frac{1}{2}\sigma_n^2 \sum_{r,s \neq i}^n D_{rs} H_n(x) R_r R_s \\ &\quad + \frac{1}{6}\sigma_n^3 \sum_{r,s,t \neq i}^n D_{rst} H_n(x + Z_n) R_r R_s R_t \end{aligned} \tag{20}$$

with some random variable Z such that $\|Z_n\| \leq \|\sigma_n W\|$

Proof. Let $F(dy) = q(y)dy$ be the distribution function of R_1. The generator of the n-th chain is

$$\begin{aligned} A_n f(x) &= \mathbb{E}\Big[f(x + \sigma_n R) - f(x); 1 \wedge \pi_n(x + \sigma_n R)/\pi_n(x)\Big] \\ &= \mathbb{E}\Bigg[\sum_{i=1}^n D_i f(x) \sigma_n R_i + \frac{1}{2}\sum_{i,j=1}^n D_{ij} f(x) \sigma_n^2 R_i R_j \\ &\qquad + o(\sigma_n^2 \|R\|^2); 1 \wedge e^{-\Big(H_n(x+\sigma_n R) - H_n(x)\Big)}\Bigg] \\ &= \sigma_n \sum_{i=1}^n D_i f(x) \int y \cdot \gamma_n^i(y) dF(y) \\ &\quad + \frac{1}{2}\sigma_n^2 \sum_{i,j=1}^n D_{ij} f(x) \iint yz \cdot \gamma_n^{ij}(y,z) dF(y) dF(z) + o(n\sigma_n^2) \end{aligned}$$

Here the coefficients are

$$\begin{aligned}\gamma_n^i(y) &= \mathbb{E}\Big[1 \wedge e^{-\left(H_n(x+\sigma_n R)-H_n(x)\right)} \mid R_i = y\Big] \\ &= \mathbb{E}\Big[1 \wedge e^{-\left(H_n(x+\sigma_n R \mid R_i=y)-H_n(x)\right)}\Big]\end{aligned}$$

where the notation $H_n(x+\sigma_n R \mid R_i = y)$ means that the variable R_i is replaced by y, i.e.

$$H_n(x+\sigma_n R \mid R_i = y) = H_n(x_1+\sigma_n R_1, \dots, x_i+\sigma_n y, \dots, x_n+\sigma_n R_n)$$

Similarly,

$$\begin{aligned}\gamma_n^{ij}(y,z) &= \mathbb{E}\Big[1 \wedge e^{-\left(H_n(x+\sigma_n R)-H_n(x)\right)} \mid R_i = y, R_j = z\Big] \\ &= \mathbb{E}\Big[1 \wedge e^{-\left(H_n(x+\sigma_n R \mid R_i=y,R_j=z)-H_n(x)\right)}\Big]\end{aligned}$$

To express the generator in terms of σ_n, we begin by expanding the functions γ_n^i, γ_n^{ij} in powers of y, z. Recall that if g is some differentiable function on $\mathbb{R}$, the the function $z \mapsto 1 \wedge \exp(-g(z))$ is also differentiable, except at a countable number of points, with derivative given Lebesgue almost everywhere by the function

$$\frac{d}{dz} 1 \wedge e^{-g(z)} = \begin{cases} -g'(z)e^{-g(z)} & \text{if } g(z) > 0 \\ 0 & \text{if } g(z) \le 0 \end{cases}$$

Now take $g(z) = H_n(x+\sigma_n R \mid R_i = z)$; for almost every $x_i \in \mathbb{R}$, as $\sigma_n \to 0$,

$$\begin{aligned}\gamma_n^i(y) = \mathbb{E}[1 \wedge e^{-\left(H_n(x+\sigma_n R \mid R_i=0) H_n(x)\right)}] \\ - (\sigma_n y)\mathbb{E}\Big[D_i H_n(x+\sigma_n R \mid R_i = 0) \cdot e^{-\left(H_n(x+\sigma_n R \mid R_i=0)-H_n(x)\right)}; \\ H_n(x+\sigma_n R \mid R_i = 0) - H_n(x) > 0\Big] + o(\sigma_n^2)\end{aligned}$$

Also

$$\gamma_n^{ij}(y,z) = \mathbb{E}[1 \wedge e^{-\left(H_n(x+\sigma_n R \mid R_i=y,R_j=z)-H_n(x)\right)}] + o(1)$$

Since (R_i) is iid with zero mean and unit variance, we then get (18), (19) and (20) as $\sigma_n \to 0$. □

Note that (18) gives the acceptance probability for a proposed move from x (compare with (11)).

5. Identifying the diffusion limit

Recall the Taylor series for $K_n^i(x)$ given in (20). We shall show in this section that, if $\sigma_n^2 = \ell/n$, we have $K_n^i(x) \Rightarrow \ell s(x)N + \frac{1}{2}\ell^2 a(x)$, where N is a standard normal variable. This will allow the identification of the limiting diffusion coefficients (18) and (19). We shall do this for the scaling problem using Lemma 1, with $H_n(x) = H_{V_n}(z_{V_n^c}, x_{V_n})$ and a sequence $R = (R_k : k \in \mathbb{Z}^d)$.

Recall the definition $H_V(x) = \sum_{v\in V} h_v(x)$ in (6) For each $\xi \in G_\oplus(\Pi)$, we have by (H4, H5)

$$s(\xi) = \left(\int \big(D_{x_0} H_V(x)\big)^2 \xi(dx)\right)^{1/2} < \infty. \tag{21}$$

Lemma 2. *For every extremal Gibbs distribution $\lambda \in G_\oplus(\Pi)$ and z [a.e. ξ], if (H1,H2,H4,H5) holds, then*

$$G_n(z,x) := \frac{1}{s(\lambda)\sqrt{n}} \sum_{k\in V_n} D_{x_k} H_{V_n}(z_{V_n^c}, x_{V_n}) \cdot R_k \Rightarrow \mathcal{N}(0,1), \quad x \text{ [a.e. } \lambda], \tag{22}$$

Moreover, if $F_n^{z,x}(u) = \mathbb{P}(G_n(z,x) \le u)$, set

$$B_n(x) = \sum_{k\in V_n} |D_{x_k} H(x)|^2, \quad I_n(x) = \sum_{k\in V_n} |D_{x_k} H(x)|^3,$$

and $\Gamma_n(x) = [B_n(x)/ns(\lambda)^2]^{1/2}$. Let δ be as in (H4), and set

$$M_n(z,x) = \left(\frac{1}{n}\sum_{k\in V+V_n} |z_k|^{2\delta}\right)^{1/\delta} + \left(\frac{1}{n}\sum_{k\in V+V_n} |x_k|^{2\delta}\right)^{1/\delta}.$$

If we take $0 < \theta < \alpha(1-1/\delta)/2$ (which is always possible, α begin given in (H2)), then there are constants C_1 and C_3 such that

$$\begin{aligned} \sup_u |F_n^{z,x}(u) - \Phi(u)| \le \Big(\Gamma_n(x) \vee 1\Big) \cdot \big|\Gamma_n(x)^{-1} - 1\big| + C_1 I_n(x)/B_n(x)^{3/2} \\ + C_3 M_n(x,z) n^{2\theta - \alpha(1-1/\delta)} + n^{-\theta}/\sqrt{2\pi}. \end{aligned} \tag{23}$$

Proof. If $k \in V_n \backslash \partial V_n$ then

$$\begin{aligned} D_{x_k} H_{V_n}(z_{V_n^c}, x_{V_n}) &= D_{x_k} \sum_{j \in V_n} h(x_j, x_{j+v^1} \dots, x_{j+v^m}) \\ &= D_{x_0} h_k(x_k, x_{k+v^1}, \dots, x_{k+v^m}) \\ &\quad + \dots + D_{x_{v^m}} h_{k-v^m}(x_{k-v^m} x_{k+v^1-v^m}, \dots, x_k) \\ &= \sum_{i \in V} D_{x_i} h_{k-i}(x) \\ &= (\sum_{i \in V} D_{x_i} h_{-i}) \circ \oplus_k(x) \\ &= D_{x_0} H \circ \oplus_k(x) = D_{x_k} H(x) \end{aligned}$$

independently of the chosen boundary configuration z. Define

$$S_n(x) = B_n(x)^{-1/2} \sum_{k \in V_n} D_{x_k} H(x) R_k$$

and

$$Q_n(z, x) = \frac{1}{s(\lambda)\sqrt{n}} \sum_{k \in \partial V_n} [D_{x_k} H_{V_n}(z_{V_n^c}, x_{V_n}) - D_{x_k} H(x)] R_k,$$

so that

$$\frac{1}{s(\lambda)\sqrt{n}} \sum_{k \in V_n} D_{x_k} H_{V_n}(z_{V_n^c}, x_{V_n}) R_k = \Gamma_n(x) S_n(x) + Q_n(z, x).$$

Then since $\Gamma_n(x) > 0$, the triangle inequality gives

$$\begin{aligned} &\sup_u \left| \mathbb{P}\Big(\Gamma_n(x) S_n(x) \le u \Big) - \Phi(u) \right| \\ &\qquad \le \sup_u \left| \mathbb{P}\Big(S_n(x) \le u/\Gamma_n(x) \Big) - \Phi(u/\Gamma_n(x)) \right| + \sup_u |\Phi(u/\Gamma_n(x)) - \Phi(u)| \\ &\qquad \le \sup_v |\mathbb{P}(S_n(x) \le v) - \Phi(v)| + (1 \vee \Gamma_n(x)) \left| 1 - \Gamma_n(x)^{-1} \right| \end{aligned}$$

By the Ergodic Theorem (Standard Fact (iv), Section 2) and (H4,H5) which ensure integrability, we have $\Gamma_n \to 1$ [*a.e.* λ]. Moreover, $S_n(x) \Rightarrow \mathcal{N}(0,1)$ by the Central Limit Theorem. Applying the Esseen bound (Petrov, 1995, Theorem 5.7, p.154) we get a constant C_1 such that

$$\sup_u |\mathbb{P}(S_n(x) \le u) - \Phi(u)| \le C_1 I_n(x)/B_n(x)^{3/2}.$$

Now using (Petrov, 1995, Lemma 1.9, p.20), we have

$$\sup_u |F_n^{z,x}(u) - \Phi(u)| \le \sup_u \left| \mathbb{P}\Big(\Gamma_n(x) S_n(x) \le u \Big) - \Phi(u) \right| \\ + \mathbb{P}(|Q_n(z,x)| > n^{-\theta}) + n^{-\theta}/\sqrt{2\pi},$$

whereas the Chebyshev inequality gives

$$\mathbb{P}\Big(|Q_n(z,x)| > n^{-\theta} \Big) \le n^{2\theta} \mathbb{E}\Big(\frac{1}{\sqrt{n}} \sum_{k \in \partial V_n} [D_{x_k} H_{V_n}(z_{V_n^c}, x_{V_n}) - D_{x_k} H(x)] R_k \Big)^2 \\ = \frac{n^{2\theta}}{|V_n|} \sum_{k \in \partial V_n} \left| D_{x_k} H_{V_n}(z_{V_n^c}, x_{V_n}) - D_{x_k} H(x) \right|^2,$$

as the sequence (R_k) is IID with zero mean and unit variance. Using the Lipschitz bound from (H5), and since $|V_n| = n$, we find a constant C_2 such that

$$\mathbb{P}(|Q_n(z,x)| > n^{-\theta}) \le \frac{n^{2\theta}}{n} \sum_{k \in V + \partial V_n} C_2 (z_k^2 + x_k^2).$$

From Hölder's inequality, taking $\delta > 1$ as above,

$$\sum_{k \in V + \partial V_n} z_k^2 = \sum_{k \in V + \partial V_n} 1_{V + \partial V_n}(k) z_k^2 \\ \le \Big(\sum_{k \in V + V + V_n} |1_{V + \partial V_n}(k)|^{\delta/(\delta - 1)} \Big)^{1 - 1/\delta} \Big(\sum_{k \in V + V + V_n} |z_k|^{2\delta} \Big)^{1/\delta} \\ = |V + \partial V_n|^{1 - 1/\delta} \Big(\sum_{k \in V + V + V_n} |z_k|^{2\delta} \Big)^{1/\delta}.$$

Consequently by (H2), there exists a constant C_3 such that

$$\mathbb{P}(|Q_n(z,x)| > n^{-1}) \le C_2 n^{2\theta} \Big(\frac{|V + \partial V_n|}{|V_n|} \Big)^{1 - 1/\delta} M_n(z,x) \\ < C_3 n^{2\theta - \alpha(1 - 1/\delta)} M_n(z,x).$$

Furthermore, by the Ergodic Theorem, (H4) and $|V_n + V| / |V_n| \to 1$ due to (H2), we get for x $[a.e.\ \lambda]/L^1(d\lambda)$ and z $[a.e.\ \xi]/L^1(d\xi)$,

$$\lim_{n \to \infty} M_n(z,x) = \Big(\int |x_0|^{2\delta} \, \xi(dx) \Big)^{1/\delta} + \Big(\int |x_0|^{2\delta} \, \lambda(dx) \Big)^{1/\delta}.$$

Thus we see that $\mathbb{P}(|Q_n(z,x)| > n^{-1})$ tends to zero, and consequently it follows both the weak convergence in (22) and the bound (23). □

The second sum in the expansion of K_n^i can also be treated by the ergodic theorem. Due to the finite range condition in (H1), we have $D_{x_u x_v} H_{V_n} = 0$ for "most" pairs $u, v \in V_n$.

Lemma 3. *For each extremal Gibbs distribution $\lambda \in G_\oplus(\Pi)$, if (H1,H2,H3) holds there exists a real number $a(\lambda)$ such that, as $n \to \infty$, for any $z \in S$,*

$$V_n(z,x) := \frac{1}{n} \sum_{i,j \in V_n} D_{x_i x_j} H_{V_n}(z_{V_n^c}, x_{V_n}) R_i R_j \to a(\lambda), \quad x \text{ [a.e. } \lambda], \text{ [a.s. } \mathbb{P}].$$

Moreover, there exists a constant C_4 such that

$$\mathbb{P}\Big(|V_n(z,x) - a(\lambda)| > \epsilon\Big) \leq C_4/\epsilon^2 n \text{ on } \Big\{x : \Big|\frac{1}{n}\sum_{i \in V_n} D^2_{x_0} H(x) - a(\lambda)\Big| < \epsilon/2\Big\}.$$

Proof. Since R_i is independent of R_j when $i \neq j$, only the diagonal terms will actually matter. Moreover, by the bounded range assumption (H1), we can write

$$V_n(z,x) := \sum_{j \in V} \frac{1}{n} \sum_{i \in V_n} D_{x_i x_{i+j}} H_{V_n}(z_{V_n^c}, x_{V_n}) R_i R_{i+j}.$$

Note that the variance of each term in the second (inner) sum is bounded, uniformly in x, $z \in S$ by (H3). When $j \neq 0$, the Strong Law of Large Numbers holds and Chebyshev's inequality gives

$$\begin{aligned}\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i \in V_n} D_{x_i x_{i+j}} H_{V_n}(\cdot) R_i R_{i+j}\Big| > \epsilon\Big) &\leq \frac{1}{\epsilon^2 n^2} \sum_{i \in V_n} \big|D_{x_i x_{i+j}} H_{V_n}(\cdot)\big|^2 \\ &\leq C_1/\epsilon^2 n,\end{aligned}$$

where C_1 is some generic constant. For $j = 0$, we write

$$\begin{aligned}\frac{1}{n}\sum_{i \in V_n} D^2_{x_i} H_{V_n}(z_{V_n^c}, x_{V_n}) R_i^2 &= \frac{1}{n}\sum_{i \in V_n} D^2_{x_i} H(x) \\ &+ \frac{1}{n}\sum_{i \in V_n} D^2_{x_i} H(x)(R_i^2 - 1) \\ &+ \frac{1}{n}\sum_{i \in \partial V_n} \Big(D^2_{x_i} H_{V_n}(z_{V_n^c}, x_{V_n}) - D^2_{x_i} H(x)\Big) R_i^2,\end{aligned}$$

of which both the second and third sums tend a.s. to zero, by the Strong Law of Large Numbers for independent random variables. Moreover, Chebyshev's inequality gives a similar estimate as before. For the first sum, the Ergodic Theorem gives

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i\in V_n} D^2_{x_i}H(x) = \langle\lambda, D^2_{x_0}H\rangle =: a(\lambda), \quad x\ [a.e.\ \lambda]/L^2(d\lambda). \tag{24}$$

□

Finally, let us consider the third term in the definition of K^i_n given in (20).

Lemma 4. *If (H3) holds, then for any sequence (Z^i_n), as $n\to\infty$*

$$J_n(z,x) := \frac{1}{n^{3/2}}\sum_{r,s,t\neq i} D_{x_rx_sx_t}H_{V_n}(z_{V_n^c}, x_{V_n}+Z_n)\cdot R_rR_sR_t \to 0, \quad [a.s.\ \mathbb{P}].$$

Moreover, $\mathbb{P}(|J_n(\cdot)|>\epsilon) \le C_5/\epsilon^2 n$ for some constant C_5.

Proof. By the nature of H_{V_n}, for fixed n the sum has at most $6\,|V|^2\,n$ nonzero terms. Hence by (H3), when divided by $n^{3/2}$, it tends to zero by the Law of Large Numbers. The estimate comes again from Chebyshev's inequality. □

For a given extreme Gibbs distribution $\lambda \in G_\oplus(\Pi)$, we shall now relate the numbers $s(\lambda)^2$ and $a(\lambda)$.

Lemma 5. *For every extremal Gibbs distribution $\lambda\in G_\oplus(\Pi)$, the numbers $s(\lambda)^2$ and $a(\lambda)$ defined by* (21) *and* (24) *are equal.*

Proof. Since the function h_k only depends on x_{k+V}, we shall condition on $\mathcal{F}_{W^c}$, where $W = V+V$.

$$\begin{aligned} s(\lambda)^2 &= \langle\lambda, \Big(D_{x_0}H\Big)^2\rangle \\ &= \langle\lambda, \Big(\sum_{i\in V} D_{x_i}h_{-i}\Big)^2\rangle \\ &= \int\lambda(dz)\int_{\mathbb{R}^W}\pi_W(z,dx)\Big(\sum_{i\in V} D_{x_i}h_{-i}(x)\Big)^2 \\ &= \int\lambda(dz)\int_{\mathbb{R}^W} C^{-1}_{z,W}e^{-H_W(z_{W^c},x_W)}\Big(\sum_{i\in V} D_{x_i}h_{-i}(x)\Big)^2 dx_W \end{aligned}$$

Now using integration by parts, since

$$\Big|e^{-H_W(z_{W^c},x_W)}D_{x_k}h_{-k}(x)\Big| \to 0 \quad \text{as } |x|\to\infty,$$

we find that

$$
\begin{aligned}
s(\lambda)^2 &= -\int \lambda(dz) \int_{\mathbb{R}^W} C_{z,W}^{-1} D_{x_0}\Big(e^{\sum_{j\in W} h_j(z_{W^c},x_W)}\Big)\cdot \sum_{i\in V} D_{x_i} h_{-i}(x) dx_W \\
&= \int \lambda(dz) \int_{\mathbb{R}^W} C_{z,W}^{-1} e^{-H_W(z_{W^c},x_W)}\Big(\sum_{i\in V} D_{x_i}^2 h_{-i}(x)\Big) dx_W \\
&= \int \lambda(dz) \int_{\mathbb{R}^W} \pi_W(z,dx) \sum_{i\in V} D_{x_0}^2 H(x) \\
&= a(\lambda)
\end{aligned}
$$

□

Combining the last four lemmas, the limiting diffusion coefficient becomes

$$
v(x) = \lim_{n\to\infty} a_n^{ii}(z_{V_n^c}, x_{V_n}) = \mathbb{E}\Big(1 \wedge e^{-\frac{1}{2}\ell^2 a(\lambda) - \ell s(\lambda) N}\Big) \quad x\ [a.e.\ \lambda]. \tag{25}
$$

We shall need a bound on the rate of convergence:

Lemma 6. *Let $v(x)$ be defined as in* (25)*, and choose any sequence $\epsilon_n \downarrow 0$. Then*

$$
\lim_{n\to\infty} \sup_{(z,x)\in E_n(\epsilon_n)} \big|a_n^{ii}(z_{V_n^c}, x_{V_n}) - v(x)\big| = 0,
$$

where the set $E_n(\epsilon) \subset S \times S$ satisfies

$$
E_n(\epsilon) = E_n^{1,x,\lambda}(\epsilon) \cap E_n^{2,x,\lambda}(\epsilon) \cap E_n^{3,x,\lambda}(1) \cap E_n^{3,z,\xi}(1) \cap E_n^{4,x,\lambda}(\epsilon),
$$

and using notation defined in Lemmas 2 and 3,

$$
\begin{aligned}
E_n^{1,x,\lambda}(\epsilon) &= \Big\{x : \Big|\frac{1}{n} B_n(x) - s(\lambda)^2\Big| < \epsilon\Big\}, \\
E_n^{2,x,\lambda}(\epsilon) &= \Big\{x : \Big|\frac{1}{n} I_n(x) - \langle \lambda, |D_{x_0} H|^3\rangle\Big| < \epsilon\Big\}, \\
E_n^{3,x,\lambda}(\epsilon) &= \Big\{x : \Big|\frac{1}{n} \sum_{k\in V_n+V} |x_k|^{2\delta} - \langle \lambda, |x_k|^{2\delta}\rangle\Big| < \epsilon\Big\}, \\
E_n^{4,x,\lambda}(\epsilon) &= \Big\{x : \Big|\frac{1}{n} \sum_{k\in V_n} D_{x_0}^2 H(x) - a(\lambda)\Big| < \epsilon\Big\}.
\end{aligned}
$$

Proof. We work with a fixed λ and put $\sigma = \ell s(\lambda)$, $\mu = \frac{\ell^2}{2} a(\lambda)$. In the notation defined in Lemmas 2, 3 and 4, we have

$$
K_n^i(z,x) = \sigma\Big(G_n(z,x) + \frac{\ell}{2s(\lambda)}[V_n(z,x) - a(\lambda)] + \frac{\ell^2}{6s(\lambda)} J_n(z,x)\Big) + \mu.
$$

Letting $H_n^{z,x}$ denote the CDF of $(K_n^i(z,x)-\mu)/\sigma$, it follows through integration by parts that

$$\begin{aligned}\left|a_n^{ii}(z_{V_n^c}, x_{V_n}) - v(x)\right| &= \left|\int 1 \wedge e^{-\sigma u+\mu}\big(dH_n^{z,x}(u) - d\Phi(u)\big)\right| \\ &= \left|\int_{\{u>-\mu/\sigma\}} \big(H_n^{z,x}(u) - \Phi(u)\big)\sigma e^{-\sigma u+\mu} du\right| \\ &\leq \sigma \sup_u \left|H_n^{z,x}(u) - \Phi(u)\right|.\end{aligned}$$

From (Petrov, 1995, p.20) we get for any choice of $\beta > 0$

$$\begin{aligned}&\sup_u \left|H_n^{z,x}(u) - \Phi(u)\right| \\ &\qquad \leq \sup_u \left|F_n^x(u) - \Phi(u)\right| + \mathbb{P}\Big(\left|V_n(z,x) - a(\lambda)\right| > 2\beta s(\lambda)/\ell\Big) \\ &\qquad\quad + \mathbb{P}\Big(\left|J_n(z,x)\right| > 6\beta s(\lambda)/\ell^2\Big) + s(\lambda)\frac{2\beta\ell^{-1} + 6\beta\ell^{-2}}{\sqrt{2\pi}}.\end{aligned}$$

Now applying the convergence estimates from Lemmas 2, 3 and 4,

$$\begin{aligned}&\sup_u \left|H_n^{z,x}(u) - \Phi(u)\right| \\ &\quad \leq \Big(\Gamma_n(x) \vee 1\Big) \cdot \left|\Gamma_n(x)^{-1} - 1\right| + C_1 I_n(x)/B_n(x)^{3/2} + C_3 M_n(x,z) n^{2\theta - \alpha(1-1/\delta)} \\ &\quad + n^{-\theta}/\sqrt{2\pi} + C_4\frac{\ell^2}{4\beta^2 s(\lambda)^2 n} + C_5\frac{\ell^4}{36\beta^2 s(\lambda)^2 n} + s(\lambda)\beta\frac{2\ell^{-1} + 6\ell^{-2}}{\sqrt{2\pi}},\end{aligned}$$

provided $x \in E_n^{4,x,\lambda}(\beta s(\lambda)\ell^{-1})$. Assuming furthermore that $x \in E_n^{1,x,\lambda}(\beta s(\lambda)\ell^{-1})$, we have $|\Gamma_n(x) - 1| < \text{const}\cdot\beta$. If also $x \in E_n^{2,x,\lambda}(\beta s(\lambda)\ell^{-1})$, then

$$\begin{aligned}I_n(x)/B_n(x)^{3/2} &= n^{-1/2}\Big[\frac{1}{n}I_n(x)\Big]\Big/\Big[\frac{1}{n}B_n(x)\Big]^{3/2} \\ &\leq n^{-1/2}\frac{\langle\lambda, |D_{x_0}H|^3\rangle + \beta s(\lambda)\ell^{-1}}{a(\lambda) - \beta s(\lambda)\ell^{-1}}.\end{aligned}$$

Finally, when $x \in E_n^{3,x,\lambda}(1)$ and $z \in E_n^{3,z,\xi}(1)$,

$$M_n(x,z) \leq \Big(\langle\lambda, |x_0|^{2\delta}\rangle + 1\Big)^{1/\delta} + \Big(\langle\xi, |x_0|^{2\delta}\rangle + 1\Big)^{1/\delta}.$$

Hence, for given $\epsilon_n \downarrow 0$, we can take $\beta = \epsilon_n\ell/s(\lambda)$, and we shall have the stated uniform convergence. □

To get the limiting drift, we can proceed in exactly the same way.

Lemma 7. *For any sequence $\epsilon_n \downarrow 0$, define the sets $E_n(\epsilon)$ as in Lemma 6, then*

$$\lim_{n\to\infty} \sup_{(z,x)\in E_n(\epsilon_n)} \left| b_n^i(z_{V_n^c}, x_{V_n}) - \frac{1}{2} D_{x_i} H(x) v(x) \right| = 0.$$

Proof. Consider the expression for $b_n^i(z_{V_n^c}, x_{V_n})$ in (19), and note first that

$$\begin{aligned}
&\lim_{n\to\infty} D_{x_i} H_{V_n}(z_{V_n^c}, x_{V_n} + \ell n^{-1/2} R \,|\, R_i = 0) \\
&\qquad = \lim_{n\to\infty} \Bigg(D_{x_i} H_{V_n}(z_{V_n^c}, x_{V_n}) + \ell n^{-1/2} \sum_{r\neq i} D_{x_i x_r} H_{V_n}(z_{V_n^c}, x_{V_n}) R_r \\
&\qquad + \frac{\ell^2}{n} \sum_{r,s\neq i}^{n} D_{x_i x_r x_s} H_{V_n}(z_{V_n^c} x_{V_n} + Z) R_r R_s \Bigg) \\
&\qquad = D_{x_i} H(x),
\end{aligned}$$

because the last two sums have respectively at most $|V|$ and $|V|^2$ nonzero terms. Moreover, the convergence is uniform in (z, x) due to (H3). By Lemmas 2, 3 and 4, we find that

$$\lim_{n\to\infty} b_n^i(z_{V_n^c}, x_{V_n}) = D_{x_i} H(x) \mathbb{E}(e^{-M}; M > 0),$$

where $M \sim \mathcal{N}(\mu, \sigma^2)$ and $\mu = (\ell^2/2)a(\lambda) = \sigma^2/2$. Using a similar technique as in Lemma 6, the limit is uniform in (z, x), along any sequence $E_n(\epsilon_n)$ with $\epsilon_n \downarrow 0$. Finally, by a standard calculation, the relation between μ and σ implies that

$$\mathbb{E}\Big(1 \wedge e^{-M}\Big) = 2\mathbb{E}\Big(e^{-M}; M > 0\Big) = 2\Phi\Big(-\sigma/2\Big).$$

Comparing this with (25) gives

$$\lim_{n\to\infty} b_n^i(z_{V_n^c}, x_{V_n}) = \frac{1}{2} D_{x_i} H(x) v(x).$$

□

Lemmas 6 and 7 immediately yield the form of the limiting diffusion operator we were aiming for:

$$A_\lambda f(x) = 2\ell^2 \Phi(-\ell s(\lambda)/2) \Big(-\frac{1}{2} \langle \nabla H(x), \nabla \rangle f(x) + \frac{1}{2} \Delta f(x) \Big), \quad x \text{ [a.e. } \lambda]. \tag{26}$$

Here $\nabla H(x)$ denotes the vector in S with components $k \mapsto D_{x_k} H(x)$. The above is true for any extreme Gibbs distribution $\lambda \in G_\oplus(\Pi)$. If we take an arbitrary

$\xi \in G_{\oplus}(\Pi)$, then according to Standard Facts (ii) and (iii) in Section 2, we get ξ a.e.

$$A_\xi f(x) = 2\ell^2 \Phi\Big(-\frac{\ell}{2}\sqrt{\xi(D_{x_0}H \,|\, \mathcal{I})(x)}\Big)\Big(-\frac{1}{2}\langle \nabla H(x), \nabla\rangle f(x) + \frac{1}{2}\Delta f(x)\Big) \tag{27}$$

We summarise with a theorem.

Theorem 8. *Suppose that ξ is a Gibbs distribution belonging to $G_{\oplus}(\Pi)$. Given (π_n) a scaling family (2) with boundary condition z and satisfying (H1, H2, H3, H4, H5), let $A^{V_n,z}$ be the generator of the Metropolis algorithm $X_t^{V_n,z}$ associated with π_n. If $f : S \to \mathbb{R}$ is any bounded differentiable test function which depends on at most a finite number of coordinates, then*

$$\lim_{n\to\infty} (n/\ell^2) A^{V_n,z} f(x) = A_\xi f(x), \quad (x, z) \ [a.e.\ \xi \otimes \xi].$$

Example 2 revisited. Let $\mu(dx) = f(x)dx$; equivalently, $h_k(x_k) = -\log f(x_k)$, and hence

$$D_{x_k} H(x) = -D_{x_k} \sum_j \log f(x_j) = f'(x_k)/f(x_k).$$

Since there is no phase transition, the σ-algebra $\mathcal{I}$ is trivial. Thus

$$\begin{aligned} s(\xi) &= \Big(\int (D_{x_0}H)^2 d\xi\Big)^{1/2} \\ &= \Big(\int \xi(dx) \int f'(x_0)^2/f(x_0)^2 \cdot f(x_0)dx_0\Big)^{1/2} \\ &= \Big(\int \frac{f'(x_0)^2 dx_0}{f(x_0)}\Big)^{1/2} \end{aligned}$$

We can write this in terms of a random variable $X \sim f(x)dx$, and then we get $s^2 = \mathbb{E}\big(f'(X)/f(X)\big)^2$. This agrees with (Roberts *et al.*, 1997).

Example 3 revisited. For the Gaussian specification (14), we have $D_{x_0}H(x) = q_0$, a constant. Consequently we must have $s(\xi) = q_0^{1/2}$, and the speed measure v of (10) is also constant, given by

$$v(x) \equiv 2\ell^2 \Phi\Big(-\frac{\ell}{2} q_0^{1/2}\Big)$$

The speed is maximised for $\hat{\ell} \approx 2.38/q_0^{1/2}$, which is one quarter the variance of the Gaussian product measure. Interestingly, this is independent of the existence or nonexistence of phase transitions.

Example 4 revisited. Recall that $M_t \sim m$ is a Markov chain with transition probability $p(x,y)dy$. When $p(x,y)$ is sufficiently smooth, we have

$$D_{x_0}H(x) = -\Big(D_2p(x_{-1},x_0)/p(x_{-1},x_0) + D_1p(x_0,x_1)/p(x_0,x_1)\Big).$$

Then

$$s^2 = \mathbb{E}\Big(D_{x_0}H(M_{-1},M_0,M_1)\Big)^2.$$

6. The limiting diffusion

In this section, we shall study properties of the diffusions associated with the operators A_λ given by (26), and more generally the operator A_ξ of (27). We first show the existence of such processes.

Let $\rho = (\rho_k : k \in \mathbb{Z}^d)$ be the probability measure on $\mathbb{Z}^d$ satisfying

$$\rho(\{k\}) := \rho_k = e^{-|k|} / \sum_{j\in\mathbb{Z}^d} e^{-|j|}$$

We shall write $E = L^2(\mathbb{Z}^d, \rho)$, and denote the corresponding Hilbert space norm by $\|\cdot\|_\rho$. The space E is a separable Hilbert space, and will be taken as the state space for our diffusions. Accordingly, we will focus only on Gibbs distributions which satisfy Hypothesis (H4) of Section 2. Note that $\mathbb{R}^{V_n} \subset E$ in a natural way, so that the Metropolis chains $(X^{V_n,z})$ can all be taken to evolve on E.

Since E is a Hilbert space, it admits the existence of a Brownian motion (B_t) Consequently, by the Hilbert space version of Ito's formula (e.g. Métivier, 1982), if the SDE

$$Z_t = Z_0 + B_t - \frac{1}{2}\int_0^t \nabla H(Z_t)dt \tag{28}$$

has a solution, it can be taken as a Markov process on E with generator

$$Lf(x) = \frac{1}{2}\sum_{k\in\mathbb{Z}^d} D^2_{x_k}f(x) - \frac{1}{2}\sum_{k\in\mathbb{Z}^d} D_{x_k}H(x)D_{x_k}f(x), \quad f \in C^2_b(E)$$

We prove this under Hypothesis (H5).

Proposition 9. *Under Hypothesis (H5), if $\mathbb{E}(Z_0^2) < \infty$, there exists a unique continuous, nonexplosive Markov process Z_t on E which satisfies* (28)*, and for each finite T, $\sup_{s\le T}\mathbb{E}\,\|Z_s\|^2_\rho < \infty$.*

Proof. The result follows from (Da Prato and Zabczyk, 1996, Theorem 5.3.1, p. 66) provided we show that the function $k \mapsto D_{x_k} H(x)$ satisfies the linear growth condition

$$\|\nabla H(x)\|_\rho \leq C_1(1 + \|x\|_\rho) \tag{29}$$

for some constant C_1, and the Lipschitz condition

$$\|\nabla H(x) - \nabla H(y)\|_\rho \leq C_2 \|x - y\|_\rho \tag{30}$$

for some constant C_2. To prove (31), note that the measure ρ satisfies

$$M_\rho := \sup_{k \in \mathbb{Z}^d} \sup_{v \in V+V} \rho_k/\rho_{k+v} < \infty.$$

Then we have

$$\begin{aligned}
\|\nabla H(x)\|_\rho &= \Big(\sum_{k \in \mathbb{Z}^d} \rho_k \left| D_{x_k} H(x) \right|^2 \Big)^{1/2} \\
&= \Big(\sum_{k \in \mathbb{Z}^d} \rho_k \Big| \sum_{v \in V} D_{x_v} h_{k-v}(x) \Big|^2 \Big)^{1/2},
\end{aligned}$$

so that by (H5), for some constants C', C''. Recalling that $V = \{0, v^1, \dots, v^m\}$,

$$\begin{aligned}
&\leq \Big(\sum_{k \in \mathbb{Z}^d} \rho_k \cdot C \, |V| \sum_{v \in V} (1 + \|(x_{k-v}, x_{k-v+v^1}, \dots, x_{k-v+v^m})\|)^2 \Big)^{1/2} \\
&\leq C' \Big(\sum_{k \in \mathbb{Z}^d} \rho_k \sum_{v \in V} \Big[1 + \|(x_{k-v}, x_{k-v+v^1}, \dots, x_{k-v+v^m})\|^2 \Big] \Big)^{1/2} \\
&= C'' \Big(1 + \sum_{k \in \mathbb{Z}^d} \rho_k \sum_{v \in V} x_{k-v}^2 + x_{k-v+v^1}^2 + \dots + x_{k-v+v^m}^2 \Big)^{1/2},
\end{aligned}$$

and setting $V - V = \{v - v' : v, v' \in V\}$, this can be written

$$\begin{aligned}
&\leq C'' \Big(1 + |V| \sum_{r \in V-V} \sum_{k \in \mathbb{Z}^d} \rho_k x_{k+r}^2 \Big)^{1/2} \\
&\leq C'' \Big(1 + M_\rho \, |V|^3 \sum_{k \in \mathbb{Z}^d} \rho_k x_k^2 \Big)^{1/2} \\
&\leq C_1 (1 + \|x\|_\rho),
\end{aligned}$$

as required in (31). The proof of (30) is virtually identical and is left to the reader. □

Below, we shall need to identify a core for the generator L.

Lemma 10. *Let* $C^{3,b}(\mathbb{R}^W) = \{f : \mathbb{R}^W \to \mathbb{R}, \left\|D^k f\right\|_\infty < \infty, k = 1,2,3\}$. *For any extremal Gibbs distribution* $\lambda \in G_\oplus(\Pi)$, *the set*

$$\begin{aligned}\mathcal{D} &= \{f \in C^{3,b} : \ f \textit{ depends only on a finite number of coordinates}\} \\ &= \bigcup_{W \textit{ finite}} C^{3,b}(\mathbb{R}^W)\end{aligned}$$

is dense in $L^2(E, d\lambda)$, *and is a core for the strong infinitesimal generator of* Z *acting on* $L^2(E, d\lambda)$.

Note that $\mathcal{D}$ separates points in E. Indeed, given two configurations x, y such that $\|x\|_\rho$, $\|y\|_\rho < \infty$, if $x \neq y$ there exists some function $g \in \mathcal{D}$ such that $g(x) \neq g(y)$.

Proof. We show first that $\mathcal{D}$ is dense. For a given finite set V_n, let λ_{V_n} be the restriction of λ to $\mathbb{R}^{V_n}$. The space $C^{3,b}(\mathbb{R}^{V_n})$ (which is contained in $\mathcal{D}$) is dense in $L^2(\mathbb{R}^{V_n}, d\lambda_{V_n})$. Note that $L^2(\mathbb{R}^{V_n}, d\lambda_{V_n})$ is naturally imbedded into the space $L^2(E, d\lambda)$. If a function f belongs to the latter, the martingale convergence theorem implies that $f_n = \lambda(f \,|\, \mathcal{F}_{V_n})$ converges to f in $L^2(E, d\lambda)$ whenever $V_n \uparrow \mathbb{Z}^d$. But f_n also belongs to $L^2(\mathbb{R}^{V_n}, d\lambda_{V_n})$, and hence can be approximated by a function in $\mathcal{D}$. Thus f itself can be approximated by a function in $\mathcal{D}$ and therefore this set is dense.

To prove that $\mathcal{D}$ is a core for L, let (T_t) denote the operator semigroup of Z, acting on $L^2(E, d\lambda)$, and suppose that $\mathcal{L}$, with domain $\mathbb{D}(\mathcal{L})$, is the associated strong infinitesimal generator. We will show that $\mathcal{D} \subset \mathbb{D}(\mathcal{L})$ and that $T_t : \mathcal{D} \to \mathbb{D}(\mathcal{L})$, which establishes the claim by Ethier and Kurtz (1986), , p.17, prop. 3.3.

It is known by Ito's formula that the process

$$M_t = f(Z_t) - f(Z_0) - \int_0^t Lf(Z_s)ds$$

is a local martingale for each $f \in \mathcal{D}$. Moreover, the solution process Z satisfies for each finite T:

$$\sup_{s \le T} \mathbb{E}_x \left\|Z_s\right\|_\rho^2 < \infty \text{ if } \left\|x\right\|_\rho^2 < \infty. \tag{31}$$

Let $T_n = \inf\{s > 0 : \|Z_s\|_\rho \geq n\}$; by the continuity of the sample paths of Z, $M_{t\wedge T_n}$ is a bounded martingale, and for $\|x\|_\rho < \infty$, we therefore have

$$\mathbb{E}_x f(Z_{t\wedge T_n}) = f(x) + \mathbb{E}_x \int_0^t Lf(Z_s) 1_{(T_n > s)} ds.$$

Now f is bounded and continuous on E, and in case f depends only on the coordinates x_W, we have by (H5) some constant C_f such that

$$\|Lf(x)\|_\rho \leq C_f(1 + \|x\|_{\mathbb{R}^{W+V}}). \tag{32}$$

Thus we have $\left|Lf(Z_s)1_{(T_n>s)}\right| \leq C_f(1 + \|Z_s\|_{\mathbb{R}^{W+V}})$, which by (31) is integrable. Dominated convergence now gives

$$\mathbb{E}_x f(Z_t) = f(x) + \int_0^t \mathbb{E}_x[Lf(Z_s)]ds.$$

Moreover, from this it follows easily, by (31), (32) and the continuity of $x \mapsto Lf(x)$ on E, that

$$\lim_{t\to 0} \int \lambda(dx) \left|t^{-1}(\mathbb{E}_x f(Z_t) - f(x)) - Lf(x)\right|^2 = 0,$$

which establishes that $\mathcal{D} \subset \mathbb{D}(\mathcal{L})$. Similarly, it is clear that $T_t f(x) = \mathbb{E}_x f(Z_t)$ also belongs to the domain of $\mathcal{L}$. This ends the proof. □

As a direct implication of the above lemma, we get

Proposition 11. *Let $\xi \in G_\oplus(\Pi)$ be a Gibbs distribution satisfying Hypothesis (H4); then ξ is a stationary distribution for the process Z_t solving* (28).

Proof. It suffices to consider the case when $\xi = \lambda$ is extreme. A simple calculation shows that $\langle \lambda, Lf \rangle = 0$ for all $f \in \mathcal{D}$, and this implies the result by Ethier and Kurtz, 1986 (Proposition 9.2, p. 239). □

By the continuity of the sample paths of Z, we have in fact also

$$\mathbb{P}_\xi(Z_t \in \mathrm{supp}(\xi) \text{ for all } t > 0) = 1.$$

Note that in the above, the support of ξ is taken in the topology of $E = L^2(\mathbb{Z}^d, \rho)$. A consequence is that if λ and λ' are two distinct extremal Gibbs distributions, then the hitting time of $\mathrm{supp}(\lambda) \cap \mathrm{supp}(\lambda')$ is a.s. infinite.

Having constructed a diffusion solving (28), that is with generator L, it is straightforward to construct a solution to the equation

$$Z_t = Z_0 + \int_0^t v(Z_t)^{1/2} dB_t - \frac{1}{2}\int_0^t v(Z_t)\nabla H(Z_t) dt, \tag{33}$$

by a time change associated with the additive functional $A_t = \int_0^t v(Z_t)dt$ of the process solving (28). This process will have as generator the operator A_ξ defined in (27).

7. Weak convergence

We now come to the main result of this paper. The implications of the theorem have already been discussed in the introduction.

We shall need the following lemma:

Lemma 12. *Let $\xi \in G_\oplus(\Pi)$ be the limit of the measures π_n as before, and suppose that (H6) holds. Then for each of the functions on S, $g_1(x) = |D_{x_0}H(x)|^2$, $g_2(x) = |D_{x_0}H(x)|^3$, $g_3(x) = |x_0|^{2\delta}$ and $g_4(x) = D^2_{x_0}H(x)$, there are constants C_i such that, for some $p > 2$,*

$$\int \left| \frac{1}{|V_n|} \sum_{k\in V_n} g_i \circ \oplus_k(x) - \langle \xi, g_i\rangle \right|^p \xi(dx) \le C_i n^{-p/2}, \quad i = 1,2,3,4.$$

Proof. Observe that for each i, $g_i \in L^{2+\epsilon}(d\xi)$ by (H3)-(H5), for sufficiently small $\epsilon > 0$. Moreover, g_i is $\mathcal{F}_V$ measurable. Fix now some $i \le 4$, and consider the centered random field $Y_k(x) = g_i \circ \oplus_k(x) - \langle \lambda, g_i\rangle$, $k \in \mathbb{Z}^d$ under the probability measure ξ. It is easy to see that the strong mixing coefficient of Y given by (17) satisfies

$$\alpha_Y(\mathcal{F}_U, \mathcal{F}_W) \le \alpha_\xi(\mathcal{F}_{U+V}, \mathcal{F}_{W+V}),$$

where α_ξ is the mixing coefficient of ξ. Let $p > 2$, and recall that d is the dimension of $\mathbb{Z}^d$. By (H6) we have, with $u = 2$,

$$\sum_{r=1}^\infty (r+1)^{d(4-u+1)-1} |\alpha_Y(r; u, v)|^{\epsilon/(4+\epsilon)} < \infty,$$

where

$$\begin{aligned}\alpha_Y(r;u,v) &= \sup\{\alpha_Y(\mathcal{F}_A,\mathcal{F}_B) : \mathrm{dist}(A,B)\geq r,\\ &\qquad 2\leq |A|\leq u, 2\leq |B|\leq v, u+v\leq 4\}\\ &\leq \sup\{\alpha_\xi(\mathcal{F}_{A+V},\mathcal{F}_{B+V}) : \mathrm{dist}(A,B)\geq r, |A|=|B|=2\}\\ &= \alpha_\xi(r).\end{aligned}$$

Consequently, by (Doukhan, 1994, p.26, Theorem 1), since Y_k is translation invariant and satisfies $\sup_{k\in\mathbb{Z}^d}\|Y_k\|_{L^{p+\epsilon}(\xi)}<\infty$,, there is a constant C such that

$$\int\left|\sum_{k\in V_n} Y_k(x)\right|^p \xi(dx)\leq Cn^{p/2}.$$

Dividing both sides by n^p finishes the proof. ☐

Theorem 13. *Let (H1)-(H6) hold, and suppose given a Gibbs distribution $\xi\in G_\oplus(\Pi)$. For ξ-almost every boundary condition z (fixed once chosen), let*

$$\pi_n(dx)=\xi(dx\,|\,\mathcal{F}_{V_n^c})(z)$$

be a corresponding scaling family of probability distributions on $\mathbb{R}^{V_n}$, and suppose that $X^{V_n,z}$, starting at π_n, is a stationary Random Walk Metropolis algorithm for π_n with proposal variance $\sigma_n^2=\ell^2/n$; then as $n\to\infty$,

$$(X^{V_n,z}_{[tn/\sqrt{\ell}]} : t\geq 0)\Rightarrow (Z_t : t\geq 0)\ \textit{on}\ E,z\ [a.e.\ \xi], \tag{34}$$

where Z is the diffusion solving (28) *with $Z_0\sim\xi$.*

Proof. Consider the operator A_ξ, with restricted domain $\mathcal{D}$. The closure of this operator generates a continuous contraction semigroup on $L^2(E,d\xi)=\bar{\mathcal{D}}$, namely the semigroup associated with the solution of (33). By Lemmas 6 and 7, we have for any sequence $\epsilon_n\downarrow 0$,

$$\lim_{n\to\infty}\sup_{(z,x)\in E_n(\epsilon_n)}\left|A^{V_n,z}f(x)-A_\xi f(x)\right|=0,\quad \forall f\in\mathcal{D}.$$

By Ethier and Kurtz, 1986 (Corollary 8.9, p.233), we have (34), provided we can choose the sequence (ϵ_n) such that

$$\lim_{n\to\infty}\mathbb{P}_{\pi_n}\left(X^{V_n,z}_{[tn/\sqrt{\ell}]}\in E_n(\epsilon_n) : 0\leq t\leq T\right)=1. \tag{35}$$

We shall do this as follows: since $X^{V_n,z}$ is stationary under π_n, we have

$$\mathbb{P}_{\pi_n}\left(X^{V_n,z}_{[tn/\sqrt{\ell}]} \notin E_n(\epsilon_n) \text{ for some } 0 \le t \le T\right) \le (nT/\sqrt{\ell})\mathbb{P}_{\pi_n}\left(X^{V_n,z}_0 \notin E_n(\epsilon_n)\right).$$

Using the definition of $E_n(\epsilon_n)$ (Lemma 6) and the fact that $X^{V_n,z} \sim \pi_n$, this last probability satisfies the bound

$$\mathbb{P}_{\pi_n}\left(X^{V_n,z}_0 \notin E_n\right) \le \sum_{i=1}^{4} \pi_n\Big[x : x \notin E^{i,x,\lambda}_n(\epsilon_n)\Big] + 1_{S \setminus E^{3,z,\xi}_n(1)}(z).$$

Now recall that $\pi_n(dx) = \xi(dx \,|\, \mathcal{F}_{V_n^c})(z)$, hence for $i = 1, 2, 3$, the functions

$$\tau^i_n(z) = (Tn/\sqrt{\ell})\pi_n\Big[x : x \notin E^{i,x,\lambda}_n(\epsilon_n)\Big]$$

are $\mathcal{F}_{V_n^c}$ measurable. The limit $\tau^i(z) = \overline{\lim}_{n\to\infty}\tau^i_n(z)$ is therefore measurable with respect to the tail σ-algebra $\mathcal{T} = \cap_W \mathcal{F}_{W^c}$, where the intersection is over all finite subsets $W \subset \mathbb{Z}^d$. However, whenever λ is ergodic, $\mathcal{T}$ is trivial. It follows that $\tau^i \ge 0$ is constant λ-almost everywhere. Now compute the estimate, by using Markov's inequality and Jensen's inequality for conditional expectations

$$\begin{aligned}\lambda(z : \left|\tau^i_n(z)\right| > c) &\le \frac{1}{c}\int \left|\tau^i_n(z)\right| \lambda(dz) \\ &\le \frac{1}{c}(Tn/\sqrt{\ell})\lambda\left(x : x \notin E^{i,x,\lambda}_n(\epsilon_n)\right).\end{aligned}$$

Each of the sets $E^{i,x,\lambda}_n(\epsilon_n)$ is of the form

$$E^{i,x,\lambda}_n(\epsilon_n) = \left\{x : \left|\frac{1}{n}\sum_{k\in V_n} g_i \circ \oplus_k(x) - \langle\lambda, g_i\rangle\right| < \epsilon_n\right\},$$

where $g_i \in L^1(d\lambda)$. Markov's inequality implies therefore that

$$\begin{aligned}\lambda(z : \left|\tau^i_n(z)\right| > c) &\le \frac{1}{c}(Tn/\epsilon^p_n\sqrt{\ell})\int \left|\frac{1}{n}\sum_{k\in V_n} g_i \circ \oplus_k(x) - \langle\lambda, g_i\rangle\right|^p \lambda(dx). \\ &\le \text{const}(n/\epsilon^p_n n^{p/2}), \quad p > 2,\end{aligned}$$

by Lemma 12. If we now choose the sequence $\epsilon^p_n = n^{-\gamma}$, where $0 < \gamma < p/2 - 1$, then we shall have that $\tau^i_n \to 0$ in λ measure, and hence that $\tau^i = 0$, λ-almost everywhere. Because ξ is a mixture of ergodic measures λ, the same holds ξ-almost everywhere. Finally note that, except on a ξ-null set, we have $z \in E^{3,z,\xi}_n(1)$ for all sufficiently large n, by the Ergodic Theorem (n depends on z). We conclude that

$$\overline{\lim}_{n\to\infty}\mathbb{P}_{\pi_n}\left(X^{V_n,z}_0 \notin E_n\right) = 0, \quad z \ [a.e.\ \xi],$$

and this establishes (35) as required. □

We end with a remark on the strong mixing condition (H6). As the proof of the previous theorem makes clear, Hypothesis (H6) was used (via Lemma 12) solely to guarantee that the partial sums $\frac{1}{|V_n|}\sum_{k\in V_n} g_i \circ \oplus_k$ converge in $L^p(d\lambda)$ at a rate faster than $n = |V_n|$, for each ergodic Gibbs distribution λ. The weak convergence conclusion of Theorem 13 therefore holds whenever such a claim can be made, irrespective of the presence or absence of phase transitions.

8. Acknowledgement

The research was supported by the European Union TMR network ERB-FMRX-CT96-0096.

References

[1] Doukhan, P. (1994) *Mixing. Properties and Examples.* Lecture Notes in Statistics 85. Springer-Verlag.

[2] Ethier, S.N. and Kurtz, T.G. (1986) *Markov Processes. Characterization and Convergence.* John Wiley and Sons.

[3] Frigessi, A., Martinelli, F., Stander (1996), J. Computational complexity of Markov chain Monte Carlo methods for finite Markov random fields. *Biometrika* **84**, no. 1, 1-18.

[4] Georgii, H.O. (1988) *Gibbs Measures and Phase Transitions* de Gruyter Studies in Mathematics 9.

[5] Nguyen, X.X. and Zessin, H. (1979) Ergodic theorem for general dynamic systems. Z. Wahrscheinlichkeitstheorie verw. Gebiete **48**, 159-176.

[6] Petrov, V. (1995) *Limit Theorems of Probability Theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press.

[7] Da Prato, G. and Zabczyk, J. (1996) *Ergodicity for Infinite Dimensional Systems.* Cambridge University Press.

[8] Roberts, G.O., Gelman, A. and Gilks, W.R. (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. Preprint.

[9] Smith A.F.M. and Roberts G.O (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *J. Roy. Stat. Soc. B* **55**, 1, 1993, 3-23.

Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, 16, Mill Lane, Cambridge CB2 1SB

E-mail address: L.Breyer@statslab.cam.ac.uk

Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, 16, Mill Lane, Cambridge CB2 1SB

E-mail address: G.O.Roberts@statslab.cam.ac.uk