Propp–Wilson algorithms and finitary codings for high noise Markov random fields

Olle Häggström* Jeffrey E. Steif†

Chalmers University of Technology

September 2, 1998

Abstract

In this paper, we combine two previous works, the first being by the first author and K. Nelander, and the second by J. van den Berg and the second author, to show (1) that one can carry out a Propp–Wilson exact simulation for all Markov random fields on $\mathbf{Z}^d$ satisfying a certain high noise assumption, and (2) that all such random fields are a finitary image of a finite state i.i.d. process. (2) is a strengthening of the previously known fact that such random fields are so-called Bernoulli shifts.

1 Introduction

A random field with finite state space S indexed by the integer lattice $\mathbf{Z}^d$ is a random mapping $X : \mathbf{Z}^d \to S$, or it can equivalently be seen as a random element of $S^{\mathbf{Z}^d}$. Here we focus on so-called **Markov random fields**, characterized by having a dependency structure which only propagates via interactions between nearest neighbors in $\mathbf{Z}^d$. We specialize further to Markov random fields satsifying a certain **high noise** assumption, which says that these interactions should be sufficiently weak in a way that will be made precise in Definition 1.2 below. Our results for such random fields are twofold: First, we devise a so-called **Propp–Wilson algorithm** for computer simulation of the random field. Second, we use the existence and some properties of such an algorithm to prove that the high noise assumption implies a certain rather strong ergodic property, known as **finitary codability**, of the random field.

The rest of this introductory section is organized as follows. In Sections 1.1 and 1.2 we provide the context by giving careful descriptions of the Markov random field property and the high noise assumption, supplemented by a couple of examples. In Section 1.3, we describe our results concerning simulation algorithms, and in Section 1.4 we state our main result concerning finitary codability.

1.1 Markov random fields

We briefly introduce Markov random fields here, and refer to Kindermann and Snell [8], Georgii [5] and Georgii et al. [6] for more background and detail.

*Research supported by a grant from the Swedish Natural Science Research Council.
†Research supported by a grant from the Swedish Natural Science Research Council.

The Markov random field property is defined relative to some neighborhood structure in $\mathbf{Z}^d$. Here we restrict, for simplicity and concreteness, to the case where $x, y \in \mathbf{Z}^d$ are considered nearest neighbors, denoted $x \sim y$, if and only if their (Euclidean) distance is 1. For $\Lambda \subset \mathbf{Z}^d$, we define the boundary $\partial\Lambda$ of Λ as

$$\partial\Lambda = \{x \in \mathbf{Z}^d \setminus \Lambda : \exists y \in \Lambda \text{ such that } x \sim y\}.$$

With a slight abuse of notation, we will, for $x \in \mathbf{Z}^d$, write ∂x for $\partial\{x\}$.

Let X be a random field on $\mathbf{Z}^d$ with finite state space S, and let μ be the corresponding probability measure on $S^{\mathbf{Z}^d}$.

Definition 1.1 *The random field X with distribution μ is said to be a* **Markov random field** *if μ admits conditional probabilities such that for all finite $\Lambda \subset \mathbf{Z}^d$, all $\xi \in S^\Lambda$, and all $\eta \in S^{\mathbf{Z}^d \setminus \Lambda}$ we have*

$$\mu(X(\Lambda) = \xi \mid X(\mathbf{Z}^d \setminus \Lambda) = \eta) = \mu(X(\Lambda) = \xi \mid X(\partial\Lambda) = \eta(\partial\Lambda)).$$

In words, the Markov random field property says that the conditional distribution of what we see on Λ given everything else only depends on what we see on the boundary $\partial\Lambda$.

A consistent set of conditional distributions for all finite Λ and all boundary conditions η as above is called a **specification**, usually denoted by $\mathcal{Q}$. A probability measure μ on $S^{\mathbf{Z}^d}$ satisfying the prescribed conditional distributions is called a Gibbs measure for $\mathcal{Q}$. Existence of Gibbs measures for a given specification follows from a standard compactness argument. In contrast, uniqueness does not always hold, and the issue of when it does (or does not) is a central and highly intricate one which has been the subject of countless studies. Nonuniqueness of Gibbs measures is referred to as a **phase transition**.

We assume that the specification is shift invariant. To make this more precise, let T be the shift operator on $\mathbf{Z}^d$: For $x, y \in \mathbf{Z}^d$, we set $T_x y = x + y$, for $\Lambda \subset \mathbf{Z}^d$ we set $T_x\Lambda = x + \Lambda$, and for $\xi \in S^\Lambda$, define $T_x\xi \in S^{T_x\Lambda}$ by setting $T_x\xi(y) = \xi(T_{-x}y)$ for each $y \in T_x\Lambda$. A specification $\mathcal{Q}$ is said to be shift invariant if for each finite Λ, each $x \in \mathbf{Z}^d$ and each $\eta \in S^{\partial\Lambda}$, the prescribed conditional distribution of $T_{-x}X(T_x\Lambda)$ given that $X(T_x\partial\Lambda) = T_x\eta$ is the same as the prescribed conditional distribution of $X(\Lambda)$ given that $X(\partial\Lambda) = \eta$.

For concreteness, we give two well-known examples of Markov random fields and their specifications.

Example: The Potts model. Fix $q \in \{2, 3, \ldots\}$ and $J \in \mathbf{R}$, and set $S = \{1, \ldots, q\}$; the elements of S are called spins. A probability measure μ on $S^{\mathbf{Z}^d}$ is said to be a Gibbs measure for the q-state Potts model with coupling constant J if it is Markov and for each finite Λ, each $\eta \in S^{\partial\Lambda}$, and each $\xi \in S^\Lambda$ we have

$$\mu(X(\Lambda) = \xi \mid X(\partial\Lambda) = \eta) = \frac{1}{Z^{q,J}_{\Lambda,\eta}} \exp\left(2J \sum_{\substack{x \sim y \\ x, y \in \Lambda}} \mathbf{1}_{\{\xi(x) = \xi(y)\}} + 2J \sum_{\substack{x \sim y \\ x \in \Lambda, y \in \partial\Lambda}} \mathbf{1}_{\{\xi(x) = \eta(y)\}}\right).$$

Here $Z^{q,J}_{\Lambda,\eta}$ is a normalizing constant, $\mathbf{1}_A$ is the indicator function of the event A, and each nearest neighbor pair $\{x, y\}$ is counted only once. Taking $J = 0$ gives an i.i.d. random field. Taking $J > 0$ biases the i.i.d. measure in favor of configurations where many nearest neighbor pairs have the same spin, whereas a negative value of J has the opposite effect. The case $q = 2$ is the much studied Ising model. Much of the interest in Ising and Potts models in statistical mechanics comes from their phase transition behavior: for $d \geq 2$, there is a critical value $J_c = J_c(d, q) \in (0, \infty)$ such that for $J \in [0, J_c)$ there is a unique Gibbs measure whereas for $J > J_c$ there are multiple Gibbs measures.

Example: The Widom–Rowlinson lattice gas model. Fix $q \in \{2, 3, \ldots\}$ and $\lambda > 0$, and set $S = \{0, 1, \ldots, q\}$. A probability measure μ on $S^{\mathbf{Z}^d}$ is said to be a Gibbs measure for the q-state Widom–Rowlinson model with activity λ, if it is Markov and for each finite Λ, each $\eta \in S^{\partial\Lambda}$, and each $\xi \in S^{\Lambda}$ we have

$$\mu(X(\Lambda) = \xi \,|\, X(\partial\Lambda) = \eta) = \frac{\lambda^{n(\xi)} \mathbf{1}_A}{\tilde{Z}^{q,\lambda}_{\Lambda,\eta}}$$

where $\tilde{Z}^{q,\lambda}_{\Lambda,\eta}$ is a normalizing constant, $n(\xi)$ is the number of vertices $x \in \Lambda$ with $\xi(x) \neq 0$, and A is the event that no two neighboring vertices x and y (with either $x, y \in \Lambda$, or $x \in \Lambda$, $y \in \partial\Lambda$) take different values unless at least one of them is a 0. It is natural to think of 0's as empty sites and $\{1, \ldots, q\}$ as q different types of particles, with a strong repulsion between different types. The phase transition behavior in $d \geq 2$ dimensions is similar to that of the Potts model: taking λ sufficiently small gives a unique Gibbs measure whereas taking λ large enough produces multiple Gibbs measures.

1.2 The high noise assumption

A high noise assumption on a Markov random field X, or equivalently on its specification, is a condition which states that the conditional distribution of $X(x)$ for $x \in \mathbf{Z}^d$ given $X(\partial x)$ does not depend too strongly on $X(\partial x)$. The high noise assumption that we will work under in this paper was used by Häggström and Nelander [7] in a context where $\mathbf{Z}^d$ is replaced by some finite graph structure.

Let $\mathcal{Q}$ be a shift invariant specification on $\mathbf{Z}^d$ with state space S. For $x \in \mathbf{Z}^d$ and $s \in S$, set

$$\gamma_s = \gamma_s(\mathcal{Q}) = \min_{\eta \in S^{\partial x}} \mathcal{Q}(X(x) = s \,|\, X(\partial x) = \eta) \tag{1}$$

and note that by shift invariance, this quantity is independent of x. Also set

$$\gamma = \gamma(\mathcal{Q}) = \sum_{s \in S} \gamma_s\,. \tag{2}$$

For reasons to be explained in Section 2, we call γ the **multigamma admissibility** of $\mathcal{Q}$. Clearly, $\gamma \in [0, 1]$, and we shall see that it can be interpreted as a probability.

Definition 1.2 *A Markov random field on $\mathbf{Z}^d$ and its specification $\mathcal{Q}$ are said to satisfy* **Condition HN** *if*

$$\gamma(\mathcal{Q}) > \frac{2d - 1}{2d}\,.$$

HN is short for High Noise, and has nothing to do with the (purely coincidental!) fact that the authors of [7] have these initials.

Other high noise conditions have been introduced by Dobrushin [4] and by van den Berg and Maes [1], and shown to be sufficient conditions for having a unique Gibbs measure for a given specification. It is easy to check that Condition HN is stronger than the van den Berg–Maes uniqueness condition. This means that we do not need to worry about the possible nonuniqueness of Gibbs measures when working under Condition HN.

Remark. In many examples (such as the Widom–Rowlinson model), certain finite patterns have probability 0. In such cases, the minimum in (1) can be taken over the set of boundary conditions η which have positive probability, because for such a Markov random field, we may take the conditional distribution under an exceptional boundary condition to be equal to the conditional distribution under some nonexceptional boundary condition.

Let us finally see what Condition HN means for the examples in Section 1.1. A calculation shows that the Potts model with coupling constant J has multigamma admissibility

$$\gamma = \begin{cases} \frac{q}{q-1+\exp(4dJ)} & \text{for } J \geq 0 \\ \frac{q}{1+(q-1)\exp(-4dJ)} & \text{for } J \leq 0\,, \end{cases}$$

so that Condition HN is satisfied if and only if

$$\frac{1}{4d}\log\left(1-\frac{q}{2dq-2d+1}\right) < J < \frac{1}{4d}\log\left(1+\frac{q}{2d-1}\right).$$

For the Widom–Rowlinson model with activity λ, we get $\gamma = \frac{1}{q\lambda+1}$, satisfying Condition HN if and only if $\lambda < \frac{1}{q(2d-1)}$.

1.3 Propp–Wilson algorithms

It is important, both in statistical mechanics and in image analysis applications, to be able to sample from a Markov random field living on a large but finite graph structure $G = (V, E)$. Direct sampling via enumeration is not computationally feasible due to the hugeness of the sample space (for the Ising model on a 256×256 grid it has cardinality 2^{65536}). The standard way to approach this problem is the **Markov chain Monte Carlo** method, which dates back to the 1953 paper by Metropolis *et al.* [11]. The idea is to find an ergodic (usually reversible) finite state Markov chain whose unique stationary distribution is the desired Gibbs measure μ. If we start the chain in an arbitrary state, and run it for long enough, then the distribution of the final state is close to μ.

An example is the **Gibbs sampler**, also known as the **heat bath algorithm**, where at each integer time a vertex x is chosen (either at random or according to some deterministic sweeping rule) and receives a new value chosen according to the conditional distribution (under μ) of $X(x)$ given the current value of $X(\partial x)$. Such a chain is reversible, and under an irreducibility condition (which usually is easy to check) it is also ergodic with unique stationary distribution μ.

A major drawback of this approach is that it is often very difficult to determine useful bounds on the convergence rate of the Markov chain, i.e. to decide how long the chain must run to come close to equilibrium. For this reason, the mid-1990's have seen an intense effort at refining the Markov chain Monte Carlo method by designing algorithms which decide automatically how long they need to run, and whose output have *exactly* the target distribution μ. See e.g. the annotated bibliography [17]. Algorithms achieving these goals are today sometimes referred to as "exact" or "perfect" simulation algorithms. The breakthrough came in the 1996 paper by Propp and Wilson [15], who coupled several Markov chains (i.e. ran them in parallel) in an ingenious way, in order to produce an algorithm which has the desired properties and which moreover is fast enough to be useful in several nontrivial instances. They gave a particularly striking variant of the algorithm for random fields obeying a certain monotonicity property: that the state space S admits a total ordering such that the conditional probability of a large value at a vertex $x \in V$ given that $X(V \setminus \{x\}) = \eta$ is increasing (with respect to the induced partial ordering) in η. This holds for the Ising model, and also for the $q = 2$ Widom–Rowlinson model (the state space $\{0, 1, 2\}$ then needs to be equipped with the slightly unusual ordering $1 < 0 < 2$), but not for the $q \geq 3$ Potts and Widom–Rowlinson models.

An extension of the Propp–Wilson algorithm for monotone Markov random fields was given by van den Berg and Steif [2]. While Propp and Wilson only handled Markov random fields living on finite graphs, van den Berg and Steif showed how to sample from the exact distribution of a monotone Markov random field living on the infinite lattice $\mathbf{Z}^d$, provided that the specification of the Markov random field has a unique Gibbs measure. Of course, an entire configuration $\eta \in S^{\mathbf{Z}^d}$ is typically not even possible to store in a finite computer, so what their algorithm does is the following: given a finite set $\Lambda \subset \mathbf{Z}^d$ and a specification $\mathcal{Q}$ which has the required monotonicity property and which gives rise to a unique Gibbs measure μ on $S^{\mathbf{Z}^d}$, it produces a sample from the projection of μ on S^Λ. Equipping $S^{\mathbf{Z}^d}$ with the usual product topology, it is thus possible to sample from μ with arbitrary prespecified precision; this is similar in spirit to the "ϵ-accurate perfect simulation" algorithm of Møller [12].

Häggström and Nelander [7] developed the Propp–Wilson approach in a different direction: they stayed within the context of finite graphs, but showed how fast exact simulation of a Markov random field is possible when the monotonicity assumption is replaced by (a finite graph variant of) Condition HN.

In Section 2 of this paper we shall unite the approaches in [2] and [7] in order to devise an algorithm which achieves the following. Given any finite $\Lambda \subset \mathbf{Z}^d$ and any specification $\mathcal{Q}$ satisfying Condition HN, we can sample from the projection on S^Λ of the (unique) Gibbs measure μ for the specification $\mathcal{Q}$. The computational complexity of this algorithm is moderate. In fact, the expected running time grows only linearly in the cardinality of Λ (as is the case for the algorithm in [2] for the Ising model with $J \in [0, J_c)$) – a fact which may be of substantial value for the practitioner of the algorithm, and which moreover is crucial to the application in ergodic theory to be described in the next subsection.

1.4 Finitary codings

In this section, we give a very brief description of the notion of a finitary coding. While all definitions will later be given in complete detail, the reader is referred to [2] for a much more complete description of known results and how these concepts fit into the general ergodic theory framework.

Loosely speaking, we say that a stationary process $Y = \{Y_n\}_{n \in \mathbf{Z}^d}$ is a factor of a stationary process $X = \{X_n\}_{n \in \mathbf{Z}^d}$ if there is a map which maps the realizations of the X process into new realizations such that these new realizations are essentially the Y process and such that this mapping is spatially invariant. If in this "coding procedure", to determine the image process at a fixed location x "downstairs", you only need to look at a sufficiently large (random) window around x in the upstairs process, then one calls this mapping **finitary**. Precise definitions are given in Section 3.

One of the main results of [2] is that there exists (resp. does not exist) a finitary mapping from a finite state i.i.d. process to the so-called plus state of the Ising model (see [2] for the relevant definitions) for $J \in [0, J_c)$ (resp. $J > J_c$). Monotonicity for the Ising model played an important role in this result as well as in the known exponential convergence results (see [10]) of the corresponding dynamics. Here, we are able to drop the monotonicity at the cost of only being able to work under condition HN (in the Ising Model, this means working well below the critical interaction parameter).

Theorem 1.3 *For any Markov random field Y on $\mathbf{Z}^d$ satisfying Condition HN, there exists an i.i.d. finite-state process X on $\mathbf{Z}^d$ and a finitary coding f such that Y and $f(X)$ are equal in distribution.*

This is a strengthening of the well known fact that all high noise Markov random fields are Bernoulli shifts, i.e. that they are a factor (not necessarily finitary) of a finite state i.i.d. process. This fact can be obtained from the standard ergodic theory literature together with any of the following papers concerning Markov random fields: [9], [16], [1] or [3].

2 The algorithm

Fix the dimension d, the finite state space S, and a specification $\mathcal{Q}$ satisfying Condition HN. Let μ be the unique Gibbs measure on $S^{\mathbf{Z}^d}$ for $\mathcal{Q}$.

As a first step towards the algorithm described in the last paragraph of Section 1.3, we first define a so-called **probabilistic cellular automaton** (PCA) $\{X_i\}_{i\in\mathbf{Z}}$, which in our case is effectively a Gibbs sampler for μ with massive parallel updating. It is, however, only a theoretical construction, not aimed at running on a computer, because at each integer time an infinite number of vertices are updated. Starting with an initial configuration $X_0 \in S^{\mathbf{Z}^d}$, what this cellular automaton does is to select a set $Y_0 \subset \mathbf{Z}^d$ of vertices, and to update these according to the conditional distribution prescribed by $\mathcal{Q}$, given $X_0(\mathbf{Z}^d \setminus Y_0)$. This gives a new configuration $X_1 \in S^{\mathbf{Z}^d}$. A new set Y_1 of vertices is then selected, and these are updated according to $\mathcal{Q}$ given $X_1(\mathbf{Z}^d \setminus Y_1)$, producing another configuration X_2, and so on. We require that each set Y_i of vertices is spread out, in the sense that it does not contain any nearest neighbor pair $x \sim y$. The point of this is that all the vertices then update independently of each other, conditional on $X_i(\mathbf{Z}^d \setminus Y_i)$. We propose two ways of selecting Y_i.

Updating scheme A: For some $p \in (0,1)$, let $\{I_{i,x}\}_{i\in\mathbf{Z},x\in\mathbf{Z}^d}$ be i.i.d. Bernoulli variables with expectation p. For each i and x, take

$$x \in Y_i \quad \text{if and only if} \quad I_{x,i} = 1 \text{ and } \sum_{y\in\partial x} I_{y,i} = 0\,.$$

In other words, each $x \in \mathbf{Z}^d$ is independently selected to be a candidate for membership in Y_i with probability p, and then all candidates that have another candidate among its nearest neighbors are dismissed. The probability that a given vertex x is included in Y_i is $p(1-p)^{2d}$. It makes sense to try to maximize this probability, and to this end we will always take $p = \frac{1}{2d+1}$.

Updating scheme B: Partition $\mathbf{Z}^d$ into two sets $\mathbf{Z}^d_{even}$ and $\mathbf{Z}^d_{odd}$, by letting each $x = (x_1, \ldots, x_d)$ belong to $\mathbf{Z}^d_{even}$ if its coordinate sum $x_1 + \ldots + x_d$ is even, and to $\mathbf{Z}^d_{odd}$ otherwise. Let $Y_i = \mathbf{Z}^d_{even}$ for i even, and $Y_i = \mathbf{Z}^d_{odd}$ for i odd.

When it comes to doing simulation in practice using algorithms based on this PCA, the most natural choice is probably to use updating scheme B. On the other hand, updating scheme A has the mathematically more pleasing property of space-time homogeneity, which in fact is crucial to the construction of finitary codings in Section 3.

Now let $U_{i,x}$ be i.i.d. random variables, uniformly distributed on $[0,1]$. Also let $\mathbf{0}$ denote the origin in $\mathbf{Z}^d$, and define a function

$$\phi : S^{\partial \mathbf{0}} \times [0,1] \to S$$

with the property that for each $\eta \in S^{\partial \mathbf{0}}$ and each $s \in S$ we have

$$\int_0^1 \mathbf{1}_{\{\phi(\eta,t)=s\}} dt = \mathcal{Q}(X(\mathbf{0}) = s \,|\, X(\partial \mathbf{0}) = \eta) . \tag{3}$$

The updating in the PCA can then be realized from ϕ and $\{U_{i,x}\}_{i \in \mathbf{Z}, x \in \mathbf{Z}^d}$ as follows. Whenever $x \in Y_i$ and $X(\partial x) = \eta$, we set $X_{i+1}(x) = \phi(T_{-x}\eta, U_{i,x})$. That this gives the dynamics of the PCA the desired distribution is clear from (3).

This gives a natural way of coupling several copies of the PCA with different initial configurations: just expose them to the same $U_{i,x}$ variables (and, in case of updating scheme A, also the same $I_{i,x}$ variables). If the single-site conditional probabilities in $\mathcal{Q}$ depend only weakly on neighbors' values, then it may be possible to define ϕ in such a way that the evolutions of PCAs starting with different initial conditions tend to rapidly coalesce, so that in other words the initial configurations are quickly forgotten. Here is where the multigamma admissibility γ defined in (2) comes in. To propose a concrete choice for ϕ, we first need some more definitions.

Define subprobability measures Q and $\{Q_\eta : \eta \in S^{\partial \mathbf{0}}\}$ on S as follows. For each $s \in S$ and $\eta \in S^{\partial \mathbf{0}}$, set

$$Q(s) = \gamma_s(\mathcal{Q})$$

and

$$Q_\eta(s) = \mathcal{Q}(X(\mathbf{0}) = s \,|\, X(\partial \mathbf{0}) = \eta) - \gamma_s(\mathcal{Q}) .$$

Note that $Q + Q_\eta$ is a probability measure for each $\eta \in S^{\partial \mathbf{0}}$, and that it equals the conditional distribution of $X(\mathbf{0})$ given $X(\partial \mathbf{0}) = \eta$. Equip S with a total ordering $<$, and denote the elements of S by $s_1, \ldots, s_k$, in such a way that $s_1 < \cdots < s_k$. For all $\eta \in S^{\partial \mathbf{0}}$ and $u \in [0,1]$, set

$$\phi(\eta, u) = \begin{cases} \min\{s_j \in S : Q(\{s_1, \ldots, s_j\}) \geq u\} & \text{for } u \leq \gamma \\ \min\{s_j \in S : Q_\eta(\{s_1, \ldots, s_j\}) \geq u - \gamma\} & \text{for } u > \gamma . \end{cases}$$

A little thought shows that this is a valid choice of ϕ, i.e. that (3) holds. The main point of this choice of ϕ is that whenever $x \in Y_i$ and $U_{i,x} \leq \gamma$, the value of $X_{i+1}(x)$ does not depend on X_i. Hence, if we couple the PCA with itself with two or more initial configurations using ϕ and the same $U_{i,x}$, then the values at a vertex after it is updated coincide for all the PCAs with probability at least γ. One can check that no other choice of ϕ can guarantee a higher such coalescence probability. This way of coupling several random objects to maximize coalescence probability is called **multigamma coupling** (hence the term multigamma admissibility for γ), and was introduced in the study of Propp–Wilson algorithms by Murdoch and Green [13].

Now define a more complicated PCA $\{\Xi_i\}_{i\in\mathbf{Z}}$, in which vertices in $\mathbf{Z}^d$ are assigned values in $\mathcal{S}$ rather then in S, where $\mathcal{S}$ denotes the set of all nonempty subsets of S. To avoid confusion between $\{\Xi_i\}$ and $\{X_i\}$, we call $\{\Xi_i\}$ a **super-PCA**, and $\{X_i\}$ is called an **ordinary PCA**. For $\xi \in \mathcal{S}^{\mathbf{Z}^d}$ and $\Lambda \subset \mathbf{Z}^d$, we identify $\xi(\Lambda)$ with the set

$$\{\eta \in S^\Lambda : \eta(x) \in \xi(x) \text{ for all } x \in \Lambda\}.$$

The evolution of the super-PCA is as follows. If at time i we have a configuration $\Xi_i \in \mathcal{S}^{\mathbf{Z}^d}$, then we obtain Ξ_{i+1} by selecting a set $Y_i \in \mathbf{Z}^d$ according to updating scheme A or B, and for each $x \in \mathbf{Z}^d$ setting

$$\Xi_{i+1}(x) = \begin{cases} \{s \in S : \exists \eta \in \Xi_i(\partial x) \text{ such that } \phi(T_{-x}\eta, U_{i,x}) = s\} & \text{if } x \in Y_i \\ \Xi_i(x) & \text{otherwise.} \end{cases} \tag{4}$$

Given Ξ, we define the $\{0,1\}$-valued random variables $\{Z_i(x)\}_{i\in\mathbf{Z},x\in\mathbf{Z}^d}$ by letting $Z_i(x) = 1$ if the cardinality of $\Xi_i(x)$ is at least 2, and letting $Z_i(x) = 0$ otherwise. In other words, $Z_i(x) = 0$ if and only if $\Xi_i(x)$ consists of a single element of S.

Suppose that we start the super-PCA $\{\Xi_i\}$ at time 0 in the state $S^{\mathbf{Z}^d}$, that we start an ordinary PCA $\{X_i\}$ at time 0 with some arbitrary configuration $\eta \in S^{\mathbf{Z}^d}$, and that we expose $\{\Xi_i\}$ and $\{X_i\}$ to the same randomness (i.e. to the same $I_{i,x}$- and $U_{i,x}$-variables). We write P for the probability measure on an appropriate probability space underlying the PCA and the super-PCA.

We claim that

$$X_i(x) \in \Xi_i(x) \tag{5}$$

for all $i \geq 0$ and all $x \in \mathbf{Z}^d$. To see this, note that it holds trivially for $i = 0$ and that it is preserved under the dynamics of $\{\Xi_i\}$ and $\{X_i\}$.

In particular, (5) tells us that if $Z_i(x) = 0$, then we know that we would see the same value of $X_i(x)$ irrespective of the starting configuration X_0. It is therefore reasonable to argue that if for some given i and $\Lambda \subset \mathbf{Z}^d$ we have $Z_i(x) = 0$ for all $x \in \Lambda$, then the values on Λ at time i for the ordinary PCA have forgotten the initial configuration. The following result tells us that the random field $\{Z_i(x) : x \in \mathbf{Z}^d\}$ quickly turns into "mostly zeroes" as i increases. It is precisely here that Condition HN does its job.

Proposition 2.1 *Suppose that $\mathcal{Q}$ satisfies Condition HN with multigamma admissibility γ, and that we run the super-PCA starting at time 0 with $\Xi_0 = S^{\mathbf{Z}^d}$. Then, for any $x \in \mathbf{Z}^d$, the probability that $Z_i(x) = 1$ tends to 0 exponentially fast as $i \to \infty$. Quantitatively:*

(a) If updating scheme A is used, then

$$P(Z_i(x) = 1) \leq \left(1 - \frac{(2d)^{2d}(1 - 2d(1-\gamma))}{(2d+1)^{2d+1}}\right)^i \tag{6}$$

for all $i \geq 0$ and all $x \in \mathbf{Z}^d$.

(b) If updating scheme B is used, then

$$P(Z_i(x) = 1) \leq (2d(1-\gamma))^{i-1}$$

for all $i \geq 0$ and all $x \in \mathbf{Z}^d$.

Proof: We prove (a); (b) follows by a similar argument. Set

$$p^* = \frac{(2d)^{2d}}{(2d+1)^{2d+1}}$$

and note that p^* is the probability that $x \in Y_k$ for given x and k. The proof is by induction over i. For $i = 0$, (6) holds trivially. Assuming that (6) holds for $i = k$, the following calculation shows that it holds also for $i = k + 1$:

$$\begin{aligned}
&P(Z_{k+1}(x) = 1) = \\
&\quad = \; P(x \in Y_k)P(Z_{k+1}(x) = 1 \,|\, x \in Y_k) + P(x \notin Y_k)P(Z_{k+1}(x) = 1 \,|\, x \notin Y_k) \\
&\quad \leq \; p^* P(Z_{k+1}(x) = 1 \,|\, x \in Y_k) + (1 - p^*)\left(1 - p^*(1 - 2d(1-\gamma))\right)^k && (7) \\
&\quad \leq \; p^* P(Z_k(y) = 1 \text{ for some } y \in \partial x,\, U_{k,x} > \gamma) + (1 - p^*)\left(1 - p^*(1 - 2d(1-\gamma))\right)^k && (8) \\
&\quad = \; p^*(1-\gamma) P(Z_k(y) = 1 \text{ for some } y \in \partial x) + (1 - p^*)\left(1 - p^*(1 - 2d(1-\gamma))\right)^k && (9) \\
&\quad \leq \; p^*(1-\gamma) \sum_{y \in \partial x} P(Z_k(y) = 1) + (1 - p^*)\left(1 - p^*(1 - 2d(1-\gamma))\right)^k \\
&\quad \leq \; p^*(1-\gamma) 2d \left(1 - p^*(1 - 2d(1-\gamma))\right)^k + (1 - p^*)\left(1 - p^*(1 - 2d(1-\gamma))\right)^k && (10) \\
&\quad = \; \left(1 - p^*(1 - 2d(1-\gamma))\right)^{k+1} .
\end{aligned}$$

Here (7) and (10) use the induction hypothesis, (8) uses the fact that $Z_{k+1}(x) = 0$ unless $U_{k,x} > \gamma$ and some neighbor y of x has $Z_k(y) = 1$, and (9) uses independence between X_k and $U_{k,x}$. □

We now show how these theoretical constructions can be turned into actual algorithms for sampling from μ. We make the usual assumption that we have access to an unlimited sequence of i.i.d. uniform $[0,1]$ random variables (this is of course unrealistic, since, at least to our knowledge, only pseudo-random number generators are available), to be used for generating elements from the arrays $\{U_{i,x}\}_{i \in \mathbf{Z}, x \in \mathbf{Z}^d}$ and $\{I_{i,x}\}_{i \in \mathbf{Z}, x \in \mathbf{Z}^d}$, sequentially as we need them. Our task is the following: for a fixed finite $\Lambda \subset \mathbf{Z}^d$, we want to sample from the projection of μ on S^Λ.

We described the PCAs as starting at time 0 and running forwards in time. The Propp–Wilson approach to simulation is a somewhat different scheme: running from negative times until time 0, and trying earlier and earlier starting times until eventually it can be verified that the starting configuration has become irrelevant. In our context, the idea amounts to the following where we write ${}_{k_1}\Xi_{k_2}$ for the state of the super-PCA at time k_2, having started at time k_1 in state $S^{\mathbf{Z}^d}$, and similarly adopt the notation ${}_{k_1}Z_{k_2}$. First run the super-PCA from some negative time K_1 (starting with $\Xi_{K_1} = S^{\mathbf{Z}^d}$) until time 0, and check whether we have coalescenced at time 0, i.e. whether ${}_{K_1}Z_0(x) = 0$ for each $x \in \Lambda$. If yes, then output the unique element of S^Λ corresponding to ${}_{K_1}\Xi_0(\Lambda)$. Otherwise, try an earlier starting time K_2 and see if we have coalescenced at time 0, and so on. Here $(K_1, K_2, \ldots)$ is some sequence of earlier and earlier starting times. For simplicity, we shall take $(K_1, K_2, \ldots) = (-1, -2, \ldots)$ even though more sophisticated choices have been advocated, e.g. in [15].

Define sets $\Lambda_0, \Lambda_1, \ldots \subset \mathbf{Z}^d$ by setting $\Lambda_0 = \Lambda$, and, for $j = 1, 2, \ldots$, setting $\Lambda_j = \Lambda_{j-1} \cup \partial(\Lambda_{j-1})$. In other words, Λ_i is the set of vertices that are at distance at most i from some vertex of Λ.

It is important to note that to generate ${}_K\Xi_0(\Lambda)$ from ${}_K\Xi_{-1}$, we only need to know the values of ${}_K\Xi_{-1}$ on Λ_1, because of the locality (spatial Markovianness) of the update rule (4). Next, ${}_K\Xi_{-1}(\Lambda_1)$ only depends on ${}_K\Xi_{-2}$ through ${}_K\Xi_{-2}(\Lambda_2)$, and so on. So to get

${}_K\Xi_0(\Lambda)$ we start with ${}_K\Xi_K(\Lambda_{-K}) = S^{\Lambda_{-K}}$ and sequentially calculate ${}_K\Xi_{K+1}(\Lambda_{-K-1})$, ${}_K\Xi_{K+2}(\Lambda_{-K-2})$ and so on. In pseudocode, the algorithm is as follows; we have taken the liberty to abridge the trivial calculations of Y_k and ${}_KZ_0$.

$K := 0$
repeat
  $K := K - 1$
  for all $x \in \Lambda_{-K}$
    ${}_K\Xi_K(x) := S$
  $k := K$
  repeat
    compute $\Lambda_{-k-1} \cap Y_k$
    for all $x \in \Lambda_{-k-1} \cap Y_k$
      ${}_K\Xi_{k+1}(x) := \{s \in S : \exists \eta \in {}_K\Xi_k(\partial x) \text{ such that } \phi(T_{-x}\eta, U_{k,x}) = s\}$
    $k := k + 1$
  until $k = 0$
  compute ${}_KZ_0(\Lambda)$
until ${}_KZ_0 = 0$ for all $x \in \Lambda$
output ${}_K\Xi_0(\Lambda)$

Note that the algorithm sometimes uses $U_{i,x}$ for the same i and x more than once; we stress that the same value of $U_{i,x}$ must be used each time (rather than generating a new random number each time). We need to verify the correctness of this algorithm, and that is done in the following theorem. We write μ_Λ for the projection on S^Λ of the Gibbs measure μ.

Theorem 2.2 *For any finite $\Lambda \subseteq \mathbf{Z}^d$, the above algorithm (with either updating scheme A or updating scheme B) terminates with probability* 1, *and outputs an unbiased sample X from μ_Λ.*

Proof: Again we give the proof for updating scheme A only; the other case follows similarly.

Proposition 2.1 in conjunction with space-time homogeneity of the dynamics of the PCA imply that

$$P({}_KZ_0(x) = 1) \leq \left(1 - \frac{(2d)^{2d}(1 - 2d(1-\gamma))}{(2d+1)^{2d+1}}\right)^{-K}$$

for each $K < 0$. Writing $|\Lambda|$ for the cardinality of Λ, we thus have

$$P({}_KZ_0(x) = 0 \text{ for all } x \in \Lambda) \geq 1 - |\Lambda| \left(1 - \frac{(2d)^{2d}(1 - 2d(1-\gamma))}{(2d+1)^{2d+1}}\right)^{-K}$$

which tends to 1 as $K \to -\infty$, and the a.s. termination of the algorithm follows.

To prove unbiasedness of the output, define the negative integer-valued random-variable $\tilde{K}(\Lambda)$ to be the earliest starting time needed before the algorithm terminates. Let X^* be a random element of $S^{\mathbf{Z}^d}$, which has distribution μ and which furthermore is independent of $\{I_{i,x}\}_{i \in \mathbf{Z}, x \in \mathbf{Z}^d}$ and $\{U_{i,x}\}_{i \in \mathbf{Z}, x \in \mathbf{Z}^d}$. For $J < 0$, write ${}_JX_0^*$ for the $S^{\mathbf{Z}^d}$-valued random element obtained at time 0 by taking $X_J = X^*$ as starting configuration for the ordinary PCA (using the same $I_{i,x}$'s and $U_{i,x}$'s) running from time J to time 0. Since the dynamics of the PCA preserves μ, we have that ${}_JX_0^*$ has distribution μ for each J.

On the other hand, an obvious extension of (5) shows that ${}_JX_0^*(\Lambda) = X(\Lambda)$ for each $J \le \tilde{K}(\Lambda)$, so that ${}_JX_0^*(\Lambda)$ converges a.s. to $X(\Lambda)$ as $J \to -\infty$. But since ${}_JX_0^*(\Lambda)$ has distribution μ_Λ for each J, we get that $X(\Lambda)$ also has distribution μ_Λ. □

A variant of the above algorithm, worth considering for large $|\Lambda|$, is to run it separately for each $x \in \Lambda$ (again reusing the random variables $U_{i,x}$ and $I_{i,x}$ whenever the same (i,x) is encountered more than once). The advantage of this approach is that the expected time needed to simulate the value at a single vertex is easily shown to be finite, whence the same holds for any number of vertices, and moreover the expected time to simulate $X(\Lambda)$ grows only linearly in $|\Lambda|$. Indeed, the number of updates needed to simulate the value at a single vertex x is (in the case of updating scheme A) no more than

$$\sum_{K=0}^{\infty} K(2K+1)^d P(\tilde{K}(x) < -K) \le \sum_{K=0}^{\infty} K(2K+1)^d \left(1 - \frac{(2d)^{2d}(1-2d(1-\gamma))}{(2d+1)^{2d+1}}\right)^K < \infty .$$

The same idea can be useful even in the context of Markov random fields of finite graphs. For large finite subgraphs Λ of $\mathbf{Z}^d$, we get an expected running time which is $O(|\Lambda|)$, improving on an algorithm in [7] which (still under Condition HN) has an expected running time which is $O(|\Lambda| \log |\Lambda|)$. The algorithm in [7] is simpler and may be preferable for moderately sized Λ.

3 The finitary coding

Recall from Section 1.4 the processes X and Y. If the X process takes its values in the finite set A and the Y process takes its values in the finite set B, then the distributions of these processes are translation invariant probability measures μ and ν on $A^{\mathbf{Z}^d}$ and $B^{\mathbf{Z}^d}$ respectively. If there exists a measure preserving map f from $(A^{\mathbf{Z}^d}, \mu)$ to $(B^{\mathbf{Z}^d}, \nu)$ which is defined a.e. and which commutes with shifts, then we say that Y is a factor of X (or ν a factor of μ). Here measure preserving means that for all measurable $U \subset B^{\mathbf{Z}^d}$, $\nu(B) = \mu(f^{-1}(B))$ or equivalently the processes $f(X)$ and Y are equal in distribution. To say that f commutes with shifts means that $f(T_x(\eta)) = T_x(f(\eta))$ for all $x \in \mathbf{Z}^d$ and μ-a.e. η.

The coding or mapping f is called **finitary** if it is continuous after removing some set of measure 0. There is another more natural equivalent definition of finitary (which also explains the word finitary). f is a finitary coding if and only if there exists a set $\mathcal{N} \subseteq A^{\mathbf{Z}^d}$ of μ–measure 0 such that for all $\zeta \in A^{\mathbf{Z}^d} \backslash \mathcal{N}$, there exists an integer r (depending on ζ) so that if $\eta \in A^{\mathbf{Z}^d} \backslash \mathcal{N}$ and $\eta(i) = \zeta(i)$ for all i with $\|i\| \le r$, then $f(\eta)(0) = f(\zeta)(0)$. (By translation invariance, the analogous thing holds at locations other than 0.) In words, after a large enough *finite* box of the ζ configuration is revealed, we know the 0th coordinate of $f(\zeta)$. If this occurs, we say that Y is a finitary factor of X.

The finitary codings for our high noise Markov random fields is based on the Propp–Wilson algorithm of the previous section. We first observe that although the random variables $U_{i,x}$ were uniform variables, it is clear, since there are only finitely many possibilities for the relevant conditional probabilities, that we could take the random variables $U_{i,x}$ to take values in some finite set $\mathcal{U}$ and proceed *precisely* as we did before. We therefore now assume that. We now let $W_{i,x} := (U_{i,x}, I_{i,x})$ for $i \in \mathbf{Z}, x \in \mathbf{Z}^d$.

The algorithm in the previous section immediately implies the following.

Proposition 3.1 *Letting $Z_x = \{W_{i,x}, i \in \mathbf{Z}, i < 0\}$, we have that $Z = \{Z_x\}_{x \in \mathbf{Z}^d}$ is an i.i.d. process and that the Markov random field $X = \{X_x\}_{x \in \mathbf{Z}^d}$ with distribution μ arises as a finitary coding of Z.*

(Of course, finitary was not defined in the case where a process is not finite-valued but it is still clear what the proposition means.)

The point now which remains to be resolved before we have a proof of Theorem 1.3, is that $\{Z_x\}_{x \in \mathbf{Z}^d}$ is not a finite-valued process. The idea is now to modify the process $\{Z_x\}_{x \in \mathbf{Z}^d}$ and the simulation algorithm so that we can obtain a finitary coding from a *finite-valued* i.i.d. process to μ.

This argument is essentially the same as that carried out in [2]. We therefore only explain intuitively why this should be possible.

The fast (exponential) convergence of the algorithm easily yields the fact that for each $x \in \mathbf{Z}^d$ the *expected* number of $i \in \{\cdots, -2, -1\}$ for which $W_{i,x}$ is actually used to simulate X (or μ) is finite. If we let M be any integer larger than this expectation, N_1 (N_2) be the number of $i \in \{-M, \cdots, -2, -1\}$ ($i \in \{\cdots, -M-1\}$) for which $W_{i,x}$ is used to simulate X (or μ) is finite, then we have $E[M - N_1] > E[N_2]$ which means that the expected number of unused $W_{i,x}$'s in $\{-M, \cdots, -2, -1\}$ is larger than the expected number of needed $W_{i,x}$'s in $\{\cdots, -M-1\}$. Therefore, it seems reasonable that if we only had $\{W_{i,x}\}_{i \in \{-M, \cdots, -2, -1\}, x \in \mathbf{Z}^d}$ available to us, this might be enough for us to simulate X since if, for some x, we need a $W_{i,x}$ with $i < -M$, we can transport some unused $W_{i',y}$ from elsewhere with $i' \in \{-M, \cdots, -2, -1\}$ and $y \in \mathbf{Z}^d$. Of course, this procedure must be done in a translation-invariant finitary manner.

At this stage, one can carry over the proof in [2] exactly and so in order to save space, we just refer to the somewhat lengthy argument there.

4 Relaxing the assumptions

A fair amount of what we have done can be carried out in somewhat greater generality. For example, we assumed throughout that the random fields that we dealt with were Markov, meaning that the conditional distribution of the state at a given location given everything else only depended on the values of its (usual) neighbors. One can instead study finite range Gibbs measures where the above conditional distribution is required to depend only on the values of locations which are at most r away where r is some fixed number. Everything we have done can easily be extended to this case, under appropriate modifications of Condition HN. However, extending to infinite range Gibbs measures (which is an interesting area in itself) appears to be a more difficult matter.

One could also deal with specifications which are not shift invariant. In this case, provided that the system is “high noise uniformly everywhere”, one can carry out the algorithm similarly as we did in the shift invariant case. This may be useful e.g. for simulating certain high temperature spin glasses (see e.g. [14]). On the other hand, since ergodic theory and the notion of finitary coding only deal with stationary processes, it does not make much sense to say anything in this direction for the non-shift invariant case.

Another direction is to try to replace Condition HN by weaker conditions, either in general or in special cases. The Ising model results in [2] show that Condition HN is far from being a necessary condition for either the simulation algorithm or the finitary coding result, and some of the computer simulations in [7] are also an indication in this direction.

References

[1] van den Berg, J. and Maes, C. (1994) Disagreement percolation in the study of Markov fields, *Ann. Probab.* **22**, 749–763.

[2] van den Berg, J. and Steif, J.E. (1997) On the existence and non-existence of finitary codings for a class of random fields, preprint.

[3] Burton, R. and Steif, J.E. (1995) Quite weak Bernoulli with exponential rate and percolation for random fields, *Stoch. Proc. Appl.* **58**, 35–55.

[4] Dobrushin, R.L. (1968) The description of a random field by means of conditional probabilities and conditions of its regularity, *Theory Probab. Appl.* **13**, 197–224.

[5] Georgii, H.-O. (1988) *Gibbs Measures and Phase Transitions*, de Gruyter, New York.

[6] Georgii, H.-O., Häggström, O. and Maes, C. (1998) The random geometry of equilibrium phases, preprint.

[7] Häggström, O. and Nelander, K. (1998) On exact simulation of Markov random fields using coupling from the past, *Scand. J. Statist.*, to appear.

[8] Kindermann, R. and Snell J.L. (1980) *Markov Random Fields and their Applications*, American Mathematical Society, Providence, Rhode Island.

[9] Liberto, F. di, Gallavotti, G., and Russo, L., (1973) Markov processes, Bernoulli schemes, and Ising model, *Commun. Math. Phys.* **33**, 259–282.

[10] Martinelli, F. and Olivieri, E., (1994) Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Commun. Math. Phys.* **161**, 447–486.

[11] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equations of state calculations by fast computing machine, *J. Chem. Phys.* **21**, 1087–1091.

[12] Møller, J. (1998) Perfect simulation of conditionally specified models, *J. Royal Statist. Soc. B*, to appear.

[13] Murdoch, D.J. and Green, P.J. (1998) Exact sampling from a continuous state space, *Scand. J. Statist.*, to appear.

[14] Newman, C.M. (1997) *Topics in Disordered Systems*, Birkhäuser, Basel.

[15] Propp, J. and Wilson, D. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures Algorithms* **9**, 223–252.

[16] Steif, J.E. (1991) $\overline{d}$–convergence to equilibrium and space–time Bernoullicity for spin systems in the $M < \epsilon$ case, *Ergodic Theory Dynam. Systems* **11**, 547–575.

[17] Wilson, D. (1998) Annotated bibliography of perfectly random sampling with Markov chains, *Microsurveys in Discrete Probability* (D. Aldous and J. Propp, eds), pp 209–220, American Mathematical Society. See also
`http://dimacs.rutgers.edu/~dbwilson/exact/` .

Department of Mathematics
Chalmers University of Technology
S-412 96 Gothenburg
Sweden
olle@math.chalmers.se , steif@math.chalmers.se