A Primer on Perfect Simulation

Elke Thönnes
(Mathematical Statistics, Chalmers University of Technology)

Abstract

Markov Chain Monte Carlo has long become a very useful, established tool in statistical physics and spatial statistics. Recent years have seen the development of a new and exciting generation of Markov Chain Monte Carlo methods: perfect simulation algorithms. In contrast to conventional Markov Chain Monte Carlo, perfect simulation produces samples which are guaranteed to have the exact equilibrium distribution. In the following we provide an example-based introduction into perfect simulation focussed on the method called Coupling From The Past.

1 Introduction

A model that is sufficiently realistic and flexible often leads to a distribution over a high-dimensional or even infinite-dimensional space. Examples for such complex distributions include Markov random fields in statistical physics and Markov point processes in stochastic geometry. For many of these complex distributions direct sampling is not feasible. However, there is a very useful tool which may produce (approximate) samples, Markov Chain Monte Carlo (MCMC).

MCMC methods base the sampling of a distribution on a Markov chain. An ergodic Markov chain whose equilibrium distribution is the target distribution is sampled after it has run for a long time. There are many standard methods, like the Metropolis-Hastings algorithm or the Gibbs Sampler, see Gilks et al. (1996), which allow the construction of Markov chains whose distribution, under regularity conditions, converges to the target distribution. A notoriously difficult problem however remains: when has the chain run for long enough to be sufficiently close to equilibrium? The MCMC literature refers to the initial time the Markov chain is run until it is assumed to be close enough to stationarity as the *burn-in period*.

In the last years a new variant of MCMC methods have been developed, so-called *perfect simulation algorithms*. These are algorithms which automatically ensure that the Markov chain is only sampled after equilibrium has been reached. Thus they produce samples which are guaranteed to have the target distribution and solve the problem of choosing an adequate burn-in period.

The aim of this paper is to give the reader a detailed introduction to the ideas of perfect simulation. We concentrate on one particular method called *Coupling From The Past* (CFTP) and its extensions. This algorithm was developed by Propp and Wilson (1996) and, at the moment, is the more widely used method. However, we would like to point out that there is an alternative general perfect

simulation method, Fill's perfect rejection sampling algorithm, see Fill (1998a). In contrast to Coupling From The Past, this method is interruptible, that is the state sampled by the algorithm and its runtime are independent. Thus the algorithm is also known as Fill's interruptible algorithm. The interested reader is referred to Fill (1998a), Fill et al. (1999) as well as Murdoch and Rosenthal (1998) for a presentation of the method. Further applications and extensions of Fill's algorithm may be found in Fill (1998b), Møller and Schladitz (1998) and Thönnes (1999).

In the first section of this paper we will motivate the problem of choosing a burn-in period using the example of a random walk. The next section then discusses couplings for Markov chains which are a basic tool in perfect simulation. In Section 4 we present Coupling From The Past as developed in Propp and Wilson (1996). This is followed by the discussion of two very useful extensions of the method, Dominated Coupling From The Past and perfect simulation in space.

Before we embark on our journey into the world of perfect simulation, let us introduce some assumptions which we make throughout the paper. We consider Markov chains which live on a state space E which is equipped with a separable σ-algebra $\mathcal{E}$. We assume that the Markov chain of interest is ergodic, that is irreducible, aperiodic and positive recurrent. For a general introduction into Markov chain theory the reader may consult Norris (1997) or Meyn and Tweedie (1993). Standard Markov chain theory tells us that the distribution of an ergodic Markov chain converges towards the limit distribution, see for example Meyn and Tweedie (1993). This distribution is called the equilibrium or stationary distribution and is denoted by π throughout this paper. Our aim is to produce an exact sample from the distribution π.

2 Conventional Markov Chain Monte Carlo

As a simple introductory example let us consider the following urn model which leads to a random walk on the four integers $\{0, 1, 2, 3\}$. Urn models, like for example the Ehrenfest urn model, are useful tools as they provide simple models which may describe the movements of molecules. Because of their simplicity and amenability to the method, simple random walks are also often used to introduce the ideas of perfect simulation, see Kendall (1997), Kendall and Thönnes (1999).

Example 2.1 A random walk: *Suppose we have three balls which are distributed over two urns. With probability* $1/2$ *we pick a ball from the left urn and put it into the right urn. Alternatively, we take a ball from the right urn and put it into the left urn. If we find a chosen urn empty we do nothing. What is the long-run average number* M *of balls in the right urn?*

We may describe the number of balls in the right urn as a Markov chain X whose state-flow diagram is shown in Figure 1. Suppose P denotes the transition matrix of X then, in this example, it is straightforward to compute the equilibrium distribution π by solving the linear equation system $\pi P = \pi$. Then we can determine M as the mean of π. However, let us assume that we would like to estimate M using simulation. We can do so by simulating the chain X for s steps and by estimating M as the average $\frac{1}{s}\sum_{n=1}^{s} X_n$. The chain X may be simulated by flipping a fair coin. Everytime the coin comes up heads we go

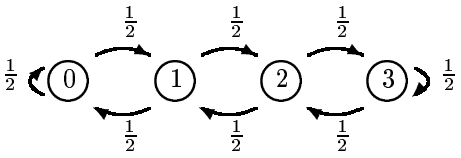

Figure 1: State-flow diagram for the Markov chain in Example 2.1.

a step upwards or if we are in state 3 we stay in state 3. If the coin comes up tails we go a step downwards or if we are in state 0 we stay in state 0.

But how do we choose the initial configuration X_0? Suppose we start in 3, that is we assume that in the beginning all 3 balls are in the right urn. Then the first samples $X_1, X_2, \ldots$ will be slightly higher than we expect in the long run. This is called the initialisation bias and it is due to choosing an initial state which is not sampled from the equilibrium distribution. Nevertheless, for any initial state the distribution of the Markov chain converges towards π. So a common procedure is to simulate the chain for a while without using the initial samples in the estimate. We choose a time $m \in \mathbb{N}$ and estimate M by $\frac{1}{s}\sum_{n=m}^{m+s} X_n$. Thus the samples we produce before time m, during the burn-in period, are ignored. But how long should we choose the burn-in period? We would like the effect of the initial state to wear off, but when can we assume this?

The choice of an appropriate burn-in period is a difficult problem which may be approached in different ways. One possibility is to try to examine analytically the convergence properties of the chain and thus to assess how fast the chain approaches equilibrium. It is usually a hard task to find bounds on the convergence rate and often the resulting bounds are not tight enough to be of any practical value. There is a vast literature on convergence rate computations and the interested reader is referred to a very incomplete selection: Roberts and Polson (1994), Rosenthal (1995) and Saloff-Coste (1999).

Another approach to determine an adequate burn-in period is to use convergence diagnostics. These are methods which observe the output of the MCMC algorithm and warn if convergence has not been reached yet. However, although these diagnostics may increase our confidence in that the Markov chain has converged, they do not guarantee convergence. For an overview on the large variety of convergence diagnostics see for example the reviews in Brooks and Roberts (1999), Cowles and Carlin (1996).

A recent development are a new variant of MCMC methods which automatically decide whether the chain has reached equilibrium. These methods have become known as *perfect simulation algorithms* and have been particularly successful for models in statistical physics and stochastic geometry. The basis of perfect simulation are couplings and the next section is devoted to a detailed introduction into the coupling method.

3 Coupling

In the last section we encountered the problem of determining the length of the burn-in period. During this burn-in we would like the effect of choosing an initial state, which was not drawn from the equilibrium distribution, to wear off.

A reasonable idea seems to start a path of the chain from each possible state and then to wait until they all produce the same results. The intuition is that then the results are no longer influenced by the starting value of the chain. To give the paths started from different initial states a chance to agree we adopt a method which is called the *coupling method*. A coupling specifies a joint distribution for given marginals. The couplings considered in our setting are of a more restrictive nature: two stochastic processes are coupled if their paths coincide after a random time, the coupling time. Couplings are an extremely useful tool in probability theory and are often used to determine convergence properties of Markov chains. For an introduction into the coupling method we refer the interested reader to the book by Lindvall (1992). As we will see at the end of this section, "forward" couplings as described here are not sufficient to produce a sample with the exact equilibrium distribution. Nevertheless, couplings are an essential tool for perfect simulation and thus we will discuss in greater detail how we may couple paths of a Markov chain which are started in different initial states.

3.1 Random walks

Let us first consider the random walk from the previous section. Here we would like to start a path of the chain in each of the possible initial states $\{0, 1, 2, 3\}$. Recall that we can use a fair coin to produce paths of the chain. This also provides us with a simple way of coupling paths from different initial states. Whenever the coin comes up heads all paths go a step upwards or stay in 3 if in 3. Alternatively, if the coin comes up tails then all paths move a step downwards or stay in 0 if in 0. Figure 2 illustrates the procedure. Each of the resulting paths behaves like a path of the random walk started in the corresponding initial state. The coupling is such that once the state of two paths coincides subsequent states of the two paths also coincide. In other words, if paths meet then they merge, we say they *coalesce*. As we continue evolving the paths they all merge eventually and we reach complete coalescence. At this point the current state of the chain is the same regardless in which state it was started. Note from Figure 2 that the paths started from the intermediate states 1 and 2 always lie between the path started in state 0 and the path started in state 3. This is due to the fact that we use a *monotone transition rule* to make the updates.

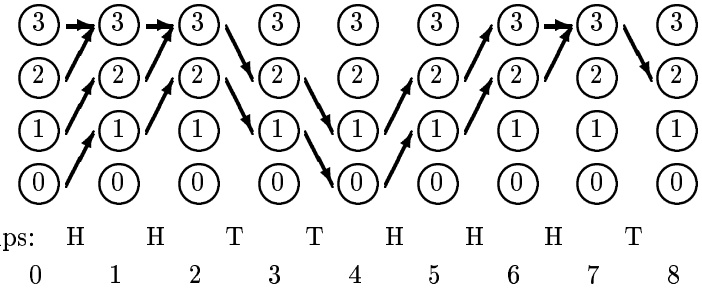

Figure 2: Coupled paths of the random walk in Example 2.1 produced by applying the same outcomes of coin flips to all paths. Note how paths coalesce as they meet. Complete coalescence is achieved after 7 steps.

A transition rule is a random map which specifies a transition for each state according to the transition kernel of the chain. If all realisations of the random map are monotone functions then we call it a monotone transition rule. In our example the transition rule is given by

$$f(n,C) \quad = \quad \begin{cases} \min(n+1,3) & \text{if } C = H \\ \max(n-1,0) & \text{if } C = T \end{cases} \qquad n \in \{0,1,2,3\} \qquad (3.1)$$

where C describes whether the coin comes up heads (H) or tails (T). The transition rule is thus a random map whose realisations are specified by the realisations of the coin toss. We achieve the coupling of paths by applying the same realisation of the random map f to all paths.

Notice that $\mathbb{P}(f(n,C) = m) = p_{n,m}$ for $n, m \in \{0,1,2,3\}$ where $p_{n,m}$ are the transition probabilities of the target chain X. Furthermore, observe that

$$\mathbb{P}\Big(f(n,C) = f(m,C)\Big) \quad \geq \quad \sum_{j=0}^{3} p_{n,j}\, p_{m,j} \qquad \text{for all } n, m \in \{0,1,2,3\}. \quad (3.2)$$

This means that at each step of our coupling the probability of two paths merging when using the transition rule f is greater or equal than the probability of two paths merging in an independent coupling. An independent coupling is achieved if we use an independent coin for each path. This does not hold for every coupling. For example consider the simple symmetric random walk on the vertices of a square as given in Figure 3. If we take a fair coin and move from each state clockwise if it comes up heads and anti-clockwise if it comes up tails, then paths started from different initial states will never meet. The perfect simulation algorithm, which is presented in the next section, assumes that we use a transition rule for which the analogue of (3.2) holds. This can always be satisfied as we can choose an independent coupling of paths. However, the speed of the algorithm is greatly increased if we choose a coupling such that paths coalesce quickly.

The realisations of the random map f in (3.1) are monotone and so f is a monotone transition rule. Thus the use of f leads to paths which maintain the initial order between the starting states. A necessary requirement for the existence of such a monotone transition rule is the stochastic monotonicity of the transition kernel of the Markov chain, for a definition see Lindvall (1992) or Stoyan (1983). Due to the monotonicity of f we can determine the time of complete coalescence simply by monitoring the path started in state 3 and the path started in state 0. Complete coalescence occurs if and only if these two paths merge and this occurs in finite time almost surely. For our random walk

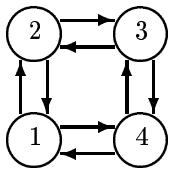

Figure 3: For this random walk we may define a coupling such that paths started in different states do not meet.

this may seem not such a big improvement. However, for many Markov chains on large state spaces the determination of complete coalescence is not practical if we have to monitor the paths from all initial states. One example for such a chain is the Gibbs Sampler for the Ising model, which will be presented in Section 3.2. But first let us discuss another urn model.

Example 3.1 Another random walk: *As before, we assume we have three balls distributed over two urns. However, in this example if we choose the left urn and it is empty, then we take a ball from the right urn and put into the left one. If we find the right urn empty we do nothing. Below is the state flow diagram of the resulting random walk. It only differs from the previous example in the type of moves the chain can make from state* 3.

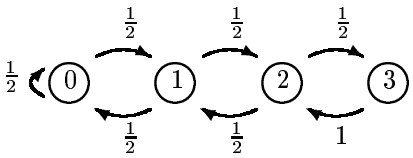

Figure 4: State-flow diagram for the Markov chain in Example 3.1.

Again we can simulate the chain by flipping a fair coin. We choose the following strategy. Everytime the coin comes up heads we remain in state 0 if we are in 0, we move a step up if we are in state 1 or 2 and we move a step down if we are in state 3. Alternatively, if the coin comes up tails, we move a step up from state 0 and we move a step down if we are in state 1, 2 or 3.

Similar to the previous example we may produce a coupling of paths by using the same realisation of a coin flip when updating the paths. Unfortunately the resultant coupling is not monotone. (This is easily verified by drawing some sample paths). However, by using a cross-over trick we may still determine the time of complete coalescence by keeping track of two paths only. This cross-over technique was first used in Kendall (1998) and is further examined for Markov random fields in Häggström and Nelander (1998).

As before, we may describe the coupling using a transition rule given by

$$f(n,C) \quad = \quad \begin{cases} n+1 & \text{if } C = H \text{ and } n \in \{1,2\} \\ n-1 & \text{if } C = T \text{ and } n \in \{1,2\} \\ 0 & \text{if } C = H \text{ and } n = 0 \\ 1 & \text{if } C = T \text{ and } n = 0 \\ 2 & \text{if } n = 3 \end{cases} \tag{3.3}$$

where C is the realisation of the coin flip. Suppose we impose the following partial order on the state space: $2 \preceq 0 \preceq 1 \preceq 3$. Then for fixed C and $n \preceq m$ the above transition rule satisfies $f(n,C) \succeq f(m,C)$. For example if $C = H$, then $0 = f(0,H) \succeq f(1,H) = 2$. Thus the transition rule is *anti-monotone*. In Figure 5 we have ordered the states according to $\preceq$ and drawn some coupled sample paths of the chain. The anti-monotonicity of our coupling can easily be seen in the figure. For example in the first update, the highest state 3 moves to the lowest state 2 and the lowest state 2 moves to the highest state 3.

The anti-monotonicity of the transition rule allows us to monitor complete coalescence by evolving two paths only. We denote the two paths by $X^{\min}$ and

$X^{\max}$. We start the minimal path in the minimal state and the maximal path in the maximal state, that is

$$X_0^{\min} = 2 \quad \text{and} \quad X_0^{\max} = 3.$$

We then evolve the two paths as follows

$$X_{k+1}^{\min} = f\Big(X_k^{\max}, C_k\Big) \qquad X_{k+1}^{\max} = f\Big(X_k^{\min}, C_k\Big),$$

where C_k is the kth coin toss. Hence the two paths evolve as a two-component Markov chain in which the update of one component is made according to the current state of the other component. The two components are not individually Markov and, as long as they differ, they do not evolve according to the transition probabilities of our random walk. However, once the two components coincide they do evolve like our random walk. Most importantly, the minimal and maximal path sandwich between them all paths of our random walk if evolved using the same coin flip realisations. Thus coalescence of the minimal and maximal path implies complete coalescence of the paths started from all initial states. In Figure 5 we have drawn the maximal and the minimal path as dotted lines. In this realisation the minimal and the maximal path $(X^{\min}, X^{\max})$ start in $(2,3)$ respectively and then evolve as $(2,3),(2,3),(2,3),(2,1),(0,1),(2,0)$ and finally coalesce after 6 steps in state 1. In the seventh step they jointly reach state 2.

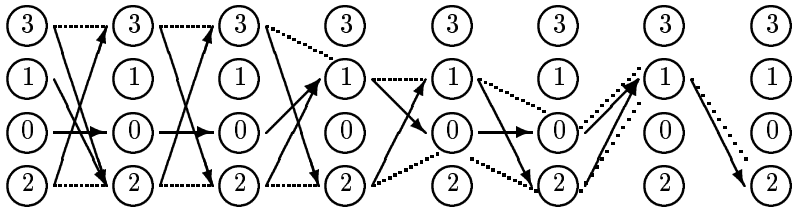

Figure 5: Coupled sample paths of the random walk in Example 3.1. Note that we reordered the initial states and that the coupling is anti-monotone with respect to the new ordering. The dotted lines show the maximal and the minimal path.

3.2 The Ising model

The next example, the *Ising model* as described for example in Winkler (1991), is taken from statistical physics and was one of the first models to be considered for perfect simulation, see Propp and Wilson (1996), Fill (1998a). The Ising model is probably the simplest form of a *Markov random field*. Markov random fields are defined on a discrete lattice Λ of sites where each site may take a value from a finite set of states S. The distribution of a Markov random field is given by the expression

$$\pi(x) = \frac{1}{Z} \exp\Big(-H(x)\Big),$$

where Z is the normalizing constant known as the *partition function* and $H(x)$ is the *energy function*. For most Markov random fields there is no closed form

expression for the partition function and therefore direct sampling of these models is not feasible. However, we may produce (approximate) samples of these models using MCMC.

Example 3.2 Ising model:
The Ising model has energy function

$$H(x) \quad = \quad -\frac{1}{KT}\Big[\, J\sum_{j\sim k} x_j x_k - Bm\sum_k x_k \,\Big]$$

where j, k are sites on a square lattice Λ. Here $j \sim k$, that is site j is a neighbour of site k, if j and k are sites at Euclidean distance one. We may imagine the Ising model as a lattice which at each site has a small dipole or spin which is directed either upwards or downwards. Thus each site j may take a value $x_j \in \{-1, +1\}$ representing a downward respectively an upward spin.

A standard method of constructing a Markov chain whose distribution converges to the target Markov random field is the *Gibbs Sampler*, see Geman and Geman (1984). The Gibbs Sampler is based on sampling from the full conditional distributions of a multi-dimensional Markov chain. For Markov random fields these full conditional distributions reduce to the local characteristics of the model. The following Gibbs Sampler, in the statistical physics literature also known as the *heat bath algorithm*, see for example Creutz (1979), produces a Markov chain whose distribution converges to the Ising model. We start by choosing some initial configuration on the sites of a finite lattice Λ. Then, step by step we go from one site to the next and update its spin. At site n we assign an upward spin with probability

$$\mathbb{P}\Big(x_n = +1 \;\Big|\; x_{-n}\Big) \quad = \quad \frac{\pi(x_n = +1,\ x_{-n})}{\pi(x_n = +1,\ x_{-n}) + \pi(x_n = -1,\ x_{-n})}.$$

Here x_{-n} denotes the configuration x on Λ excluding the site n and so the above is the conditional probability of the Ising model having an upwards spin at site n given the current spin configuration on all other sites. The reader may verify that

$$\mathbb{P}\Big(x_n = +1 \;\Big|\; x_{-n}\Big) \quad = \quad \Big(1 + \exp\Big(-\frac{2}{KT}\Big[\, mB + J\sum_{j\sim n} x_j\Big]\Big)\Big)^{-1},$$

which, notably, does not depend on the partition function Z.

More specifically, at each step k we independently draw a random number U_k which is uniform on the interval $(0, 1)$ and a random number N_k which is uniform on the lattice Λ. We then assign an upward spin to the site N_k if

$$U_k \quad \leq \quad \mathbb{P}\Big(x_{N_k} = +1 \;\Big|\; x_{-N_k}\Big),$$

otherwise, we assign a downward spin.

As the lattice Λ is finite we need to specify how we treat sites which are on the boundary of Λ. One possibility is to impose periodic boundary conditions (also called the torus condition). Here the lattice is mapped onto a torus by identifying opposite boundaries. However, edge-effects may occur, that is the

sample we draw may not behave exactly like a finite lattice sample of an Ising model defined on an infinite grid. In Section 6 discuss how these edge-effects may be avoided but for now let us assume the torus condition.

We can couple paths of the Gibbs Sampler started from different initial states by reusing the sampled random variates N_k and $U_k, k \in \mathbb{N}$. At time k we update the same site N_k in each path using the same realisation of U_k for all paths. As in the previous example we may describe our updating procedure using a transition rule. We set

$$f(x, U, N) \;=\; \begin{cases} \{x_N = +1, x_{-N}\} & \text{if } U \le \mathbb{P}(x_N = +1 \mid x_{-N}) \\ \{x_N = -1, x_{-N}\} & \text{otherwise} \end{cases} \tag{3.4}$$

where $\{x_N = +1, x_{-N}\}$ is the configuration which we obtain by setting $x_N = +1$ and leaving the spins of all other sites in x unchanged.

One problem we encounter is that the set of all initial states of the Ising model is usually very large. Thus it may be prohibitively expensive to monitor all paths. However, as for the random walk examples, we may exploit the monotonicity or anti-monotonicity of the transition rule to determine efficiently the time of complete coalescence. Let us have a closer look at the update rule which we are using. We assign an upward spin to site N if

$$U \;\le\; \mathbb{P}\Big(x_N = +1 \;\Big|\; x_{-N}\Big) \;=\; \Big(1 + \exp\Big(-\frac{2}{KT}\Big[mB + J\sum_{j\sim N} x_j\Big]\Big)\Big)^{-1}.$$

First consider the case when $J > 0$, that is the ferromagnetic Ising model. Then the probability $\mathbb{P}(x_N = +1 \mid x_{-N})$ is the greater the more neighbours of N have an upward spin. We may exploit this fact by equipping the state space of the Ising model with an appropriate partial order $\preceq$. We say the spin configuration x is smaller than y, that is $x \preceq y$ if

$$x_j \quad \le \quad y_j \qquad \text{for all } j \in \Lambda.$$

This partial order, which was used in Propp and Wilson (1996), may seem counter-intuitive from a physical point of view as a larger state may not necessarily have smaller energy. However, we do not attempt to attach any physical meaning but simply define a partial order for which our transition rule is monotone. Figure 6 shows a triple of configurations which are ordered with respect to $\preceq$.

Now, if $x \preceq y$ then

$$\mathbb{P}\Big(x_N = +1 \;\Big|\; x_{-N}\Big) \;\le\; \mathbb{P}\Big(y_N = +1 \;\Big|\; y_{-N}\Big) \quad \text{for any } N \in \Lambda.$$

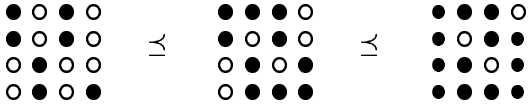

Figure 6: Three spin configuration on a 4×4 lattice. Upward spins are represented as black sites and downward spins as white sites. We ordered the three configurations with respect to $\preceq$.

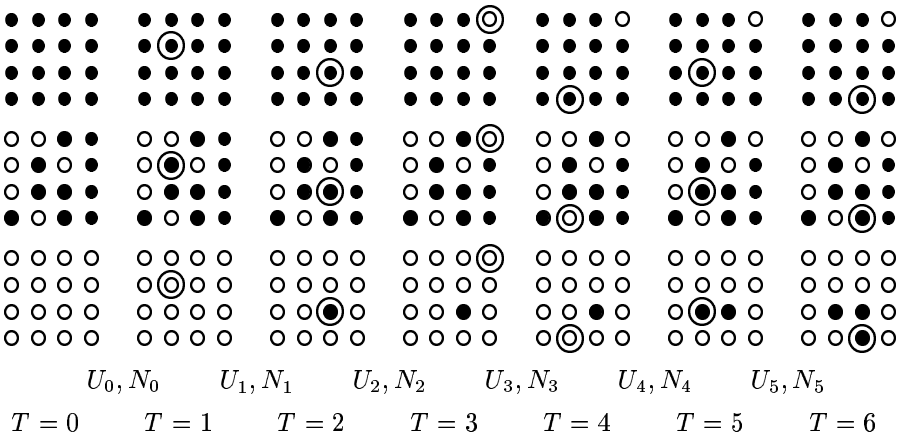

Figure 7: Coupled paths of the Gibbs Sampler for the ferromagnetic Ising model on a 4×4 lattice. Black sites have an upward spin, white sites a downward spin. The site that has just been updated is encircled. The uppermost path is started in $x_{\max}$, the lowermost path in $x_{\min}$. Observe how the path started from the intermediate configuration is sandwiched between the two paths started in the maximal and minimal state.

It follows that, for fixed U and N, whenever f assigns an upwards spin to x_N then it also assigns an upwards spin to y_N as

$$U \quad \leq \quad \mathbb{P}\Big(x_N = +1 \,\Big|\, x_{-N}\Big) \quad \leq \quad \mathbb{P}\Big(y_N = +1 \,\Big|\, y_{-N}\Big).$$

Analogously, for fixed U and N, whenever f assigns a downward spin to y_N then it assigns a downward spin to x_N. Thus if $x \preceq y$ then the updated configurations maintain their partial ordering, that is $f(x, U, N) \preceq f(y, U, N)$ for fixed U and N. It follows that our transition rule is monotone with respect to $\preceq$ and so the partial ordering between paths is preserved. The state space has a maximal state $x_{\max}$ with respect to $\preceq$ which is the configuration consisting of upward spins only. Similarly, the minimal state $x_{\min}$ is given by the configuration consisting of downward spins only. Due to the monotonicity of the transition rule f the paths are coupled in such a way that all paths lie between the path started in the maximal state and the path started in the minimal state, see also Figure 7. Complete coalescence occurs if and only if these two paths coalesce which will occur in almost surely finite time.

Now let us discuss the anti-ferromagnetic case, that is if $J < 0$. Careful inspection of the transition rule leads to the observation that f is anti-monotone because for $x \preceq y$

$$\mathbb{P}\Big(x_N = +1 \,\Big|\, x_{-N}\Big) \quad \geq \quad \mathbb{P}\Big(y_N = +1 \,\Big|\, y_{-N}\Big).$$

Moreover, if we start two paths in two states which are comparable with respect to $\preceq$ then after some updates the states of the two paths may no longer be comparable. However, we can still monitor complete coalescence by monitoring a minimal and a maximal path which are evolved according to a cross-over. We start a path $X^{\max}$ in $x_{\max}$ and another path $X^{\min}$ in $x_{\min}$. We then update the

two paths according to

$$X_{k+1}^{\min} \quad = \quad f\Big(X_k^{\max}, U_k, N_k\Big) \quad \text{and} \quad X_{k+1}^{\max} \quad = \quad f\Big(X_k^{\min}, U_k, N_k\Big).$$

This leads to a maximal path $X^{\max}$ and a minimal path $X^{\min}$ which sandwich between them the paths which are started from all initial states and evolved according to the transition rule f and same realisations of U_k and N_k, $k \in \mathbb{N}$. Thus we may determine complete coalescence by monitoring whether $X^{\max}$ and $X^{\min}$ coalesce. This can be shown to occur in almost surely finite time. Figure 8 illustrates the coupling and cross-over procedure.

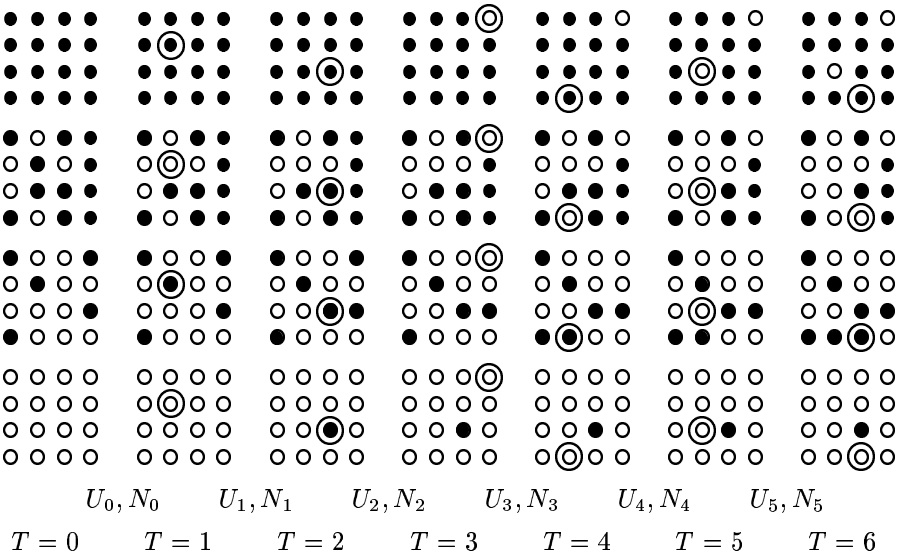

Figure 8: Coupled paths of the Gibbs Sampler for the anti-ferromagnetic Ising model on a 4×4 lattice. The uppermost and lowermost path are the maximal and minimal path respectively and evolved according to a cross-over. The two intermediate paths in the middle are evolved according to the standard coupling construction. Note that although these two paths are comparable at time $T = 0$, they are no longer comparable at time $T = 6$. However they are still comparable to the states of the minimal and the maximal path.

3.3 Immigration-death process

Our final example is an immigration-death process on the natural numbers and thus a Markov chain on an infinite state space.

Example 3.3 Immigration-Death process:
Consider the number of dust particles contained in a given small volume. If there are N particles in the system, then new particles enter at a rate $\lambda\ (N+1)/(N+2)$, where λ is some positive constant. Thus the immigration rate increases the more particles are already in the system. Particles stay in the system for an exponential amount of time with unit mean.

The number of particles N in the given volume is an immigration-death process with transition rates

$$\begin{aligned} N &\to N+1 \qquad \text{at rate } \lambda\,(N+1)/(N+2), \\ N &\to N-1 \qquad \text{at rate } N \qquad\quad \text{for } N \in \mathbb{N}_0\,. \end{aligned} \tag{3.5}$$

We may simulate N as follows. We start by simulating an immigration-death process D with transition rates

$$\begin{aligned} D &\to D+1 \qquad \text{at rate } \lambda \\ D &\to D-1 \qquad \text{at rate } D \qquad \text{for } D \in \mathbb{N}_0\,. \end{aligned} \tag{3.6}$$

Observe that D and N have the same death rate, but the immigration rate of D is larger than for N. In mathematical terms, D stochastically dominates N. A description on how to simulate a constant rate immigration-death process may be found in Ripley (1987).

Given a realisation of D we may derive a realisation of the process N. As the initial configuration N_0 at time 0 we choose a number from $\{0, 1, \ldots, D_0\}$. Now, whenever a particle arrives in D say at time t, it enters the given volume with probability $(N_{t-}+1)/(N_{t-}+2)$ where N_{t-} is the number of particles in the system immediately before the arrival of the new particle. We may achieve this by marking every immigration time t of D with an independent random variable U_t which is uniform on the unit interval. The new particle enters the system if and only if

$$U_t \quad \leq \quad \frac{N_{t-}+1}{N_{t-}+2}.$$

Finally consider a death in D at time t which leads to a death in N at time t with probability N_{t-}/D_{t-}. Again we may achieve this by marking any death time t in D with a mark U_t which is uniform on the unit interval. A death occurs in N at time t if

$$U_t \quad \leq \quad \frac{N_{t-}}{D_{t-}}.$$

In the following whenever we speak of the dominating process D we implicitly mean the process D and its jump time marks. By coupling N to D we not only define a simulation procedure, but, as we will see in Section 5, we also make the process N amenable to a perfect simulation algorithm. Perfect simulation of birth-and-death processes on the natural numbers like the one above was first described in Kendall (1997).

The coupling procedure may be illustrated using a Hasse diagram as in Figure 9. It consists of a sequence of horizontal levels which stand for the states $\{0, 1, 2, \ldots\}$. On each level we have arrows representing the jump times of D. An arrow pointing upwards indicates an immigration time and an arrow pointing downwards a death time. The arrow corresponding to the jump time t is marked with U_t. Now, for each level we may delete arrows according to the rules described above. For example, on level 3 corresponding to state 3 we delete any upwards arrow whose associated mark U_t exceeds $(N_{t-}+1)/(N_{t-}+2) = 4/5$. We delete any downwards arrow whose associated mark U_t exceeds $N_{t-}/D_{t-} = 3/D_{t-}$. The process N started in some state $j \in \{0, \ldots, D_0\}$ may now be constructed as follows. We start on level j and move from left to right. Whenever we come across an upward arrow we go a level upwards. Alternatively,

whenever we come across a downward arrow then we go a level downwards, see Figure 9.

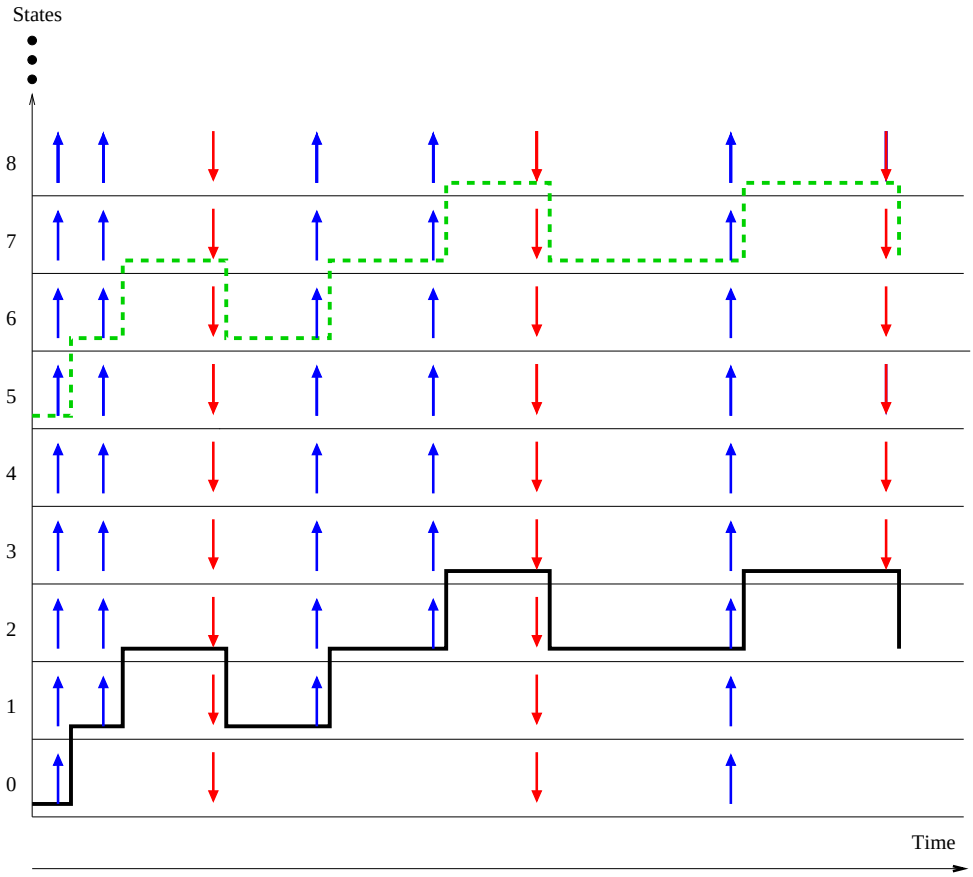

Figure 9: The Hasse diagram for the immigration-death process in Example 3.3. The dashed line shows D which was started at time 0 in state 5. Some of the jump times were deleted according to our decision rules. Notice that the more jump times of D are deleted the lower the level. The solid line shows a path of N started in state 0 and evolved coupled to D.

For the immigration-death process in Example 3.3 we may produce a coupling by using the same realisation of D and the associated jump time marks $U_t, t \geq 0$, and applying the above procedure to all paths started in a state $j \in \{0, \ldots, D_0\}$. From Figure 9 we may see that two paths can never cross each other but can only meet and then merge. Thus the paths maintain the partial ordering of their initial states and so the coupling is monotone. It follows that complete coalescence of paths started from every state in $\{0, \ldots, D_0\}$ occurs if and only if the path started in state 0 and the path started in state D_0 coalesce. The state space of the target process N are the natural numbers $\mathbb{N}_0$ which, of course, is much larger than the finite set $\{0, \ldots, D_0\}$. However, as we will see in Section 5, to produce a perfect sample we only need complete coalescence for the set bounded above by D. Complete coalescence occurs in almost surely finite time as a sufficient event is that D hits zero, in which case all relevant

paths of N also hit zero.

3.4 Forward Coupling and exact sampling

We started this section with the motivation of finding a time when the effect of the initial state of the chain has worn off. We may argue now that this has happened when complete coalescence occurs as in this case the chain is in the same state regardless of its starting value. Let S be the time of complete coalescence which is a random stopping time. We may think that X_S has the equilibrium distribution, however, this intuition is flawed! Even if the chain had been started in equilibrium we cannot conclude that X_S has the equilibrium distribution. This is due to the fact that S is not a fixed but a random time. As an illustration let us consider again our random walk from Example 2.1. From Figure 2 we can see that at the time of complete coalescence, that is at time S, the chain is necessarily either in state 0 or in state 3. Clearly, this is not a sample of the equilibrium distribution which is uniform on the integers $\{0, 1, 2, 3\}$.

Fortunately, a rather simple but very effective modification enables us to sample X in equilibrium. This modification is called Coupling From The Past and is due to Propp and Wilson (1996).

4 Coupling From The Past

The last section showed how to couple paths of a Markov chain from different initial states such that after a random time S, the time of complete coalescence, all paths have merged into one. Although the state of any of these paths at time S does not depend on its starting value we have seen that sampling the chain at time S may give a biased sample which is caused by the fact that S is not a fixed time but a random stopping time. In the following we discuss an alternative approach called *Coupling From The Past* (CFTP) which was introduced by Propp and Wilson (1996). It is also based on coupling and complete coalescence but samples the chain at a fixed time, namely time 0.

Recall from the previous section that we produced coupled sample paths of the chain X started from every initial state by sampling transition rules. At each time step $k \in \mathbb{N}$ we independently sampled a transition rule f_k and produced a sample path starting in $x \in E$ by setting

$$X_k(x) \quad = \quad f_k(X_{k-1}(x)) \quad = \quad f_k \circ f_{k-1} \circ \cdots \circ f_1(x).$$

For instance, for our random walk in Example 2.1 we sampled an independent coin C_{k-1} and set

$$X_k(x) \quad = \quad f(X_{k-1}(x), C_{k-1})$$

where f is defined as in (3.1). We have also seen in the previous section that an adequate choice of transition rules eventually leads to complete coalescence of the sample paths. Or, in other words, the image of the composite map defined as

$$F_{0,k} \quad = \quad f_k \circ f_{k-1} \circ \cdots \circ f_1 \tag{4.7}$$

eventually becomes a singleton as k approaches infinity. Let S be the time of complete coalescence then, unfortunately, the unique image of $F_{0,S}$ does not have the equilibrium distribution in general.

In 1995, Propp and Wilson made a simple but ingenious observation: if we *reverse* the order in which we compose the sampled transition rules and continue sampling until the image of the composite map becomes a singleton then this unique image has the equilibrium distribution!

Let us look at this in more detail. At time k we sample the transition rule f_k but now we define a composite map by

$$\tilde{F}_{0,k} \quad = \quad f_1 \circ \cdots \circ f_k.$$

Thus we have reversed the order of composition compared to (4.7). The above is equivalent to the following procedure. We go backwards in time and sample at time $-k$ the transition rule f_{-k} and now define a composite map by

$$F_{-k,0} \quad = \quad f_{-1} \circ \cdots \circ f_{-k} \qquad k \in \mathbb{N}. \tag{4.8}$$

A nice illustration of the difference between forward coupling as in (4.7) and backward coupling as in (4.8) on the example of Matheron's dead leaves model can be found in Kendall and Thönnes (1999). It is also illustrated in an animated simulation on `http://www.warwick.ac.uk/statsdept/Staff/WSK/dead.html`.

How can we interpret the composite map in (4.8)? If we set

$$X_j^{-k}(x) \quad = \quad f_{-j-1} \circ \cdots \circ f_{-k}(x)$$

then $\{X_j^{-k}(x), -k \leq j \leq 0\}$ behaves like a path of X started at time $-k$ in state x. Thus $F_{-k,0}(x)$ is the state at time 0 of a path of X started at time $-k$ in state x. It follows that $F_{-k,0}$ has a singleton image if and only if the corresponding coupled paths of X started at time $-k$ in all initial states achieve complete coalescence by time 0. Thus the above procedure produces coupled paths started in all initial states at earlier and earlier times until they achieve complete coalescence by time 0. The time when all paths coalesce is not necessarily time 0 but can occur earlier. Nevertheless, we only ever sample at the fixed time 0. If complete coalescence is achieved then the common state at time 0 is an exact sample from π.

We may describe the procedure using the pseudo-code notation from computer science. Suppose we have an algorithm `RandomMap(`$-k$`)` which samples the transition rule f_{-k}. Then we can describe the CFTP algorithm as follows:

```
CFTP:
   k ← 0
   F_0 ← identity map
   Repeat
      k ← k − 1
      f ← RandomMap(k)
      F_k ← F_{k+1} ∘ f
      until image of F_k is a singleton
   return image of F_k
```

We will first give a heuristical argument which provides an intuitive explanation why this method produces an exact sample. This is followed by some examples. A rigorous proof for the correctness of the procedure is given at the end of the section.

Let $-T$ be the first time when the image of $F_{-k,0}$ becomes a singleton, that is $T = \min\{k : F_{-k,0} \text{ has a singleton image}\}$. Suppose we could start the chain X at time $-\infty$ and run it up to time $-T$. As the chain is ergodic and has run for an infinite amount of time, the heuristic suggests that the chain is in equilibrium at time $-T$. Suppose now that the value of the infinite time simulation at time $-T$ is x, then x is a sample from π. The transition rules $f_{-k}, k \in \mathbb{N}$ describe transitions according to the transition kernel P of X. As $\pi P = \pi$ the transition rules preserve the equilibrium distribution and so it follows that $F_{-T,0}(x)$ is also a sample from π. Of course, we do not know x but, as the image of the composite map is unique, it does not matter which value x the infinite time chain takes at time $-T$. Hence, if the image of the composite map becomes a singleton then we may deduce the state of the infinite time simulation at time 0. Thus, by extending backwards in time until the image of the composite map becomes a singleton, we reconstruct the path of the infinite time simulation in the recent past. In other words, we create a "virtual simulation from time $-\infty$".

Let us now apply the above algorithm to our examples. For the random walk in Example 2.1 the procedure runs as follows. Recall that we may simulate the random walk using fair coin flips. We go step by step backwards in time and perform the following routine.

1. At time $-k, k \in \mathbb{N}$, we independently flip a fair coin C_{-k}.
2. We then start a path from all initial states $\{0, 1, 2, 3\}$ and evolve them from time $-k$ till time 0 according to the coin flips $C_{-k}, C_{-k+1}, \ldots, C_{-1}$ (note the order of the coins!).
3. If all paths coalesce at time 0 then we return their common state as a sample from π.

If the paths do not coalesce, then we go a step further backwards in time and repeat the above steps. Thus we independently flip another biased coin C_{-k-1}, and again evolve the paths started in all initial states from time $-k-1$ to time 0 using the coin flips $C_{-k-1}, C_{-k}, \ldots, C_{-1}$. We continue going successively further backwards in time until we finally reach complete coalescence at time 0.

Remark 4.1 *It is essential that in the kth iteration the coins are used in the order $C_{-k}, C_{-k+1}, \ldots C_{-1}$ and that we reuse all previously sampled coin flips in the appropriate order. Only then is the sample guaranteed to have the equilibrium distribution.*

We can make the above procedure more efficient by noting the following. It is not necessary to evolve paths started in *all* initial states till time 0. As discussed in the previous section, due to the monotonicity of the transition rule, complete coalescence occurs if and only if the path started in state 0 and the path started in state 3 coalesce. Thus we only need to monitor these two paths for coalescence.

Neither is it necessary to check for coalescence at each time step. Recall that complete coalescence of paths started from time $-T$ means that the map $F_{-T,0}$ has a singleton image. But then the composite map $F_{-T-1,0} = F_{-T,0} \circ f_{-T-1}$ has exactly the same singleton image and, by induction, so has any $F_{-S,0}$ with $S > T$. Thus we may proceed as follows. Let $0 = T_0 < T_1 < T_2 \ldots$ be an increasing sequence of time points. Then for $k = 1, 2, \ldots$, we perform the following steps until we reach complete coalescence:

1. Sample independent coin flips $C_{-T_k}, C_{-T_k+1}, \ldots, C_{-T_{k-1}-1}$.

2. Evolve one path started in state 0 and one started in state 3 from time $-T_k$ to time 0 using the coin flips $C_{-T_k}, C_{-T_k+1}, \ldots, C_{-1}$.

3. Check for coalescence at time 0.

A recommended choice for the sequence of time points is $T_k = 2^{k-1}$, which in Propp and Wilson (1996) is shown to be close to optimal.

For simulation purposes we do not need to store all coin toss realisations. Instead we can reproduce them by resetting the seed of a seeded pseudo-random number generator.

Figure 10 illustrates the CFTP algorithm for the random walk from Example 2.1.

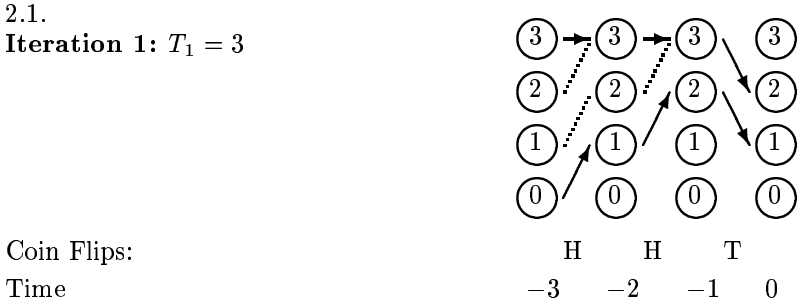

Figure 10: CFTP for Example 2.1. The paths started in state 0 and in state 3 are shown as solid lines. The dotted lines are the paths started from intermediate states. However we do not need to monitor these to determine complete coalescence. Note how the coin toss realisations of the previous iteration are reused! Complete coalescence occurs at time -1, however we continue till time 0 and sample state 2.

For the random walk in Example 3.1 we may proceed in a similar fashion.

As discussed earlier, in this setting we may use a cross-over to detect complete coalescence of all paths by monitoring two paths only, a minimal and a maximal path. At time $-T_k$ we start a path in the minimal state 2 and a path in the maximal state 3. (Recall that we chose a partial ordering $\preceq$ which differs from the natural partial ordering on the integers.) We then evolve the minimal and the maximal path by updating the minimal path according to the current configuration of the maximal path and vice versa. If coalescence of these two paths occurs by time 0, then their common state at time 0 has the equilibrium distribution.

For the Ising model we may proceed as follows. In iteration k we sample independent random variables $U_{-T_k}, U_{-T_k+1}, \ldots, U_{-T_{k-1}-1}$ which are uniform on the unit interval and random variables $N_{-T_k}, N_{-T_k+1}, \ldots, N_{-T_{k-1}-1}$ which are uniform on the lattice Λ. In the ferromagnetic case, we start a path in $x_{\max}$, that is the configuration consisting of only upwards spins, and a path in $x_{\min}$, that is the configuration with only downward spins. The two paths are evolved from time $-T_k$ to time 0 using the transition rule f as defined in (3.4) together with the realisations of $U_{-T_k}, U_{-T_k+1}, \ldots U_{-1}$ and $N_{-T_k}, N_{-T_k+1}, \ldots N_{-1}$. If the two paths coalesce then we output their common state at time 0 as a sample from the equilibrium distribution. If coalescence has not been achieved yet then we extend further backwards in time.

In the anti-ferromagnetic case we adopt the following procedure. In iteration k we also sample independent random variables $U_{-T_k}, U_{-T_k+1}, \ldots, U_{-T_{k-1}-1}$ and $N_{-T_k}, N_{-T_k+1}, \ldots N_{-T_{k-1}-1}$. We again start two paths at time $-T_k$, one in $x_{\max}$ and one in $x_{\min}$. However, we now evolve the two paths according to a cross-over as described in the previous section. As before, if coalescence of these two paths occurs then we may deduce complete coalescence of the paths started in all initial values. The unique state at time 0 is then a perfect sample. If coalescence has not occurred yet then we extend further backwards in time.

Remark 4.2 *The heat bath algorithm is known to mix slowly for temperatures close to criticality. To produce samples of the Ising model close to the critical temperature Propp and Wilson (1996) apply Coupling From The Past to a Gibbs Sampler for the random cluster model. (This type of Gibbs Sampler is also known as single bond heat bath.) By assigning random colours to the obtained clusters a realisation of a random cluster model is turned into a realisation of an Ising or Potts model. Propp and Wilson (1996) produced samples of the Ising model at critical temperature on a* 512×512 *toroidal grid in about 20 seconds on a Sparcstation. CFTP needed to go back only to about time -30 to produce such a sample. The authors also show how to produce samples from the Ising model simultaneously for a range of temperature values.*

We now give a rigorous proof for the correctness of the CFTP algorithm for finite state spaces, see also Propp and Wilson (1996).

Theorem 4.3 *Let X be an ergodic Markov chain with transition matrix P and stationary distribution π. Coupling From The Past as presented above produces an exact sample of the target equilibrium distribution π.*

Proof : For $k \in \mathbb{N}$ consider the composite map

$$F_{-k,-j} \quad = \quad f_{-j-1} \circ \cdots \circ f_{-k} \qquad \text{where } j \leq k.$$

We will first show that the image of $F_{-k,0}$ almost surely becomes a singleton as k approaches infinity. We then proceed to show that this unique image has the distribution π.

Let $z \in E$ be an arbitrary state in the state space of X. As the chain X is irreducible and aperiodic there is a finite $N > 0$ such that $P^N(y, z) > 0$ for all $y \in E$. Here P^N denotes the $N-$step transition matrix of X, that is $P^N(y, z) = \mathbb{P}(X_N = z | X_0 = y)$. It follows the existence of a constant $\epsilon > 0$ such that for any $k \in \mathbb{N}$ we have

$$\mathbb{P}\Big(\text{image of } F_{-kN,-(k-1)N} \text{ is a singleton}\Big) > \epsilon.$$

Now, the events

$$\Big\{\text{image of } F_{-kN,-(k-1)N} \text{ is a singleton}\Big\}, \qquad k \in \mathbb{N}$$

are independent and have a probability of at least ϵ. Thus by Borel-Cantelli

$$S \quad = \quad \min\Big\{k \in \mathbb{N} : \text{ image of } F_{-kN,-(k-1)N} \text{ is a singleton}\Big\}$$

is almost surely finite. But then the composite map

$$F_{-SN,0} \quad = \quad F_{-SN,-(S-1)N} \circ F_{-(S-1)N,0}$$

has also a unique image.

Like the "forwards" composite map

$$F_{0,k} \quad = \quad f_k \circ \cdots \circ f_1$$

the map $F_{-k,0}$ is composed of k independent transition rules and thus $X_0^{-k}(x) = F_{-k,0}(x)$ has the same distribution as $X_k(x) = F_{0,k}(x)$. Now, $X_k(x)$ for $k \in \mathbb{N}_0$ is a path of X started at time 0 in x and so, due to the ergodicity of X, the distribution of $X_k(x)$ converges to π as k approaches infinity. As $X_0^{-k}(x)$ has the same distribution as $X_k(x)$ it follows that its distribution also converges to π. Moreover, by the definition of X_0^{-k} we have

$$\lim_{k\to\infty} X_0^{-k}(x) \quad = \quad \lim_{k\to\infty} F_{-k,0}(x) \quad = \quad \text{unique image of } F_{-SN,0}$$

where S is defined as above. As $X_0^{-k}(x)$ tends in distribution to π it follows that the unique image of $F_{-SN,0}$ must also have the equilibrium distribution π.

□

5 Dominated Coupling From The Past

In this section we discuss an extension of the original CFTP algorithm as in Propp and Wilson (1996). This extension is called *Dominated Coupling From The Past*, or Coupling Into And From The Past, and is due to Kendall (1997), see also Kendall (1998). Suppose $T \geq 0$ is the smallest random time such that coupled paths of the target chain started in *all* initial states at time $-T$ have coalesced by time 0. Foss and Tweedie (1998) showed that T is almost surely finite if and only if the chain is uniformly ergodic. Thus CFTP as described in the previous section only applies to uniformly ergodic Markov chains. However, many Markov chains of interest, in particular chains which converge to point process distributions, are not uniformly, but only geometrically ergodic. Fortunately, Dominated Coupling From The Past may enable us to sample the stationary distribution of these Markov chains.

Dominated CFTP essentially specifies a time-evolving bounded random set $\Theta_t, t \in \mathbb{R}$, such that

1. there exists an almost surely finite time T such that paths started at time $-T$ from all initial states in Θ_{-T} coalesce at time 0,

2. if coalescence as in 1. occurs then the unique state at time 0 has equilibrium distribution.

Heuristically we may think of Θ_t as a random set which provides a stochastically varying upper and lower bound on the values at time t of an infinite time simulation. This is best explained using an example, so consider our immigration-death process from Example 3.3. In Section 3 we showed how to couple paths of the process started from different initial states. We have seen that this coupling is monotone with respect to the natural ordering on the integers. The state space has a minimal state with respect to this partial order, the state 0, but it does not have a maximal state. However, due to our coupling construction, we do know that the process N is bounded above at any time by the process D. Thus, although we do not have a fixed bound on N_t for any $t \in \mathbb{R}$, we do have a random bound given by D_t and we may set $\Theta_t = \{0, \ldots, D_t\}$.

How can we exploit this in a CFTP-type algorithm? Let us first use the heuristic of an infinite time simulation to provide the intuition; a formal proof follows later. Consider an infinite time simulation of the target Markov chain N started at time $-\infty$. We denote the infinite time simulation by $N^{-\infty}$. Our aim is to reconstruct the path of $N^{-\infty}$ in the recent past. Clearly, $N^{-\infty}$ is bounded below by 0. But how about an upper bound? If we assume that the infinite time simulation $N^{-\infty}$ is started in state 0 and coupled to an infinite time simulation of D then $N^{-\infty}$ is bounded above by $D^{-\infty}$. So our first task is to reconstruct the path of $D^{-\infty}$ in a finite interval $[-T, 0]$. A little thought shows that this is easily done. If D was started at time $-\infty$ then heuristics suggest that it is in equilibrium at time $-T$. Now, the stationary distribution of D is a Poisson distribution which is easy to sample. Thus, if we start D at time $-T$ in its equilibrium and simulate it till time 0 then we may interpret this realisation as the path of $D^{-\infty}$ on $[-T, 0]$.

Because $N^{-\infty}$ is coupled to and thus bounded above by D, we may deduce that the path of $N^{-\infty}$ on $[-T, 0]$ lies below the given realisation of D on the same time interval. In particular we have that $N_{-T}^{-\infty} \leq D_{-T}$. Suppose that the paths of N started at time $-T$ from all initial states in $\{0, 1, \ldots, D_{-T}\}$ and coupled to the realisation of D on $[-T, 0]$ coalesce by time 0. Then, according to our heuristic, their common state at time 0 is also the state of the infinite time simulation $N^{-\infty}$ at time 0. Therefore this state is a sample from the equilibrium distribution. Recall that the coalescence of the path of N started in state 0 and the path of N started in state D_{-T} implies the coalescence of all paths started in $\{0, \ldots, D_{-T}\}$. Thus it is sufficient to monitor only these two paths for coalescence.

If the paths started from $\{0, 1, \ldots, D_{-T}\}$ do not coalesce by time 0, then we need to extend backwards in time and repeat the above procedure. To do this we need to extend the realisation of D on $[-T, 0]$ backwards in time, that is we need to produce a realisation of D on $[-T - S, 0]$ which coincides with the previous realisation on $[-T, 0]$. We can do this by exploiting the time-reversibility of D. We start D at time 0 in equilibrium and simulate it up to time T. Then we set

$$\widetilde{D}_{-t} \quad = \quad D_t \qquad \text{for } t \in [0, T]$$

that is we reverse the path of D in time. This produces a path of D on $[-T, 0]$. If we extend backwards then we just continue our simulation of D from time T to time $T + S$ and again reverse the resulting path in time.

Here is a pseudo-code description of the algorithm, which is also illustrated in Figure 11. The algorithm `Extend`$(D, -T)$ extends a given path of D to a path on $[-T, 0]$ and assigns marks to any new jump times. The algorithm `Evolve`$(D, -T)$ then starts path of N in 0 and one in D_{-T} and evolves them coupled to the path D till time 0.

`DominatedCFTP:`
  $T \leftarrow 0$
  $D \leftarrow \emptyset$
  `Repeat`
    $T \leftarrow T - 1$
    $D \leftarrow$ `Extend`$(D, -T)$
    $N_0 \leftarrow$ `Evolve`$(D, -T)$
    `until` N_0 `is a singleton`
  `return` Y_0

What are the characteristics of the above procedure? Firstly, we started with a time-homogeneous process D which stochastically dominated N. This process D had a standard stationary distribution which is easy to sample. Furthermore, we made use of the fact that D was time-reversible.

Paths of the target process N were derived as an adapted functional of D. The coupling between D and N ensured that if we started N at time $t \in \mathbb{R}$ in a configuration bounded above by D then the path of N was bounded by the path of D at any later time.

We determined coalescence by starting a path of N at time $-T$ in D_{-T} and one in state 0. Let the "upper" path started at time $-T$ in D_{-T} be denoted by U^{-T} and the "lower" path started at time $-T$ in state 0 by L^{-T}. The two processes U and L have the following properties, some of which can also been seen in Figure 11.

1. Conditional on a realisation of D on $[-T - S, 0]$ we have a *funneling* property, that is

$$L_t^{-T} \leq L_t^{-T-S} \leq U_t^{-T-S} \leq U_t^{-T} \quad \text{for } t \in [-T, 0]. \tag{5.9}$$

   Thus the earlier we start U and L the closer the two paths get.

2. If $L_t^{-T} = U_t^{-T}$ for $t \in [-T, 0]$ then $L_u^{-T} = U_u^{-T}$ for $u \in [t, 0]$. Hence once the upper and lower path coalesce they remain coalesced. We call this the *coalescence* property.

3. Suppose we start a path of N at time $-T$ in some $j \in \{0, \ldots, D_{-T}\}$ and evolve it coupled to the same realisation D as the lower and upper path, then we have the following *sandwiching property:*

$$L_t^{-T} \quad \leq \quad N_t^{-T} \quad \leq \quad U_t^{-T} \qquad \text{for } t \in [-T, 0]. \tag{5.10}$$

   Therefore a path of N started at time $-T$ in some state bounded above by D_{-T} and evolved according to D lies between the upper and lower path started at time $-T$ and coupled to the same realisation of D. Together with the funneling property it follows that any path of N started at time $-T - S$ in a state bounded above by D_{-T-S} lies on $[-T, 0]$ between the

lower and upper path started at time $-T$ and evolved according to the same realisation of D.

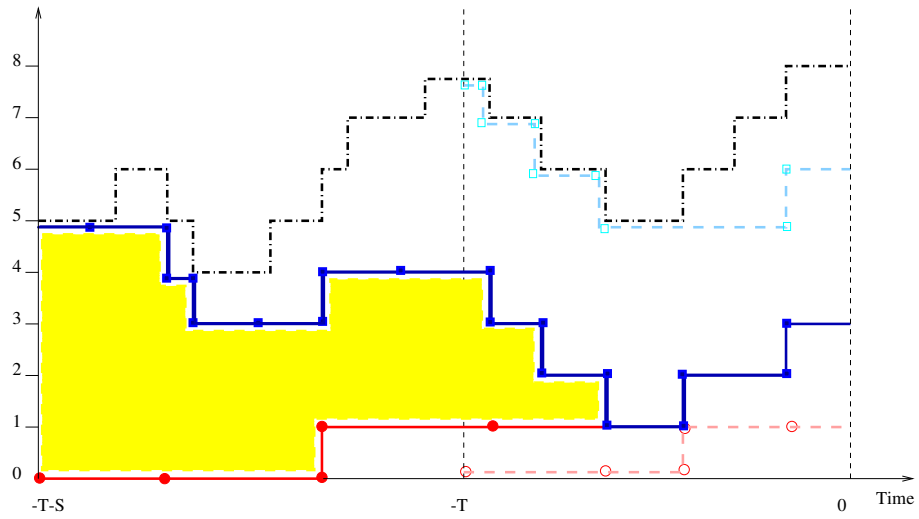

Figure 11: The immigration-death process example 3.3. The dashed-dotted line is the immigration-death process D. The solid line marked with squares shows the maximal path and the solid line marked with disks the minimal path started at time $-T-S$. The two paths have coalesced by time 0. The shaded area is the area in which any path of N started at time $-T-S$ in some state smaller than D_{-T-S} lies. The dashed lines show an earlier CFTP iteration started from time $-T$ in which coalescence of the lower and upper path did not occur. Note how the process D has been extended backwards in time. Observe also that the lower process lies below the upper process and how both processes satisfy the funneling property.

We now prove rigorously that our Dominated CFTP algorithm does in fact sample the desired equilibrium distribution. The proof is a special case of the proof in Kendall and Møller (1999) for general Dominated CFTP algorithms.

Theorem 5.1 *Suppose U and L are defined as above and satisfy the funneling, coalescence and sandwiching properties. If*

$$T_C \quad = \quad \inf\left\{T \geq 0 : \; U_0^{-T} = L_0^{-T}\right\}$$

is almost surely finite then $U_0^{-T_C}$ has the distribution π.

Proof : As T_C is almost surely finite the funneling property implies that the limit $\lim_{T\to\infty} U_0^{-T}$ exists and that

$$\lim_{T\to\infty} U_0^{-T} \quad = \quad \lim_{T\to\infty} L_0^{-T} \quad = \quad U_0^{-T_C}.$$

Now, let N_0^{-T} be the state at time 0 of a path of N started at time $-T$ in state 0. Then the distribution of N_0^{-T} is the same as of N_T, that is a path of N

started at time 0 in state 0 and run up to time T. Due to the ergodicity of N, it follows that the distribution of N_0^{-T} converges to π as $T \to \infty$. The limit of N_0^{-T} may be interpreted as our infinite time simulation. The sandwiching property ensures that

$$\lim_{T\to\infty} L_0^{-T} \quad \leq \quad \lim_{T\to\infty} N_0^{-T} \quad \leq \quad \lim_{T\to\infty} U_0^{-T}$$

and so

$$\lim_{T\to\infty} N_0^{-T} \quad = \quad U_0^{-T_C}$$

which implies that $U_0^{-T_C}$ has the equilibrium distribution π. □

Dominated Coupling From The Past was originally developed for locally stable Markov point processes, see Kendall (1997), Kendall (1998), Kendall and Møller (1999). Markov point processes are usually specified by a density π with respect to a unit rate homogeneous Poisson point process on a bounded window W. For example the Strauss process, which is described in Stoyan and Stoyan (1994) and which models repulsive point patterns, is given by

$$\pi(x) \quad = \quad \alpha\, \beta^{n(x)}\, \gamma^{t(x)} \qquad x \subset W,$$

where $\beta > 0$ and $0 < \gamma < 1$. Here $n(x)$ counts the number of points in x and $t(x)$ the number of pairs of neighbour points, that is points which are less than the interaction range R apart. Like for many other point process models, the normalizing constant α cannot be computed in closed form. From the density of a Markov point process we may derive its Papangelou conditional intensity

$$\lambda(x,\xi) \quad = \quad \begin{cases} \frac{\pi(x\cup\{\xi\})}{\pi(x)} & \text{if } \pi(x) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $x \subset W$ is a point pattern and $\xi \in W$ an individual point. For example, for the Strauss process the Papangelou conditional intensity is given by

$$\lambda(x,\xi) \quad = \quad \beta\, \gamma^{t(x,\xi)},$$

where $t(x,\xi)$ counts the number of neighbours of ξ in x. More information on Papangelou conditional intensities may be found in Daley and Vere-Jones (1988). A Markov point process that is locally stable has a Papangelou conditional intensity which is uniformly bounded above by some constant λ^*. For example, for the Strauss model $\lambda^* = \beta$.

Spatial birth-death processes may converge in distribution to point processes. These birth-and-death processes are Markov jump processes whose states are point patterns and which evolve in time through births and deaths of individual points. For an introduction see Stoyan et al. (1995) or Møller (1999a). A spatial birth-death process is specified by its birth rate and its death rate. If we choose a unit death rate and a birth rate which is equal to the Papangelou conditional intensity of a Markov point process then, under regularity conditions specified for example in Møller (1999a), the resultant spatial birth-death process Y converges to the distribution of the point process.

Similar to the immigration-death process example, we can produce exact samples for locally stable Markov point processes using Dominated CFTP. Suppose the ergodic spatial birth-and-death process Y converges to the distribution

of such a Markov point process with density π. Note that we may derive a realisation of Y from a realisation of a spatial birth-and-death process Z with the same death rate and a higher birth rate. In our setting we may choose Z to have unit death rate and birth rate λ^*. Then Z is time-reversible and has a Poisson point process as its equilibrium. We mark every birth time t of Z with a mark V_t which is uniform on $(0,1)$. We then can derive a path of Y from a path of Z as follows. A birth of a point ξ in Z at time t leads to the birth of the same point at the same time in Y if

$$V_t \quad \leq \quad \lambda(Y_{t-}, \xi)/\lambda^*.$$

Thus the acceptance rule for births is very similar to our acceptance rule for births in the immigration-death process example. The acceptance rule for deaths differs from the procedure for the immigration-death process. For spatial birth-and-death processes we can distinguish the individual elements of a configuration. Thus we may adopt the following simple procedure. Whenever a point η dies in Z we check whether this point exists in Y and if so, let it die at the same time in Y. The reader may verify that this coupling construction leads to the correct birth and death rate for Y.

The above coupling is very similar to the coupling we chose for our immigration-death process. Careful inspection leads to the observation that Y_t is always a subset of Z_t if we start it in a configuration which is a subset of Z at the starting time. We can use the set-up for a Dominated CFTP algorithm as follows. We produce a stationary path of Z on $[-T, 0]$ and mark all birth times. Then we start a path of Y at time $-T$ in every point pattern which is a subset of Z_{-T}. We evolve the paths according to the above coupling till time 0 and check for complete coalescence. If all paths have coalesced at time 0 their common state has the distribution π. If they have not coalesced, then we need to extend backwards in a similar manner as for the immigration-death process.

We can detect complete coalescence more efficiently if $\lambda(x, \xi)$ is monotone, that is $\lambda(x, \xi) \leq \lambda(y, \xi)$ if $x \subseteq y$. Then we only need to monitor a path started at time $-T$ in the empty set and a path started at time $-T$ in the point pattern Z_{-T}. If the two paths coalesce by time 0 then their common state at time 0 is a perfect sample. If $\lambda(x, \xi)$ is anti-monotone, as for example for the Strauss process, then we may use a cross-over to monitor coalescence efficiently.

For a more detailed introduction into the perfect simulation of locally stable Markov point processes see Kendall and Møller (1999). Further examples may found in Kendall (1997) and Kendall (1998). The method is extended to random set processes in Kendall and Thönnes (1999). It may also be applied to general distributions, see for example Mira et al. (1998), Møller (1999b) or Murdoch and Green (1998).

6 Perfection in Space

When introducing the Ising model we promised the reader a simulation method which avoids edge-effects. A perfect sample in space may be achieved by extending not only backwards in time but also in space. This idea was first presented in Kendall (1997) and for Markov random fields is discussed in more detail in Häggström and Steif (1999) or in van den Berg and Steif (1999). We explain the method using the example of a ferromagnetic Ising model.

Suppose we would like to produce a perfect sample on the $m \times m$ lattice $\Lambda = \Lambda_0$. For $k \in \mathbb{N}$ let $\Lambda_{-k-1} = \Lambda_{-k} \cup \partial(\Lambda_{-k})$ be the lattice we achieve by adding the neighbours of the boundary sites of Λ_{-k} to the lattice.

In the kth iteration of the CFTP algorithm we now perform the following procedure.

1. Sample independent random variables $U_{-T_k}, U_{-T_k+1}, \ldots, U_{-T_{k-1}-1}$ which are uniform on the unit interval and random variables N_{-T_k}, N_{-T_k+1}, $\ldots, N_{-T_{k-1}-1}$ which are uniform on the lattice Λ_{-k+1}.

2. Start one path of the Gibbs Sampler in $x_{\max}$ and one in $x_{\min}$ on the lattice Λ_{-k}. Evolve the paths from time $-T_k$ to time 0 using the transition rule h as defined in (3.4) together with the realisations of $U_{-T_k}, U_{-T_k+1}, \ldots U_{-1}$ and $N_{-T_k}, N_{-T_k+1}, \ldots N_{-1}$.

If the two paths coalesce at time 0 on Λ_0, then we output their common state at time 0 as a sample from the equilibrium distribution. If coalescence has not been achieved yet we extend further backwards in time and space. Figure 12 further illustrates the procedure. The reader is invited to compare this algorithm to the standard CFTP procedure on page 18.

The algorithm is set up such that from time $-T_j$ until time $-T_{j-1} - 1$ only sites on the interior of the lattice Λ_{-j} are updated but taking into account the configuration on the boundary sites.

If the two paths coalesce at time 0 on Λ then their common state is not only independent from the starting configuration on Λ but also independent from any starting configuration on the infinite lattice $\mathbb{Z}^2$. For a formal proof see Häggström and Steif (1999). Of course, it can happen that we do not achieve coalescence in finite time. This crucially depends on the strength of interaction between sites. In Häggström and Steif (1999) a bound on the strength of interaction is given such that coalescence in almost surely finite time is ensured.

Many refinements of the algorithm presented here are possible. If we use a cross-over then we may also apply the method to the anti-ferromagnetic Ising model. An application to general Markov random fields can be found in Häggström and Steif (1999), van den Berg and Steif (1999). Furthermore, the point process setting is discussed in Kendall (1997).

7 Conclusion

Our aim was to introduce the reader to the concept of Coupling From The Past, one of the new perfect simulation methods which allow the exact sampling of the equilibrium distribution of Markov chains.

We have seen that CFTP is based on couplings of paths of the target Markov chain started in different initial states such that after an almost surely finite time the paths coincide. Crucial for the practicality of CFTP algorithms is a way of determining complete coalescence in an efficient manner. We have shown how monotonicity or anti-monotonicity of the chosen coupling may provide us with a practical procedure of checking for complete coalescence. Many models in spatial statistics and statistical physics are amenable to the method as modelling spatial interaction may lead to (anti-)monotonicity of a Markov chain converging to these distributions. (Anti-)Monotonicity is with respect to a partial order

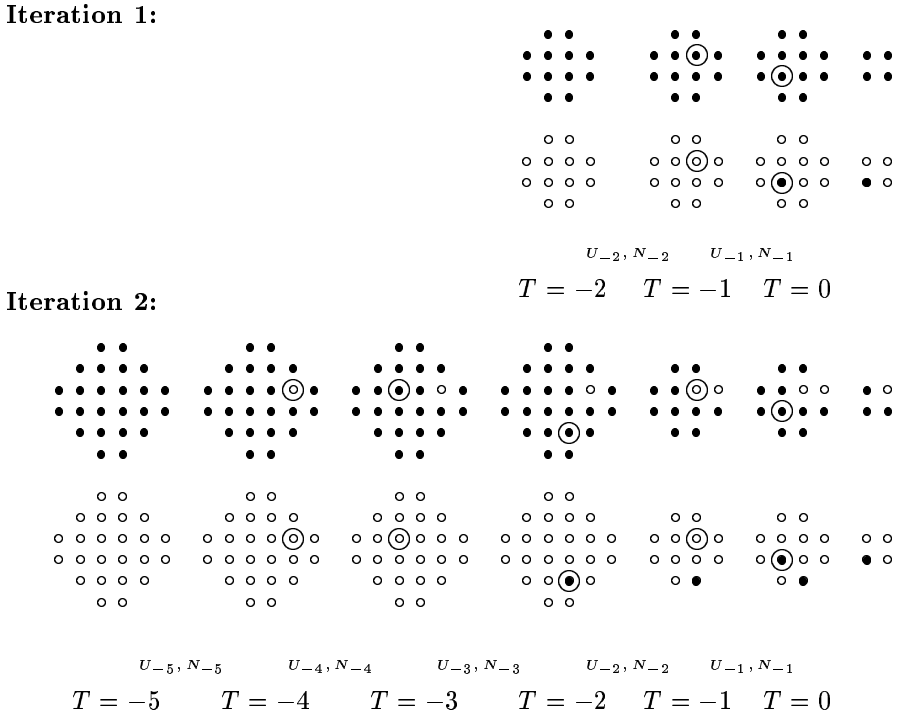

Figure 12: Two iterations of the "perfect in space" CFTP algorithm for the ferromagnetic Ising model on a 4×4 lattice. Sites with upward spins are black and sites with downward spins white. The upper path is started in $x_{\max}$ and the lower in $x_{\min}$. The encircled site has just been updated. Only sites in the interior of each lattice are updated, taking into account the configuration on the boundary sites. We need to achieve coalescence on the centre 4×4 lattice, which for clarity we have drawn at the end of each path. In the second iteration, we reduce the site of the lattice after time $T = -2$ because from then onwards only sites of the centre 4×4 lattice are updated.

and, as was illustrated in the examples, we can choose any partial ordering for our purposes. However, there are also CFTP algorithms which determine coalescence efficiently without exploiting (anti-)monotonicity, see for example Green and Murdoch (1999), Murdoch and Green (1998). They are usually based on a coupling where paths started in all initial states coalesce very quickly into a small number of paths which can then be monitored efficiently.

If the target chain is uniformly ergodic, then we may proceed as in Section 4. If not, then we have to use Dominated CFTP as in Section 5. In the (anti-)monotone setting, we define an upper and a lower path such that if they coalesce at time 0, their common state is an exact sample from the equilibrium distribution. Note that the lower and upper path do not need to evolve like the target chain if they have not coalesced yet. Before coalescence they may even live on an augmented state space, for an example see Kendall and Thönnes (1999). However, once they coalesce, they behave like the target chain.

All in all, there is a lot of freedom in setting up a CFTP algorithm. The

challenge is to construct an algorithm which is efficient enough to be practical. An annotated bibliography which contains a multitude of examples of perfect simulation algorithms may be found on the perfect simulation website `www.dimacs.rutgers.edu/~dbwilson/exact`.

Acknowledgement:
The author would like to acknowledge the support by the EU TMR network ERB-FMRX-CT96-0095 on "Computational and statistical methods for the analysis of spatial data".

References

S.P. Brooks and G.O. Roberts. Assessing convergence of Markov Chain Monte Carlo algorithms. *Statist. Comput.*, 8:319 – 335, 1999.

M.K. Cowles and B.P. Carlin. Markov chain convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.*, 91:883–904, 1996.

M. Creutz. Confinement and critical dimensionality of space-time. *Phys. Rev. Lett.*, 43:553 – 556, 1979.

D.J. Daley and D. Vere-Jones. *Introduction to the Theory of Point Processes.* Springer, New York, 1988.

J. A. Fill. An interruptible algorithm for exact sampling via Markov chains. *Ann. Appl. Probab.*, 8:131–162, 1998a.

J. A. Fill, M. Machida, D. J. Murdoch, and J.S. Rosenthal. Extension of Fill's perfect rejection sampling algorithm to general chains, 1999. Preprint available on `www.mts.jhu.edu/~fill/`.

J.A. Fill. The Move-To-Front rule: A case study for two perfect sampling algorithms. *Probab. Engrg. Inform. Sci.*, 12:283–302, 1998b.

S.G. Foss and R.L. Tweedie. Perfect simulation and backward coupling. *Stochastic Models*, 14:187–203, 1998.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. PAMI*, 6:721–741, 1984.

W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo In Practice.* Chapman & Hall, London, 1996.

P.J. Green and D.J. Murdoch. Exact sampling for Bayesian inference: towards general purpose algorithms. In J. M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 6*, pages 301 – 321, Oxford, 1999. Oxford University Press.

O. Häggström and K. Nelander. Exact sampling from anti-monotone systems. *Statist. Neerlandica*, 52:360 – 380, 1998.

O. Häggström and J.E. Steif. Propp-Wilson algorithms and finitary codings for high noise random fields, 1999. To appear in *Combin. Probab. Comput.*

W.S. Kendall. Perfect simulation for spatial point processes. In *Proc. ISI* 51^{st} *session, Istanbul (August 1997)*, volume 3, pages 163–166, 1997.

W.S. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C.C. Heyde, editors, *Probability Towards 2000*, pages 218–234, New York, 1998. Springer.

W.S. Kendall and J. Møller. Perfect Metropolis-Hastings simulation of locally stable point processes. Research report 347, University of Warwick, 1999.

W.S. Kendall and E. Thönnes. Perfect simulation in stochastic geometry. *Pattern Recognition*, 32:1569 – 1586, 1999.

T. Lindvall. *Lectures On The Coupling Method.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1992.

S.P. Meyn and R.L. Tweedie. *Markov Chains And Stochastic Stability.* Springer Verlag, New York, 1993.

A. Mira, J. Møller, and G. O. Roberts. Perfect slice samplers, 1998. Preprint, available on `www.dimacs.rutgers.edu/~dbwilson/exact`.

J. Møller. Markov Chain Monte Carlo and spatial point processes. In W.S. Kendall O.E. Barndorff-Nielsen and M.N.M. van Lieshout, editors, *Stochastic Geometry: Likelihood And Computation*, Boca Raton, 1999a. Proceedings Séminaire Européen de Statistique, Chapman & Hall/CRC. 141-172.

J. Møller and K. Schladitz. Extensions of Fill's algorithm for perfect simulation, 1998. Preprint, to appear in *J. Roy. Statist. Soc.* **B**.

Jesper Møller. Perfect simulation of conditionally specified models. *J. Roy. Statist. Soc.*, B 61:251 – 264, 1999b.

D.J Murdoch and P.J. Green. Exact sampling from a continuous state space. *Scand. J. Statist.*, 25:483 – 502, 1998.

D.J. Murdoch and J.S. Rosenthal. An extension of Fill's exact sampling algorithm to non-monotone chains, 1998. Preprint.

J.R. Norris. *Markov Chains.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.

J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms*, 9:223–252, 1996.

B.D. Ripley. *Stochastic Simulation.* John Wiley & Sons, New York, 1987.

G.O. Roberts and N.G. Polson. On the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc.*, B 56:377–384, 1994.

J.S Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 90:558–566, 1995.

L. Saloff-Coste. Simple examples of the use of Nash inequalities for finite Markov chains. In O. Barndorff-Nielsen, W.S. Kendall, and M.N.M van Lieshout, editors, *Stochastic Geometry: Likelihood And Computation.* Chapman & Hall/CRC, Boca Raton, 1999. 365 - 400.

D. Stoyan. *Comparison Methods For Queues And Other Stochastic Models.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Chichester, 1983.

D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry And Its Applications.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Chichester, second edition, 1995.

D. Stoyan and H. Stoyan. *Fractals, Random Shapes And Point Fields.* John Wiley & Sons, Chichester, 1994.

E. Thönnes. Perfect simulation of some point processes for the impatient user. *Adv. Appl. Probab.*, 31:69 – 87, 1999.

J. van den Berg and J.E. Steif. On the existence and non-existence of finitary codings for a class of random fields, 1999. To appear in *Ann. Probab.*

G. Winkler. *Image Analysis, Random Fields and Monte Carlo Methods.* Applications of Mathematics. Springer, Berlin, 1991.