Bounding the convergence time of the Gibbs sampler in Bayesian image restoration

By Alison L. Gibbs
Department of Mathematics and Statistics, York University
Toronto, Ontario, Canada M3J 1P3
agibbs@mathstat.yorku.ca

Summary

This paper shows how coupling methodology can be used to give precise, a priori bounds on the convergence time of Markov chain Monte Carlo algorithms for which a partial order exists on the state space which is preserved by the Markov chain transitions. This methodology is applied to give a bound on the convergence time of the random scan Gibbs sampler used in the Bayesian restoration of an image of N pixels. For our algorithm, in which only one pixel is updated at each iteration, the bound is a constant times N^2. The proportionality constant is given and is easily calculated. These bounds also give an indication of the running time of coupling from the past algorithms.

Some key words: Bayesian image restoration; Convergence; Coupling from the past; Gibbs sampler; Ising model; Markov chain Monte Carlo; Total variation distance.

1 Introduction

Markov chain Monte Carlo algorithms, first used in statistical physics and later used in the statistics community for problems in spatial statistics, including image processing, are now widely used, particularly in Bayesian analysis, for exploring complicated probability distributions; see for example Gelfand & Smith (1990), Besag & Green (1993), Smith & Roberts (1993), Besag et al. (1995) and Gilks, Richardson & Spiegelhalter (1996). An important issue in the implementation of Markov chain Monte Carlo algorithms is whether or not they actually converge to the distribution of interest and, if so, how quickly. For a discussion of these issues see for example Tierney (1994) and Roberts & Rosenthal (1998). Convergence diagnostics (Cowles & Carlin, 1996; Brooks & Roberts, 1997) have been developed to monitor convergence of the algorithm while it is running, but none is completely satisfactory (Cowles, Roberts & Rosenthal, 1997). There has been much work on developing rigorous, a priori, quantitative bounds on the convergence time; see for example Sinclair & Jerrum (1989), Diaconis & Stroock (1991), Frieze, Kannon & Polson (1994), Ingrassia (1994), Meyn & Tweedie (1994), Rosenthal (1995), Mengersen & Tweedie (1996), Polson (1996) and Frigessi, Martinelli & Stander (1997). Application of many of these results is difficult in practice except to the simplest of problems, and the constant of proportionality is not always available.

In this paper, we show how coupling methodology can be used to give precise, a priori bounds on the convergence time of Markov chain Monte Carlo algorithms. We consider monotone Markov chains, for which a partial order exists on the state space which is preserved by the Markov chain transitions. In particular, we develop convergence time bounds for a simplified problem in Bayesian image restoration which involves sampling from a Gibbs distribution using the Gibbs sampler. The case of image synthesis, where observed data are absent, is equivalent to what is referred to in the mathematical physics literature as Glauber dynamics for the stochastic Ising model.

Using coupling and martingale techniques, for an image of N pixels we are able to obtain precise upper bounds on the convergence time in total variation distance which are equal to an easily computable constant times N^2 for the random scan version of the Gibbs sampler, where each iteration involves the update of only one randomly chosen pixel. While we believe that similar arguments will lead to a similar bound on the convergence time for the systematic scan algorithm, the fact that the values of neighbouring pixels change at each iteration makes analysis of the systematic scan algorithm more difficult. The general methodology outlined in § 3.2 can be applied to any monotone Markov chain Monte Carlo algorithm.

In the mathematical physics literature, it is well known that the convergence rate for the stochastic Ising model is $O(N \log N)$ for appropriate values of the parameter which include those for which our results hold (Frigessi et al., 1997). However the constant of proportionality for these results is not known. In a future paper, we will show how some of the calculations used in this paper can be applied to give precise bounds that are $O(N \log N)$.

Our results are presented as follows. The model and Gibbs sampler algorithm are described in § 2. The coupling methodology used to derive our bounds is described in § 3.2. The application of these bounds to the running time of Propp & Wilson's (1996) coupling from the past algorithm is discussed in § 6. Results for sampling from the Ising model without data and from the posterior distribution with data are presented in § 4 and § 5.

2 Image Restoration using the Gibbs Sampler

We consider the Bayesian restoration of images where the prior consists of a probability model for the true image and the posterior is formed from the prior conditional on the data, which in our cases are the values of the observed image. These observed data are obtained from the true image through a known random process; see Geman & Geman (1984), Besag (1986) and Green (1996). The Gibbs sampler is used to produce samples from the posterior distribution. We also consider the use of the Gibbs sampler for simulation of the prior distribution since this is of interest on its own.

Our model of the image is a Markov random field X of pixels each taking the value $+1$ or -1, with the value of each pixel affected by its nearest neighbours in an attractive manner. Equivalently,

$$\mathrm{pr}(X = x) = \frac{1}{Z}\exp\left\{-U(x)\right\}, \tag{1}$$

where $x = (x_1, \ldots, x_N)$ is a configuration of the colours at the N pixels, the energy function U reflects the neighbourhood structure within which configurations with pixels having like neighbours are favoured, and Z is the normalising constant, called the partition function in mathematical physics. The particular prior probability model we place on the configuration is the Ising model for which

$$U(x) = -\beta \sum_{<i,j>} x_i x_j, \tag{2}$$

where the sum is taken over pairs of sites (i, j) which are nearest neighbours, and β is a positive parameter. For a discussion of the physical significance of the Ising model, see Cipra (1987). The presence of data results in a posterior distribution that is equivalent to a Gibbs distribution with the presence of an external field. With the Ising model prior, the posterior distribution of x given the data, y, is of the form

$$\pi_P(x|y) = \frac{1}{Z_P}\exp\left\{\beta \sum_{<i,j>} x_i x_j + \sum_i f(x_i, y_i)\right\} \tag{3}$$

where Z_P is the normalising constant, and the function f depends on the random distortion mechanism.

Our data are an observed distortion of the true image. We consider two distortion mechanisms. In § 5.1 we consider y to be obtained from the true image by, with a constant probability, independently switching the sign of each pixel, and in § 5.2 we consider the case where independent, normal noise is added to the value of each pixel. Examples of the form of the function f from (3) are available in equations (16) and (17).

Even for the simple models studied here, it is impractical to examine both the prior and posterior by calculating the probability of each configuration, because of the large configuration space. The Gibbs sampler is used to produce a sample from the distribution of interest. At each iteration one randomly chosen pixel is updated according to its conditional distribution given the value of all of the other pixels.

For our problem, it is easy to calculate and to sample from these conditional distributions. For the case of no data where the distribution of interest is the Ising model (1) and (2), the full conditionals are

$$\pi(x_i|x_{-i}) = \frac{e^{\beta \sum_{j\in\partial i} x_i x_j}}{e^{\beta \sum_{j\in\partial i} x_j} + e^{-\beta \sum_{j\in\partial i} x_j}}, \tag{4}$$

where x_{-i} denotes the configuration of all pixels except the ith and ∂i is the set of pixels that are neighbours of pixel i. For the case with data, the full conditionals are

$$\pi_P(x_i|x_{-i}, y) = \frac{e^{\beta \sum_{j\in\partial i} x_i x_j + f(x_i, y_i)}}{e^{\beta \sum_{j\in\partial i} x_j + f(1, y_i)} + e^{-\beta \sum_{j\in\partial i} x_j + f(-1, y_i)}}. \tag{5}$$

The iterations continue until the current configuration can be considered to be a sample from the posterior distribution, independent of the initial configuration. We are concerned with the number of iterations required.

The Markov chain whose state space is the space of all possible configurations and whose transition probabilities are $\frac{1}{N}$ times the full conditional probability, with transitions only possible between configurations which differ at only one site, is an irreducible, aperiodic Markov chain with stationary distribution π.

3 Bounding the Convergence Time of the Algorithm

3.1 Introduction

Convergence is measured by the total variation distance D. For a Markov chain with probability transition matrix P, stationary distribution π, countable state space Ω and initial configuration $x^0 \in \Omega$, the total variation distance at time t is

$$\begin{aligned} D_{x^0}(t) &= \sup_{A \in \Omega} \left| P^t(x^0, A) - \pi(A) \right| \qquad (6) \\ &= \frac{1}{2} \sum_{x \in \Omega} \left| P^t(x^0, x) - \pi(x) \right|, \end{aligned}$$

where $P^t(x^0, x)$ is the probability that the Markov chain with initial state x^0 is in state x at iteration t, and A is any set. The convergence time of the Markov chain used by the Gibbs sampler is defined as

$$\tau(\epsilon) = \max_{x^0} \min\{t : D_{x^0}(t') \leq \epsilon \text{ for all } t' \geq t\}, \qquad (7)$$

where ϵ is a pre-specified error tolerance, chosen at the user's discretion. Propp & Wilson (1998) use the arbitrary value $1/e$ as the value of ϵ which gives their mixing time threshold. The first definition of the total variation distance (6) leads to perhaps the clearest interpretation of the choice of ϵ; for every possible set A in the state space, convergence to within ϵ in total variation distance guarantees that the difference between the probability that our Markov chain is in A and the probability of A for the stationary distribution is at most ϵ. The relationship between the value of ϵ and the number of iterations required is further explored through simulation of the stochastic Ising model in § 4.

Requiring that the total variation distance is less than ϵ gives an immediate tolerance on the error caused by lack of convergence in the estimation of the expectation of bounded functions because of the following equivalent formulation of D:

$$D_{x^0}(t) = \frac{1}{2} \max_{|h| \le 1} \left| \int_\Omega h P^t(x^0, dx) - \int_\Omega h\pi(dx) \right|$$

where the maximum is taken over functions $h : \mathcal{X} \to R$ satisfying $\sup_x |h(x)| \le 1$.

In this paper, we are concerned with the number of iterations required to achieve convergence for a given algorithm, and not with other important issues such as the variance of estimates of expectations; see for example Green & Han (1992). We recommend that our results be used in determining the number of iterations required to achieve stationarity; the simulation of the Markov chain can then be continued beyond this and these additional values used for purposes such as estimating expectations.

Our results are an application of coupled Markov chains. The coupling methodology is presented in § 3.2. In § 6 we discuss how these results can be applied to perfect simulation algorithms involving coupling from the past. Other methods for achieving a bound on (7) are discussed in § 3.3.

3.2 Using coupling to bound the convergence time

One method of bounding the convergence time, $\tau(\epsilon)$, of a Markov chain is through monitoring two coupled Markov chains. Suppose X_t^1 and X_t^2 are two Markov chains on the same state space, with the same transition probabilities and with initial values x^1 and x^2 respectively. At each iteration, the same uniform random number is used to determine the transition for both chains. They are said to be coupled at time T^{x^1,x^2} if

$$T^{x^1,x^2} = \min\{t : X_t^1 = X_t^2 | X_0^1 = x^1, X_0^2 = x^2\}. \tag{8}$$

Our bound on $\tau(\epsilon)$ will be in terms of the maximum mean coupling time

$$T = \max_{x^1, x^2} E(T^{x^1, x^2}), \tag{9}$$

where the maximum is taken over all possible initial states x^1 for X_t^1 and x^2 for X_t^2.

As shown in Aldous (1983),

$$\tau(\epsilon) \leq 2eT(1 + \log \epsilon^{-1}). \tag{10}$$

The method used here was inspired by that of Luby, Randall & Sinclair (1995), whose Markov chains were lattice routings in order to generate a random tiling of a planar lattice structure in studying the combinatorics of tiling two-dimensional lattices. They use coupling to get bounds on the convergence time of their Markov chains that are polynomial in the size of the lattice.

For our model for binary images, a partial ordering exists on the set of all configurations. One configuration is greater than another if each pixel of the larger configuration is greater than or equal to the corresponding pixel of the smaller configuration. We set the initial configurations of the two chains to be all $+1$ and all -1. We label these configurations $x^{\max}$ and $x^{\min}$ respectively. Our process will preserve this order; the chain that starts in the maximal state will always be greater than the chain that starts in the minimal state. This is because at each iteration our algorithm will use the same random number to determine the transition for both chains and, as will be seen, the probability distributions used to determine the new value of the pixel assign higher probability to pixels which are like their neighbours.

As argued in Propp & Wilson (1996) for monotone Markov chains such as this, it suffices to consider the case where the initial configurations are the extreme states. Chains started in any other initial states x^1, x^2 ($x^{\min} \le x^1 \le x^2 \le x^{\max}$) must couple in a time less than or equal to the coupling time for $x^{\min}$ and $x^{\max}$ for the same set of random numbers determining the transitions.

Let $\Phi(t)$ be a function that assigns a positive integer to the difference between the configurations at time t of the Markov chains started in the maximal and minimal states. This Φ should be defined such that $\Phi(0) = N$, the number of sites, and $0 \le \Phi(t) \le N$ for all t. Two chains will have coupled at time t if $\Phi(t) = 0$. Once coupled, they will remain so. Define the coupling time

$$T^{x^{\max}, x^{\min}} = \inf\{t : \Phi(t) = 0\}.$$

Then

$$T = E(T^{x^{\max}, x^{\min}}). \tag{11}$$

Let $\Delta\Phi(t) = \Phi(t+1) - \Phi(t)$ denote the change in the value of Φ after one iteration of the random scan Gibbs sampler. Suppose a region of the parameter space for β can be found such that $E\{\Delta\Phi(t)|X_t^1, X_t^2\} < 0$ for all t for which $X_t^1 \neq X_t^2$; say $E\{\Delta\Phi(t)|X_t^1, X_t^2\} \le -a_\beta$ where $-1 < -a_\beta < 0$. Then for these values of β, as shown in the proof of Theorem 1, the quantity $E(T^{x^1, x^2})$ can be bounded above by $N a_\beta^{-1}$.

Theorem 1. *Suppose there exist two coupled realisations, X_t^1, X_t^2, of a Markov chain where $X_0^1 = x^1$ and $X_0^2 = x^2$. Also suppose a constant $a > 0$ can be found such that $E\{\Delta\Phi(t)|X_t^1, X_t^2\} < -a$ for all t for which $X_t^1 \neq X_t^2$, and the distance between the initial states is $\Phi(0) = N$. Then the following bound exists on the mean coupling time* (8):

$$E(T^{x^1, x^2}) \le \frac{N}{a}. \tag{12}$$

Proof. Define the stochastic process $Z_t = \Phi(t) + at$. Then Z_t is a supermartingale up to time T^{x^1,x^2} since

$$E(Z_{t+1}|X_t^1, X_t^2) - Z_t = E(\Delta\Phi|X_t^1, X_t^2) + a \leq 0.$$

Also, T^{x^1,x^2} is a stopping time. Since Z_t is non-negative, we can apply the Optional Stopping Theorem (Durrett, 1996, Theorem 7.6, p. 274) giving

$$E(Z_{T^{x^1,x^2}}) \leq E(Z_0),$$

which gives

$$aE(T^{x^1,x^2}) \leq N. \qquad \square$$

In our examples, a is of the form $N^{-1}f(\beta)$ where, as will be seen, it is straightforward to compute $f(\beta)$ and the range of possible values of β which guarantees that the distance function is decreasing on average. Combining (12) with (10) gives

$$\tau_\beta(\epsilon) \leq \frac{2eN^2}{f(\beta)}(1 + \log \epsilon^{-1});$$

we have introduced the subscript β to make explicit the dependence of τ on the model parameter.

3.3 Other convergence results

In mathematical physics, the case without data is known as the stochastic Ising model with Glauber dynamics and it is well-known that its convergence rate, asymptotically in N, is $O(N \log N)$. In dimensions higher than one, this result holds for values of β below a critical value at which a phase transition occurs. For the Ising model with an external field, if we apply the log Sobolev inequality, the convergence rate can be shown to be $O(N \log N)$ for all β in two dimensions, and for small enough β and large enough external field in higher dimensions; see for example Martinelli and Olivieri (1994). However, it may be impossible to calculate a precise upper bound using this method, so it is difficult to apply these results in practice. Frigessi et al. (1997) present the $O(N \log N)$ results in the context of Bayesian image restoration. While our results are $O(N^2)$, we are able to give the proportionality constant. In a future paper employing another metric, we show how some of the calculations of this paper can be used to get a bound in total variation distance that is $O(N \log N)$.

The total variation distance can also be bounded above by a simple function of the eigenvalue of the Markov chain transition matrix which is second largest in absolute value. Poincaré and Cheeger inequalities can be used to get simple bounds on this eigenvalue in terms of a set of canonical paths on a graph associated with the Markov chain. The vertices of the graph are the states of the Markov chain, and an edge set is chosen between states such that an edge exists between states x^1 and x^2 only if there is a positive probability of moving from state x^1 to x^2 in one iteration; see for example Diaconis & Stroock (1991) and Sinclair (1992). While these approaches seem promising in providing precise bounds, for our image restoration problem we were only able to find canonical paths that gave convergence $O(e^N)$, even for the one-dimensional model.

Using a path bounds approach, Jerrum & Sinclair (1993) develop a Markov chain algorithm for estimating the partition function of the Ising model that they show takes polynomial time. They achieve their convergence results by transforming the problem to a graph model, but it is not clear how their results can be translated to calculating other features of the Ising model.

For the case with no data, corresponding to the stochastic Ising model with no external field, our results apply for small values of β in dimensions higher than one, corresponding to large temperature when the model is considered in thermodynamic terms. Frigessi et al. (1993) consider the question of which Markov chain Monte Carlo algorithm provides fastest convergence for this problem, comparing them via their eigenvalues. They show that, for high temperature, the Metropolis algorithm gives the slowest convergence of any random scan updating dynamics. While the Gibbs sampler is better, they also show that convergence can be improved by considering dynamics which include the current value of the site being updated. In the case of simulation of Gaussian random fields, which have applications to image restoration on a continuous state space, Barone & Frigessi (1990) propose a class of such algorithms. Green & Han (1992) consider this algorithm in the context of both speed of convergence and precision of estimation of functionals of interest.

4 The Case of No Data: the Stochastic Ising Model

4.1 One dimension

In one dimension, with each interior site equally influenced by its two nearest neighbours,

$$\pi(x) = \frac{1}{Z} \exp\left(\beta \sum_{i=1}^{N-1} x_i x_{i+1}\right) \tag{13}$$

and the full conditionals (4) for interior sites are

$$\pi(x_i | x_{-i}) = \frac{\exp\{\beta(x_{i-1}x_i + x_i x_{i+1})\}}{\exp\{\beta(x_{i-1} + x_{i+1})\} + \exp\{\beta(-x_{i-1} - x_{i+1})\}}. \tag{14}$$

Throughout, we ignore the sites on the edge since only the sites with the most extreme probabilities need be considered, and these are interior sites.

Recall that our bound on the convergence time requires a bound on the mean time to couple for Markov chains started in the maximal state, where each pixel is $+1$, and the minimal state, where each pixel is -1. Define the distance function between the current two states of these Markov chains to be $\Phi_s = N - c$ where N is the total number of pixels and c is the number of sites at the right end that have coupled. For example, suppose that, at time t, the configurations of the Markov chains started in the maximal and minimal states are

$$\begin{array}{lcccccccc} X_t^{\max}: & + & + & \ldots & + & - & + & + \\ X_t^{\min}: & + & - & \ldots & - & - & + & + \,; \end{array}$$

then $\Phi_s(t) = N - 3$. Note that $\Phi_s(0) = N$ and $\Phi_s(T) = 0$. We call this distance function the 'sweep distance function'. The following upper bound on the convergence time exists for sampling from the one-dimensional Ising model at all values of β.

THEOREM 2. *For sampling via the random scan Gibbs sampler from the one-dimensional Ising model with N sites given by* (13), *the convergence time* (7) *can be bounded above by*

$$\tau_\beta(\epsilon) \leq 2eN^2 \left(\frac{e^{2\beta} + e^{-2\beta}}{2e^{-2\beta}} \right) (1 + \log \epsilon^{-1})$$

for all values the Ising model parameter β, where ϵ is the specified tolerance for convergence in total variation distance.

For all proofs in §§4–5, see the Appendix.

For example, if $\epsilon = 0.01$ and $\beta = 0.5$, then $\tau \leq 128N^2$. Using a value such as $\beta = 1.5$ gives more influence to the smoothing inherent in the prior distribution and gives the convergence bound $\tau \leq 6162N^2$.

Note that if we take the distance function to be the total number of sites less the number of sites coupled at both ends, the mean coupling time can be reduced by a factor of two.

There is no phase transition in the one-dimensional Ising model (Cipra, 1987), so a result such as this that holds for all β should exist. However, in higher dimensions, convergence is known to change at the critical value of β at which phase transition occurs. Convergence is known to be slow for β above this value. Our results for higher dimensions hold for small β, below this critical value.

4.2 Extension to higher dimensions and larger neighbourhood systems

In two and higher dimensions, there is no simple distance function analogous to the sweep distance function of one dimension. The immediately obvious analogue, where the number of sites coupled at an endpoint is replaced by the size of a corner that is coupled, is not appropriate since, on any future step of the algorithm, any of the sites along the coupled boundary may change, destroying the structure. An irregular boundary around the coupled sites can change in many ways, including losing contact with any corner or edge sites, making it very complex to keep track of the size of the coupled corner. Considering a cluster of coupled sites seems to be too complex to be useful.

For a systematic scan Gibbs sampler it may be possible to define a distance function like this, since at each iteration all pixels are updated and the number coupled in a corner structure can be maintained.

We address this problem by defining a different distance function, which will lead to restrictions on the values of β.

Define the distance function Φ_d as the number of sites where the two chains differ. Then $\Phi_d(0) = N$ and $\Phi_d(T) = 0$. This distance function can be used in any dimension. We call

Φ_d the 'number-of-sites-different distance function'. A change in Φ_d may now occur for any site chosen for updating, unless it is in the middle of a string of at least three coupled sites.

Our result is stated in terms of n, the number of nearest neighbours that are equally influential; n is typically 2 in one dimension, either 4 or 8 in two dimensions and so on. Our upper bound on the convergence time is still a simple function of the model parameter β times N^2, where N is the total number of sites; however, it now holds only for a restricted range of β. As the number of influential neighbours increases, the range of admissible values of β decreases.

THEOREM 3. *For sampling with the random scan Gibbs sampler from the Ising model* (1) *and* (2) *in arbitrary dimension with N sites, where each site is influenced by its n nearest neighbours, the convergence time* (7) *can be bounded above by*

$$\tau_\beta(\epsilon) \le 2eN^2 \left\{ \frac{e^{n\beta} + e^{-n\beta}}{(n+2)e^{-n\beta} - ne^{n\beta}} \right\} \left(1 + \log \epsilon^{-1}\right)$$

for

$$0 \le \beta \le \frac{1}{2n} \log\left(\frac{n+2}{n}\right),$$

where β is the Ising model parameter and ϵ is the specified tolerance for convergence in total variation distance.

For example, if $n = 4$ as with two dimensional first-order neighbourhoods, $\epsilon = 0.01$ and $\beta = 0.05$, then $\tau \le 2322N^2$. Reducing β to 0.01 gives an improvement in our upper bound on τ to $38N^2$. In the case of $n = 8$, as would occur in two dimensions with second-order neighbourhoods, $\beta = 0.01$ gives $\tau \le 108N^2$.

The results for this distance function give larger bounds on the convergence time than those obtained with the sweep distance function in one dimension, in addition to introducing restrictions on the values of β. However, the result using Φ_d gives results that are applicable in any dimension.

As an indication of the role of the error tolerance, ϵ, we simulated 1000 coupled pairs of Markov chains, started in the maximal and minimal states. We use the following characterisation of the total variation distance between two probability measures μ and ν:

$$D(\mu, \nu) = \inf \mathrm{pr}(X \neq Y),$$

where the infimum is over all random variables X and Y for which $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$; see for example Lindvall (1992, p. 19). In Fig. 1, we have plotted the number of iterations versus the probability the Markov chains have not coupled, our approximation to the total variation distance. Tight requirements on ϵ require increasing numbers of iterations, while fewer than 7000 iterations do not give a randomised chain. Note that, while this forward coupling time gives an indication of the time required for convergence to stationarity, we cannot wait until the chains have coupled and use the resulting state as a sample from the stationary distribution. Doing so would bias our results in favour of states at which the probability of coupling is greater (Propp & Wilson, 1996).

5 The Case with Observed Data

5.1 True image with random flips

Suppose the observed configuration y consists of the true configuration x with each spin site flipped independently with probability α, i.e. for each i,

$$\mathrm{pr}(Y_i = y_i | X_i = x_i) = \begin{cases} \alpha & \text{if } y_i \neq x_i \\ 1-\alpha & \text{if } y_i = x_i. \end{cases}$$

Then

$$\mathrm{pr}(Y = y | X = x) = \alpha^{\sum_{j=1}^{N} I(y_j \neq x_j)} (1-\alpha)^{N - \sum_{j=1}^{N} I(y_j \neq x_j)}. \tag{15}$$

For illustration, we give the posterior distribution and the full conditionals in one dimension since it is notationally simplest. Higher-dimensional calculations are completely analogous. If we combine (15) with the prior (1) and (2), the posterior distribution for the true configuration given the observed configuration is

$$\pi_P(x|y) = \tag{16}$$
$$\frac{1}{Z_P} \exp \left[\beta \sum_{i=1}^{N-1} x_i x_{i+1} + \sum_{j=1}^{N} I(y_j \neq x_j) \log(\alpha) + \left\{ N - \sum_{j=1}^{N} I(y_j \neq x_j) \right\} \log(1-\alpha) \right].$$

The posterior full conditionals for interior spin sites can then be calculated to be

$$\pi_P(x_i | x_{-i}, y) =$$
$$\frac{\exp\left\{\beta(x_{i-1}x_i + x_i x_{i+1}) + I(y_i \neq x_i)\log\left(\frac{\alpha}{1-\alpha}\right)\right\}}{\exp\left\{\beta(x_{i-1} + x_{i+1}) + I(y_i \neq 1)\log\left(\frac{\alpha}{1-\alpha}\right)\right\} + \exp\left\{\beta(-x_{i-1} - x_{i+1}) + I(y_i \neq -1)\log\left(\frac{\alpha}{1-\alpha}\right)\right\}}$$

We now state our convergence bound for arbitrary dimension.

THEOREM 4. *Suppose we have observed, in arbitrary dimension, an image of N pixels taking the values $+1$ or -1 where it is known that each pixel is incorrectly observed with probability*

α. *For sampling from the posterior Gibbs distribution with our prior distribution the Ising model,* (1) *and* (2), *with the random scan Gibbs sampler, the convergence time* (7) *can be bounded as*

$$\tau_\beta(\epsilon) \leq 2eN^2 \left\{ \frac{k_\alpha + e^{2n\beta} + e^{-2n\beta}}{k_\alpha - ne^{2n\beta} + (n+2)e^{-2n\beta}} \right\} \left(1 + \log \epsilon^{-1}\right)$$

for

$$0 \leq \beta \leq \frac{1}{2n} \log \left[\frac{k_\alpha + \{k_\alpha^2 + 4n(n+2)\}^{\frac{1}{2}}}{2n} \right],$$

where $k_\alpha = (1-\alpha)/\alpha + \alpha/(1-\alpha)$, β *is the Ising model parameter,* n *is the number of nearest neighbours of interior pixels, and* ϵ *is the specified tolerance for convergence in total variation distance.*

For example, if $\alpha = 0.05$, $n = 4$, $\beta = 0.05$ and $\epsilon = 0.01$, $\tau \leq 38N^2$, improving the bound from the case with no data by a factor greater than 60. Moreover, for $n = 4$ and $\alpha = 0.05$, the range of admissible values of β is four times as great as that in the case with no data. Smaller values of α increase the range of β and decrease the convergence time bound, reflecting the increased reliability of the observed image.

If $\alpha = 0$ the observed image is correct and if $\alpha = 1$ the observed image is completely incorrect. In these cases, our result holds for all β. If $\alpha = 1/2$ the observed image gives no information. Our result then coincides with the no data case of Theorem 3.

5.2 True image with additive normal noise

We will consider the simple case where the normal noise is additive at each pixel, i.e. $y = x + \mathcal{N}$, where $\mathcal{N}$ is a vector with each entry an independent realisation from a $N(\mu, \sigma^2)$

distribution. In one dimension, this leads to the posterior density

$$\pi_P(x|y) = \frac{1}{Z^P} \exp\left\{\beta \sum_{i=1}^{N-1} x_i x_{i+1} - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (\mu - y_i + x_i)^2 \right\}, \tag{17}$$

and for interior points ($i = 2, \ldots, N-1$) the full conditionals are

$$\begin{aligned} &\pi_P(x_i|x_{-i}, y) \\ &= \frac{\exp\left\{\beta(x_{i-1}x_i + x_i x_{i+1}) - \frac{1}{2\sigma^2}(\mu - y_i + x_i)^2\right\}}{\exp\left\{\beta(x_{i-1} + x_{i+1}) - \frac{1}{2\sigma^2}(\mu - y_i + 1)^2\right\} + \exp\left\{\beta(-x_{i-1} - x_{i+1}) - \frac{1}{2\sigma^2}(\mu - y_i - 1)^2\right\}}. \end{aligned}$$

We consider the case where $\mu = 0$.

Theorem 5. *Suppose we have observed, in arbitrary dimension, an image of N pixels where it is known that pixel i, $i = 1, \ldots, N$ should have a value of $+1$ or -1 but has been observed as a random sample from a $N(x_i, \sigma^2)$ distribution, where x_i is the true value of the ith pixel. For sampling from the posterior Gibbs distribution with our prior distribution the Ising model,* (1) *and* (2), *with the random scan Gibbs sampler, the convergence time* (7) *can be bounded as*

$$\tau_\beta(\epsilon) \le 2eN^2 \left\{ \frac{k_{\sigma, y_{\min}} + e^{2n\beta} + e^{-2n\beta}}{k_{\sigma, y_{\min}} + (n+2)e^{-2n\beta} - ne^{2n\beta}} \right\} \left(1 + \log \epsilon^{-1}\right)$$

for

$$0 \le \beta \le \frac{1}{2n} \log\left[\frac{k_{y_{\min}} + \left\{k_{y_{\min}}^2 + 4n(n+2)\right\}^{\frac{1}{2}}}{2n} \right].$$

Here $y_{\min} = \min_i\{|y_i|\}$, the smallest of the observed pixels in absolute value, $k_{y_{\min}} = e^{2y_{\min}/\sigma^2} + e^{-2y_{\min}/\sigma^2}$, β is the Ising model parameter, n is the number of nearest neighbours of interior pixels, and ϵ is the specified tolerance for convergence in total variation distance.

For the case where $\sigma = 0.3$, if $n = 8$, a value of $y_{\min}$ such as 0.65 gives $\beta \le 0.773$. As a guide to what is an appropriate value of β, we consider the work of Besag (1986). In his two-colour model formulation, his parameter is twice ours for a two-dimensional image restoration with pixels affected by their 8 nearest neighbours. He found that a parameter value of 1.5 worked well in practice, corresponding to a value of $\beta = 0.75$ in our model.

Note that smaller values of the variance of the normal noise increase the range of possible values of β for which our results hold and decrease the upper bound on the convergence time, reflecting the increased reliability of the observed image. In the limit as $\sigma \to \infty$, our observed image gives no information. In this case, our result coincides with the no data case of Theorem 3.

Fig. 2 gives an example of the image restoration process. Fig. 2(a) shows the original image, drawn on a 32×32 grid. It was randomly degraded with $N(0, 0.4^2)$ noise, added independently to each pixel; see Fig. 2(b). Our prior parameter, β, was set at 0.05, and first-order neighbourhoods were used. The specified error tolerance for randomisation in total variation distance was 0.01. The algorithm was run for the number of iterations our theory specifies, taking $y_{\min}$ to be 0, from the initial state with every pixel black. Our approximate sample from the posterior distribution is shown in Fig. 2(c).

6 The Expected Number of Steps Required for Exact Sampling

Propp & Wilson's (1996) coupling from the past algorithm can be applied to our examples by running two coupled realisations of the Markov chain, starting in the maximal and minimal states at some time $-t$ in the past. If the two chains have coupled at time zero, the state at time zero is an exact sample from the Markov chain's stationary distribution.

Our results give an indication of the expected value of $-t$ required for coalescence at time zero in Propp & Wilson's algorithm. As Propp & Wilson note, the random variables T^*, the smallest t such that chains started in the maximal and minimal state have coupled state at time t, and T_*, the smallest t such that chains started in the maximal and minimal state at time $-t$ will be in the same state at time zero, have the same probability distribution.

By Theorem 1

$$E(T_*) = E(T^*) \leq \frac{N}{a},$$

where a is the positive constant with $E(\Delta\Phi|X_t, Y_t) < -a$ for all t. For example, in the case of no data

$$E(T^*) \leq N^2 \left\{ \frac{e^{n\beta} + e^{-n\beta}}{(n+2)e^{-n\beta} - ne^{n\beta}} \right\}.$$

Application of Markov's inequality to our result gives an upper bound on the probability that the coupling from the past algorithm will take a very large number of runs as follows

$$\mathrm{pr}(T^* > B) < \frac{E(T^*)}{B} \leq \frac{N}{aB}.$$

Acknowledgement.

I wish to thank Jeffrey S. Rosenthal, under whose direction this work was done as part of my doctoral research at the University of Toronto. Thanks are also due to Radford Neal, Neal Madras, two anonymous referees and the editor whose comments and suggestions greatly improved the clarity and scope of this paper.

Appendix

Proofs of theorems from §§4 and 5

Proof of Theorem 2. Consider all possible configurations of a site and its two neighbours, ignoring endpoints, in which a change in the sweep distance function Φ_s may occur. Since we are considering the random scan Gibbs sampler, one site is updated at each iteration. At each step in the algorithm, Φ_s will change by $+1$ if the $(N-c+1)$th site changes and by -1 or more if the $(N-c)$th site changes. We call sites which can contribute to an increase in Φ_s 'bad' sites, and sites which can contribute to a decrease in Φ_s 'good' sites. Updating a good site may result in a change in c of more than one if sites to the left of the site being updated have already coupled. However, we will create our bound by considering worst-case scenarios, so we will consider good updates which only decrease Φ_s by 1. Note that, because sites to the right of the $(N-c+1)$th site have the same neighbours in both configurations, they will change in the same manner, so they cannot affect the value of the distance function. If a site to the left of the $(N-c)$th site is chosen for updating, the value of the distance function cannot change.

The configurations of three sites in Table 1 will possibly result in a change in Φ_s. The site being updated is to the left of the boundary for good sites, the $(N-c)$th site, and to the right of the boundary for bad sites, the $(N-c+1)$th site. The top row indicates the current configuration of $X_t^{\max}$, the chain started in the maximal configuration, and the bottom row indicates the current configuration of $X_t^{\min}$, the chain started in the minimal configuration. The update probabilities are calculated from (14). As an example, suppose the site to the left of the boundary in the first configuration shown has been selected for updating. Then

$$\begin{aligned}
&\text{pr}(\Delta\Phi_s \text{ occurring}) \\
&\quad = \text{pr}\left(\text{configuration becomes} \quad \begin{array}{cccc} X_t^{\max}: & + & + & + \\ \\ X_t^{\min}: & - & + & + \end{array} \quad \text{or} \quad \begin{array}{cccc} X_t^{\max}: & + & - & + \\ \\ X_t^{\min}: & - & - & + \end{array}\right) \\
&\quad = \min\left\{\frac{e^{2\beta}}{e^{2\beta}+e^{-2\beta}}, \frac{1}{2}\right\} + \min\left\{\frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}, \frac{1}{2}\right\}
\end{aligned}$$

where the minima are being taken over the probabilities for $X_{t+1}^{\max}$ and $X_{t+1}^{\min}$, respectively, to be the given states given their configurations at iteration t.

In order to obtain an upper bound on $E(\Delta\Phi_s)$ that holds for all configurations, we assume that the site to the left of the boundary is a good site with the smallest probability of changing Φ_s and that results in a change of Φ_s of only 1. At each iteration, the site to be updated is chosen uniformly from the N pixels. Thus for all t where $X_t^{\max} \neq X_t^{\min}$

$$\begin{aligned}
&E(\Delta\Phi_s | X_t^{\max}, X_t^{\min}) \\
&\quad \leq \frac{1}{N}\left\{\left(\frac{1}{2} - \frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}\right) - \left(\frac{1}{2} + \frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}\right)\right\} \\
&\quad < 0 \text{ for all } \beta.
\end{aligned}$$

If we apply Theorem 1 and equation (11) with

$$a = \frac{1}{N}\frac{2e^{-2\beta}}{e^{2\beta}+e^{-2\beta}},$$

the mean coupling time can be bounded above by

$$T \leq N^2\left(\frac{2e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}\right)^{-1},$$

where T has been defined in (9), and applying (10) gives the result. □

Proof of Theorem 3. Consider all possible configurations of a site and its n neighbours, ignoring edge and corner sites, in which a change in Φ_d, the number-of-sites-different distance function, may occur. Since we are using the random scan Gibbs sampler, at each iteration Φ_d can change by at most 1. A site which can lead to a change in Φ_d of -1 is considered a 'good' site and by $+1$ a 'bad' site.

For ease of presentation, the possible configurations are illustrated in one dimension with two influential neighbours. The argument in higher dimensions and with more influential neighbours is completely analogous. The configurations in one dimension of three sites in Table 2, where the middle site is the one randomly chosen for updating, will possibly result in a change in Φ_d. The top row indicates the current configuration of $X_t^{\max}$, the chain started in the maximal configuration, and the bottom row indicates the current configuration of $X_t^{\min}$, the chain started in the minimal configuration.

Note that each bad site has has at least one good site as a neighbour. There are therefore at most two bad sites for each good site. If a bad site is chosen, a change of $+1$ in Φ_d occurs with probability at most

$$\frac{e^{2\beta} - e^{-2\beta}}{e^{2\beta} + e^{-2\beta}}.$$

If a good site is chosen, a change of -1 in Φ_d occurs with probability at least

$$\frac{2e^{-2\beta}}{e^{2\beta} + e^{-2\beta}}.$$

In the general case, where each site is influenced by its n nearest neighbours, there are at most n bad sites for each good site. If a bad site is chosen, a change of $+1$ in Φ_d occurs with probability at most

$$\frac{e^{n\beta} - e^{-n\beta}}{e^{n\beta} + e^{-n\beta}}.$$

If a good site is chosen, a change of -1 in Φ_d occurs with probability at least

$$\frac{2e^{-n\beta}}{e^{n\beta} + e^{-n\beta}}.$$

At each iteration, a particular site is chosen with probability $\frac{1}{N}$ for updating. Thus for all t where $X_t^{\max} \neq X_t^{\min}$

$$
\begin{aligned}
&E(\Delta\Phi_d | X_t^{\max}, X_t^{\min}) \\
&= \frac{1}{N}\left\{\sum_{\text{bad sites}} \text{pr(this change occurs)} - \sum_{\text{good sites}} \text{pr(this change occurs)}\right\} \\
&\leq \frac{1}{N}\left\{(\text{Number of bad sites})\frac{e^{n\beta}-e^{-n\beta}}{e^{n\beta}+e^{-n\beta}} - (\text{Number of good sites})\frac{2e^{-n\beta}}{e^{n\beta}+e^{-n\beta}}\right\} \\
&\leq \frac{1}{N}\left\{n(\text{Number of good sites})\frac{e^{n\beta}-e^{-n\beta}}{e^{n\beta}+e^{-n\beta}} - (\text{Number of good sites})\frac{2e^{-n\beta}}{e^{n\beta}+e^{-n\beta}}\right\} \\
&= \frac{(\text{Number of good sites})}{N}\left\{n\frac{e^{n\beta}-e^{-n\beta}}{e^{n\beta}+e^{-n\beta}} - \frac{2e^{-n\beta}}{e^{n\beta}+e^{-n\beta}}\right\}.
\end{aligned}
$$

This is negative for $\beta < (2n)^{-1}\log(\frac{n+2}{n})$. The number of good sites is Φ_d. The chain has coupled when Φ_d reaches 0, so at each iteration the number of good sites is at least 1. Thus, the mean coupling time T can be bounded above by

$$
T \leq N^2\left\{\frac{(n+2)e^{-n\beta} - ne^{n\beta}}{e^{n\beta}+e^{-n\beta}}\right\}^{-1},
$$

and applying (10) gives the result. □

Note that the result of Theorem 3 is not sharp. The limiting configuration of n high-probability bad sites for each low probability good site cannot occur in isolation.

Proof of Theorem 4. As in the case of no data, there are at most n good sites, contributing to a decrease in Φ_d, for every bad site, contributing to an increase in Φ_d. Regardless of the observed value at the site being updated, the good configuration with least probability of coupling is all $+1$ and all -1, with update probability

$$
1 - \frac{e^{n\beta}}{e^{n\beta}+e^{-n\beta+\log\left(\frac{\alpha}{1-\alpha}\right)}} + \frac{e^{-n\beta}}{e^{-n\beta}+e^{n\beta+\log\left(\frac{\alpha}{1-\alpha}\right)}}.
$$

The bad configuration with greatest probability of uncoupling has update probability

$$\frac{e^{n\beta}}{e^{n\beta}+e^{-n\beta+\log\left(\frac{\alpha}{1-\alpha}\right)}}-\frac{e^{-n\beta}}{e^{-n\beta}+e^{n\beta+\log\left(\frac{\alpha}{1-\alpha}\right)}}.$$

Thus,

$$\begin{aligned}
&E(\Delta\Phi|X_t^{\max},X_t^{\min})\\
&\leq \quad \frac{(\text{Number of good sites})}{N}\left(n\frac{e^{-\log\left(\frac{\alpha}{1-\alpha}\right)}\left(e^{2n\beta}-e^{-2n\beta}\right)}{\left\{e^{n\beta}+e^{-n\beta-\log\left(\frac{\alpha}{1-\alpha}\right)}\right\}\left\{e^{-n\beta}+e^{n\beta-\log\left(\frac{\alpha}{1-\alpha}\right)}\right\}}\right.\\
&\qquad\left.-\left[1-\frac{e^{-\log\left(\frac{\alpha}{1-\alpha}\right)}\left(e^{2n\beta}-e^{-2n\beta}\right)}{\left\{e^{n\beta}+e^{-n\beta-\log\left(\frac{\alpha}{1-\alpha}\right)}\right\}\left\{e^{-n\beta}+e^{n\beta-\log\left(\frac{\alpha}{1-\alpha}\right)}\right\}}\right]\right).
\end{aligned}$$

For this to be negative we must have

$$ne^{4n\beta}-k_\alpha e^{2n\beta}-(n+2)<0,$$

where $k_\alpha=(1-\alpha)/\alpha+\alpha/(1-\alpha)$. Note that the result is the same when the flip rate is $1-\alpha$ as when it is α, so the values of β that guarantee convergence in $O(N^2)$ time are the same for a flip rate of, for example, 0.05 as for 0.95.

Application of Theorem 1 and (10) gives our result. □

Proof of Theorem 5. Suppose site i is being updated. It can be shown that, regardless of the value of y_i, the good configuration which has the smallest probability of becoming coupled is all +1 and all −1. The probability of the middle site becoming the same in the two chains, given the data value, is

$$\frac{2e^{-2n\beta}+e^{2y_i/\sigma^2}+e^{-2y_i/\sigma^2}}{e^{2n\beta}+e^{-2n\beta}+e^{2y_i/\sigma^2}+e^{-2y_i/\sigma^2}}.$$

Similarly, regardless of the value of y_i, the bad configuration which has the greatest probability of becoming uncoupled has probability

$$1-\frac{2e^{-2n\beta}+e^{2y_i/\sigma^2}+e^{-2y_i/\sigma^2}}{e^{2n\beta}+e^{-2n\beta}+e^{2y_i/\sigma^2}+e^{-2y_i/\sigma^2}}$$

of the middle site becoming different. The data value that minimises the least probable good probability and maximises the most probable bad is $\min_i\{|y_i|\}$. Substituting this for y_i and using the same argument as in the proof of Theorem 4 gives our result. □

References

ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/82*, Lecture Notes in Mathematics, Vol. **986**. Ed. J. Azema and M. Yor, pp. 243–97. Berlin; New York: Springer-Verlag.

BARONE, P. & FRIGESSI, A. (1990). Improving stochastic relaxation for Gaussian random fields. *Prob. Eng. Info. Sci.* **4**, 369–89.

BESAG, J. (1986). On the statistical analysis of dirty pictures (with Discussion). *J. R. Statist. Soc.* B **48**, 259–302.

BESAG, J. & GREEN, P.J. (1993). Spatial statistics and Bayesian computation. *J. R. Statist. Soc.* B **55**, 25–37.

BESAG, J., GREEN, P.J., HIGDON, D. & MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with Discussion). *Statist. Sci.* **10**, 3–66.

BROOKS, S.P. & ROBERTS, G.O. (1997). Assessing convergence of Markov chain Monte Carlo algorithms. *Statist. Comp.* **8**, 319–35.

CIPRA, B.A. (1987). An introduction to the Ising model. *Am. Math. Mon.* **94**, 937–59.

COWLES, M.K. & CARLIN, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Assoc.* **91**, 883–904.

Cowles, M.K., Roberts, G.O. & Rosenthal, J.S. (1999). Possible biases induced by MCMC convergence diagnostics. *J. Statist. Comp. Simul.* **64**, 87–104.

Diaconis, P. & Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1**, 36–61.

Durrett, R. (1996). *Probability: Theory and Examples*, second edition. Belmont, California: Wadswoth & Brooks/Cole.

Frieze, A., Kannan, R. & Polson, N.G. (1994). Sampling from log-concave distributions. *Ann. Appl. Prob.* **4**, 812–34.

Frigessi, A., di Stefano, P., Hwang, C.-R. & Sheu, S.-J. (1993). Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. R. Statist. Soc.* B **55**, 205–19.

Frigessi, A., Martinelli, F. & Stander, J. (1997). Computational complexity of Markov chain Monte Carlo methods for finite Markov random fields. *Biometrika* **84**, 1–18.

Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pat. Anal. Mach. Intell.* **6**, 721–41.

Gilks, W., Richardson, S. & Spiegelhalter, D.J., Eds.(1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

GREEN, P.J. (1996). MCMC in image analysis. In *Markov Chain Monte Carlo in Practice*. Ed. W. Gilks, S. Richardson and D.J. Spiegelhalter, pp. 381–400. London: Chapman and Hall.

GREEN, P. & HAN, X.L. (1992). Metropolis methods, Gaussian proposals and antithetic variables. In *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*. Ed. P. Barone, A. Frigessi and M. Piccioni, pp. 142–64. Berlin: Springer-Verlag.

INGRASSIA, S. (1994). On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds. *Ann. Appl. Prob.* **4**, 347–89.

JERRUM, M. & SINCLAIR, A. (1993). Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comp.* **22**, 1087-116.

LINDVALL, T. (1992). *Lectures on the Coupling Method.* New York: Wiley.

LUBY, M., RANDALL, D. & SINCLAIR, A. (1995). Markov chain algorithms for planar lattice structures (extended abstract). *Foundat. Comp. Sci.*, 150–9.

MARTINELLI, F. & OLIVIERI, E. (1994). Approach to equilibrium of Glauber dynamics in the one phase region I: the attractive case. *Commun. Math. Phys.* **161**, 447–86.

MENGERSEN, K.L. & TWEEDIE, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101–21.

MEYN, S.P. & TWEEDIE, R.L. (1994). Computable bounds for convergence rates of Markov chains. *Ann. Appl. Prob.* **4**, 981–1011.

POLSON, N.G. (1996). Convergence of Markov chain Monte Carlo algorithms (with Discussion). In *Bayesian Statistics 5*. Ed. J.M. Bernardo, A.P. Dawid, and A.F.M. Smith, pp. 297–321. Oxford: Oxford University Press.

PROPP, J.G. & WILSON, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Algor.* **9**, 223–52.

PROPP, J.G. & WILSON, D.B. (1998). How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algor.* **27**, 170–217.

ROBERTS, G.O. & ROSENTHAL, J.S. (1998). Markov chain Monte Carlo: Some practical implications of theoretical results, with discussion. *Can. J. Statist.* **26**, 5–31.

ROSENTHAL, J.S. (1995). Minorization condition and convergence rates for Markov chain Monte Carlo. *J. Am. Statist. Assoc.* **90**, 558–66.

SINCLAIR, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinat. Prob. Comp.* **1**, 351–70.

SINCLAIR, A. & JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Info. Comp.* **1**, 93–133.

SMITH, A.F.M. & ROBERTS, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B **55**, 3–23.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with Discussion). *Ann. Statist.* **22**, 1701–62.

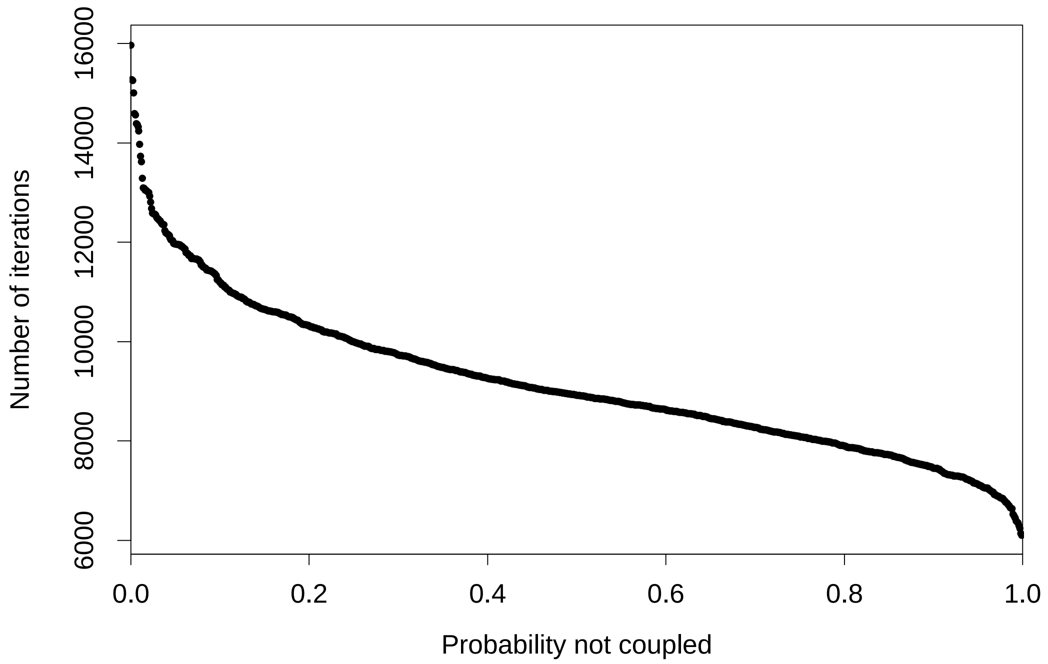

Figure 1: The number of iterations required for various error tolerances in total variation distance, indicated as the probability not coupled, based on 1000 simulations.

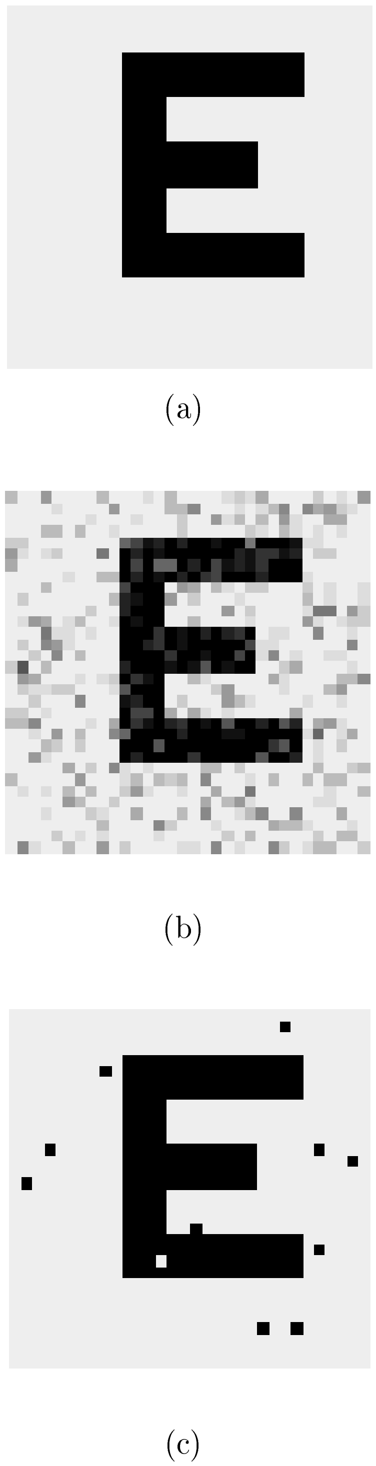

Figure 2: A simulated restoration of a 32×32 image. (a) true image; (b) observed image; (c) sample from the posterior distribution.

Configuration	**Probability of $\Delta\Phi_s$ occurring if site chosen**
Good Sites ($\Delta\Phi_s = -1$):	
$\begin{array}{l} X_t^{\max}: \\ X_t^{\min}: \end{array} \begin{array}{cc\|c} + & - & + \\ - & - & + \end{array} , \begin{array}{cc\|c} + & + & - \\ - & - & - \end{array}$	$\frac{1}{2} + \frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$
Good Sites ($\Delta\Phi_s < -1$):	
$\begin{array}{l} X_t^{\max}: \\ X_t^{\min}: \end{array} \begin{array}{cc\|c} + & + & + \\ + & - & + \end{array} , \begin{array}{cc\|c} + & + & - \\ + & - & - \end{array} , \begin{array}{cc\|c} - & + & + \\ - & - & + \end{array} , \begin{array}{cc\|c} - & + & - \\ - & - & - \end{array}$	1
Bad Sites ($\Delta\Phi_s = +1$):	
$\begin{array}{l} X_t^{\max}: \\ X_t^{\min}: \end{array} \begin{array}{c\|cc} + & + & - \\ - & + & - \end{array} , \begin{array}{c\|cc} + & + & + \\ - & + & + \end{array} , \begin{array}{c\|cc} + & - & - \\ - & - & - \end{array} , \begin{array}{c\|cc} + & - & + \\ - & - & + \end{array}$	$\frac{1}{2} - \frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$

Table 1: Possible configurations that may lead to a change in the sweep distance function in one dimension.

Configuration	**Probability of $\Delta\Phi_d$ occurring if site chosen**
Good Sites ($\Delta\Phi_d = -1$):	
$\begin{array}{llll} X_t^{\max}: & + & + & + \\ X_t^{\min}: & - & - & - \end{array}$	$\frac{2e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$
$\begin{array}{llll} X_t^{\max}: & + & + & + \\ X_t^{\min}: & + & - & - \end{array} , \begin{array}{lll} + & + & - \\ - & - & - \end{array}$	$\frac{1}{2} + \frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$
$\begin{array}{llll} X_t^{\max}: & + & + & + \\ X_t^{\min}: & + & - & + \end{array} , \begin{array}{lll} + & + & - \\ + & - & - \end{array}$	1
Bad Sites ($\Delta\Phi_d = +1$):	
$\begin{array}{llll} X_t^{\max}: & + & - & + \\ X_t^{\min}: & - & - & + \end{array} , \begin{array}{lll} + & - & - \\ - & - & - \end{array}$	$\frac{1}{2} - \frac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$
$\begin{array}{llll} X_t^{\max}: & + & - & + \\ X_t^{\min}: & - & - & - \end{array}$	$\frac{e^{2\beta}-e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$

Table 2: Possible configurations that may lead to a change in the number-of-sites-different distance function in one dimension.