# Phase transition and simulation for a penalized Ising model with applications in image analysis

Rasmus Waagepetersen

**Abstract**

In this paper phase transition and simulation is studied for a penalized Ising model of potential use in Bayesian image analysis. The penalized Ising model appears by penalizing presence of small connected components of pixels in the realizations of the Ising model. Also estimation of the normalizing constant is considered.

**Keywords:** Bayesian image analysis, Ising model, Markov chain Monte Carlo, Markov connected component field, normalizing constant, partition function, phase transition.

## 1 Introduction

Markov connected component fields (MCCF's) is a class of discrete valued random fields introduced in Møller & Waagepetersen (1996). In this paper mathematical properties of MCCF's were studied, and also specific parametric MCCF models were applied for modelling of binary images. One of these models appeared as a modification of the well-known Ising model (see section 2) where realizations with presence of small connected regions of constant colour were penalized. This model appeared to be a much more realistic model for images than the Ising model which has been used extensively as a prior for maximum a posteriori estimation of images. Experiments showed that the Ising model is not suitable as a prior when the full posterior distribution is required for the inference, but promising results were obtained with the penalized Ising model.

Another interesting issue was that phase transition as it appears for the Ising model did not seem to occur for the penalized model. Phase transition exhibits itself for the Ising model as a very dramatic and abrupt change from short to infinite range correlations and this behaviour is not convenient for applications in spatial statistics. In section 3 it is demonstrated that phase transition in fact also occurs for the penalized Ising model. Simulations in Møller & Waagepetersen (1996) however suggest that phase transition does not occur for parameter values which are relevant when the model is applied to typical image data.

For both the Ising model and the penalized Ising model the constant black or white images constitute the modes of the distributions. In Møller & Waagepetersen (1996) a single-site Metropolis chain was used for the simulations of the MCCF's. Even though this chain in theory converges to the right stationary distribution, it is clear that for the penalized Ising model it will often in practice be impossible to sample the two modes, so that in effect a simulation of the model conditioned away from the modes is obtained. In section 4 a Markov chain Monte Carlo (MCMC) algorithm which may be better suited for moving between constant and nonconstant images is constructed. For many practical purposes the problem related to the single-site algorithm may not be very important if the probability of the modes is negligible. This probability is estimated in section 5. It is assumed that the reader is familiar with the methods of MCMC. Background material on MCMC can be found in e.g. Hastings (1970), Tierney (1994) or Besag, Green, Higdon & Mengersen (1995).

## 2 Definitions

Let $I$ be a subset of $\mathbf{Z}^2$, and let $V = \{0, 1\}$, so that $S = V^I$ is the statespace of binary images. Let further $\sim$ be a symmetric and reflexive relation on $I$. Then a connected component $K$ is a nonempty subset of $I$ so that for all $i, j \in K$ there exists $i_1, \ldots, i_n \in K$ with $i = i_1 \sim \cdots \sim i_n = j$. The set of connected components is denoted $\mathcal{K}$, and for each $x = (x_i)_{i \in I} \in S$ and any $A \subseteq I$, the connected component relation $\underset{x_A}{\sim}$ is defined by

$$\forall\, i, j \in A : i \underset{x_A}{\sim} j \Leftrightarrow \exists\, i_1, \ldots i_n \in A : \; i = i_1 \sim \cdots \sim i_n = j \text{ and } x_{i_1} = \cdots = x_{i_n}.$$

Hence the set $\mathcal{K}(x_A)$ of nonempty maximal cliques with respect to $\underset{x}{\sim}$ constitutes the set of maximal connected components of pixels with the same colour in the image $x_A = (x_i)_{i \in A}$. Let further $\mathcal{K}^{(l)}(x_A) = \{K \in \mathcal{K}(x_A) \mid \forall j \in K : x_j = l\}$ be the set of connected components of colour $l \in V$ in the image $x_A \in V^A$, and for $i \in A$ and $B \subseteq A$, write $i \underset{x_A}{\sim} B$ if there is a $j \in B$ with $i \underset{x_A}{\sim} j$.

Assume that $I$ is finite. A random field $X = (X_i)_{i \in I}$ is then a Markov connected component field if the probability density function of $X$ is given by

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \Psi_K\big(l(x_K)\big), \; x \in S,$$

where $\Psi$ is some nonnegative funtion on $\mathcal{K} \times V$ and $l(x_K)$ is the common colour of the pixels in the connected component $K \in \mathcal{K}(x)$. Let for $K \in \mathcal{K}$, the area be given by $a(K) = |K|$ (the cardinality of $K$) and let $u(K) = \#\{(i, j) \in K \times I \setminus K | i \sim j\}$ be the length of the boundary of $K$ measured in terms of neighbouring pairs of pixels where one pixel is in $K$ and the other is outside $K$. The Ising model given by

$$p(x) \propto \exp\big(-\beta \sum_{\substack{\{i,j\} \subseteq I: \\ i \sim j}} 1(x_i \neq x_j)\big) = \exp\big(-\frac{\beta}{2} \sum_{K \in \mathcal{K}(x)} u(K)\big),\ x \in S \qquad (1)$$

is one example of a Markov connected component field. The penalized Ising model is given by

$$p(x) \propto \exp\Big(-\big( \sum_{K \in \mathcal{K}(x)} \frac{\beta}{2} u(K) + \frac{\gamma}{a(K)}\big)\Big),\ x \in S, \qquad (2)$$

so that small components are penalized when $\gamma$ is positive.

The $k$'th order relation $\sim_k$ is given by $i \sim_k j \Leftrightarrow ||i-j|| < k,\ k = 1, 2, \ldots$ and $\mathcal{K}_k$ denotes the set of $\sim_k$-connected components . Further, let $\partial A = \{i \in \mathbf{Z}^2 \setminus A \,|\, \exists j \in A : i \sim j\}$ and $\bar{A} = A \cup \partial A$ for any $A \subseteq I$, and write $B \uparrow A$ if $A \cap B \neq \emptyset$, $B \subseteq I$. In the following $\sim$ will be the second order relation (so that $\mathcal{K} = \mathcal{K}_2$) unless other is stated.

# 3 Phase transition

In this section $I$ is equal to $\mathbf{Z}^2$. Phase transition is a phenomena related to extensions of finite lattice distributions to infinite lattices and is said to occur when such extensions are nonunique. Let $\mathcal{G}$ be the set of nonempty finite subsets of $I$. In Møller & Waagepetersen (1996) the existence of a random field $X = (X_i)_{i \in I}$ with conditional probabilities

$$P(X_\Lambda = x_\Lambda | X_{\Lambda^c} = x_{\Lambda^c}) = p_\Lambda(x_\Lambda | x_{\Lambda^c}) \propto$$
$$\exp\big(-\beta/2 \sum_{\substack{\{i,j\} \subseteq \mathbf{Z}^2: \\ i \sim j \text{and} \{i,j\} \uparrow \Lambda}} 1(x_i \neq x_j) - \sum_{\substack{K \in \mathcal{K}(x): \\ K \uparrow \bar{\Lambda}, |K| < \infty}} \gamma / a(K)\big),\ \Lambda \in \mathcal{G},\ x \in S, \qquad (3)$$

was established. For $x \in S$ for which $K \uparrow \bar{\Lambda}$ implies $a(K) < \infty$ for $K \in \mathcal{K}(x)$, (3) is identical to the conditional probability of the model (2) defined on a finite lattice including all finite $K \in \mathcal{K}(x)$ with $K \uparrow \bar{\Lambda}$, and (3) can furthermore be obtained as the limit of conditional probabilities of models (2) defined on an increasing sequence of finite lattices (for further details see Møller & Waagepetersen, 1996). The random field $X$ can therefore be considered as an extension to the infinite lattice of the model (2).

For a given "boundary condition" $x \in S$, extensions of finite lattice Gibbs distributions are obtained as limits of probability functions (like (3)) along a sequence $\{\Lambda_n\}_{n \geq 1}$ of strictly increasing subsets $\Lambda_n \in \mathcal{G}$ (see e.g. Preston, 1976). For the Ising model and for values of $\beta$ greater than the socalled critical value

$\beta_c$ the limiting distribution depends on the chosen boundary condition whereby phase transition occurs. The technique which Peierls (1936) used to demonstrate the presence of phase transition for the Ising model can immediately be applied also for the model specified by (3) as explained below. More recent accounts of Peierls argument can be found in e.g. Griffiths (1964) or Georgii (1988).

Denote by $P^0$ and $P^1$ the distributions on the infinite lattice with common local characteristics given by (3), and where $P^0$ and $P^1$ are obtained by using the boundary conditions $x^0 = (0)_{i\in I}$ and $x^1 = (1)_{i\in I}$, respectively. When applying Peierls argument the distribution of the random field $X$ only enters when the probability $P(K \in \mathcal{K}(X), l(X_K) = 0 | X_{\Lambda^c} = x^1_{\Lambda^c})$ is estimated for a given finite $K \in \mathcal{K}$. Suppose that such a $K$ is given, and consider an $x \in S$ with $K \in \mathcal{K}(x)$ and $l(x_K) = 0$. For any such $x$ define the state $z(x)$ by $z(x)_i = 1$, $i \in K$, and $z(x)_i = x_i$, $i \in I \setminus K$. Choose further $N$ such that $K \subseteq \Lambda_n$ for each $n \geq N$. Then for any $n \geq N$ and $\gamma > 0$,

$$
\begin{aligned}
&p_{\Lambda_n}(z(x)_{\Lambda_n} | x^1_{\Lambda_n^c}) = \\
&\qquad p_{\Lambda_n}(x_{\Lambda_n} | x^1_{\Lambda_n^c}) \exp\big(\beta u(K) + \gamma/a(K) - \gamma/a(\tilde{K}) + \sum_{\substack{K' \in \mathcal{K}(x)\setminus K: \\ K' \uparrow K}} \gamma/a(K')\big) \geq \\
&\qquad\qquad p_{\Lambda_n}(x_{\Lambda_n} | x^1_{\Lambda_n^c}) \exp(\beta u(K)) \quad (4)
\end{aligned}
$$

where $\tilde{K} = (\cup_{\substack{K' \in \mathcal{K}(x)\setminus K: \\ K' \uparrow K}} K') \cup K$. Thereby

$$
\begin{aligned}
&P_{\Lambda_n}(K \in \mathcal{K}(X), l(X_K) = 0 | X_{\Lambda_n^c} = x^1_{\Lambda_n^c}) = \\
&\qquad \frac{\sum_{\substack{x \in S: x_{\Lambda_n^c} = x^1_{\Lambda_n^c}, \\ K \in \mathcal{K}(x), l(x_K) = 0}} p_{\Lambda_n}(x_{\Lambda_n} | x^1_{\Lambda_n^c})}{\sum_{x \in S: x_{\Lambda_n^c} = x^1_{\Lambda_n^c}} p_{\Lambda_n}(x_{\Lambda_n} | x^1_{\Lambda_n^c})} \leq \frac{\sum_{\substack{x \in S: x_{\Lambda_n^c} = x^1_{\Lambda_n^c}, \\ K \in \mathcal{K}(x), l(x_K) = 0}} p_{\Lambda_n}(x_{\Lambda_n} | x^1_{\Lambda_n^c})}{\sum_{\substack{x \in S: x_{\Lambda_n^c} = x^1_{\Lambda_n^c}, \\ K \in \mathcal{K}(x), l(x_K) = 0}} p_{\Lambda_n}(z(x)_{\Lambda_n} | x^1_{\Lambda_n^c})} \\
&\qquad\qquad \leq \exp(-\beta u(K)) \quad (5)
\end{aligned}
$$

which is precisely the estimate used in Griffiths (1964) version of Peierls proof of phase transition for the Ising model. By following Peierls argument one obtains that $P^0(X_0 = 1) < P^1(X_0 = 1)$ when $\beta$ is sufficiently large, whereby it is demonstrated that phase transition occurs. For large $\beta$ where phase transition occurs, $P^0$ and $P^1$ may be considered as perturbations of the point mass measures on $x^0$ and $x^1$, respectively, in the sense that typical realizations of $P^0$ and $P^1$ are dominated by zeros or ones, respectively, with only a small fraction of scattered pixels equal to the other colour.

For a given value of $\gamma > 0$ it is of course of interest to estimate how large $\beta$ needs to be in order that phase transition occurs. It would also be interesting to know whether there as for the Ising model is a sharp transition from short to longe range correlations when $\beta$ passes from subcritical values to supercritical values. These questions will however not be pursued here.

# 4 Simulation: A Metropolis-Hastings algorithm with simultaneous updating of pixels in connected components

In Møller & Waagepetersen (1996) a single-site updating Metropolis chain was used for the simulation of the MCCF's applied as models for binary images. In theory this chain converges to the right stationary distribution but for the penalized Ising model (2) the modes which are the constant black or white images cause problems. For large values of $\gamma$ and positive values of $\beta$ a "typical" realization $y$ of the penalized Ising model contains a few large components. Suppose that $y_1, \dots, y_n$, $n \geq 2$, is a sequence of images such that $y_1$ is equal to the typical image $y$, that the image $y_n$ is one of the modes, and that for each $i = 1, \dots, n-1$, $y_i$ and $y_{i+1}$ differs at most at one pixel. Then some of the $y_i$'s will contain small components and thus have a very small probability under the model (2). This makes it difficult for a single-site updating chain to move between the nonconstant images and the modes.

The Metropolis-Hastings kernel designed in this section should be better suited for making transitions between the modes and the nonconstant images than single-site updating kernels. Suppose that $x \in S$ is the current image. With a certain probability $q$ it is proposed to remove a randomly chosen component in $\mathcal{K}(x)$ and otherwise it is proposed to insert a randomly generated component $K$ within another randomly chosen component $\tilde{K} \in \mathcal{K}(x)$. First the stochastic procedure used to generate a connected component $K \in \mathcal{K}$ within another connected component $\tilde{K} \in \mathcal{K}$ is described.

## 4.1 A stochastic procedure to generate connected components

The stochastic procedure must be deviced so that it is easy to calculate the probability of generating a given component $K \in \mathcal{K}$, since this probability is needed for the calculation of the Metropolis-Hastings ratio. Let $x \in S$ be the current image, and let $z$ be a realization of a binary white noise $Z = (Z_i)_{i \in I}$ (i.e. $Z_i$ *iid* and $P(Z_i = 1) = 1 - P(Z_i = 0) = r$ for some $0 \leq r \leq 1$). One possibility is to sample uniformly a pixel $i \in I$, let $\tilde{K} \in \mathcal{K}(x)$ be determined by $i \in \tilde{K}$, and finally return the component $K \in \mathcal{K}^{(1)}(z_{\tilde{K}})$ containing $i$, if such a component exists, and return $\emptyset$ otherwise. It is then easy to see that the probability of generating a nonempty $K \subseteq \tilde{K} \in \mathcal{K}$ is

$$\frac{a(K)}{a(I)} r^{a(K)} (1-r)^{\#\{(i,j) \in K \times \tilde{K} \setminus K \mid i \sim j\}}.$$

A component generated this way will however typically contain many small holes corresponding to small connected components, so that the Metropolis-Hastings acceptance probability for inserting such a component becomes very small. It is therefore necessary to refine the recipe so that components are generated for which the resulting proposal images are more likely to get accepted.

First some definitions are needed: for any $A \subseteq \mathbf{Z}^2$ define $z(A) \in V^{\mathbf{Z}^2}$ by

$$z(A)_l = \begin{cases} 1 & l \in A, \\ 0 & \text{otherwise} \end{cases}, \; l \in \mathbf{Z}^2,$$

and let $A^c = \mathbf{Z}^2 \setminus A$. For any finite $A \subseteq \mathbf{Z}^2$ and $j \in \mathbf{Z}^2$ write $j \underset{z(A)}{\sim} \infty$ if $j \underset{z(A)}{\sim} \partial([-n,n]^2 \cap \mathbf{Z}^2)^c$ for any $n \geq 0$ with $A \subseteq ([-n,n]^2 \cap \mathbf{Z}^2) \setminus \partial([-n,n]^2 \cap \mathbf{Z}^2)^c$. Here $\partial A^c$ should be read as $\partial(A^c)$ for $A \subseteq \mathbf{Z}^2$. Define further,

$$S(A) = \{i \in A \,|\, \exists\, j \in (\partial\{i\}) \cap A^c : j \underset{z(A)}{\sim} \infty\} \tag{6}$$

$$Int(A) = Int\big(S(A)\big) = \{j \in S(A)^c | j \underset{z(S(A))}{\not\sim} \infty\} \tag{7}$$

$$Ext(A) = Ext\big(S(A)\big) = \big(S(A) \cup Int(A)\big)^c \tag{8}$$

$$C(A) = S(A) \cup Int(A). \tag{9}$$

For a component $K$, $S(K)$ consists of those pixels in $K$ which are $\sim$-neighbours to pixels "outside" $K$, while $Int(K)$ consists of pixels in the "interior" given by $K$, see Figure 1.

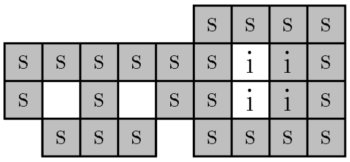

Figure 1: Pixels marked with "s" belongs to $S(K)$ for the component $K$ consisting of the gray pixels. The pixels in $Int(K)$ are marked with "i" ($\sim$ is the second order relation).

Given an $x \in S$ I then propose the following procedure:

1. A pixel $i \in I$ is generated from the uniform distribution on $I$.

2. The component $\tilde{K} \in \mathcal{K}(x)$ is determined by $i \in \tilde{K}$.

3. A realization $z$ of a binary white noise with rate $r$ is generated on $\tilde{K}$, and if $z_i = 0$ then the value of $z_i$ is changed to 1.

4. The component $K' \in \mathcal{K}^{(1)}(z_{\tilde{K}})$ is determined by $i \in K'$.

5. $K = C(K')$ is returned if $i \in S(K)$, otherwise $K = \emptyset$ is returned.

A component $K \in \mathcal{K}$ generated by this procedure will contain no holes, so that an insertion of $K$ is more likely to get accepted. For any $x \in S$ and any component $K \in \mathcal{K}$ with $K \subseteq \tilde{K}$ for a $\tilde{K} \in \mathcal{K}(x)$, the probability of generating $K$ given $x$ is

$$Q_0(x, K) = 1\big(K = C(K)\big)\frac{a(S(K))}{a(I)} r^{a(S(K))-1}(1-r)^{\#\{(i,j)\in K\times\tilde{K}\setminus K | i\sim j\}}. \tag{10}$$

A proof of this result is given in the appendix. Note that $S(K) \in \mathcal{K}$ by Lemma 4 in the appendix, so that any $K \in \mathcal{K}$ with $K = C(K)$ can be generated by the stochastic procedure described in this section. For Lemma 4 to hold it is important that $\sim = \sim_2$. Figure 1 is e.g. a counter example when $\sim = \sim_1$.

## 4.2 The proposal kernel

The proposal kernel for the Metropolis-Hastings algorithm with simultaneous updating of pixels in connected components can now be specified. Let $x \in S$ be the current image.

(a) With a probability $0 \leq q < 1$ choose a component $K \in \mathcal{K}(x)$ according to a probability distribution $(p_x(K))_{K\in\mathcal{K}(x)}$.

(b) With probability $1-q$ generate a component $K$ from the kernel $Q_0(x,\cdot)$ given by (10).

The proposal image $y$ is then given by

$$y_l = \begin{cases} (x_l+1) \mod 2, & l \in K, \\ x_l, & l \in I \setminus K. \end{cases}$$

In case (a) this means that a component is removed, while in case (b) a component is inserted. However, if in case (b), $K = \tilde{K}$, then the effect of (b) is indistinguishable from (a) (this explains why we can take $q = 0$ and still maintain reversibility). The probability distribution $(p_x(K))_{K\in\mathcal{K}(x)}$ may e.g. be chosen so that

$$p_x(K) = 1/|\mathcal{K}(x)|,$$

i.e. so that each component is equally likely to be chosen, but $p_x(K)$ may perhaps more naturally depend on the size of $K$ so that e.g.

$$p_x(K) \propto a(I) - a(K),\ K \in \mathcal{K}(x),$$

whereby small components are most likely to be selected.

Consider $x, y \in S$ and assume that $K = \{i \in S | x_i \neq y_i\} \neq \emptyset$. The proposal kernel $Q$ is then given by

$$Q(x,y) = \begin{cases} qp_x(K) + (1-q)Q_0(x,K), & K \in \mathcal{K}(x), \\ (1-q)Q_0(x,K), & K \subsetneq \tilde{K} \in \mathcal{K}(x),\ K \in \mathcal{K}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $p(\cdot)$ denote the density of the target distribution for the simulation. In the case where $K \in \mathcal{K}(x)$ and $K = C(K)$, the Metropolis-Hastings ratio becomes

$$mh(x,y) = \frac{p(y)Q(y,x)}{p(x)Q(x,y)} = \frac{p(y)(1-q)\frac{a(S(K))}{a(I)} r^{a(S(K))-1}(1-r)^{\#\{(i,j)\in K\times\tilde{\tilde{K}}\setminus K | i\sim j\}}}{p(x)(qp_x(K)+(1-q)\frac{a(S(K))}{a(I)} r^{a(S(K))-1})}, \quad (11)$$

where $\tilde{\tilde{K}} = \cup_{K'\in\mathcal{K}(x):K'\uparrow\tilde{K}} K'$. If $C(K) \neq K$ then $Q_0(y,K) = 0$ so that also $mh(x,y) = 0$.

In the case where $K \subsetneq \tilde{K} \in \mathcal{K}(x)$, $K \in \mathcal{K}$, and $C(K) = K$ there are two possibilities. In the first case $K$ is an "isolated" component within $\tilde{K}$, i.e. there does not exist $(i,j) \in K \times I \setminus \tilde{K}$ with $i \sim j$. Then

$$mh(x,y) = \frac{p(y)(qp_y(K)+(1-q)\frac{a(S(K))}{a(I)} r^{a(S(K))-1})}{p(x)(1-q)\frac{a(S(K))}{a(I)} r^{a(S(K))-1}(1-r)^{\#\{(i,j)\in K\times\tilde{K}\setminus K | i\sim j\}}}. \quad (12)$$

Otherwise if there exists $(i,j) \in K \times I \setminus \tilde{K}$ with $i \sim j$, then

$$mh(x,y) = \frac{p(y)(1-r)^{\#\{(i,j)\in K\times I\setminus\tilde{K} | i\sim j\}}}{p(x)(1-r)^{\#\{(i,j)\in K\times\tilde{K}\setminus K | i\sim j\}}}. \quad (13)$$

## 4.3 Experiments

By construction, the Markov chain given by the Metropolis-Hastings kernel described in section 4 will converge to the stationary distribution $p(\cdot)$ for any values of $0 \leq q < 1$ and $0 < r < 1$. In practice it turns out that the convergence properties of the Markov chain are very sensitive with respect to the choice of $r$. I have e.g. considered simulation of the target distribution given by the penalized Ising model (2) defined on a $67 \times 62$ lattice with $\beta = 0.74$ and $\gamma = 10$. This value of $\beta$ is far above the critical value $\beta_c$ for the second order Ising model for which $\beta_c \in ]0.35, 0.40[$. I used a Markov chain where each update was given by a combination of an update using the kernel constructed in section 4 and subsequently an update given by a systematic scan of single-site Metropolis-Hastings updates through all the pixels of the image. The initial image was chosen to be constant white. Figure 2 (left) shows time series for the number of black pixels after every tenth of 50000 combined updates of the Markov chain obtained with $q = 0.5$ and $r = 0.25$. The time series seems to be in equilibrium and moves between constant black and constant white images. Figure 2 (right) shows time series for the same statistic but now obtained with $q = 0.5$ and $r = 0.33$. In this case the Markov chain bounces between almost black and almost white images. It would be very interesting if the results obtained with $r = 0.25$ could be trusted, since it then appears that $\beta$ need to be very large before phase transition occurs for the penalized Ising model when $\gamma$ is large. The chain obtained with $r = 0.33$ in contrast resembles the phase transition behaviour of the Ising model.

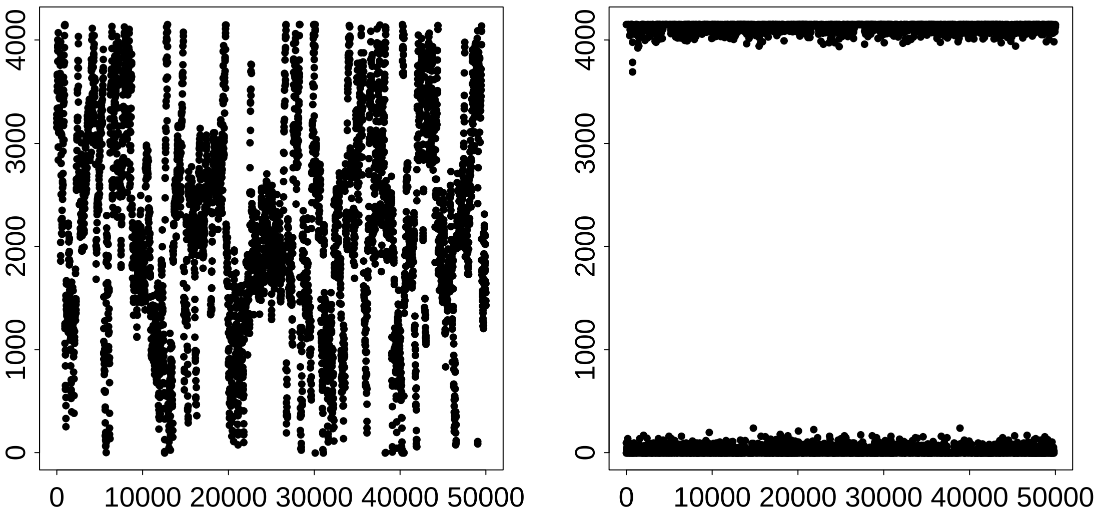


Figure 2: Timeseries for the number of black pixels. Left: $r = 0.25$. Right: $r = 0.33$.

The different behaviour of the two chains is a worry, but may be explained by considering the typical size of the components generated by stochastic procedure described in section 4.1. With $r = 0.25$ the mean size for components accepted for insertion during the run of the Markov chain is 138 pixels, but for $r = 0.33$ these components are either very small (around 8 pixels) or very big (around 4000 pixels) in which case they fill almost all of the image. This explains why the chain jumps between almost black and almost white images when $r = 0.33$, and this perhaps also makes it difficult for the chain to move to other parts of the statespace.

Another point is that the Metropolis-Hastings ratio for insertion of components is usually very large, and, correspondingly, the acceptance probability for removing components is practically zero except when it is proposed to remove very small components. In general it seems desirable that the Metropolis-Hastings ratio is of moderate size. One can e.g. consider a Metropolis-Hastings chain for sampling the uniform distribution on $\{0, 1\}$, where the proposal probabilities are given by $q(0, 1) = q(0, 0) = 0.5$, $q(1, 0) = \epsilon$, and $q(1, 1) = 1 - \epsilon$ with $1 \geq \epsilon > 0$. This chain will mix badly for very small values of $\epsilon$ for which the values of the Metropolis-Hastings ratio becomes extreme. For the chains considered here for simulation of the penalized Ising model, the Metropolis-Hastings ratios were most well-behaved when $r = 0.25$ was used.

Unfortunately it is not easy for given $\beta$ and $\gamma$ to choose a value of $r$, so that the Metropolis-Hastings ratios given by (11), (12), and (13) becomes moderate (i.e. not extremely small or extremely large), and so that furthermore the size of a typical component generated by the proposal kernel matches the size of typical components in the realizations of the target distribution. In order to simplify

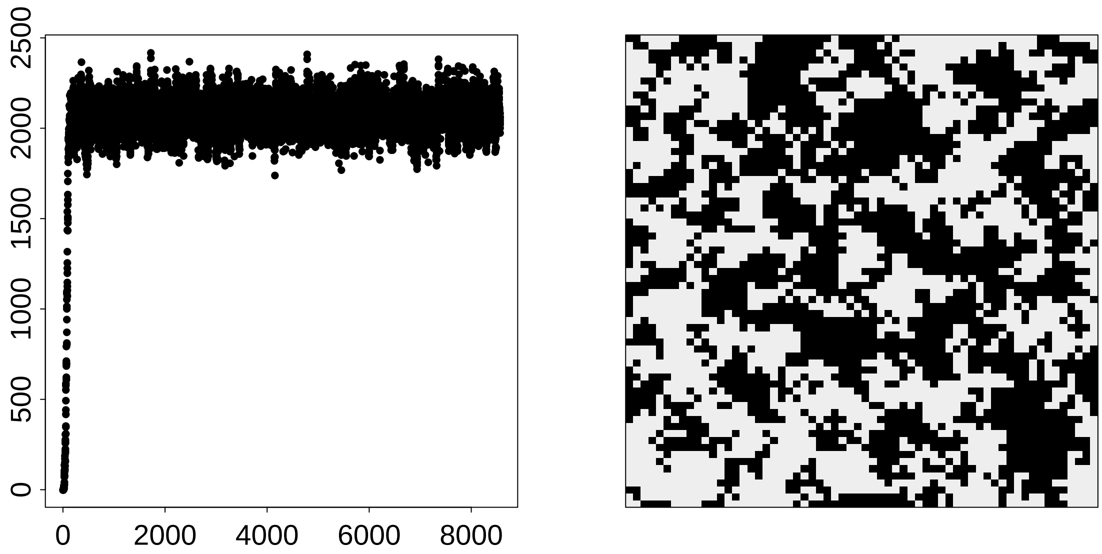


Figure 3: Left: Timeseries for the number of black pixels, $\beta = 0.288$, $\gamma = 10$, $q = 0$, and $r = 0.25$. Right: Realization of the penalized Ising model with $\beta = 0.288$ and $\gamma = 10$.

matters I have considered the case $q = 0$. Thereby the Metropolis-Hastings ratio for insertion of a component simplifies to (13). Since

$$\frac{p(y)}{p(x)} = \exp\Big(-\beta\big(\#\{(i,j) \in K \times \tilde{K} \setminus K \,|\, i \sim j\} - \#\{(i,j) \in K \times I \setminus K \,|\, i \sim j\}\big)\Big) \cdot \\ \exp\big(\gamma/a(\tilde{K}) + \sum_{\substack{K' \in \mathcal{K}(x)\setminus \tilde{K}:\\ K' \uparrow \tilde{K}}} \gamma/a(K') - \sum_{K' \in \mathcal{K}(x_{\tilde{K}\setminus K})} \gamma/a(K') - \gamma/a(\tilde{\tilde{K}})\big)$$

where $\tilde{\tilde{K}} = \cup_{K' \in \mathcal{K}(x): K' \uparrow \tilde{K}} K'$, then if $-\beta = \log(1-r)$, the Metropolis-Hastings ratio reduces to

$$\exp\big(\gamma/a(\tilde{K}) + \sum_{\substack{K' \in \mathcal{K}(x)\setminus \tilde{K}:\\ K' \uparrow \tilde{K}}} \gamma/a(K') - \sum_{K' \in \mathcal{K}(x_{\tilde{K}\setminus K})} \gamma/a(K') - \gamma/a(\tilde{\tilde{K}})\big),$$

which will be of moderate size if the components appearing in the exponent are not too small compared to $\gamma$. The choice of $r = 1 - \exp(-\beta)$ is however restrictive. If e.g. $\beta = 0.431$ then $r = 0.35$, and with this value of $r$ rather big components will be generated. Such components $K$ will often divide $\tilde{K} \setminus K$ into a number of small components, such that the proposed image has a little chance of being accepted. For values of $\beta = 0.288$, $\gamma = 10$, $r = 0.25$, $q = 0$, and $\beta = 0.356$, $\gamma = 10$, $r = 0.30$, $q = 0$, the algorithm worked satisfactory. A time series for the number of black pixels in the case $\beta = 0.288$, $\gamma = 10$, $r = 0.25$, and $q = 0$, is shown in Figure 3 together with a realization of the penalized Ising model. I also tried to use a single-site algorithm. This chain never left the initial constant image. Note that the proposed images are always accepted for the Ising model (i.e. the case $\gamma = 0$) when $q = 0$.

### 4.4 A lattice process with compact objects on a background

The ideas introduced in section 4.1 may also be used for construction of a lattice process where black connected components without holes appear on a white background and where the probabilities of the process are given explicitly. Let $Z$ be a realization of a binary white noise on $I$ with rate $0 \le r \le 1$, and let $X$ be given by $X_i = 1$ if $i \in C(K)$ for a $K \in \mathcal{K}(Z)$ and $X_i = 0$ otherwise. Then for a given $x \in S$,

$$P(X = x) = r^{a\left(\cup_{K \in \mathcal{K}(x)} S(K)\right)} (1 - r)^{a(I \setminus \cup_{K \in \mathcal{K}(x)} K)}.$$

## 5 Estimation of the normalizing constant

This section is concerned with estimation of the normalizing constant and the probability of the modes for the penalized Ising model which was fitted to the $67 \times 62$ rock sample image in Møller & Waagepetersen (1996). The maximum likelihood estimates obtained in Møller & Waagepetersen (1996) were $\hat{\beta} = 0.966$ and $\hat{\gamma} = 17$. In Møller & Waagepetersen (1996) the connected components were defined with respect to $\sim_2$ but $u(\cdot)$ was actually the first order boundary (i.e. for a component $K$, $u(K) = \#\{(i,j) \in K \times I \setminus K \mid i \sim_1 j\}$).

The method of reverse logistic regression (Geyer, 1991) is used to estimate the normalizing constant. This technique is here applied to conditional densities

$$\tilde{f}_i(x) \propto h_i(x) 1(x \text{ not constant}),\ i = 1, \ldots, m,$$

where $f_i(x) \propto h_i(x)$, $x \in S$, $i = 1, \ldots, m$, are Markov random field approximations to the penalized Ising model for varying parameter values $(\beta_i, \gamma_i)$, i.e.,

$$h_i(x) = \exp\big(-( \sum_{K \in \mathcal{K}(x)} \frac{\beta_i}{2} u(K) + 1(a(K) < \delta) \frac{\gamma_i}{a(K)} )\big),\ x \in S,\ i = 1, \ldots, m, \tag{14}$$

where $\delta$ is a positive parameter which determines the size of the neighbourhood of $f_i$. The densities $\tilde{f}_i$, $i = 1, \ldots, m$, can be simulated by ordinary single-site updating algorithms, and the Markov property of $f_i$, $i = 1, \ldots, m$, simplifies the computations.

Let $\tilde{c}_i = \big(\sum_{x \text{ not constant}} h_i(x)\big)^{-1}$ be the normalizing constant of $\tilde{f}_i$ for $i = 1, \ldots, m$. By reverse logistic regression, $\log \tilde{c}_1, \ldots, \log \tilde{c}_m$ can be estimated up to an additive constant $\tilde{k}$ from samples of $\tilde{f}_1, \dots, \tilde{f}_m$. If e.g. $\tilde{f}_1$ is the uniform density for which $\tilde{c}_1$ is known, then estimates of $\tilde{c}_1, \dots, \tilde{c}_m$ can be derived from estimates of $\tilde{k} + \log \tilde{c}_1, \dots, \tilde{k} + \log \tilde{c}_m$. The supports of $\tilde{f}_1, \dots, \tilde{f}_m$ need not overlap, but numerical problems occur in practice if the supports are too different. Note also that

$$(\frac{\tilde{c}_i}{\tilde{c}_j})^{-1} = \sum \frac{h_i(x)}{h_j(x)} \tilde{f}_j(x), \tag{15}$$

whenever $\tilde{f}_i(x) > 0 \Rightarrow \tilde{f}_j(x) > 0$, $x \in S$. This means that the ratio $\tilde{c}_j/\tilde{c}_i$ can be estimated by $(1/n)\sum_{l=1}^{n} h_i(X_l)/h_j(X_l)$ where $\{X_l\}_{l=1}^{n}$ is an ergodic sample from $\tilde{f}_j$. In order to obtain an accurate estimate by using (15) it is necessary that $\tilde{f}_j$ places much probability mass in the part of the state space where $\tilde{f}_i$ is concentrated.

Let $\tilde{c}$ be the normalizing constant of the fitted penalized Ising model conditioned away from the constant images. The normalizing constant $c$ for the fitted penalized Ising model is then estimated as follows:

1. Estimates of $\tilde{c}_1, \dots, \tilde{c}_m$ are obtained by reverse logistic regression applied to pairs $\tilde{f}_i, \tilde{f}_{i+1}$, $i = 1, \dots, m-1$.
2. An estimate $\hat{\tilde{c}}$ is obtained by using (15), $\hat{c}_m$, and by reusing the sample generated from $\tilde{f}_m$.
3. $\hat{c}^{-1} = \hat{\tilde{c}}^{-1} + 2\exp\bigl(-\hat{\gamma}/a(I)\bigr)$.

I chose $m = 20$, $\beta_1 = \gamma_1 = 0$, $\beta_{20} = \hat{\beta}$, $\gamma_{20} = \hat{\gamma}$, and intermediate values of $\beta_i, \gamma_i$, $i = 2, \dots, 19$, and the estimate of $c$ is $\hat{c} = \exp(-29.8)$, so that the estimate of the probability of a constant image is $\hat{c}\exp\bigl(-\hat{\gamma}/a(I)\bigr) = 10^{-13}$. This result should be regarded with some caution since many Monte Carlo approximations were required, but the result suggests that the probability of the constant images is very small.

# 6 Discussion

In section 4 a MCMC algorithm with simultaneous updates of connected components of pixels was constructed. The convergence properties of this algorithm should in theory be much better than those of an ordinary single-site algorithm when applied to simulation of the penalized Ising model (2). The parameters of the proposal kernel however need to be chosen such that components of a reasonable size are generated, and furthermore such that the values of the Metropolis-Hastings ratio are moderate. This restricts the practical applicability of the algorithm.

For the particular set of parameter values considered in Møller & Waagepetersen (1996) a single-site algorithm will in practice not be able to move to the modes of the penalized Ising model. From the estimates calculated in section 5 it however appears, that the probability of the modes is very small. It may therefore still be a valid conjecture, that phase transition does not occur for the fitted penalized Ising model in Møller & Waagepetersen (1996), even though this conjecture was based on single-site simulations. The result of section 3 on the other hand shows, that phase transition will always occur when the parameter $\beta$ is chosen sufficiently large.

# A Appendix

In this appendix a proof of the result (10) will be given. First some preliminary results are considered. Recall that $I \subseteq \mathbf{Z}^2$ and $\sim = \sim_2$, so that $\mathcal{K}$ is the set of second order connected components. Define for any $n \geq 1$,

$$R_n = [-n, n]^2 \cap \mathbf{Z}^2$$

and define for every finite $A \subseteq I$, the number $N(A) = \min\{n \in \mathbf{N} | A \subseteq R_n \backslash \partial R_n^c\}$. It is assumed that $I$ is finite so that $a(K) < \infty$ for any $K \in \mathcal{K}$. A tuple $(i_1, \ldots, i_n) \in I^n$, $n \geq 1$, is said to be a path if $i_l \sim i_{l+1}$, $l = 1, \ldots, n-1$. A path $(i_1, \ldots, i_n) \in I^n$, $n \geq 1$, is said to intersect a set $A \subseteq I$ if $A \cap \{i_1, \ldots, i_n\} \neq \emptyset$, and the path is contained in $A$ if $\{i_1, \ldots, i_n\} \subseteq A$.

**Lemma 1** *Suppose $K \in \mathcal{K}$. Then $K \subseteq C(K)$.*

*Proof:* Suppose $j \in K$. If $j \in S(K)$ then clearly $j \in C(K)$, so suppose that $j \notin S(K)$, that is, $\nexists l \in (\partial\{j\}) \cap K^c : l \underset{z(K)}{\sim} \infty$. Suppose further that $(i_1, \ldots i_n)$, $n \geq 2$, is an arbitrary path with $j \sim i_1$ and $i_n \in \partial R^c_{N(K)}$. Then $K \cap \{i_1, \ldots, i_n\} \neq \emptyset$ and $i_n \notin K$. Let $m = \min\{q \in \{1, \ldots, n\} \mid i_k \in K^c \forall\, k = q, \ldots, n\}$. Then $i_{m-1} \in S(K)$ so that $j \in Int(K) \subseteq C(K)$. $\square$

**Lemma 2** *Suppose $K \in \mathcal{K}$. Then S(K)=S(C(K)).*

*Proof:* We show first that $S(C(K)) \subseteq S(K)$. Assume that $j \in S(C(K))$. This means that $\exists l \in (\partial\{j\}) \cap C(K)^c : l \underset{z(C(K))}{\sim} \infty$. Note that $S(C(K)) \subseteq C(K) = S(K) \cup Int(S(K))$. Assume in order to obtain a contradiction that $j \in Int(S(K))$, that is, $j \in S(K)^c$ and $j \underset{z(S(K))}{\not\sim} \infty$. Then any path $(i_1, \ldots, i_n)$, $n \geq 2$, with $j \sim i_1$ and $i_n \in \partial R^c_{N(K)}$ intersects $S(K)$ and thereby also $C(K)$. This contradicts $j \in S(C(K))$.

To show that $S(K) \subseteq S(C(K))$ assume that $j \in S(K)$ so that there exists $l \in (\partial\{j\}) \cap K^c : l \underset{z(K)}{\sim} \partial R^c_{N(K)}$. This means that there exists a path

$(\{i_1,\dots,i_n)n \geq 1$, contained in $K^c$ with $l = i_1$ and $i_n \in \partial R^c_{N(K)}$. Since $i_p \underset{z(K)}{\sim} \partial R^c_{N(K)} \Rightarrow i_p \underset{z(S(K))}{\sim} \partial R^c_{N(K)}$, $p = 1,\dots,n$, we get that $\{i_1,\dots,i_n\} \subseteq (Int(K) \cup S(K))^c = C(K)^c$. Thereby $l \underset{z(C(K))}{\sim} \partial R^c_{N(K)}$, $l \in C(K)^c$, and since $j \in S(K) \subseteq K \subseteq C(K)$ by Lemma 1, we get $j \in S(C(K))$. $\square$

Before we consider Lemma 4 some new definitions and a Lemma 3 is needed. For any $A \subseteq I$ define

$$\partial_{ext}(A) = \{l \in \partial A | l \underset{z(A)}{\sim} \infty\},$$

and for a $z = (z_1, z_2) \in \mathbf{Z}^2$ let $h(z) = z_1$ and $v(z) = z_2$. Suppose $i \sim_1 j$. We then write $i \sim_1^h$ if $i$ and $j$ are "horizontal" neighbours, i.e. if $v(i) = v(j)$. If $v(i) \neq v(j)$ then $i \sim_1^v$. A path $(i_1,\dots,i_n)$ is further said to be a cycle if $i_1 \sim i_n$. Suppose that $A \in \mathcal{K}_3$ is a $\sim_3$-connected component (the $\sim_3$-neighbours of a pixel $i$ is $\cup_{j\in\partial\{i\}}\partial\{j\}$; an example of a $\sim_3$-connected component is given in Figure 4). The following Lemma 3 is concerned with the existence of a cycle $C = (i_1,\dots,i_n)$ which obeys the following conditions:

i) $i_l \sim_1 i_{l+1}$, $l = 1,\dots,n$ (here and in the following addition is mod $n$). That is, $C$ is a first order path.

ii) $C$ surrounds $A$, i.e. any path $(l_1,\dots,l_m)$ with $l_1 \in A$ and $l_m \in \partial R^c_{N(A)}$ intersects $\{i_1,\dots,i_n\}$.

iii) $C$ is contained in $\partial_{ext}(A)$.

iv) Suppose that $i_l \sim_1^v \dots \sim_1^v i_{l+k}$ for some $l = 1,\dots,n$, and $k = 0, 1$ or $2$. Then $k_1 \sim_1^h \{i_l,\dots,i_{l+k}\}$ and $k_2 \sim_1^h \{i_l,\dots,i_{l+k}\}$ implies $h(k_1) = h(k_2)$.

v) Same as in iv) but with $v$ and $h$ replaced by $h$ and $v$, respectively.

vi) Suppose $i_l \sim_1^h i_{l+1}$ for some $l = 1,\dots,n$. Then $\exists j \in A$ where $j \sim_1^v \{i_l, i_{l+1}\}$, $v(j) = v(i_l) - 1$ implies $h(i_{l+1}) = h(i_l) - 1$, and $v(j) = v(i_l) + 1$ implies $h(i_{l+1}) = h(i_l) + 1$.

vii) Suppose $i_l \sim_1^v i_{l+1}$ for some $l = 1,\dots,n$. Then $\exists j \in A$ where $j \sim_1^h \{i_l, i_{l+1}\}$, $h(j) = h(i_l) - 1$ implies $v(i_{l+1}) = v(i_l) + 1$, and $h(j) = h(i_l) + 1$ implies $v(i_{l+1}) = v(i_l) - 1$.

An example of a set with a surrounding cycle for which i)-vii) hold is given in Figure 4. The condition iv) e.g. means, that the pixels $i_6$ and $i_{23}$ can not be consecutive members of the cycle surrounding the set $A$ in Figure 4.

**Lemma 3** *Suppose $A \in \mathcal{K}_3$. Then there exists a cycle $C$ for which the conditions i)-vii) holds.*

*Proof:* The proof is by induction in $a(A)$. If $a(A) = 1$ the result clearly holds. Suppose now that $a(A) = n \geq 2$ and that the result holds for all $A' \in \mathcal{K}_3$ where $a(A') < n$. Let $t_2(A) \in \mathbf{Z}$ be given by $t_2(A) = \min\{k \in \mathbf{Z} \,|\, \exists l \in \mathbf{Z} : (l,k) \in A\}$ and let $t_1(A) = \max\{l \in \mathbf{Z} \,|\, (l, t_2(A)) \in A\}$. Then the pixel $t(A) = (t_1(A), t_2(A))$ is the rightmost pixel in the lowest row in $A$ and either $A \backslash t(A) \in \mathcal{K}_3$ or $A \backslash t(A) = A_1 \cup A_2$ where $A_i \in \mathcal{K}_3$, $i = 1, 2$. By considering the possible pixel configurations in the $\sim_3$-neighbourhood of $t(A)$ and by using the induction hypothesis the result follows. $\square$

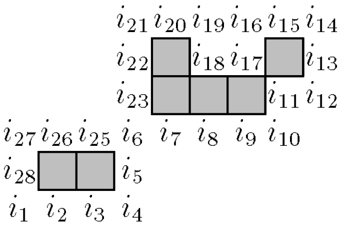


Figure 4: $A$ consists of the gray pixels. The surrounding cycle is $(i_1, \ldots, i_{28})$ where $i_6 = i_{24}$.

Lemma 1 and 2 are valid both when $\mathcal{K}$ is defined with respect to $\sim_1$ and $\sim_2$. In order to obtain (10) it is however crucial that $S(K) \in \mathcal{K}$ and this result stated in Lemma 4 does not hold if $\mathcal{K} = \mathcal{K}_1$, cf. the last remark in section 4.1.

**Lemma 4** *If $K \in \mathcal{K}$ then $S(K) \in \mathcal{K}$.*

*Proof:* Suppose $S(K) \notin \mathcal{K}$. Then $S(K) = S_1 \cup \cdots \cup S_m$ where $m \geq 2$, $S_p \in \mathcal{K}$, $p = 1, \ldots, m$, the $S_p$'s are disjoint, and $(\partial S_p) \cap S_q = \emptyset$ for each $p \neq q$. Define $t_1(K)$ and $t_2(K)$ as in the proof of Lemma 3, let $\tilde{t}(K) = (t_1(K), t_2(K) - 1)$, and let $S_l$ be determined by $\tilde{t}(K) \in C_l$ where $C_l$ is a cycle surrounding $S_l$ as in Lemma 3. There is then a $k \in \{1, \ldots, m\} \setminus \{l\}$, and a path $(i_1, \ldots, i_n)$, $n > 2$ such that $i_1 \in S_l$, $i_n \in S_k$ and $\{i_2, \ldots, i_{n-1}\} \subseteq K \setminus S(K)$. There is further a path $(l_1, \ldots, l_{n'})$ contained in $C_l$ where $i_2 = l_1$ and $\tilde{t}(K) = l_{n'}$, and where $(l_1, \ldots, l_{n'})$ does not intersect $S(K)$ (for this it is important that $\sim = \sim_2$). Let $M = \min\{q \in \{2, \ldots, n'\} \,|\, l_k \in K^c \forall k = q, \ldots, n'\}$. Then $l_{M-1} \in K$, $l_M \in K^c$, $l_M \sim l_{n'} = \tilde{t}(K) \underset{z(K)}{\sim} \infty$ and this is a contradiction since $l_{M-1} \notin \cup_{i=1}^{m} S_i = S(K)$. $\square$

Let for given $x \in S$, $\mathbf{K}$ denote the outcome of the stochastic procedure given in section 4.1, and let $J$ be the random variable corresponding to the initially sampled pixel in the stochastic procedure. For a connected component $K$ with

$C(K) = K$ and $K \subseteq \tilde{K} \in \mathcal{K}(x)$ the probability of generating $K$ is

$$P(\mathbf{K} = K) = \sum_{i \in S(K)} P(\mathbf{K} = K | J = i) P(J = i) = \frac{1}{a(I)} \sum_{i \in S(K)} P(\mathbf{K} = K | J = i) =$$

$$\frac{1}{a(I)} \sum_{i \in S(K)} \sum_{\substack{z \in S : z_i = 1 \text{ and} \\ \exists K' \in \mathcal{K}^{(1)}(z_{\tilde{K}}) : C(K') = K}} p_i(z_{-i}),$$

where $z_{-i} = (z_j)_{j \in I \setminus \{i\}}$ and $p_i(z_{-i}) = P(Z_{-i} = z_{-i})$ for $Z$ a binary white noise with rate $r$, i.e.,

$$p_i(z_{-i}) = r^{\#\{j \in I \setminus \{i\} | z_j = 1\}} (1 - r)^{\#\{j \in I \setminus \{i\} | z_j = 0\}}.$$

By Lemma 2 it is easily seen that

$$P(\mathbf{K} = K | J = i) = \sum_{\substack{z \in S : z_i = 1 \text{ and} \\ \exists K' \in \mathcal{K}^{(1)}(z_{\tilde{K}}) : C(K') = K}} p_i(z_{-i}) = \sum_{\substack{z \in S : z_i = 1 \text{ and} \\ \exists K' \in \mathcal{K}^{(1)}(z_{\tilde{K}}) : S(K') = S(K)}} p_i(z_{-i}).$$

The equality (10) therefore follows by showing

$$\exists K' \in \mathcal{K}^{(1)}(z_{\tilde{K}}) : S(K') = S(K) \Leftrightarrow z_l = 1,\ l \in S(K) \text{ and } z_l = 0,\ l \in (\partial K) \cap \tilde{K}.$$

Assume that $\exists K' \in \mathcal{K}^{(1)}(z_{\tilde{K}}) : S(K') = S(K)$. Since $S(K) = S(K') \subseteq K'$ it follows that $z_l = 1$, $l \in S(K)$. Furthermore, $l \in (\partial C(K')) \cap \tilde{K} \Rightarrow z_l = 0$, and $\partial C(K') = \partial K$ so that $z_l = 0$ for all $l \in (\partial K) \cap \tilde{K}$.

To show the other implication assume that $z_l = 1$, $l \in S(K)$ and $z_l = 0$, $l \in (\partial K) \cap \tilde{K}$. By Lemma 3, $S(K) \in \mathcal{K}$ so there exists $K' \in \mathcal{K}^{(1)}(z_{\tilde{K}})$ with $S(K) \subseteq K'$. We now need to show that $S(K') = S(K)$ and consider first $S(K) \subseteq S(K')$. Assume that $j \in S(K)$. This implies that $j \in K'$ and $\exists l \in (\partial\{j\}) \cap K^c$ : $l \underset{z(K)}{\sim} \partial R^c_{N(K)}$. Note that $K' \subseteq K$. This is easily seen because if there exists a $l \in K' \cap K^c$ then there is a path $(i_1, \ldots, i_n)$, $n \geq 2$ contained in $\tilde{K}$ such that $l = i_1$ and $i_n \in S(K)$ and where $i_1, \ldots, i_n \subseteq K'$. By letting $q = \max\{p = 1, \ldots, n | \, i_k \in K^c \, \forall \, k = 1, \ldots, p\}$ we get that $1 \leq q < n$ and $i_q \in \partial K$ which contradicts $z_{i_q} = 0$. Therefore $l \underset{z(K)}{\sim} \partial R^c_{N(K)} \Rightarrow l \underset{z(K')}{\sim} \partial R^c_{N(K')}$ so that $j \in S(K')$.

To show that $S(K') \subseteq S(K)$ assume that $j \in S(K')$ and remember that $S(K') \subseteq K' \subseteq K$. If there exists an $l \in \partial R^c_{N(K)}$ with $j \sim l$ then $j \in S(K)$. So assume that $j \not\sim \partial R^c_{N(K)}$ and assume in order to obtain a contradiction that $j \in Int(K)$. This means that for any path $(i_1, \ldots, i_n)$, $n \geq 2$, with $j \sim i_1$ and $i_n \in \partial R^c_{N(K)}$ we have that $\{i_1, \ldots, i_n\} \cap S(K) \neq \emptyset$ which implies $\{i_1, \ldots, i_n\} \cap K' \neq \emptyset$ whereby a contradiction is obtained.

## References

Besag, J. E., Green, P. J., Higdon, D. & Mengersen, K. (1995), 'Bayesian computation and stochastic systems', *Stat. Sci.* **10**, 3–66.

Georgii, H.-O. (1988), *Gibbs measures and phase transitions*, Walter de Gruyter, Berlin.

Geyer, C. J. (1991), Reweighting Monte Carlo mixtures, Technical Report 568, School of Statistis, University of Minnesota.

Griffiths, R. B. (1964), 'Peierls proof of spontaneous magnetization in the two dimensional Ising ferromagnet', *Phys. Rev. A* **136**, 437–438.

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

Møller, J. & Waagepetersen, R. (1996), Markov connected component fields, Research Report 341, Dept. of Theoretical Statistics, University of Aarhus. To appear in Adv. Appl. Prob.

Peierls, R. E. (1936), 'On Ising's ferromagnet model', *Proc. Camb. Phil. Soc.* **32**, 477–481.

Preston, C. (1976), *Random Fields*, Springer-Verlag, Berlin.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions', *Ann. Statist.* **22**, 1701–1728.