Boundary detection through dynamic polygons

A. Pievatolo* Peter J. Green†
CNR-IAMI, Italy. University of Bristol, UK.

Revised version, 30 May 1997

Abstract

A method for the Bayesian restoration of noisy binary images portraying an object with constant greylevel on a background is presented. The restoration, performed by fitting a polygon with any number of sides to the object outline, is driven by a new probabilistic model for the generation of polygons in a compact subset of $\mathcal{R}^2$, which is used as a prior distribution for the polygon. The simulation from the prior and the calculation of the *a posteriori* mean of greylevels are carried out through reversible jump MCMC computation. Two examples of restoration of synthetic images are presented.

Keywords and phrases: Bayesian object restoration, probability distribution of polygons, reversible jump MCMC computation.

1 Introduction

A difficult task in the restoration of an image portraying objects is the recovery of their boundaries. This can be done at a low level (e.g. using an edge detector, see Rosenfeld and Kak, 1982) or at a high level, trying to embed into the restoration process more complex information regarding the shape of objects, a form of pattern restoration.

In this article we will try to recover the boundary of a single object starting from a noisy version of it, relying on the prior information that the boundary is a closed and non-intersecting curve in the plane, that is, fitting a polygon to the object. In doing this, we can use more information than that required by a simple edge detector (such as the gradient of greylevel intensity, whose nature is essentially local), because a polygon, besides this information, conveys the idea of connectivity between the pairs of its vertices (not depending on their distance), and the global idea of closed curve.

Related approaches to this problem have been attempted by several authors. In all cases some kind of curve in the plane is used to introduce prior information about the outline of objects in a Bayesian image analysis framework.

The deformable template model of Grenander *et al.* (1991) is based upon the deformation of a template, a polygon with a fixed number of sides of variable length, to find the optimum

*CNR-IAMI, via Ampère 56, 20131 Milano, Italy.
E-mail: marco@iami.mi.cnr.it
†Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK.
E-mail: P.J.Green@bristol.ac.uk

fit to the object. Grenander and Miller (1994) generalize this to a variable number of objects and use jump-diffusion sampling to explore the state space of deformations.

In Qian and Mardia (1995) and in Mardia *et al.* (1997), a template model is employed for the recognition of multiple objects with occlusions.

Mardia and Qian (1993), and Qian *et al.* (1996) use a star-shaped polygon, that is, one for which the line segment joining each vertex to some fixed internal point lies entirely within the polygon. The main difference between these two papers is that in the first the restoration is chosen as the best possible fit to a known template belonging to a library of templates, while in the second the prior knowledge is contained in some smoothness properties of the object outline. In this respect, our paper is closer in spirit to the second of these.

Storvik (1994) reverses the problem, by considering a polygon with a variable number of sides with a fixed length, located on the pixel sides of the digitised image, thus obtaining a more flexible curve, but with a possibly unnecessarily high number of sides (such as stair-like configurations).

Our approach aims to achieve flexibility in shape modelling and parsimony at the same time, by considering polygons with any number of sides of any length. The choice of the number of sides of the polygon can be viewed as a choice between different models, for which the reversible jump MCMC algorithm of Green (1995) is particularly suitable.

A different way of modelling object contours makes use of marked point processes as priors, with the advantage that one can exploit the well known properties of these processes, but the disadvantage given by the limitations in the shape of objects one can consider (Baddeley and Van Lieshout, 1993). Rue and Syversveen (1995) and Rue and Hurn (1997) find a way around this difficulty by combining the template and marked point process approach to handle scenes containing variable numbers of objects of different types.

Other papers in which the emphasis is on boundary detection are the papers by Gavin and Jennison (1997), and by Helterbrand *et al.* (1994), but their perspective is more in the tradition of the line site process proposed by Geman and Geman (1984), with a constraint of connectedness imposed on the edges and no limitations on the number of objects. Gavin and Jennison's edges lie inside the pixels and are to be joined together to form closed curves, while Helterbrand *et al.* locate the edges on the pixel themselves and use the classical clique-based Markov random field models for lattices (see e.g. Besag, 1986) to put zero probability on non-connected configurations.

Finally, Clifford and Nicholls (1994) developed a method for the restoration of polygonal images, based on a model for random polygonal patterns in the plane by Arak *et al.* (1993) which will be sketched in the next Section. With the same theory, Nicholls (1996) builds a model for image analysis characterized by mosaic-like partitioning of the image domain with triangular tiles.

In the following Section we describe a statistical model that can produce polygons with any number of sides, and we show the connection between our model and the template model of Grenander *et al.* (1991). In Section 3 we review the reversible jump algorithm developed by Green (1995) which will be used for the restoration process and we show how to implement the algorithm in practice for this problem. Finally, in Section 4 we describe the result of an experiment into the restoration of two synthetic pictures. A brief discussion concludes the paper.

2 The model

We will tackle the problem of the restoration of an object via the Bayesian approach (see e.g. Besag, 1986) that is classical in statistical image analysis, focussing our attention on the outline of the object only, so that we can also regard our procedure as a binary image restoration.

We are given a noisy image digitised into $M \times N$ square pixels, whose centers have coordinates (i, j), $i = 1, \ldots, M$, $j = 1, \ldots, N$. The rectangle containing the image can thus be defined as $D = [0.5, M + 0.5] \times [0.5, N + 0.5]$. The object splits D into two sets, O and O^c, the set of the points inside the object and its complement. The set O is assumed to be simply connected. If we decide to regard a pixel as being inside O when this is true for its central point, we obtain the corresponding partition for the pixels; note that there is loss of information in discretization, so the same partition of the pixels can be produced by many different objects.

We assume that the observed image (data) has been produced by the following model:

$$z_{ij} = \nu_{ij}(O) + \varepsilon_{ij} \qquad i = 1, \ldots, M, \quad j = 1, \ldots, N \tag{1}$$

where

$$\nu_{ij}(O) = \begin{cases} \nu_1 & \text{if} \quad (i,j) \in O \\ \nu_0 & \text{if} \quad (i,j) \notin O \end{cases} \qquad \varepsilon_{ij} \ i.i.d. \ N(0, \sigma^2).$$

The values of ν_1, ν_0 and σ are assumed known in this article. This model has also been used in this form by Storvik (1994) and in a more general form by Qian *et al.* (1996), with two generic functions $f_1(i, j)$ and $f_0(i, j)$, whose value had to be estimated, in the place of ν_1 and ν_0.

Our aim is to recover the object boundary by fitting to it a polygon v. Following the Bayesian approach, we will define a prior distribution on the set of all polygons entirely contained in D, and for a restoration use some functional of the posterior distribution of v:

$$p(v \mid z) \propto p(z \mid v)\, \pi(v)$$

where $z = \{z_{ij},\ i = 1, \ldots, M,\ j = 1, \ldots, N\}$, $p(z \mid v)$ is the likelihood from model (1) when v is the true polygon, and $\pi(v)$ is the prior distribution of the polygon.

The steps needed to properly carry out the Bayesian analysis are to define a space of closed polygonals, construct a measure space based on it and then to take the same space restricted to polygons as a representation space.

We call *polygonal* with k sides ($k \geq 3$) a vector $v = (v_1, \ldots, v_k) \in D^k$, where all the pairs of vertices $(v_i, v_{(i \bmod k)+1})$, $i = 1, \ldots, k$, are connected. There are $2k$ vectors in D^k which represent the same polygonal, corresponding to k clockwise and k anticlockwise equivalent relabellings; we denote the set of these $2k$ labellings by L_k and its elements by $l = (l_1, \ldots, l_k)$. We also introduce the equivalence relationship $\simeq$ and declare $v \in D^k$ to be equivalent to $v' \in D^k$ ($v \simeq v'$) if there exist $l \in L_k$ such that $v_i = v'_{l_i}$, $i = 1, \ldots, k$. Our representation space for polygonals is now the quotient space $E_k = D^k / \simeq$ with elements $[v]$.

A measure space based on E_k is obtained as follows. Let us introduce the function $f : D^k \to E_k$ which maps v into $[v]$. Letting $\mathcal{B}(D^k)$ be the Borel σ-algebra on D^k, and letting $\mathcal{B}(E_k)$ be the σ-algebra generated by the class of sets $\{f(A)\}$, where $A \in \mathcal{B}(D^k)$, then it is easy to show that f is a measurable function from $(D^k, \mathcal{B}(D^k))$ to $(E_k, \mathcal{B}(E_k))$, it being sufficient to prove that, for any given $A \in \mathcal{B}(D^k)$, $f^{-1}(f(A))$ belongs to $\mathcal{B}(D^k)$. Thus we take as a measure space for k-sided polygonals the space $(E_k, \mathcal{B}(E_k), \mu_k)$, where

$$\mu_k(A) = \frac{1}{2k} \int_{D^k} I(v \in f^{-1}(A))\, dv_1 \ldots dv_k \qquad A \in \mathcal{B}(E_k). \tag{2}$$

Here $I(\cdot)$ is the indicator function of an event.

If we set $C_k = \{[v] \in E_k : [v] \text{ is a polygon}\}$ and prove that C_k is measurable, then a measure space of polygons is immediately defined as $(C_k, \mathcal{F}_k, \mu_k)$, where $\mathcal{F}_k = \{A \in \mathcal{B}(E_k) : A \subset C_k\}$. To prove this, we need the following notations:

$$g_{qq'}(v) = \min_{(i,j), i \neq j} d(q\, v_i + (1-q)\, v_{i+1}, q'\, v_j + (1-q')\, v_{j+1})$$

where $d(\cdot,\cdot)$ is the Euclidean distance, and $q, q' \in (0,1)$; for an integer m, $\{1/m, \ldots, 1-1/m\}$ is the set of rational numbers ranging between these two extremes by a step of $1/m$. The fact that C_k is measurable comes from considering that

$$f^{-1}(C_k) = \bigcap_{m \geq 1} \bigcup_{n \geq 1} \bigcap_{q,q' \in \{\frac{1}{m}, \ldots, 1-\frac{1}{m}\}} \left\{ v \in D^k : g_{qq'}(v) > \frac{1}{n} \right\}.$$

Indeed, being $g_{qq'}(\cdot)$ a continuous function from D^k into $\mathcal{R}$, $g_{qq'}^{-1}(1/n, +\infty)$ is open, so it is Borel, and $f^{-1}(C_k)$ itself is Borel, having been obtained by countable unions and intersection of Borel sets.

The measure space for all polygons with any number of vertices is thus $(C, \mathcal{F}, \mu)$, where $C = \bigcup_{k=3}^{\infty} C_k$, $\mathcal{F} = \biguplus_{k=3}^{\infty} \mathcal{F}_k$, and, for a given $A \in \mathcal{F}$, $\mu(A) = \sum_{k=3}^{\infty} \mu_k(A)$. We say that $A \in \mathcal{F}$ if $A = \cup_{k=3}^{\infty} A_k$, $A_k \in \mathcal{F}_k$ for all $k \geq 3$.

We do not know if $\mu(C)$ is finite, but we can define a prior density on C with respect to $\mu(\cdot)$ which we can show to be proper:

$$\pi([v]) \propto \exp\{-\beta k^\gamma - \frac{\alpha}{k} \sum_{i=1}^{k} (\theta_i(v) - \pi)^2\} \tag{3}$$

where α and β are non-negative, γ is a parameter to be specified later, and $\theta_i(v)$ represents the angle interior to the i-th vertex of $[v]$ in radians.

This density is invariant with respect to the position, the orientation and the size of the polygon, so it is a shape density. The term containing the angles penalises strongly irregular polygons and its minimum points, for fixed k, are all the regular polygons in D with k sides. It is also evident that the minimised value of this term tends to zero as $k \to \infty$, so, if it was the only term in the distribution, the upper limit points of the distribution would be all the circles in D. If γ was positive, the term βk^γ would increasingly penalise an increasing number of vertices, contributing to assigning a higher probability to "parsimonious" polygons.

We now have to look for some sufficient conditions on γ for the prior distribution to be proper, because otherwise in this particular case the posterior distribution will be improper. ¿From now on we will make a constant use of integrations with respect to μ and with respect to Lebesgue measure; in order to keep the notation simple, we will always use v as integration variable even when $[v]$ would be needed, it being clear which space one is integrating over. Denoting the right member of (3) by $\exp\{-U(v)\}$, we have

$$
\begin{aligned}
\int_C \exp\{-U(v)\}\,\mu(dv) &= \sum_{k=3}^{\infty} \frac{1}{2k} \int_{D^k} I(v \in C_k)\, \exp\{-U(v)\}\, dv_1 \ldots dv_k \\
&\le \sum_{k=3}^{\infty} \frac{1}{2k} \int_{D^k} \max_{v \in D^k} \exp\{-U(v)\}\, dv_1 \ldots dv_k \\
&\le \sum_{k=3}^{\infty} \frac{1}{2k} \int_{D^k} \exp\{-\beta k^{\gamma} - \frac{\alpha 4\pi^2}{k^2}\}\, dv_1 \ldots dv_k \\
&\le \sum_{k=3}^{\infty} \left(\frac{MN \exp\{-\beta k^{\gamma-1}\}}{\sqrt[k]{k}} \right)^k \qquad (4)
\end{aligned}
$$

As $MN > 1$, the series (4) converges if and only if $\gamma \geq 1$, provided $\exp(\beta) > MN$ when $\gamma = 1$. Keeping in mind (1) and (3), the posterior distribution of the polygon turns out to be

$$
p(v \mid z) \propto \exp\{-\beta k^{\gamma} - \frac{\alpha}{k} \sum_{i=1}^{k} (\theta_i(v) - \pi)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{M} \sum_{j=1}^{N} (z_{ij} - \nu_{ij}(v))^2\} \qquad (5)
$$

where we have substituted the actual object v to the generic object O in the argument of $\nu_{ij}(\cdot)$.

Our prior model appears to be related to the construction derived by Arak *el al.* (1993). In that case one simulates two independent Poisson point processes, the first in the interior of the image rectangle and the second on its perimeter, and creates a graph by connecting pairs of vertices subject to the constraint that V, Y, T and I-shaped vertices only are allowed. Provided an appropriate measure on all point patterns has been defined, it is possible to explicitly calculate the normalisation constant and some moments. Moreover, a consistency property holds: the process one observes in a window is defined on all $\mathcal{R}^2$.

The process we use is similar to one that could be defined as an Arak process conditioned on the fact that no points have been observed on the perimeter, V-vertices only are allowed and only the graphs with non-intersecting edges are admissible.

The density (3) can also be rewritten by extending the theoretical framework used by Grenander, Chow and Keenan (1991) to define a probability distribution on the deformations of a given simple closed curve (template) belonging to a space of fixed dimension (such as C_k with k fixed). Every polygon belonging to a set C_k is obtainable by deforming a regular polygon $\bar{v}^{(k)}$ with k vertices and fixed labelling (template) by applying to each of its sides k functions $(s_1, \ldots, s_k)$ belonging to a group of transformations S. If we predefine a sequence of templates $\{\bar{v}^{(k)}, k \geq 3\}$, we can define the prior distribution for the transformations themselves instead of that for the vertices, thus modifying the support of the prior distribution to the set $S = \bigcup_{k=3}^{\infty} S_k$, where $S_k = \{s \in S^k : s\,\bar{v}^{(k)} \text{ is a polygon}\}$, and $s\,\bar{v}^{(k)} = (s_1\,\bar{v_1}^{(k)}, \ldots, s_k\,\bar{v_k}^{(k)})$. We can then rewrite (3) as a product of *acceptor functions* depending on s as follows:

$$
\pi(s; \bar{v}^{(k)}) \propto \exp\{-\beta k^{\gamma}\} \prod_{i=1}^{k} A(s_{i-1}, s_i; \bar{v}^{(k)}) \qquad (6)
$$

where

$$
A(s_{i-1}, s_i; \bar{v}^{(k)}) = \exp\{-\frac{\alpha}{k}\,[\theta(s_{i-1}\,\bar{v}_{i-1}^{(k)},\ s_{i-1}\,\bar{v}_i^{(k)},\ s_i\,\bar{v}_{(i \bmod k)+1}^{(k)}) - \pi]^2\},
$$

$s_0 = s_k$, $\bar{v}_0^{(k)} = \bar{v}_k^{(k)}$ and $\theta(\cdot)$ measures the interior angle formed by the vertex identified by its three arguments. The densities (3) and (6) have the Markovian property that the distribution of a side (identified by its vertices in (3) and by the relative s function in (6)) conditioned to the other sides and to the number of sides depends only on its adjacent sides.

In what follows we will use neither this approach nor that of Clifford and Nicholls for the synthesis and the restoration of objects, but we will use model (3), which is simpler and more immediate for the kind of application we are considering here.

3 Application of the reversible jump algorithm to object restoration

The posterior distribution (5) allows one to calculate different kinds of restoration, such as the point of maximum *a posteriori* (MAP) probability density

$$\max_v p(v \mid z)$$

or the posterior mean of the state of every pixel with respect to the polygon

$$\begin{aligned} & \sum_{k=3}^{\infty} \frac{1}{2k} \int_{D^k} \nu_{ij}(v)\, I(v \in C_k)\, p(v \mid z)\, dv_1 \ldots dv_k \\ = \;& \nu_1 \ \Pr(v \ni i) + \nu_0 \ \Pr(v \not\ni i) \quad i = 1, \ldots, M, \ j = 1, \ldots, N, \end{aligned} \tag{7}$$

which corresponds to the posterior probability that the polygon includes a given pixel when $\nu_1 = 1$ and $\nu_0 = 0$. The posterior mean of v cannot be defined usefully, because it is impossible to average a variable which ranges over spaces of different dimension, and the MAP restoration does not carry a very meaningful information, so we have chosen to use (7) for our restoration, because it has the desirable feature of not being dependent on the dimension of the space in which the polygon lies.

Since we cannot calculate (7) directly, it is necessary to simulate a Markov chain with state space C and with (5) as equilibrium distribution, and then calculate

$$\frac{1}{n_{\text{sweep}} - n_0} \sum_{n=(n_0+1)}^{n_{\text{sweep}}} \nu_{ij}(v(n)) \tag{8}$$

as an approximation of (7). The sequence $\{v(n)\}$ is the sequence of polygons $v(n)$ obtained during the simulation, n_{sweep} is the number of sweeps, and n_0 is the length of the *burn-in* period.

The problem we have just stated cannot be solved with ordinary MCMC methods (e.g. the Metropolis-Hastings algorithm, see Tierney, 1994), because the chain must visit spaces of different dimension and switch from one to another. The reversible jump algorithm proposed by Green (1995), gives a solution by suitably modifying the Metropolis-Hastings algorithm to work in general state spaces. In the remainder of this Section, we describe how a reversible jump algorithm can be constructed for our problem; a detailed measure-theoretic specification of the resulting transition kernels is deferred to the Appendix.

Assuming that the current state is $v \in C_k$, we move to different states only through slight changes, so that we can easily check whether the move has led the chain to an admissible state. We do this by choosing among two *move types* for moving to $v' \in C_{k'}$.

1. Move type 1 ("perturb"): for all $k \geq 3$, with probability m_k we propose a new position ξ for a randomly chosen vertex i of the polygon, $i = 1, \ldots, k$, sampling it from the proposal density $q(\cdot; i, v)$; we denote the modified polygon by $v^{(i,\xi)}$. This is a regular fixed-dimension Metropolis-Hastings move, so the acceptance probability is

$$\alpha_1(v, v^{(i,\xi)}) = \min\left\{1, \frac{\pi(v^{(i,\xi)})q(v_i; i, v^{(i,\xi)})}{\pi(v)q(\xi; i, v)}\right\}.$$

2. Move type 2 ("change space"). For this move type we need some additional notation: for any given vertex i, $i = 1, \ldots, k$, we denote by i^+ the vertex following i, according to either a clockwise or anticlockwise ordering of vertices. Move type 2 works as follows:

   (a) for all $k \geq 3$, with probability b_k we add a new vertex at position ξ, between two randomly chosen consecutive vertices i and i^+. The new position is defined as $\xi = (v_i + v_{i^+})/2 + u\,\sigma_b(i, v)$, where u is drawn from a standard bivariate normal density $\phi_2(\cdot)$, and $\sigma_b(i, v)$ is a standard deviation to be defined in the Appendix. The modified polygon is denoted $v^{(i,\xi)}$ as in the "perturb" move.

   (b) for all $k \geq 4$, with probability d_k we delete a randomly chosen vertex i and join its two adjacent vertices, and denote the new state by v_{-i}. When $k = 3$ and this move type is selected we do nothing and resample the move type.

The acceptance probability for the "change space" move is obtained by the application of Green's (1995) result. The general result is quite abstract, but there is a particular case which we now state in a general notation. Suppose the state space is a collection of Euclidean spaces of different dimension, and that the state variable x has a density $g(\cdot)$ with respect to the Lebesgue measure for any given dimension. A move from x to a state y in a higher dimensional space is proposed by drawing an auxiliary variable w from some density $h(\cdot)$ and setting $y = y(x, w)$, where $y(\cdot, \cdot)$ is an invertible function. The acceptance probability for this move is

$$\min\left\{1, \frac{g(y)\,t(y)}{g(x)\,t(x)\,h(w)}\left|\frac{\partial\, y(x, w)}{\partial x\, \partial w}\right|\right\},$$

where $t(x)$ is the probability of choosing the move "change space" when in state x.

Going back to our problem, we see that $g(v) = \pi(v)/2k$ is a density with respect to the Lebesgue measure, as its base measure (2) is actually calculated on elements of $\mathcal{B}(D^k)$. If we took this density as it stands in the calculation of the acceptance probability, we would lose the identifiability of the polygons, since there are $2k$ elements in D^k which correspond to a given k-side-polygon $[v]$. A way round this difficulty is to give a polygon a unique labelling, e.g., by assigning label 1 to the vertex with higher abscissa, and assigning the other labels in clockwise order. The resulting density of a k-side-polygon $[v]$ is then $2k\,g(v) = \pi(v)$ on the subset of D^k containing the label-ordered vectors.

Setting also $t(v) = b_k/k$, $t(v^{(i,\xi)}) = d_{k+1}/(k+1)$ and $h(u) = \phi_2(u)$, the acceptance probability for move type 2 is then

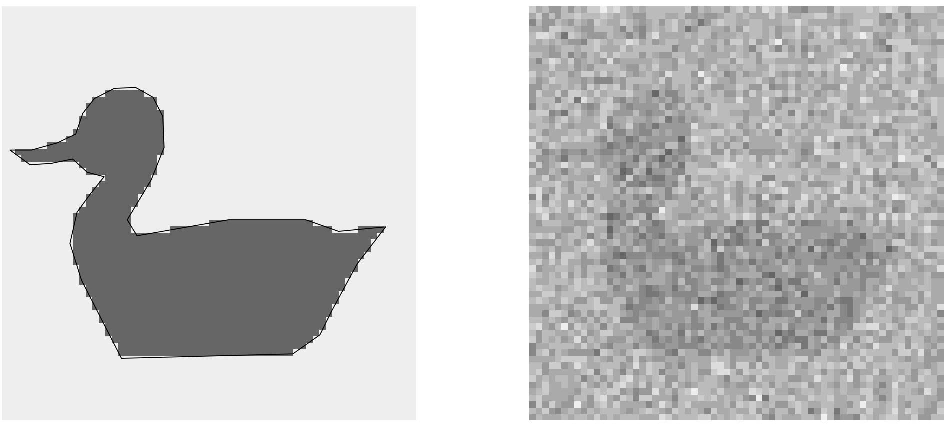
$$\alpha_2(v, v^{(i,\xi)}) = \min\left\{1, \frac{\pi(v^{(i,\xi)})\,d_{k+1}\,k}{\pi(v)\,\phi_2(u)\,b_k\,(k+1)}\left|\frac{\partial\, v^{(i,\xi)}}{\partial u\, \partial v}\right|\right\},$$

where the ratio $k/(k+1)$ comes from the random scan of vertices.

Naturally $m_k + b_k + d_k = 1 \ \forall \ k \geq 3$. Notice that we have not imposed any explicit constraint on the support of the proposal densities and this implies that the new proposed state might be a self-intersecting polygon or a polygon with some vertex out of D. The same could happen after the deletion of a vertex. However, the acceptance probability is zero in this situation, so there is not a problem. The advantage of this method is that we do not need to explicitly impose any constraint of closeness or non-intersection of polygons within the probability distribution, because the admissibility of the configuration is controlled during the execution of the reversible jump algorithm, simply refusing the illegal configurations. But there are two disadvantages: checking the admissibility of the proposed configurations is time consuming; the variance of $q(\cdot; i, v)$ and $\sigma_b^2(\cdot, \cdot)$ must be carefully specified because if they were too high too many illegal configuration would be proposed, while if they were too low, the algorithm would be too "timid", slowing down in both cases the convergence to equilibrium of the chain.

As for the convergence itself, we show in the Appendix that this chain has π as an equilibrium distribution and is also Harris-recurrent.

4 Experiments



(a) (b)

Figure 1: A 64×64 test image (a) and a noisy version of it (b) according to model (1) with $\nu_1 = 1$, $\nu_0 = 0$ and $\sigma^2 = 0.5$

As an application of the method we have tried to restore a 64×64 artificial picture with a central body, whose dimensions are not too small compared to the dimension of the entire image, and some tiny extensions. The picture is shown in figure 1a, with both the object and its defining contour; the noisy version, produced by model (1) with $\nu_1 = 1$, $\nu_0 = 0$ and $\sigma^2 = 0.5$, is in figure 1b, A method based on ordinary smoothing techniques would probably shorten or even cause the extensions to part from the central body, while our method, exploiting the additional information that the image portrays only one object and the connectivity among the polygon vertices, has succeeded in recovering these features, starting from a very poor initial configuration (a square located more or less at the centre of the object).

Three restorations of figure 1b appear in figure 2 with hyperparameters $\alpha = 3$, $\gamma = 1.6$, and β taking the values 0.3, 0.9 and 1.5 from left to right. The algorithm parameters were set to $m_k = b_k = d_k = 1/3\ \forall k \geq 3$, and the proposal variances as explained in the introductory part of the Appendix. Fifty thousand sweeps were performed with a burn-in period of 5000 sweeps. A single restoration took about 18 minutes of CPU time on a Sun 4/40 workstation at 1.7 *megaflops*. The restoration with the lowest error rate is in figure 2b. The error rates are based throughout (except for figure 5) on thresholding the restored image at 0.5, so they are the fraction of misclassified pixels. This is done primarily for purposes of comparison among different methods, whose results are described in the following. A flavour of the prior beliefs expressed by the values of hyperparameters for each restoration is given in the bottom row of figure 2 where the last realizations from a chain simulating the prior are shown.

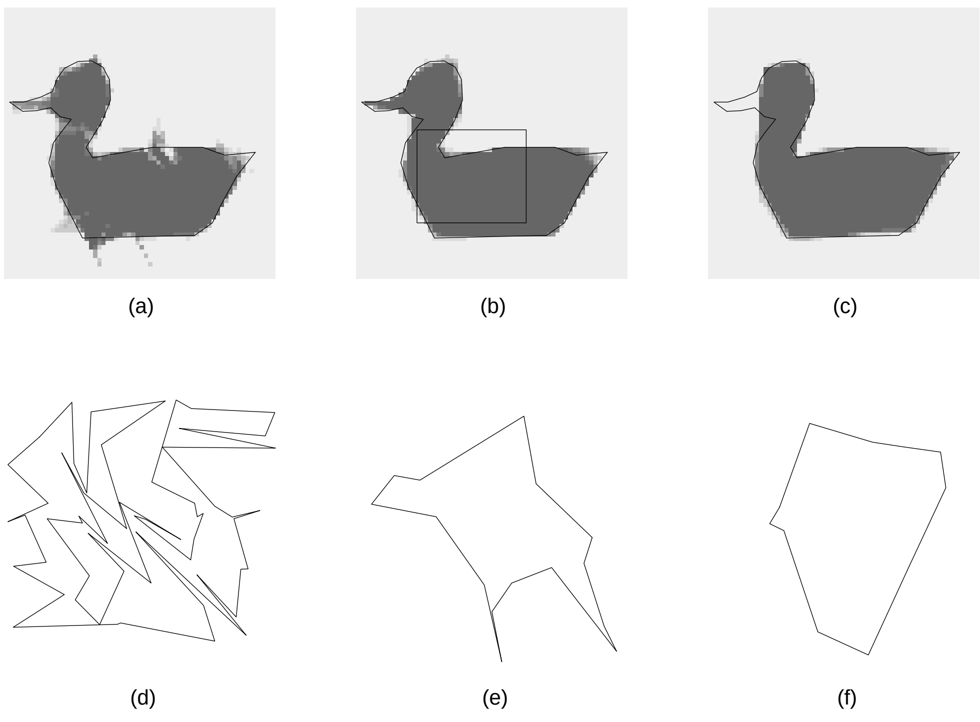


Figure 2: Three restorations of figure 1b compared to the true outline and three realizations from chains simulating the prior (d), (e), (f). Parameter values are $\alpha = 3$, $\gamma = 1.6$ for all pictures; β is 0.3, 0.9 and 1.5 in columns 1, 2 and 3. The error rates are 2.2% (a), 1.49% (b), 2.05% (c). The square in (b) is the starting configuration.

That convergence to equilibrium is satisfactory for the purpose of restoration is shown by the sequence of the number of pixels inside the polygon (figure 3) for the run of figure 2b.

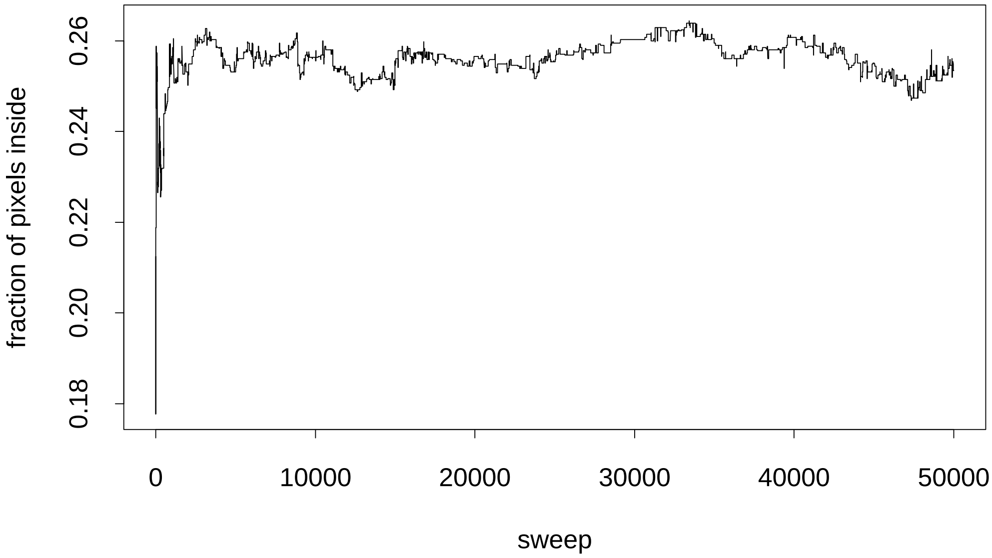


Figure 3: The sequence of the fraction of pixels inside the polygon during the run which produced figure 2b

The starting configuration was found automatically, by simply producing a smoothed version of the noisy image with a running mean performed by a 5×5 window and thresholding the result to two greylevels. Considering the resulting image as a bivariate probability distribution on D, we have calculated its two marginal means and variances (say m_x, m_y, σ_x^2 and σ_y^2) and then have taken as initial configuration the square $[m_x \pm 2\,\sigma_x] \times [m_y \pm 2\,\sigma_y]$. The same procedure without smoothing and thresholding did not produce any meaningful result.

A comparison of our method with pixel-based methods for image restoration appears in figure 4. In all cases an Ising model with parameter ρ was used as prior distribution on the true binary greylevel. The three restorations in the top row have been performed using the Gibbs sampler (Geman and Geman, 1984), whereas those in the bottom row are ICM restorations (Besag, 1986). The centre restorations with $\rho = 1.2$ have the lower error rate, whereas the other two values of ρ are 0.4 and 2.0. The parameter β in figure 2 has the same relative variation, so the pictures can be compared. It is evident that the additional information embodied in model (3) has a dramatic effect in the restoration.

The above experiments show that our method is not very sensitive to changes in the parameter β of the order of about 70%. A similar conclusion is also valid for γ and α. Of course, if we allow the hyperparameters to range in much wider intervals ($0 \le \alpha \le 10$, $0.0001 \le \beta \le 6$ and $1.0001 \le \gamma \le 6$) the range of behaviours shown in figure 2 is magnified. A value for α that is too small produces a restoration with small acute peaks protruding from and into the true object border, whereas small β and γ (giving a very diffuse prior for the number of vertices) cause the number of vertices to explode, as the polygon tends to follow the outline of every pixel with a higher greylevel. On the contrary, a high α oversmooths sharp angles, and high β and γ cut them off. Other experiments (not reported here) suggest that variations of β or γ affect the final restoration much more than variations in α, which can result into a more or less smooth object boundary, but with its essential features preserved.

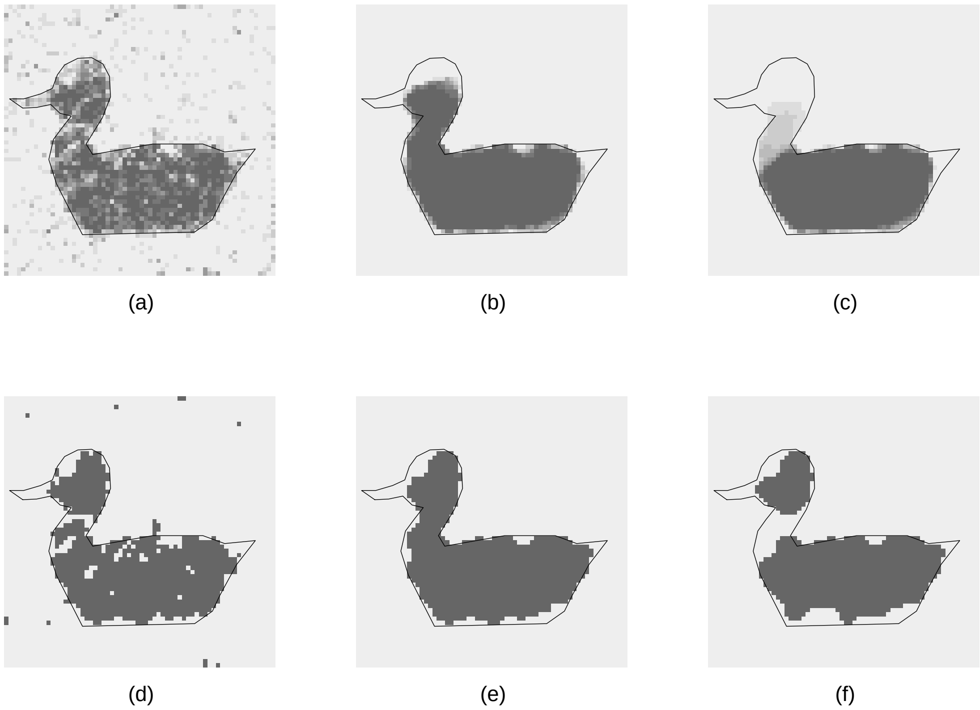


Figure 4: Restorations of figure 1b with the Ising model as prior with parameter ρ taking values 0.4, 1.2 and 2 in columns 1, 2 and 3. First row: Gibbs sampler restorations, with error rates 5.10% (a), 4.22% (b), 8.13% (c). Second row: ICM restorations with error rates 5.27% (d), 3.98% (e), 6.03% (f).

Another question of interest is to establish whether there is a real advantage in varying the number of vertices, since one could think of using a fixed number which can suit different situations. In figure 5 three different runs of our algorithm with $\alpha = 3$, $\beta = 0.9$ and $\gamma = 1.6$ (top row) are compared to three other runs of the same algorithm in space C_{14} with no change of space allowed ($m_{14} = 1$). The number 14 is the *a posteriori* mean number of vertices from the run which produced figure 2b, so it is optimal in some sense. We see that the top row restorations have a lower error rate on average and look also better visually, due to the additional "model averaging" capabilities. Here the error rate is the expected posterior number of misclassified pixels, which can be calculated from the estimated posterior probabilities of inclusion of each pixel in the polygon.

One can envisage more complicated situations than model (1) to describe the observation process which do not alter the basic message of this article. If ν_0 and ν_1 were unknown it would be fairly straightforward to define a prior distribution on them (which should be independent from that of the polygon for the problem to be tractable) and add a step to the MCMC algorithm for their updating. A texture image could be dealt with by defining two appropriate Markov random fields (see e.g. Dubes and Jain, 1989) for modelling the texture inside and outside the polygon (and conditionally on it), even if it appears to be quite a cumbersome task both in programming and computationally.

A blurred image with superimposed noise could be restored successfully simply by using

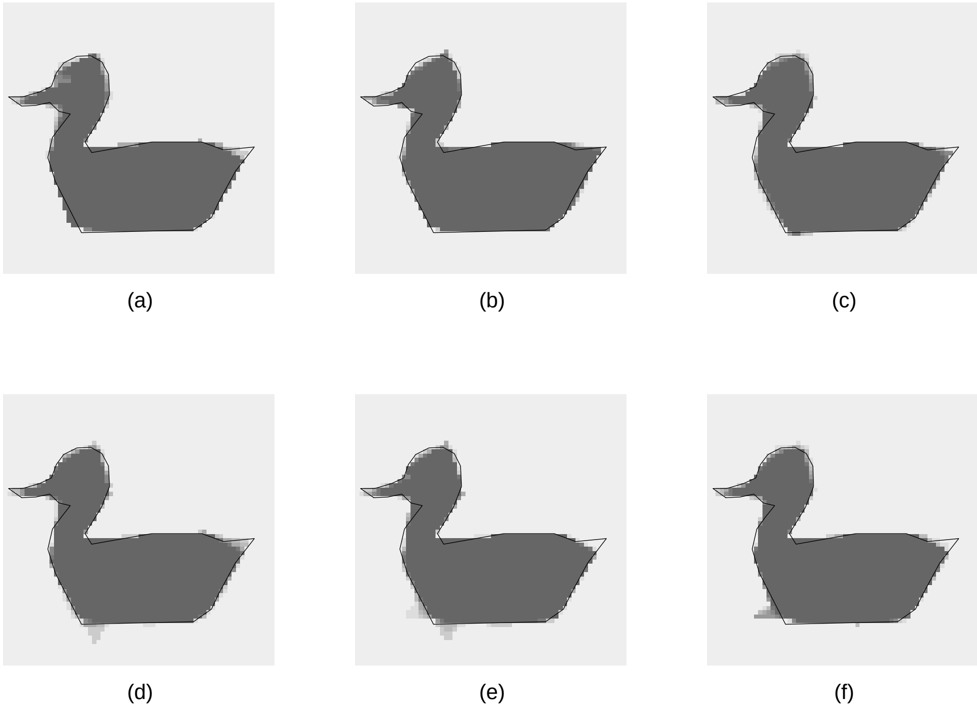


Figure 5: Top row: three runs of the reversible jump MCMC algorithm with $\alpha = 3$, $\beta = 0.9$ and $\gamma = 1.6$. Error rates: 1.82% (a), 1.56% (b), 1.63% (c). Bottom row: as the top row but with $k = 14$ and $m_{14} = 1$. Error rates: 1.75% (d), 1.95% (e), 1.87% (f)

the correct likelihood function, assuming that the blurring parameters are known (as it is often the case). In figure 6, we show instead an image blurred by a moving average with a 3×3 window and with superimposed i.i.d. Gaussian noise ($\sigma^2 = 0.5$), and its restoration made with the likelihood for the unblurred case, compared to the true outline.

A different set of problems is related to the nature of the true image, which could contain multiple objects or objects not wholly contained in the observation window. The first author is working on a generalisation of the present setting to allow multiple objects: the density for a given configuration is based upon the product of densities like (3), with an interaction term depending on the area of the model window contained in more than one object; the number of objects can be varied in a fashion similar to that of vertices. Of course, the colouring inside each object must be modelled in the same way, as the polygons are allowed to overlap.

A simple way to restore partially visible objects is to use as D a window which is bigger than the observation window, and to allow the polygon to visit the resulting outer band with just the prior information, whereas inside the window the observational information would continue to be used. In figure 7 the result of such an experiment is shown, along with the current polygon at the last sweep. Our method is evidently quite robust to the truncation of the data.

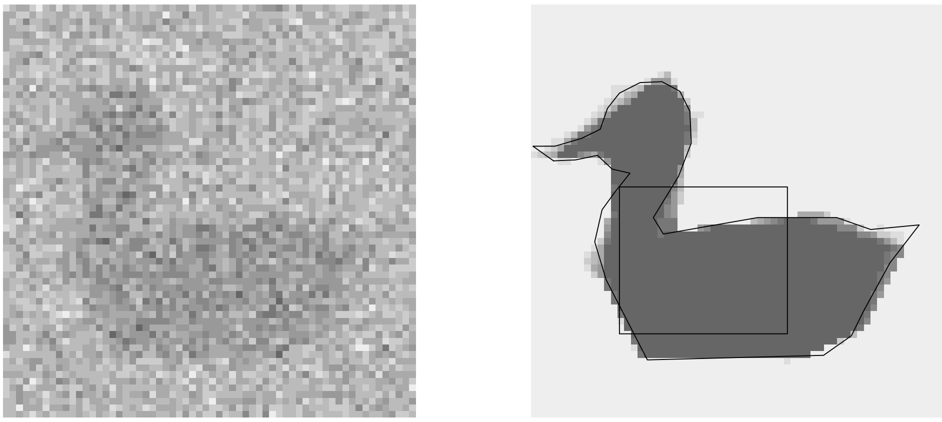


Figure 6: A blurred noisy image and its restoration

5 Discussion

In this paper we have illustrated a method for the restoration of a noisy binary image portraying a single object whose outline is entirely defined by a polygon (with an indefinite number of sides). The computing machine provided by the reversible jump MCMC algorithm was used to perform the restoration, exploiting the capability of the algorithm of switching between spaces of different dimension, that is, of drawing inference about the number of sides of the polygon to be fitted to the object.

The problem setting was as simple as possible: the values of the two greylevels present in the original image, and the noise variance were assumed known, and there was no problem of estimation of the parameters of the prior distribution. Indeed, it is not possible to find a maximum pseudo-likelihood estimate of the parameters in (3) without a preliminary guess about the true underlying polygon.

Thus many other routes should be tried. First, the likelihood model should be extended to more general situations, substituting ν_0 and ν_1 with unknown function to be estimated, as, for example, in Qian *et al.* (1996). A generalisation of the noise process would also be possible, but all the additional efforts in modelling would be justified only if motivated by a practical application, whereas the main scope of this article is to illustrate a simple way to generate polygons in the plane and to show the capabilities of the reversible jump MCMC methodology in the context of Bayesian image analysis. Second, this setting should be extended to allow multiple objects in the image, which is by no means trivial. Third, we might improve the implementation of the reversible jump algorithm to accelerate convergence to equilibrium, e.g. by choosing less blindly the vertex visiting scheme or the probabilities of the move types. Fourth, a full Bayesian approach could allow us to include in the restoration process the estimation of prior parameters, provided we are able to define some hyperpriors for the parameters which describe meaningfully our prior expectations about the shape of the polygon.

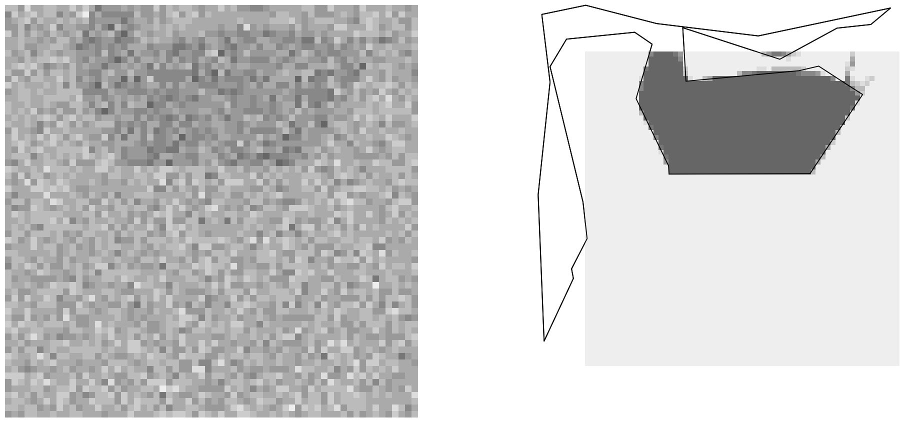


Figure 7: A partially visible noisy object and its restoration

Appendix

Transition kernels

We give here a complete measure-theoretic construction of the kernel for the Markov chain described in Section 3, for exploring the space C. In the proofs of irreducibility and Harris-recurrence we adapt to our case the method followed by Geyer (1997) for the simulation of spatial point processes.

Here we allow the multiple labelling of the same polygon, so we must be say precisely what we mean by $v^{(i,\xi)}$ and v_{-i}. Letting $v = (v_1, \ldots, v_k)$ we set $v^{(i,\xi)} = (v_1, \ldots, v_i, \xi, v_{i+1}, \ldots, v_k)$ and $v_{-i} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$, with the obvious adjustments when $i = 1$ or $i = k$. Also, i^- and i^+ are the labels of the vertices preceding and following i, according to the current ordering of labels (clockwise or anticlockwise).

Assuming that the current state is $v \in C_k$, the complete kernel for a transition to a set $A \in \mathcal{F}$ can be written as a mixture of three kernels:

$$P(v, A) = m_k\, P_1(v, A) + b_k\, P_2(v, A) + d_k\, P_3(v, A). \tag{9}$$

The components of the kernel are as follows, with zero acceptance probabilities for proposed moves to non-admissible states. The first component of the mixture is defined by

$$\begin{aligned} P_1(v, A) \;&=\; \frac{1}{k} \sum_{i=1}^{k} \int_{\mathcal{R}^2} q(\xi; i, v)\, \alpha_1(v, v^{(i,\xi)})\, I(v^{(i,\xi)} \in A \cap C_k)\, d\xi \\ &\quad + I(v \in A)\, s_1(v), \end{aligned} \tag{10}$$

where

$$s_1(v) \;=\; \frac{1}{k} \sum_{i=1}^{k} \int_{\mathcal{R}^2} q(\xi; i, v)\, [1 - \alpha_1(v, v^{(i,\xi)})]\, d\xi$$

and $q(\cdot; i, v)$ is a bivariate normal density with mean v and standard deviation $\sigma_p(i, v) = d(v_{i^-}, v_{i^+})/c$. The function $d(\cdot, \cdot)$ denotes the Euclidean distance, whereas c is a constant to be defined later.

Recalling that in the "change space" move $\xi = (v_i + v_{i^+})/2 + u\,\sigma_b(i, v)$, the second component of the kernel is

$$\begin{aligned} P_2(v, A) &= \frac{1}{k} \sum_{i=1}^{k} \int_{\mathcal{R}^2} \phi_2(u)\, \alpha_2(v, v^{(i,\xi)})\, I(v^{(i,\xi)} \in A \cap C_{k+1})\, du \\ &\quad + I(v \in A)\, s_2(v), \end{aligned} \tag{11}$$

where

$$s_2(v) = \frac{1}{k} \sum_{i=1}^{k} \int_{\mathcal{R}^2} \phi_2(u)\, [1 - \alpha_2(v, v^{(i,\xi)})]\, du$$

and $\sigma_b(i, v) = d(v_i, v_{i^+})/c$.

The third component of the kernel is

$$P_3(v, A) = \frac{1}{k} \sum_{i=1}^{k} \alpha_3(v, v_{-i})\, I(v_{-i} \in A \cap C_{k-1}) + I(v \in A)\, s_3(v), \tag{12}$$

where

$$s_3(v) = \frac{1}{k} \sum_{i=1}^{k} [1 - \alpha_3(v, v_{-i})].$$

Notice that the acceptance probabilities appearing in the kernels are unspecified. Of course, they will turn out to be those already defined.

The proposal variances $\sigma_p^2(\cdot, \cdot)$ and $\sigma_b^2(\cdot, \cdot)$ allow with high probability moves that are consistent with the local configuration of the polygon; for example, if we have to add a vertex between two vertices very near to each other, we concentrate the probability mass on a small set of possible positions, thus avoiding a too high probability of the occurrence of self-intersecting polygons. The constant c appearing above must be chosen to fulfil this condition of coherence to the local configuration. A simple criterion is to select c so that a new vertex is added between two vertices with a fixed probability (say 0.95) in the square with center $(v_i + v_{i^+})/2$ and with side $d(v_i, v_{i^+})$, in which case $c = \sqrt{20}$.

Once the acceptance probabilities $\alpha_i(\cdot, \cdot), i = 1, 3$, have been defined so that the kernel has $\pi(v)$ as an invariant distribution and it is irreducible and aperiodic, we obtain a Markov chain with $\pi(v)$ as equilibrium distribution for π-almost all starting points $v \in C$. Such a chain is positive recurrent; if it is also Harris recurrent the convergence to equilibrium occurs for all starting points (see Tierney, 1994, for a proof of these assertions).

Invariance

A sufficient condition for the invariance is the reversibility condition for all pairs of sets $A, B \in \mathcal{F}$:

$$\int_A \left[\int_B P(v, dv') \right] \pi(v)\, \mu(dv) = \int_B \left[\int_A P(v', dv) \right] \pi(v')\, \mu(dv'). \tag{13}$$

Letting $Q = b_k P_2 + d_k P_3$, a sufficient condition for (13) is that P_1 be reversible and Q be reversible. The reversibility of P_1 is immediately clear from the fact that P_1 is the mixture of k kernels, where the i-th kernel affects the i-th vertex only. Since the label i identifies a unique vertex as the chain switches from A to B and *vice versa*, the i-th kernel can be considered as a regular fixed-dimension Metropolis-Hastings kernel, with acceptance probability $\alpha_1(\cdot,\cdot)$ as in Section 3.

The check for reversibility of Q is greatly simplified if we notice that we can always write $A = A_{k_1} \cup \ldots \cup A_{k_m}$ and $B = B_{l_1} \cup \ldots \cup B_{l_n}$, where $A_{k_i} \in \mathcal{F}_{k_i}$, $i = 1, \ldots, m$, and $B_{l_j} \in \mathcal{F}_{l_j}$, $j = 1, \ldots, n$, for some integers m and n. Since all the sets A_{k_i} are disjoint as well as the sets B_{l_j}, the reversibility becomes

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\int_{A_{k_i}}\int_{B_{l_j}} \pi(v)Q(v,dv')\,\mu(dv) = \sum_{j=1}^{n}\sum_{i=1}^{m}\int_{B_{l_j}}\int_{A_{k_i}} \pi(v')Q(v',dv)\,\mu(dv').$$

The only nonzero terms in the sums above are those for which $|k_i - l_j| \le 1$, so, we just need to check that

$$\int_{A_{k_i}}\int_{B_{l_j}} \pi(v)Q(v,dv')\,\mu(dv) = \int_{B_{l_j}}\int_{A_{k_i}} \pi(v')Q(v',dv)\,\mu(dv') \tag{14}$$

for all pairs $k_i \le l_j$ that meet this condition. Since the form of the kernel does not depend on the values of k_i and l_j, we take either $k_i = l_j = k$ or $k_i = k$ and $l_j = k+1$.

The first case is easily dealt with. If there exist $k_i = l_j = k$, we obtain in both members of (14) the probability of not moving from $A_k \cap B_k$.

When $k_i = k$ and $l_j = k+1$, let us take sets $A = A_k \subset C_k$ and $B = B_{k+1} \subset C_{k+1}$. Since the term containing P_3 vanishes at the first member of (14), as well as the term containing P_2 at the second member, we get

$$\int_A\int_B \pi(v)\, b_k\, P_2(v,dv')\,\mu_k(dv) = \int_B\int_A \pi(v')\, d_{k+1}\, P_3(v',dv)\,\mu_{k+1}(dv') \tag{15}$$

Substituting to P_2 and P_3 their definition and rewriting the integrals with respect to the Lebesgue measure we get

$$\begin{aligned}
&\sum_{i=1}^{k}\int_{D^k}\int_{\mathcal{R}^2} \frac{b_k}{k}\,\frac{\pi(v)}{2k}\,\phi_2(u)\,\alpha_2(v,v^{(i,\xi)})\,I(v\in A)\,I(v^{(i,\xi)}\in B)\,du\,dv \\
= &\sum_{j=1}^{k+1}\int_{D^{k+1}} \frac{d_{k+1}}{k+1}\,\frac{\pi(v')}{2(k+1)}\,\alpha_3(v,v'_{-j})\,I(v'\in B)\,I(v'_{-j}\in A)\,dv'
\end{aligned}$$

Observe that all the terms in the sum at the first member of the above equation are equal, and the same holds for the terms at the second member. This is because in the i-th integral at the first member the index i identifies all the vertices of a given configuration as v ranges over D^k. The same holds for any j in the second member. So, for an arbitrary value of i and $j = i^+$ the above equation simplifies to

$$\begin{aligned}
&\int_{D^k}\int_{\mathcal{R}^2} b_k\,\frac{\pi(v)}{k}\,\phi_2(u)\,\alpha_2(v,v^{(i,\xi)})\,I(v\in A)\,I(v^{(i,\xi)}\in B)\,du\,dv \\
= &\int_{D^{k+1}} d_{k+1}\,\frac{\pi(v')}{k+1}\,\alpha_3(v',v'_{-i^+})I(v'\in B)\,I(v'_{-i^+}\in A)\,dv'
\end{aligned}$$

The final step is to make a change of variable in the second integral above, letting $v' = v^{(i,\xi)}$, so that $v'_{-i^+} = v$. The second integral is now

$$\int_{D^k}\int_{\mathcal{R}^2} d_{k+1}\,\frac{\pi(v^{(i,\xi)})}{k+1}\,\alpha_3(v^{(i,\xi)},v) I(v^{(i,\xi)}\in B)\, I(v\in A)\,\left|\frac{\partial\, v^{(i,\xi)}}{\partial u\,\partial v}\right|\,du\,dv$$

from which we easily see that $\alpha_2(\cdot,\cdot)$ must be as in Section 3 and $\alpha_3(\cdot,\cdot)$ is the same but with the acceptance ratio changed to its reciprocal. The Jacobian is found to be equal to $\sigma_b^2(i,v)$.

Aperiodicity and irreducibility

The kernel (9) is clearly aperiodic, since there is a positive probability that self-intersecting polygons be proposed from any current state v; these will always be rejected so the state of the chain will remain unchanged.

Turning to irreducibility, we only need to check that the kernel is irreducible with respect to some measure φ that is positive on only one set with measure π positive, for example C_3 (see Meyn and Tweedie, 1993, proposition 4.2.2 and theorem 10.4.9). In other words, we must show that from any state v the chain goes to a triangle with positive probability. Since there is at least one sequence of deletions of vertices that reduces any given polygon to a triangle, this happens with positive probability if we take d_k to be bounded away from zero by a constant d and there exists a lower bound to $\alpha_3(\cdot,\cdot)$. In fact, we take also b_k to be bounded away from zero by b, so that there is a constant η such that $b_k/d_{k+1} > \eta$.

To find the lower bound to $\alpha_3(\cdot,\cdot)$, set $v' = v^{(i,\xi)}$ and notice that

$$\pi(v')/\pi(v) < \exp\{-\beta\,(4^\gamma - 3^\gamma) + \alpha\,\pi^2\} = M$$

so

$$\alpha_3(v',v) \geq \min\left\{1, \frac{b_k\phi_2(u)\,(k+1)}{d_{k+1}\,k\,\sigma_b^2(i,v)\,M}\right\}.$$

Since both the components of u and $\sigma_b^2(i,v)$ have an upper bound for any possible move, there is a constant R such that $\phi_2(u)/\sigma_b^2(i,v) > R$ and we have

$$\alpha_3(v',v) \geq \frac{\eta\,R}{M}.$$

Harris recurrence

Proving Harris-recurrence is much more involved. A sufficient condition for it is the *drift condition* (Meyn and Tweedie, 1993, theorem 9.1.8). The drift condition says that given a π-irreducible kernel P, if there is a small set $A \in \mathcal{F}$ and a function unbounded off small sets $V : C \to \mathcal{R}_+$ such that

$$\int_C V(v')\,P(v,dv') \leq V(v) \qquad \forall v \notin A$$

then P is Harris-recurrent.

A set $A \in \mathcal{F}$ is *small* if there exist a nonzero measure φ and an integer n such that

$$P^n(v,B) \geq \varphi(B) \qquad \forall v \in A \text{ and } \forall \mathrm{B} \in \mathcal{F}.$$

A function $V : C \to \mathcal{R}_+$ is *unbounded off small sets* if for any positive number $l > 0$ the set $\{v : V(v) \le l\}$ is small.

Then we must find a small set for our transition kernel P. One such set is $C_{\le k} = \{v \in C : k(v) \le k\}$, where $k(v)$ denotes the number of vertices in v. Indeed let us take $m \ge k$ and let $v \in C_{\le k}$; then we have

$$P^m(v, C_3) \ge \int_{C_3} P^{k(v)-3}(v, dv')\, P^{m-k(v)+3}(v', C_3).$$

The probability $P(v', C_3)$ is bounded from below by d_3, when $v' \in C_3$, so $P^{m-k(v)+3}(v', C_3) > d_3^{m-k(v)+3}$. The above integral reduces then to

$$d_3^{m-k(v)+3}\, P^{k(v)-3}(v, C_3)$$

where the second factor is also bounded from below by $J^{k(v)-3}/k(v)!$. The constant J can be set to

$$J = \frac{d\,\eta\, R}{M}$$

as a consequence of the lower bounds on d_k and on $\alpha_3(\cdot,\cdot)$, whereas the term $k(v)!$ comes from the random scan of vertices. One is then allowed to state that for all $v \in C_{\le k}$

$$P^m(v, C_3) > \frac{(d_3\, J)^m}{m!}$$

(recall that $m > k(v)$) which coincides with $C_{\le k}$ being a small set when $\varphi(C_3) = (d_3\, J)^m/m!$.

Now we are in the position to prove the drift condition. First we set $V(v) = H^{k(v)}$, with $H > 1$, as a function unbounded off small sets like $C_{\le k}$. Second we notice that

$$\alpha_2(v, v^{(i,\xi)}) \le \frac{\exp\{-\beta[(k+1)^\gamma - k^\gamma] + \alpha\,\pi^2\}}{\eta\, R} = r(k).$$

Third, we calculate the drift for $H^{k(v)}$. After some calculations, this is found to be

$$\begin{aligned}
\int_{\mathbf{C}} V(v')\, P(v, dv') &= H^k\, m_k \\
&\quad + b_k \left\{ H^k + \frac{H^{k+1} - H^k}{k} \sum_{i=1}^{k} \int_{\mathcal{R}^2} \phi_2(u)\, \alpha_2(v, v^{(i,\xi)})\, du \right\} \\
&\quad + d_k \left\{ H^k + \frac{H^{k-1} - H^k}{k} \sum_{i=1}^{k} \alpha_3(v, v_{-i}) \right\} \\
&\le H^k \left\{ m_k + b_k + b_k\, r(k)\, (H-1) + \frac{d_k}{H} \right\}
\end{aligned} \tag{16}$$

Since $r(k)$ tends to zero as k goes to infinity and b_k and m_k are bounded away from 1, the term in braces in (16) can be made less than 1 for all k greater than some large enough k_0. Thus, if we take the small set C_{k_0}, the drift condition is satisfied for this set with $V(v) = H^{k(v)}$ and P is Harris-recurrent.

Acknowledgements

The authors wish to thank Carlo Gaetan for helping with the definition of (2) and Luca Pratelli for his hints about the measurability of C_k.

References

Arak, T., Clifford, P. and Surgailis, D. (1993) Point-based polygonal models for random graphs. *Adv. Appl. Prob.*, **25**, 348-372.

Baddeley, A.J. and Van Lieshout, M.N.M. (1993) Stochastic geometry models in high-level vision. *Statistics and Images* (Mardia, K. and Kanji, G.K. eds.), Vol. 1, *J. Appl. Statist.*, **20**, 233-258. Abingdon: Carfax.

Besag, J. (1986) On the statistical analysis of dirty pictures. *J. R. Statist. Soc.* B, **48**, 259-302.

Clifford, P. and Nicholls, G. (1994) A Metropolis sampler for polygonal image reconstruction. *Technical report.* University of Oxford.

Dubes, R.C. and Jain, A.K. (1989) Random fields models in Image analysis. *J. Appl. Statist.*, **16**, 131-164.

Gavin, J. and Jennison, C. (1997) A subpixel image analysis algorithm. *J. Comp. Graph. Statist.*, to appear.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE PAMI*, **6**, 721-741.

Geyer, C.J. (1997) Likelihood inference for spatial point processes. In *Stochastic Geometry, Likelihood and Computation* (Barndorff-Nielsen, O.E. Kendall, W.S. and van Lieshout M.N.M. eds.), to appear. London: Chapman & Hall.

Green, P.J. (1995) Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, **82**, 711-732.

Grenander, U., Chow, Y. and Keenan, D.M. (1991) Hands: a pattern theoretic study of biological shapes. *Research Notes on Neural Computing*, **2**.

Grenander, U. and Miller, M.I. (1994) Representations of knowledge in complex systems. *J. R. Statist. Soc.* B, **56**, 549-603.

Helterbrand, J.D., Cressie, N. and Davidson, J.L. (1994) A statistical approach to identifying closed object boundaries in images. *Adv. Appl. Prob.*, **26**, 831-854.

Mardia, K.V. and Qian, W. (1993) Compact object recognition using Bayesian method. *Technical report n. 17/c.* University of Leeds.

Mardia, K.V., Qian, W., Shah, D. and de Souza, K. (1997) Deformable template recognition of multiple occluded objects. *IEEE-PAMI*, to appear.

Meyn, S.P. and Tweedie, R.L. (1993) *Markov chains and stochastic stability.* London: Springer-Verlag.

Nicholls, G. (1996) Bayesian image analysis with Markov chain Monte Carlo and continuum triangulation models. *Technical report.* University of Auckland.

Qian, W. and Mardia, K.V. (1995) Recognition of multiple objects with occlusions. *Research Report STAT 95/01.* Department of Statistics, University of Leeds.

Qian, W., Titterington, D.M. and Chapman, J.M. (1994) An image analysis problem in electron microscopy. *J. Amer. Statist. Ass.*, **91**, 944-952.

Rue, H. and Hurn, M. (1997) Bayesian object identification. *Technical report no. 6/1997.* Norwegian Institute of Technology.

Rue, H. and Syversveen, A.R. (1995) Bayesian object recognition with Baddeley's delta loss. *Technical report.* Norwegian Institute of Technology.

Rosenfeld, A. and Kak, A.C. (1982) *Digital Picture Processing*, vol. II. New York: Academic Press.

Storvik, G. (1994) A Bayesian approach to dynamic contours. *IEEE PAMI*, **16**, 976-986.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701-1762.

List of Figures