IMPROVING THE ACCEPTANCE RATE OF REVERSIBLE JUMP MCMC PROPOSALS

FAHIMAH AL-AWADHI [*] MERRILEE HURN[†]
CHRISTOPHER JENNISON [‡]

October 1, 2003

ABSTRACT

Recent articles have commented on the difficulty of proposing efficient reversible jump moves within MCMC. We suggest a new way to make proposals more acceptable using a secondary Markov chain to modify proposed moves — at little extra programming cost.
KEYWORDS: Image analysis; Markov chain Monte Carlo; Object recognition; Reversible jump MCMC; Stochastic simulation; Tempering.

1 INTRODUCTION

Since 1970, when Hastings introduced Markov chain Monte Carlo algorithms into the statistics literature, their use has grown enormously and become an almost routine tool in some application areas (Hastings, 1970, Gilks et al, 1995). The appeal and applicability of these methods received another boost with the extension to distributions defined over variable dimensions, reversible jump MCMC (Green, 1995). Implementations of this RJMCMC methodology generally use a selection of move types, some of which explore the sample space within a fixed dimension while others make changes in dimensionality. Both types of move should be as efficient as possible

[*]Department of Statistics and OR, Kuwait University, P.O.Box 5969 Safat, Kuwait, 13060
[†]Corresponding author: Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY
[‡]Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY

in moving around the sample space; one sign of inefficiency is an unacceptably high rejection rate for the proposed changes. In fixed dimensions, small changes usually have higher acceptance rates than large ones, and proposal mechanisms can be scaled to achieve a desired acceptance rate. Brooks, Giudici and Roberts (2003) point out that this option is not always available for moves between dimensions as there may be no natural distance measure between states of different dimensions, for example these may correspond to competing non-nested models with different sets of parameters.

In this article we consider highly structured problems where jumps must be made between multi-modal distributions in different dimensional spaces. We focus on some of the severe problems that occur in sampling image distributions, and suggest an approach which uses a secondary Markov chain to modify an initial proposal, allowing it to move from a low-probability region of the new space towards a mode before comparing it with the starting state. The approach has the benefit that existing code may often be used for the secondary chain. In Section 2, we describe the reversible jump framework, the problem of slow movement between dimensions, and existing solutions. In Section 3, we describe our approach and the circumstances under which it is likely to be particularly beneficial. We present a simple example to illustrate the method's potential in Section 4 and an object recognition example in Section 5.

2 THE REVERSIBLE JUMP FRAMEWORK

Suppose we wish to sample from a distribution $\pi(\boldsymbol{x})$ in which the vector $\boldsymbol{x}$ is of variable dimension. Let $\pi_n(\boldsymbol{x})$ denote the sub-density with respect to Lebesgue measure on R^n for values of $\boldsymbol{x}$ with dimension n. Suppose $q_{mn}(\boldsymbol{x}, \boldsymbol{x}')$ is a proposal density on R^n, or a subspace of R^n, for moves from R^m to R^n and the reverse proposal for moving from R^n to R^m has density $q_{nm}(\boldsymbol{x}', \boldsymbol{x})$. (For $m < n$, transitions from $\boldsymbol{x} \in R^m$ to $\boldsymbol{x}' \in R^n$ are often defined such that $\boldsymbol{x}$ determines m elements of $\boldsymbol{x}'$ and $q_{mn}(\boldsymbol{x}, \boldsymbol{x}')$ is a density on the remaining $n - m$ dimensional subspace of R^n; in this case, the reverse step is deterministic and the kernel $q_{nm}(\boldsymbol{x},' \boldsymbol{x})$ degenerate.) Green (1995) gives the formula

$$\alpha_{mn}(\boldsymbol{x}, \boldsymbol{x}') = \min\left\{1, \frac{\pi_n(\boldsymbol{x}')q_{nm}(\boldsymbol{x}', \boldsymbol{x})}{\pi_m(\boldsymbol{x})q_{mn}(\boldsymbol{x}, \boldsymbol{x}')}\right\}, \tag{1}$$

for the acceptance probability needed to maintain detailed balance with respect to π.

In any MCMC sampler, the acceptance rate of proposals should be reasonably high. In fixed dimension problems, a proposal of the form $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a zero-mean random variable, can be calibrated by adjusting the variance of $\boldsymbol{\eta}$ to achieve a suitable acceptance rate. Unfortunately, there is often no simple parallel to this approach when $\boldsymbol{x}$ and $\boldsymbol{x}'$ have different dimensions.

Suggestions have been made for improving acceptance rates in RJMCMC. Green and Mira (2001) introduce a second stage to the proposal mechanism in their "delayed rejection sampler". Following rejection of any proposed move, a second proposal is made bearing in mind the first rejection, and possibly making use of the rejected proposal value. The method offers the option of proposing a large change in dimension and then switching to a smaller change in dimension if this is rejected. The overall acceptance probability takes into account this more complicated structure.

Brooks, Giudici and Roberts (2003) return to the scaling idea by looking for canonical jump functions which play the role of $\boldsymbol{\eta} = \mathbf{0}$ in tuning the acceptance rate, but now act across dimensions. For example, with nested models such as AR processes of unknown order, a canonical jump might be to increase the order by one, with the newly introduced coefficient equal to zero and all other coefficients unchanged. Once suitable correspondences have been identified, proposals can be constructed as scaled perturbations around these central moves. A second idea suggested by Brooks et al (2003) is that of using auxiliary variables to "pad out" the problem to a common dimension for all models under consideration. These auxiliary variables retain information about a model even when that model is not the current one. This means that when a move back to that model is proposed, the auxiliary variables may be able to guide the proposal to somewhere close to where it last left.

In another approach, Green (2003) tries to regain some of the simplicity of the random walk Metropolis sampler with a view to automating implementation of RJMCMC methods. A trial run is used to approximate the mean and variance of $\boldsymbol{x}$ in each dimension. These values are then used in constructing appropriate dimension Gaussian proposals both within and between dimensions.

Stephens (2000) suggests an alternative to RJMCMC sampling of mixture distributions based on spatial birth and death processes. Here, the mixture components correspond to individual marked points in the point process and dimensionality changes are made using birth and death processes to simulate the point process. Although this approach can be generalised to other applications, it does not necessarily offer a solution to mixing problems: Cappé, Robert and Rydén (2003) note the close relationship between RJMCMC and spatial birth and death processes, and warn that problems affecting the former are liable to arise in the latter.

3 ALTERNATIVE PROPOSAL MECHANISM

When a target distribution is highly modal within each dimension, proposals which land away from the modal region have virtually no chance of acceptance. If the locations of modes are not initially known, designing a matched pair of dimension-changing proposals to produce jumps from mode to mode requires much ingenuity, if indeed it is possible at all. Our approach is to modify the value of a simpler RJMCMC proposal, moving it closer to a mode before the accept-reject decision is taken.

Suppose the Markov chain sampler is currently in state $\boldsymbol{x} \in R^m$ and we have chosen to consider a jump to R^n. We first generate a proposal $\boldsymbol{x}'$ in R^n from the density $q_{mn}(\boldsymbol{x}, \boldsymbol{x}')$, just as before. But now we follow this by k fixed-dimension MCMC steps through states $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^k = \boldsymbol{x}^*$ say, each step satisfying detailed balance with respect to a new distribution $\pi_n^*(\boldsymbol{x})$. The final state $\boldsymbol{x}^* \in R^n$ is either accepted as the next state of the Markov chain or rejected, in which case the Markov chain remains at $\boldsymbol{x}$. In the reverse of this move, a jump from $\boldsymbol{x}^* \in R^n$ to R^m is proposed by taking k fixed dimension steps within R^n, each of which satisfies detailed balance with respect to $\pi_n^*(\boldsymbol{x})$, ending at $\boldsymbol{x}'$ say, then sampling from the density $q_{nm}(\boldsymbol{x}', \boldsymbol{x})$ to give the proposal $\boldsymbol{x}$ for possible acceptance. The role of π_n^* is to promote movement towards the modes of the original distribution; we will discuss the choice of $\pi_n^*(\boldsymbol{x})$ and k later.

The acceptance probability must be calculated to ensure detailed balance with respect to the overall target distribution π. Let P denote the transition kernel for the full sequence of k moves in R^n from $\boldsymbol{x}'$ to $\boldsymbol{x}^*$. Assuming equal probabilities of choosing the two types of move from R^m to R^n and from R^n to R^m, we require

$$\int_{(\boldsymbol{x},\boldsymbol{x}',\boldsymbol{x}^*)\in A\times B\times C} \pi_m(\boldsymbol{x})q_{mn}(\boldsymbol{x},\boldsymbol{x}')P(\boldsymbol{x}',\boldsymbol{x}^*)\alpha_{mn}(\boldsymbol{x},\boldsymbol{x}^*)d\boldsymbol{x}\; d\boldsymbol{x}'\; d\boldsymbol{x}^* =$$

$$\int_{(\boldsymbol{x},\boldsymbol{x}',\boldsymbol{x}^*)\in A\times B\times C} \pi_n(\boldsymbol{x}^*)P(\boldsymbol{x}^*,\boldsymbol{x}')q_{nm}(\boldsymbol{x}',\boldsymbol{x})\alpha_{nm}(\boldsymbol{x}^*,\boldsymbol{x})d\boldsymbol{x}\; d\boldsymbol{x}'\; d\boldsymbol{x}^*, \qquad (2)$$

for all Borel sets $A \in R^m$, $B \in R^n$ and $C \in R^n$ (or suitably defined sets $B(\boldsymbol{x})$ when q_{mn} and q_{nm} are defined on subspaces of R^n and R^m). Then, the probability under π of being in state $\boldsymbol{x}$, moving to $\boldsymbol{x}'$ and on to $\boldsymbol{x}^*$ and accepting this proposal exactly balances the probability under π of being in $\boldsymbol{x}^*$, making the reverse transitions to $\boldsymbol{x}'$ and on to $\boldsymbol{x}$ and accepting this proposal. One way to make the integrals equal is to make the integrands in (2) equal by taking

$$\alpha_{mn}(\boldsymbol{x},\boldsymbol{x}^*) = \min\left\{1, \frac{\pi_n(\boldsymbol{x}^*)P(\boldsymbol{x}^*,\boldsymbol{x}')q_{nm}(\boldsymbol{x}',\boldsymbol{x})}{\pi_m(\boldsymbol{x})q_{mn}(\boldsymbol{x},\boldsymbol{x}')P(\boldsymbol{x}',\boldsymbol{x}^*)}\right\} \qquad (3)$$

and

$$\alpha_{nm}(\boldsymbol{x}^*,\boldsymbol{x}) = \min\left\{1, \frac{\pi_m(\boldsymbol{x})q_{mn}(\boldsymbol{x},\boldsymbol{x}')P(\boldsymbol{x}',\boldsymbol{x}^*)}{\pi_n(\boldsymbol{x}^*)P(\boldsymbol{x}^*,\boldsymbol{x}')q_{nm}(\boldsymbol{x}',\boldsymbol{x})}\right\}.$$

Since the k individual transitions comprising P satisfy detailed balance with respect to π_n^*, so does P; $\pi_n^*(\boldsymbol{x}^*)P(\boldsymbol{x}^*,\boldsymbol{x}') = \pi_n^*(\boldsymbol{x}')P(\boldsymbol{x}',\boldsymbol{x}^*)$, and (3) simplifies to

$$\alpha_{mn}(\boldsymbol{x},\boldsymbol{x}^*) = \min\left\{1, \frac{\pi_n(\boldsymbol{x}^*)\pi_n^*(\boldsymbol{x}')q_{nm}(\boldsymbol{x}',\boldsymbol{x})}{\pi_m(\boldsymbol{x})\pi_n^*(\boldsymbol{x}^*)q_{mn}(\boldsymbol{x},\boldsymbol{x}')}\right\}. \qquad (4)$$

There is an asymmetry in transitions in the two directions: in moves from R^m to R^n the jump across dimensions occurs first, then the MCMC steps under π_n^*; moves in the other direction start by taking steps within R^n under π_n^* and then jump to the proposal in R^m. With $m < n$, this may be appropriate if there is only a need to improve initial proposals in the higher dimensional space. If it is also desirable to modify the value reached on jumping to R^m a spell of MCMC sampling under suitably modified target distributions could be included in both R^m and R^n, either side of the reversible jump step, with a consequent change in the acceptance probability formula.

It is instructive to compare likely values of $\alpha_{mn}(\boldsymbol{x},\boldsymbol{x}')$ given by (1) for the original method and of $\alpha_{mn}(\boldsymbol{x},\boldsymbol{x}^*)$ in (4) for the new method. Since the current state $\boldsymbol{x}$ has arisen in a Markov chain with equilibrium distribution π, we can hope $\boldsymbol{x}$ will take a typical value for the sub-distribution π_m. Assuming it is relatively easy to define a suitable transition to a lower dimension, we suppose $q_{nm}(\boldsymbol{x}',\boldsymbol{x})$ will be comparable with $\pi_m(\boldsymbol{x})$. Difficulty arises in generating the additional elements in $\boldsymbol{x}'$ in such a way that $q_{mn}(\boldsymbol{x}, \boldsymbol{x}')$ places weight on important parts of $\pi_n(\boldsymbol{x}')$: if this is not achieved and $\boldsymbol{x}'$ is a typical sample from $q_{mn}(\boldsymbol{x}, \boldsymbol{x}')$ but not from π_n, we have $\pi_n(\boldsymbol{x}') \ll q_{mn}(\boldsymbol{x}, \boldsymbol{x}')$ and the acceptance probability $\alpha_{mn}(\boldsymbol{x}, \boldsymbol{x}')$ in (1) will be extremely small. The difference in our new method is that the ratio $\frac{\pi_n(\boldsymbol{x}')}{q_{mn}(\boldsymbol{x},\boldsymbol{x}')}$ in the old acceptance probability (1) is replaced in the new $\alpha_{mn}(\boldsymbol{x}, \boldsymbol{x}^*)$ given by (4) by

$$\frac{\pi_n(\boldsymbol{x}^*)}{\pi_n^*(\boldsymbol{x}^*)} \times \frac{\pi_n^*(\boldsymbol{x}')}{q_{mn}(\boldsymbol{x}, \boldsymbol{x}')}. \tag{5}$$

These two factors are still liable to be less than 1 but a good choice of π_n^*, intermediate in some sense between q_{mn} and π_n, can ensure that (4) is substantially larger than (1). Distributions defined in terms of a temperature γ could be employed, where $\pi^* = \pi^\gamma$ is a "tempered" version of the target distribution π. Although not the only option, tempering is a natural way to create a flatter version of π for the transition process to pass through, but one which retains the same modes. There is a compromise here, between making π^* easy to sample efficiently, and ensuring that the resulting $\boldsymbol{x}^*$ do actually tend to be in modal regions of π. Sampling from a tempered version of π can also be based on the same fixed-dimensional transitions used in sampling from π, hence little additional programming is needed beyond that required to implement the original reversible jump method. (For more discussion of tempering see Neal, 1996.)

4 A SIMPLE ILLUSTRATIVE EXAMPLE

The following example demonstrates the scale of improvement our proposed method offers. Suppose π is defined on R^1 and R^2 with π_1 equal to 0.5 times a $N(0,1)$ density on R^1 and π_2 equal to 0.5 times a bivariate normal density on R^2 in which X_1 and X_2 are independent with $X_1 \sim N(0,1)$ and $X_2 \sim N(0,\sigma^2)$. In proposing transitions from $\boldsymbol{x} \in R^1$ to $\boldsymbol{x}' \in R^2$ we set $x_1' = x_1$ and generate x_2' from a $N(0,\xi^2)$ distribution where $\xi^2 \gg \sigma^2$; thus, this is an example of a proposal kernel q_{12} which is poorly matched to π_2. For transitions from $\boldsymbol{x}' \in R^2$ to $\boldsymbol{x} \in R^1$ we simply take $x_1 = x_1'$.

Consider first the standard reversible jump method when the typical value $x_2' = \xi$ is proposed. (A more thorough treatment would integrate over the $N(0,\xi^2)$ distribution of x_2' but this simple case provides a simple, and compelling, illustration.) The factor in $\pi_2(\boldsymbol{x}')$ associated with x_1' is $\pi_1(x_1')$ and the remaining factor is a $N(0, \sigma^2)$ density evaluated at $x_2' = \xi$. By definition, $q_{21}(\boldsymbol{x}', \boldsymbol{x}) = 1$. In the denominator of (1) $\pi_1(\boldsymbol{x}) = \pi_1(x_1')$ and $q_{12}(\boldsymbol{x}', \boldsymbol{x})$ is a $N(0, \xi^2)$ density evaluated at ξ. Writing $f_v(w)$ to denote a $N(0, v)$ density evaluated at w, the acceptance probability of $\boldsymbol{x}'$ is

$$\alpha_{12}(\boldsymbol{x}, \boldsymbol{x}') = \min\left\{1, \frac{\pi_2(\boldsymbol{x}')q_{21}(\boldsymbol{x}', \boldsymbol{x})}{\pi_1(\boldsymbol{x})q_{12}(\boldsymbol{x}, \boldsymbol{x}')}\right\} = \frac{f_1(x_1)f_{\sigma^2}(x_2')}{f_1(x_1)f_{\xi^2}(x_2')}$$

$$= \frac{f_{\sigma^2}(\xi)}{f_{\xi^2}(\xi)} = \frac{\xi}{\sigma}\exp\{-\frac{1}{2}(\frac{\xi^2}{\sigma^2} - 1)\} \tag{6}$$

and for $\xi^2 \gg \sigma^2$ this probability can be very small.

Now consider our method with π_2^* defined as a bivariate density in which X_1 and X_2 are independent, $X_1 \sim N(0, 1)$ and $X_2 \sim N(0, \phi^2)$ with $\sigma^2 < \phi^2 < \xi^2$. After an initial move to $\boldsymbol{x}'$ where $x_1' = x_1$ and x_2' takes the typical value ξ, the k MCMC steps will lead to a new value x_1^* of x_1 and a value x_2^* sampled from π_2^*. Using the typical value ϕ for x_2^*, we obtain

$$\alpha_{12}(\boldsymbol{x}, \boldsymbol{x}^*) = \min\left\{1, \frac{\pi_2(\boldsymbol{x}^*)\pi_2^*(\boldsymbol{x}')q_{21}(\boldsymbol{x}', \boldsymbol{x})}{\pi_1(\boldsymbol{x})\pi_2^*(\boldsymbol{x}^*)q_{12}(\boldsymbol{x}, \boldsymbol{x}')}\right\} = \frac{f_1(x_1^*)f_{\sigma^2}(x_2^*)f_1(x_1)f_{\phi^2}(x_2')}{f_1(x_1)f_1(x_1^*)f_{\phi^2}(x_2^*)f_{\xi^2}(x_2')}$$

$$= \frac{f_{\sigma^2}(\phi)f_{\phi^2}(\xi)}{f_{\phi^2}(\phi)f_{\xi^2}(\xi)} = \frac{\xi}{\sigma}\exp\{-\frac{1}{2}(\frac{\xi^2}{\phi^2} + \frac{\phi^2}{\sigma^2} - 2)\}. \tag{7}$$

This acceptance probability (7) is greater than that given by (6) for all $\sigma^2 < \phi^2 < \xi^2$, and is maximised over ϕ^2 by $\phi^2 = \xi\sigma$ which we shall assume to be used. In this case

$$\alpha_{12}(\boldsymbol{x}, \boldsymbol{x}^*) = \frac{\xi}{\sigma}\exp\{-(\frac{\xi}{\sigma} - 1)\}. \tag{8}$$

Since only the square root of the ratio ξ^2/σ^2 now appears in the exponential term, this acceptance probability decreases much more slowly than (6) as ξ^2/σ^2 increases. A few numerical values of (6) and (8) are tabulated below:

ξ^2/σ^2	$\alpha_{12}(\boldsymbol{x}, \boldsymbol{x}')$	$\alpha_{12}(\boldsymbol{x}, \boldsymbol{x}^*)$
10	0.035	0.364
25	0.00003	0.092
50	1.6×10^{-10}	0.016
100	3.2×10^{-21}	0.0012

Further analysis of this example shows there can be additional gains from introducing a succession of target distributions intermediate between q_{12} and π_2: with $r-1$ suitably chosen intermediate distributions, the rth root of ξ^2/σ^2 appears in the exponential in $\alpha_{12}(\boldsymbol{x}, \boldsymbol{x}')$. The clear conclusion is that our method can lead to a practically useful acceptance rate in situations where the original method is hopelessly slow.

The price of the improved acceptance rate for jumps between dimensions is the computational cost of the k intermediate MCMC steps. In this particular example, an independence sampler can be used for the intermediate steps and $k = 1$ suffices. In other situations, more steps may be necessary to reach states typical of π_n^*. Then, the choice of k will be a compromise, large enough for the intermediate sampler to move towards a mode but small enough to remain inexpensive computationally.

5 An example in object recognition

We consider an example which first motivated our search for an improved proposal mechanism. The data $\boldsymbol{Y}$ displayed in Figure 1 were obtained by confocal microscopy. Our aim is to identify the number of the cells, N, in the scene and approximate their shapes by a set of ellipses $\boldsymbol{X}$. Using a Bayesian approach, inference is based on the posterior distribution of N and $\boldsymbol{X}$ given the data $\boldsymbol{Y} = \boldsymbol{y}$. In line with our previous notation, we denote this posterior distribution $\pi(\boldsymbol{x})$ and let $\pi_n(\boldsymbol{x})$ denote the sub-density of the set of parameters for scenes with n cells — now with respect to Lebesgue measure on R^{6n} since each cell has 6 parameters. Thus,

$$\pi_n(\boldsymbol{x}) \;\propto\; \mathcal{L}(\boldsymbol{y}|n, \boldsymbol{x})\, p_n(\boldsymbol{x}) \tag{9}$$

where $\mathcal{L}(\boldsymbol{y}|n, \boldsymbol{x})$ is the likelihood of the observed data given N and $\boldsymbol{X}$, and p_n is the density assigned to this scene in the prior model. Following the marked pointed process approach of Baddeley and van Lieshout (1993), we take the prior model to be a hard core interaction point process which prevents cells overlapping. The data $\boldsymbol{Y}$ are assumed to be Gaussian with different means and variances assigned to observations from different cells and from the background region; for more details see, for example, Rue and Hurn (1999) or Al-Awadhi, Jennison and Hurn (2003).

In MCMC sampling from π, several types of proposal are defined, including fixed

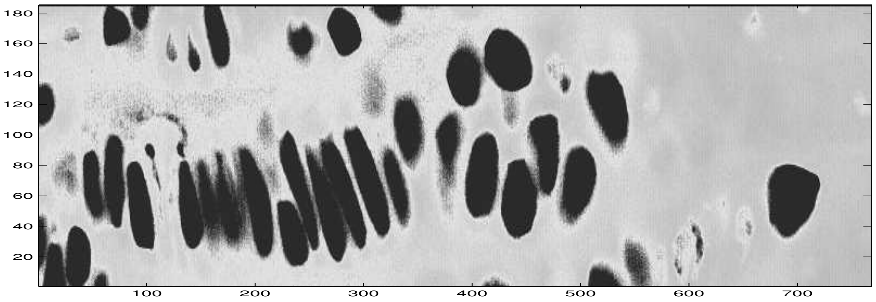

FIGURE 1: The image data.

dimensional moves which alter the size, shape or location of a single cell, as well as reversible jump moves which alter the dimension of x by adding or removing a cell. Other very useful proposals are to split one cell into two or, conversely, to merge two nearby cells into one. Such split and merge transitions are essential when there is uncertainty whether a feature in the data is one large cell or two smaller ones because moving between these two configurations by birth and death moves would require the chain to pass through intermediate states of very low probability. It is difficult to create rules for splitting or merging cells expressed in terms of the various shape, orientation and location parameters which manage to be both efficient and inexpensive to compute. Crude proposals are generally available, but these tend to suffer high rejection rates since the current configuration has had ample opportunity to match the data well while an initial proposal of merged or split cells has not. Incorporating a fit to the data into the proposal itself is hard because of the computational cost of evaluating the posterior over a range of values for the proposal and of evaluating the reverse step to compute the acceptance probability. Rue and Hurn (1999) demonstrate one approach for such "data-matching" suitable for work with deformable polygonal templates, but the restriction to ellipses prevents us from using their approach here.

Figure 2 displays some simple split and merge proposals for elliptical templates. These proposals are easily generated and use the common RJMCMC idea of averaging pairs of parameters when reducing dimension, and perturbing a parameter by a random amount in opposite directions to increase dimension. It is only sensible to attempt to merge cells which are close together and a condition formalising this is included in the proposal (and thus in the acceptance ratio). To illustrate how poorly the Markov chain can mix, we note results from several runs of 20,000 steps of the sampler incorporating all move types: these runs produced between 1 and 5 accepted splits and no merges.

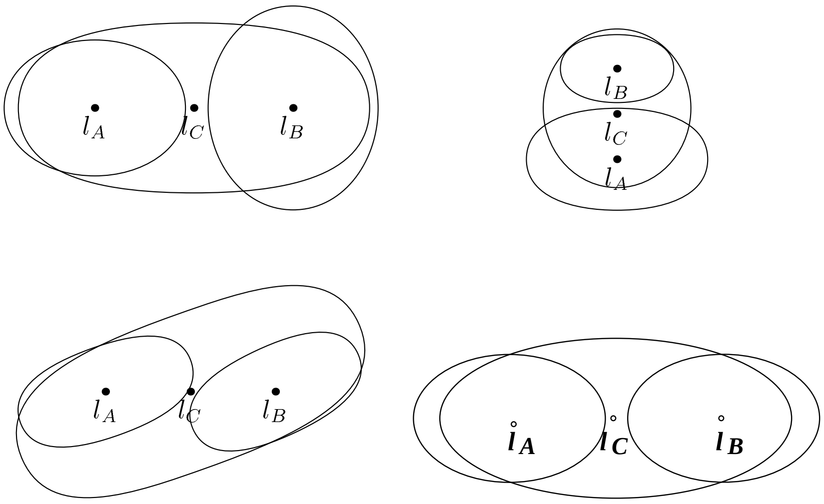

FIGURE 2: Some possible proposed split and merge moves

In this problem, we know that the high failure rate of proposals is due to their relative lack of fit to the rather strong data; our suggestion is to improve the data matching of a proposal before making the accept/reject decision. We apply tempering to the contribution to π from the likelihood term to define the modified distribution

$$\pi_n^*(\boldsymbol{x}) \propto \mathcal{L}^\gamma(\boldsymbol{y}|n, \boldsymbol{x})\, p_n(\boldsymbol{x}). \tag{10}$$

Here $\gamma \in (0, 1)$ and a good choice of γ should give sufficient weight to the data to improve the data-fit of the final proposal $\boldsymbol{x}^*$ while limiting the effect of the low probability under π of the initial proposal $\boldsymbol{x}'$ on the final acceptance probability (4). In modifying $\boldsymbol{x}'$ we use the usual fixed dimensional proposal types. Note too that only the cells affected by the RJ proposal are changed, all others are held fixed.

Selecting the tempering parameter γ and the number of steps k is a tuning problem; we have repeated experiments running 20,000 iterations of reversible jump Markov chains using various combinations of the tempering parameter γ and the number of steps, k. Results, averaged over 3 replicates, are listed in Table 1 and show an increase in the number of successful splits and merges across parameter combinations. In particular, there have been successful merge moves, something which was not achieved before. There is no significance evidence of improvement when k is increased from 100 to 300, suggesting that adequate data-matching occurs quite quickly and a low value of $k = 50$ or $k = 100$ is sufficient. Obviously this is important because the cost of the modified proposals increases with k (although at less than a factor of k since the variable dimension part of the move is more costly than the fixed dimension part). Note also that the overall cost of the sampler increases by an even smaller factor since only the merge and split moves become more expensive.

There is also welcome robustness to the choice of the tempering parameter γ: values between 0.1 and 0.001 seem effective. The success at very small values of γ is perhaps surprising but can be attributed to the strong signal in this problem which leads to a highly modal target distribution π. Some pairs of image configurations during the RJMCMC runs are shown in Figure 3 to illustrate the scale of the changes; in each pair, the arrow indicates the area about to be affected. The modified method has succeeded in improving a sampler which was really not mixing at all. Although the acceptance rates remain painfully low, we now have greater confidence that the sampler is capable of changing dimension through these essential routes. For this extremely hard sampling problem, this is significant progress.

In this example, π and π^* have the same support. In consequence, any proposal $\boldsymbol{x}'$ with zero probability under π (as a result of cell overlaps, for example) can be rejected straight away as we know $P(\boldsymbol{x}^*, \boldsymbol{x}')$ will be equal to zero in the hypothetical return route from a legal $\boldsymbol{x}^*$. It may sometimes be beneficial to extend the support of π^*, thereby allowing initial jumps to an illegal $\boldsymbol{x}'$ to revert to legitimacy before reaching the final proposal $\boldsymbol{x}^*$. In order to encourage $\boldsymbol{x}'$ to be in the state space of π, relatively little probability should be assigned to the extension of the state space. Similar ideas have been explored in a different context by Hurn, Rue and Sheehan (1999).

In conclusion, the approach suggested in this article is one way in which low acceptance rates for RJMCMC samplers could be improved. It is particularly suitable for problems where it is hard to design very efficient proposals directly, but where the modality of the target distribution gives clues as to why many proposals are rejected. As another application, consider mixture modelling where observations are allocated to an unknown number of components; proposing to split a component may only be acceptable if the observations are reasonably well allocated to the two new components. A crude split might allocate the observations at random to the two new components; our modification would allow reallocation proposals before accepting or

	$k=50$		$k=100$		$k=300$	
γ	merges	splits	merges	splits	merges	splits
0.1	6	14	5	9	5	7
0.01	3	3	10	5	10	4
0.001	2	3	8	6	10	4
0.0001	1	2	4	5	4	7

TABLE 1: The average number of acceptances for different values of γ and k.

rejecting this split. Although some extra running costs are incurred in the modified proposals, there is little additional programming outlay, as existing code may be used.

REFERENCES

Al-Awadhi, F., Jennison, C. and Hurn, M. (2003), Statistical Image Analysis for a Confocal Microscopy 2D Section of Cartilage Growth, *Applied Statistics, Journal of the Royal Statistical Society, Series C*, to appear.

Baddeley, A. and van Lieshout, M., (1993), Stochastic Geometry Models in High-level Vision, in *Statistics and Images*, eds Mardia and Kanji: Carfax Publishing, Abingdon, pp 231–256.

Brooks, S., Giudici, P. and Roberts, G., (2003), Efficient Construction of Reversible Jump MCMC Proposal Distributions (with discussion), *Journal of the Royal Statistical Society, Series B*, $\boldsymbol{B65}$, 3–56.

Cappé, O., Robert, C. and Rydén, T., (2003), Reversible Jump, Birth-and-death and More General Continuous Time Markov Chain Monte Carlo Samplers, *Journal of the Royal Statistical Society, Series B*, $\boldsymbol{B65}$, 679–700.

Gilks, W., Richardson, S. and Spiegelhalter, D., (1995), *Markov Chain Monte Carlo in Practice*, (Chapman and Hall, London).

Green, P., (1995), Reversible Jump MCMC Computation and Bayesian Model Determination, *Biometrika*, $\mathbf{82}$, 711–732.

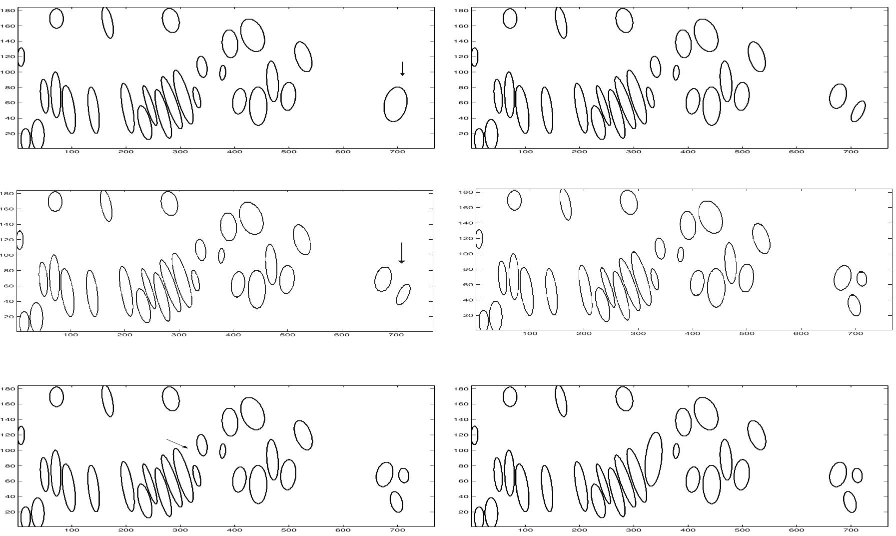

FIGURE 3: Several pairs of output configurations using the modified proposal technique (the areas affected by successful split and merge moves are pointed out using arrows, the left panel is before and the right panel is after the move).

Green, P., (2003), Transdimensional Markov Chain Monte Carlo, in *Highly Structured Stochastic Systems*, eds Green, Hjort and Richardson: OUP, Oxford, 179–198.

Green, P. and Mira, A., (2001), Delayed Rejection in Reversible Jump Metropolis-Hastings, *Biometrika*, **88**, 1035–1053.

Hastings, W., (1970), Monte Carlo Sampling Methods using Markov Chains and their Applications, *Biometrika*, **57**, 97–109.

Hurn, M., Rue, H. and Sheehan, N., (1999), Block Updating in Constrained Markov Chain Monte Carlo Sampling, *Statistics and Probability Letters*, **41**, 353–361.

Neal, R., (1996), Sampling from Multimodal Distributions using Tempered Transitions, *Statistics and Computing*, **6**, 353–366.

Rue, H. and Hurn, M., (1999), Bayesian Object Identification, *Biometrika*, **86**, 649–660.

Stephens, M., (2000), Bayesian Analysis of Mixture Models with an Unknown Number of Components: An Alternative to Reversible Jump Methods, *Annals of Statistics*, **28**, 40–74.