# AN ALTERNATIVE TO THE RJMCMC ALGORITHM

**Vincent Mazet** [*] **David Brie** [**]

[*] *LSIIT UMR 7005, Université Louis Pasteur, CNRS, Pôle API, BP 10413, 67412 Illkirch Cedex, France (e-mail: vincent.mazet@ensps.u-strasbg.fr)*
[**] *CRAN UMR 7039, Nancy University, CNRS, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France (e-mail: david.brie@cran.uhp-nancy.fr)*

Abstract: This paper studies a new approach for non-constant dimension problems, such as, for example, mixture deconvolution. In a Bayesian framework, Monte Carlo Markov chain (MCMC) algorithms provides an efficient way to optimize the problem. A particular class of these algorithms have been developed since classical MCMC methods (*e.g.* Metropolis-Hastings or Gibbs) cannot deal with systems whose order may change. The most famous of these algorithms is the so-called reversible jump MCMC which is recalled in this article. An alternative approach to reversible jump is also proposed: it consists in working with a constant dimension model and introducing a variable coding the occurrences of the objects to estimate. It provides an interesting method whose main advantage is to have a faster dynamic behavior than reversible jump MCMC.

Keywords: Model selection, Markov Chain Monte Carlo, reversible jump algorithms, Bayesian methods.

## 1. INTRODUCTION

The goal of statistical signal processing is to estimate unknown parameters from data. A lot of statistical problems face to the fact that the dimension of the object of inference (the parameter vector) is not fixed. Some examples are:

- multiple change-point problems (Green, 1995; Dobigeon *et al.*, 2005), which assume a piecewise-constant signal where unknown parameters to estimate are the time of arrival and amplitudes of changes;
- mixture analysis (Richardson and Green, 1997) consisting in decomposing the data distribution into elementary ones. In this application, the unknowns are the parameters of elementary distributions (*i.e.* locations, amplitude, width, etc.);
- mixture deconvolution (Mazet, 2005) in which the signal has to be decomposed in simple patterns, whose locations, amplitudes and form parameters are the unknown parameters;
- extraction of structures using marked point processes (Lacoste *et al.*, 2005; Perrin *et al.*, 2005), aiming at revealing any object in an image (as building extraction in remote sensing imagery). The unknown parameters are, again, the object parameters (position, orientation, size, color, ...).

Moreover, in the considered problems, the number of parameters (in other words, the object number), is unknown and then has to be estimated.

In fact, "the number of things you don't know is one of the things you don't know" (Richardson and Green, 1997).

In the sequel, we suppose that each object as a set of parameters (position, amplitude, etc.) gathered into vector $\theta_k^K$, where $k \in \{1, \ldots, K\}$ corresponds to the object index and $K$ to the object number. Every $\theta_k^K$ is then a vector whose length equals the object parameter number, and $\theta^K$ gathers all the object parameters to estimate. Finally, $y$ denotes the observed data which are a function of the unknowns:

$$y = f\left(\theta^K\right). \tag{1}$$

A Bayesian approach is attractive for such problems (see last references). Furthermore, MCMC (Markov chain Monte Carlo) methods are efficient algorithms for solving such problems (for a review, see (Gilks *et al.*, 1996; Robert, 1996)). MCMC methods are iterative algorithms which generate, at each iteration, a random variable from the parameter posteriors, yielding a Markov chain from which an estimator is computed. However, if the model dimension is not constant (that is, the object number may change from one iteration to the next), one has to deal with MCMC methods for model determination. For ten years, new MCMC techniques have been proposed, as, for example, reversible jump MCMC (RJMCMC) algorithm (Green, 1995; Richardson and Green, 1997), Carlin & Chib algorithm (Carlin and Chib, 1995), birth and death MCMC algorithm (Stephens, 2000; Cappé *et al.*, 2003), ... The most used technique is RJMCMC which will be detailed in the next section.

Nevertheless, an alternative to these algorithms is presented in this paper in a Bayesian framework. The idea is to use a model in which the system order is constant, which then allows to use the Gibbs sampler. Indeed, we consider a maximal number of objects and introduce a variable coding their presence or not (*i.e.* their occurrence). Then, the proposed approach has a faster dynamic behavior than the RJMCMC algorithm, and, as a consequence, a smaller burn-in period.

This article is structured as follows. In next section, we present the RJMCMC algorithm and in section 3 the proposed alternative method. Then, we discuss about these two methods in section 4. Finally, section 5 concludes the paper.

## 2. REVERSIBLE JUMP MCMC

(Green, 1995) proposed a new framework for the construction of reversible Markov chain samplers that jumps between spaces of different dimensionality. Its algorithm is based on a modified version of the Metropolis-Hastings algorithm.

Suppose that we propose a move from current state $S = (K, \theta^K)$ (*i.e.* $K$ objects with parameters $\theta^K$) to $S' = (K', \theta^{K'})$. Because the reversible jump sampler is based on the Metropolis-Hastings algorithm, the move is not automatically accepted. The acceptance probability is given by:

$$\alpha = \min\left(1, R(S, S')\right), \tag{2}$$

where

$$R(S, S') = \frac{p(K', \theta^{K'}|y)\tilde{q}(K, \theta^K|K', \theta^{K'})}{p(K, \theta^K|y)\tilde{q}(K', \theta^{K'}|K, \theta^K)}. \tag{3}$$

Generally, relationships between parameters of different models can be used by drawing variables $u$ and $u'$ from proposal distributions $\tilde{q}_2(u)$ and $\tilde{q}_2(u')$ and then forming $\theta^{K'}$ as deterministic functions of the form $\theta^K = g(\theta^{K'}, u)$ and $\theta^{K'} = g(\theta^K, u')$. In this way, it is straightforward to incorporate useful information from the current parameter vector $\theta^K$ into the proposal for the new parameter vector $\theta^{K'}$. Provided that the "dimension matching" condition is fulfilled (that is, $\dim(\theta^{K'}, u) = \dim(\theta^K, u')$), the acceptance probability is given by (Green, 1995):

$$R(S, S') = \frac{p(K', \theta^{K'}|y)}{p(K, \theta^K|y)} \times \frac{\tilde{q}_1(K|K')\tilde{q}_2(u)}{\tilde{q}_1(K'|K)\tilde{q}_2(u')} \times \left|\frac{\partial(\theta^{K'}, u)}{\partial(\theta^K, u')}\right|$$

which includes a Jacobian term (absolute value of the determinant of the Jacobian matrix) to account for the change between $(\theta^K, u)$ and $(\theta^{K'}, u')$.

Different moves from $S$ to $S'$ can be proposed. At each iteration of the algorithm, each move is explored either deterministically or randomly. In this paper, we propose to work with the three following popular moves (Green, 1995; Richardson and Green, 1997; Andrieu *et al.*, 2001):

- the birth of a new object (with parameters drawn at random);
- the death of an existing object chosen randomly;
- the update of all the parameters $\theta$.

The birth and death moves correspond respectively to a dimension change from $K$ to $K+1$ and from $K$ to $K-1$. Other moves may of course been proposed (as the split/merge moves (Richardson and Green, 1997)). At each iteration, one of these three moves is randomly chosen. The probabilities of these moves are $b_K$, $d_K$ and $u_K$ such that $b_K + d_K + u_K = 1$. We choose the probabilities $b_K = d_K = 0.3$ and $u_K = 0.7$ except for the cases $K = 0$ (then $d_K = 0$) and $K = K_{\max}$ (then $b_K = 0$), where $K_{\max}$ is an upper bound for the object number.

Consider the particular case of a birth move consisting in testing the proposed state $S' = (K+$

$1, \theta^{K+1})$ from the current state $S = (K, \theta^K)$. In implementation, it will often be convenient to take the advantage of any nested structure in the model or interrelationships between the parameters of different models in constructing effective proposal distributions, rather than proposing an entire new parameter vector. So, for the birth move, a fully nested model structure can easily be implemented by fixing the first $K$ parameters in both models and then the acceptance ratio (3) simplifies to

$$R_b(S, S') = \frac{p(K+1, \theta^{K+1}|y)}{p(K, \theta^K|y)} \times \frac{\tilde{q}_1(K|K+1)}{\tilde{q}_1(K+1|K)\tilde{q}_2(\theta_{K+1}^{K+1}|\theta^K)} \times 1.$$

The acceptance probability of the proposed birth move is then $\alpha = \min(1, R_b(S, S'))$. In the same way, we have the acceptance ratio for the death move equals to $R_d(S, S') = 1/R_b(S, S')$.

## 3. PROPOSED APPROACH

An alternative method to RJMCMC is presented in this section. The idea is to consider the number of object constant and equals to $K_{\max}$, so that the object parameters are gathered into vector $\theta^{K_{\max}}$ whose length is constant. This makes the use of the Gibbs sampler possible. $K_{\max}$ corresponds to an upper bound greater than the real object number and is fixed by the user. In fact, it is supposed to represent the a priori upper bound of objects. However, to avoid to obtain an estimation with $K_{\max}$ objects, we are inspired by the Bernoulli-Gaussian (BG) model (Kormylo and Mendel, 1982; Champagnat *et al.*, 1996) and introduce the vector $q \in \{0, 1\}^{K_{\max}}$ coding the object occurrences. Thus, for all $k \in \{1, \ldots, K_{\max}\}$:

- if $q_k = 1$, then the $k$th object exists;
- on the contrary, if $q_k = 0$, the $k$th object is not present: then it does not appear in the estimation because it has no signification.

With this model, the variables to estimate are then $(\theta^{K_{\max}}, q)$. Being still inspired by the BG model, the object occurrences $q_k$ are a priori distributed according to a Bernoulli distribution with parameter $\lambda \in [0; 1]$, that is to say $q_k = 1$ with probability $\lambda$, or:

$$\forall k \in \{1, \ldots, K_{\max}\}, \quad q_k \sim (1-\lambda)\delta_0(q_k) + \lambda\delta_1(q_k).$$

In all the considered applications, it exists a parameter, say $a$, which codes the object amplitude (or intensity, or height, etc.). Obviously, if $a_k = 0$, then the object does not appear in the data: it has no significance. Therefore, we propose to set $a_k$ to zero if and only if $q_k$ is zero too. Rather than estimating separately $q$ and $a$, it is often better to estimate the couple $(q, a)$. For instance, if $a$ is distributed according to a Gaussian distribution, then the couple $(q, a)$ is a Bernoulli-Gaussian process (see (Champagnat *et al.*, 1996; Kormylo and Mendel, 1982) for more information of Bernoulli-Gaussian processes).

So, since objects whose occurrence is zero have also a zero amplitude, equation 1 reads:

$$y = f\left(\theta^{K_{\max}}\right). \tag{4}$$

Note also that for object parameters which do not have a classical posterior distribution (*e.g.* the object locations), one can simulate them using a common Metropolis-Hastings algorithm. Furthermore, it may also be interesting to use a proposal distribution of the form:

$$\tilde{q}(\xi) = \begin{cases} \tilde{q}_0(\xi) & \text{if } q_k = 0, \\ \tilde{q}_1(\xi) & \text{otherwise.} \end{cases}$$

where $\xi \in \theta_k^{K_{\max}}$ is the considered object parameter. On the one hand, we advise to choose for $\tilde{q}_1$ a proposal distribution centered on the current value of $\xi$ and with small variance. Indeed, if $q_k = 1$, then the object exists and their parameters are roughly estimated. The goal is then to define precisely their parameters. This result in a random-walk Metropolis-Hastings algorithm. On the other hand, if $q_k = 0$, the object is not present: the goal is then to search the entire space to try, if possible, to make the object present. One can use, for example, a uniform distribution for $\tilde{q}_0$.

## 4. DISCUSSION

The previous methods were implemented in the case of mixture deconvolution (Mazet, 2005). Both have the same initial values and prior distributions. Moreover, only the birth and death moves have been implemented in the RJMCMC algorithm. From a general manner, the two approaches give similar results.

Our proposed approach needs to estimate the Bernoulli variable $\lambda$. This can be done easily since $\lambda$ depends only on $q$, so its posterior is:

$$\lambda|q \sim \mathcal{B}e\left(\bar{K} + 1, 2K_{\max} - \bar{K} + 1\right) \tag{5}$$

where $\bar{K} = \sum_{k=1}^{K} q_k$ corresponds to the estimated object number and $\mathcal{B}e$ denotes the beta distribution which can be easily simulated (Robert, 2001, appendixes A and B).

Besides, at each iteration of the proposed approach, $K_{\max}$ objects are considered. In other words, the algorithm generates a number of random variables equals to the parameter number of one object times the object number, plus the hyperparameters. In comparison, the RJMCMC algorithm considers $K$ objects (for the update move) or only one (for the birth and death moves).

In that case, the same number of iteration is performed faster with the RJMCMC algorithm. However, note that only a part of these iterations are concerned with the parameter updating, contrary to the proposed approach in which each variable is simulated at each iteration. Therefore, considering the same iteration number for both approaches, one has to deal with the following fact: the RJMCMC algorithm is faster but generates less variables than the proposed algorithm.

Nevertheless, the major advantage of our approach is to have a faster behavior to "jump" between spaces with very different dimensions. Indeed, while the RJMCMC is able to create or destroy only one object in one iteration (with the birth and death moves), the proposed approach is able to create or destroy many ones! Thus, the proposed approach converges faster from far initial conditions. Figure 1 shows the Markov chain of $K$ obtained with the two methods. One can see that the burn-in period of our approach (about 10 iterations) is smaller than the RJMCMC one (more than 500 iterations). As well, figures 2(a) and 2(b) illustrate the fact that Markov chains generated with the RJMCMC algorithm evolutes slower than those generated with the proposed algorithm. Indeed, with the proposed approach, a fake estimated object is present during only one iteration, contrary to the RJMCMC approach in which it can exist during several iterations before being proposed to death. The advantages of this behavior are manifold:

- if the signal-to-noise ratio is very high, then only few iterations are needed with the proposed approach to yield a good estimation. On the contrary, because an RJMCMC approach has a slower dynamic, the iteration number has to be longer. This is due to the fact that the burn-in period of the proposed approach is smaller (see figure 1);
- with the proposed approach, the estimation is computed with all the generated variables (assuming there is no burn-in period). With the RJMCMC approach, it is also common to compute the estimation with all the variables, although several are represented more than once. Indeed, because some iterations do not update variables (since only a birth or death move is proposed), the values of the Markov chain at the current iteration equal the values at the last iteration (except, maybe, for the new-born or dead object).

To withdraw with these problems, one can use other moves. Hence, the merge and split moves of the RJMCMC algorithm (Richardson and Green, 1997) consists in replacing two objects by one or to create two objects from a single one. However, these moves cannot allow to create or destroy

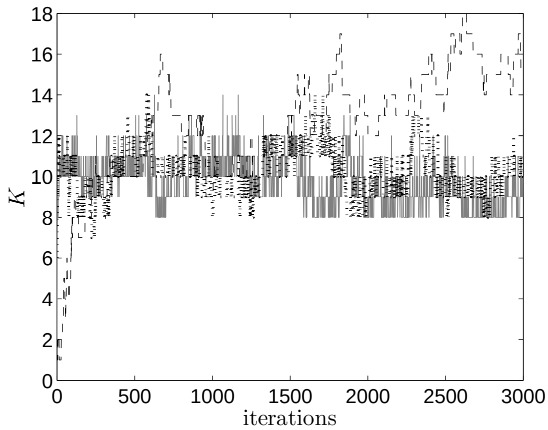
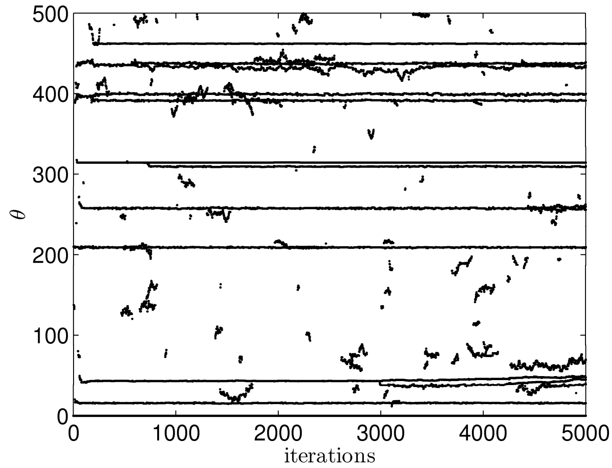
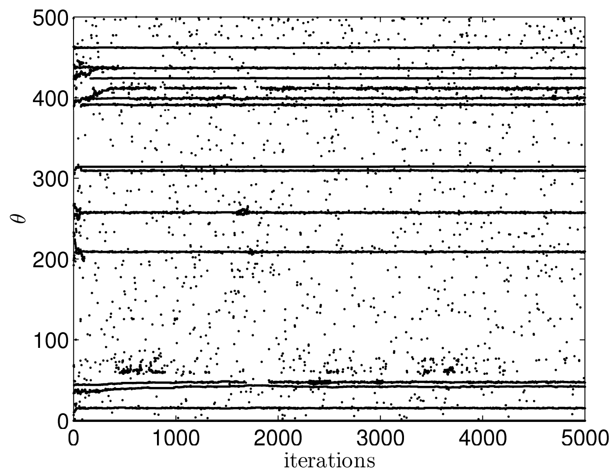



Fig. 1. Evolution of the estimated number of object (real value: 15; initial value: 1). Dashed line: RJMCMC; plain line: proposed approach; dotted line: RJMCMC with multiple births and multiple deaths moves.




(a) RJMCMC approach.




(b) Proposed approach.

Fig. 2. Markov chains of the mixture positions.

more than one object. Besides, these moves need to calculate a Jacobian which can be not so easy.

So, one can consider two new moves for the RJMCMC algorithm which are: a "multiple births" move and a "multiple deaths" move. The idea is, as in the proposed approach, to fix an upper bound $K_{\max}$ for the object number. On the one hand, the "multiple births" move consists in proposing $K_{\max} - K$ objects and accept it with an acceptance probability. On the second hand, the "multiple deaths" move proposes to destroy each object with another acceptance probability. In fact, This new version of the reversible jump algorithm performs, at each iteration, $K_{\max} - K$ births or $K$ deaths or the parameters updating. Therefore, as shown on figure 1, the dynamic behavior of this algorithm is (almost) as good as the one of the proposed approach, but, as for the common RJMCMC algorithm, only a part of the iterations are concerned with the parameters updating. Nevertheless, these alternative moves are interesting in an RJMCMC approach.

## 5. CONCLUSION

An alternative to the reversible jump algorithm has been proposed in this article. The main idea consists in considering a maximal number of objects and adding a variable which codes the object occurrence. The major advantage of the proposed approach is to allow to create or destroy more than one object at each iteration, and to perform the update of the variables in the same way. Consequently: (i) the burn-in period of the proposed approach is smaller than the RJMCMC algorithm one, (ii) the algorithm evolutes faster (*i.e.* it spend less time in a wrong configuration), and (iii) it generates more variables than a classical RJMCMC method. By replacing the birth and death moves by the "multiple births" and "multiple deaths" moves, the RJMCMC approach is able to evolves faster. Nevertheless, it generates less random variables comparing to the proposed approach. In fact, the interest of the proposed approach is to perform birth, death and update of one object in the same time.

## REFERENCES


Andrieu, C., É. Barat and A. Doucet (2001). Bayesian deconvolution of noisy filtered point processes. *IEEE Trans. Signal Processing* **49**(1), 134–146.

Cappé, O., C. Robert and T. Rydén (2003). Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *JRSSB* **65**(3), 679–700.

Carlin, B.P. and S. Chib (1995). Bayesian model choice via Markov chain Monte Carlo. *J. Roy. Stat. Soc. B* **57**, 473–484.

Champagnat, F., Y. Goussard and J. Idier (1996). Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE Trans. Signal Processing* **44**(12), 2988–2998.

Dobigeon, N., J.-Y. Tourneret and J. D. Scargle (2005). Change-point detection in astronomical data by using a hierarchical model and a bayesian sampling approach. In: *IEEE Workshop Statistical Signal Processing.* #116, Bordeaux, France.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., Eds. (1996). *Markov chain Monte Carlo in practice.* Chapman & Hall.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**(4), 711–732.

Kormylo, J.J. and J.M. Mendel (1982). Maximum likelihood detection and estimation of Bernoulli-Gaussian processes. *IEEE Trans. Inform. Theory* **28**(3), 482–488.

Lacoste, C., X. Descombes and J. Zerubia (2005). Point processes for unsupervised line network extraction in remote sensing. *IEEE Trans. Pattern Analysis and Machine Intelligence* **27**(10), 1568–1579.

Mazet, V. (2005). Développement de méthodes de traitement de signaux spectroscopiques : estimation de la ligne de base et du spectre de raies. PhD thesis. Université Henri Poincaré, Nancy 1.

Perrin, G., X. Descombes and J. Zerubia (2005). A marked point process model for tree crown extraction in plantations. In: *Proc. IEEE International Conference on Image Processing (ICIP).* Genova.

Richardson, S. and P.J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Stat. Soc. B* **59**, 731–792.

Robert, C.P. (1996). *Méthodes de Monte Carlo par chaînes de Markov.* Economica.

Robert, C.P. (2001). *The Bayesian Choice.* 2nd ed.. Springer.

Stephens, M. (2000). Bayesian analysis of mixtures with an unknown number of components — an alternative to reversible jump methods. *Annals of Statistics* **28**, 40–74.