

# Decomposition of a Chemical Spectrum using a Marked Point Process and a Constant Dimension Model

V. Mazet\*, D. Brie\* and J. Idier†

\*CRAN UMR 7039, Nancy University, CNRS,  
BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France  
(e-mail: [firstname.lastname@cran.uhp-nancy.fr](mailto:firstname.lastname@cran.uhp-nancy.fr))

†IRCCyN UMR 6597, École Centrale de Nantes, CNRS,  
1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3, France  
(e-mail: [jerome.idier@irccyn.ec-nantes.fr](mailto:jerome.idier@irccyn.ec-nantes.fr))

**Abstract.** We consider the problem of estimating the peak parameters in a spectroscopic signal, *i.e.* their locations, amplitudes and form parameters. A marked point process provides a suitable representation for this phenomenon: it consists in modeling the spectrum as a noisy sum of points lying in the observation space and characterized by their locations and some marks (amplitude and widths). A non-supervised Bayesian approach coupled with MCMC methods is retained to solve the problem. But the peak number is unknown. Rather than using a method for model uncertainty (such as RJMCMC) we propose an approach in which the dimension model is constant: consequently, the Gibbs sampler appears possible and natural. The idea consists in considering an upper bound for peak number and modelling the peak occurrence by a Bernoulli distribution. At last, a label switching method adapted to the approach is also proposed. The method is illustrated by an application on a real spectrum.

**Keywords:** sparse spike train restoration, spectroscopy, signal decomposition into elementary patterns.

**PACS:** 02.30.Zz: Inverse problems; 02.50.-r: Probability theory, stochastic processes, and statistics.

## 1. INTRODUCTION

We consider the problem of estimating the peak parameters in a spectroscopic signal (a peak spectrum), *i.e.* their locations, amplitudes and widths. The goal is to analyse a chemical spectra to provide an interpretation for physico-chemists. The proposed approach may be used for many kinds of spectra (Raman, infrared, fluorescence, etc.), needing only a parametric form to model the peaks.

This is typically an ill-posed problem, justifying the use of a Bayesian approach, as retained in previous works. In [1], the optimisation is done either with a modified version of Newton-Raphson's algorithm or an accept-reject like algorithm; yet, these methods may yield to a local minimum and the estimation of peak number is not optimal. MCMC (Monte Carlo Markov chain) methods are now preferred [2, 3, 4]. In these works, an RJMCMC (reversible jump MCMC) algorithm [5, 6] is used to estimate the variables.

In this paper, the proposed method is also set in a Bayesian approach coupled with an MCMC method. The main originality results from the use of a constant dimension model avoiding to use the RJMCMC algorithm. The paper is organised as follows. Section 2

defines the marked point process used to model the signal. In section 3, we propose the model to keep a constant variable number, which allows to use the Gibbs sampler. Then we present prior distributions for each variable and compute the posterior distributions. At last, simulation methods are proposed. Because of the label-switching problem, the estimation computation may not be straightforward. So, a new method to deal with this problem is presented in section 4. In section 5, the proposed method is supplied on a real Raman spectrum and, finally, section 6 concludes the paper.

## 2. PROBLEM FORMULATION

A first approach would be to consider the chemical spectrum as the convolution of a sparse spike train and a pattern representing the peak form. A marked point process (MPP) provides a suitable representation for the considered problem: it consists in a finite set of objects lying in a bounded space and characterized by their locations and some marks. So, the Bernoulli-Gaussian (BG) process [7, 8] is a widespread model for the sparse spike train; it is a MPP with only one mark corresponding to the amplitudes. Considering that the pattern is unknown, the problem becomes a blind deconvolution problem.

However, the common implementation of this model with MCMC methods is not efficient (see [8]) since it implies to estimate a great number of variables (each sample of the signal), though the majority is zero! Furthermore, the model suffers from two drawbacks: the peaks are inevitably located on discrete positions in the grid and they have the same shape, which is not always true in real chemical spectra.

To bypass these problems, we propose to model the spectrum  $\mathbf{y}$  as a noisy sum of  $K$  positive peaks  $\mathbf{f}$ :

$$\mathbf{y} = \sum_{k=1}^K \mathbf{f}(\mathbf{n}_k, \mathbf{w}_k, \mathbf{s}_k) + \mathbf{e} \quad (1)$$

where  $\mathbf{n}_k$ ,  $\mathbf{w}_k$  and  $\mathbf{s}_k$  ( $k \in \{1, \dots, K\}$ ) stand respectively for the location, amplitude and width of the  $k$ th peak,  $\mathbf{f}$  is a vector function of length  $N$ , and  $\mathbf{e}$  is a  $N \times 1$  vector corresponding to the noise and model errors. We consider a Lorentzian shape for the peaks which is usual in Raman spectroscopy and allows to simplify the presentation in this paper since it has only one form parameter. Then, the  $k$ th peak amplitude at wavenumber  $n$  is:

$$\mathbf{f}_n(\mathbf{n}_k, \mathbf{w}_k, \mathbf{s}_k) = \mathbf{w}_k \frac{\mathbf{s}_k^2}{\mathbf{s}_k^2 + (n - \mathbf{n}_k)^2}. \quad (2)$$

Of course, the method can be adapted to other shapes, like Gaussian or Voigt functions. There is a substantial gain when modeling peaks as a known parametric expression rather than a non-parametric shape since it introduces available information, restricting the solution space consequently. This new MPP considers the signal as a finite set of Lorentzian whose marks are the amplitudes and widths.

The considered problem has similarities with the problem of mixture analysis [6, 9], as for example the common choice of MCMC methods for optimization or the label switching problem. Nevertheless, in the considered problem, we estimate the peaks in the data themselves and not in the data distribution.

### 3. PROPOSED MODEL

#### 3.1. A Constant Dimension Model

Thanks to the likelihood factorization, the Gibbs sampler appears easy and natural to use. But the peak number is unknown, as a result of which the system order is likely to change and the variable number too. In this case, classical MCMC methods (such as Metropolis-Hastings algorithm or Gibbs sampler) can not be applied because the posterior is not stationary. For ten years, new MCMC techniques for model uncertainty have been proposed [10, 11]. The most famous is the RJMCMC algorithm [5, 6]. However, we propose another model in which the system order is constant, allowing to use Gibbs sampler.

The idea is to consider the peak number constant and equal to  $K_{\max}$  corresponding to an upper bound greater than the real peak number and fixed by the user. To avoid to obtain an estimated signal with  $K_{\max}$  peaks, we are inspired by the Bernoulli-Gaussian model and introduce the vector  $\mathbf{q} \in \{0, 1\}^{K_{\max}}$  coding the peak occurrences. Thus, for all  $k \in \{1, \dots, K_{\max}\}$ :

- if  $\mathbf{q}_k = 1$  then the  $k$ th peak is present and located at  $\mathbf{n}_k$  with amplitude  $\mathbf{w}_k$  and width  $\mathbf{s}_k$ ;
- on the contrary, if  $\mathbf{q}_k = 0$ , the  $k$ th peak is not present: it does not appear in the signal. Its amplitude  $\mathbf{w}_k$  is set to zero but not its location and width: this choice is motivated by the simulation method (see section 3.4).

Being still inspired by the BG model, the peak occurrences  $\mathbf{q}_k$  are distributed according to a Bernoulli distribution with parameter  $\lambda$ . Equation (1) reads then:

$$\mathbf{y} = \sum_{k=1}^{K_{\max}} \mathbf{f}(\mathbf{n}_k, \mathbf{w}_k, \mathbf{s}_k) + \mathbf{e}, \quad (3)$$

where, of course,  $\mathbf{w}_k = 0$  if and only if  $\mathbf{q}_k = 0$ . In matrix form, we then have:

$$\mathbf{y} = \mathbf{G}\mathbf{w} + \mathbf{e} \quad (4)$$

where  $\mathbf{G}$  is the  $N \times K_{\max}$  matrix :

$$\mathbf{G} = ( \mathbf{f}(\mathbf{n}_1, 1, \mathbf{s}_1) \quad \dots \quad \mathbf{f}(\mathbf{n}_{K_{\max}}, 1, \mathbf{s}_{K_{\max}}) ) \quad (5)$$

The variable number is then smaller than a common BG implementation (as in [8]) in which one has to calculate  $N$  variables to estimate the sparse spike train, though the present model needs only  $3K_{\max}$  variables ( $\mathbf{n}$ ,  $\mathbf{q}$ , and  $\mathbf{w}$ ). Indeed, one can reasonably suppose that  $3K_{\max} < N$ , that is there is less than  $N/3$  peaks in the signal. Consequently, the method will be faster (considering an equivalent time for each variable simulation in the Gibbs sampler) and the estimation quality will be better (there is less unknowns for the same data number).

### 3.2. Prior distributions

*Noise.* We choose the classical model of a white, Gaussian and i.i.d. noise with variance  $r_e$ :

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, r_e \mathbf{I}). \quad (6)$$

*Peak Location.* A priori, we do not have any information about the peak location. For this reason, we suppose that the peaks are uniformly distributed on  $[1, N]$ :

$$\forall k \in \{1, \dots, K_{\max}\}, \quad \mathbf{n}_k \sim \mathcal{U}_{[1, N]}. \quad (7)$$

*Peak Amplitude.* As we said in the previous section,  $(\mathbf{q}, \mathbf{w})$  is modelled as a BG process; in addition, the amplitudes are positive. Therefore, we have  $\forall k \in \{1, \dots, K_{\max}\}$ :

$$\mathbf{q}_k \sim \mathcal{Ber}(\lambda), \quad (8)$$

$$\mathbf{w}_k | \mathbf{q}_k \sim \begin{cases} \delta_0(\mathbf{w}_k) & \text{if } \mathbf{q}_k = 0, \\ \mathcal{N}^+(0, r_w) & \text{if } \mathbf{q}_k = 1. \end{cases} \quad (9)$$

where  $\mathcal{Ber}$  stands for a Bernoulli distribution,  $\delta_0$  denotes a Dirac centered in zero and  $\mathcal{N}^+$  stands for a Gaussian distribution with positive support. Of course, another prior for amplitudes could be used. For example, a Gamma distribution seems efficient, but the advantages of a normal distribution with positive support are twofold: first, it yields a normal posterior distribution which is easily simulable; second, it can be easily adapted to other applications where the peaks can have positive and negative peaks.

*Peak Width.* In Raman spectroscopy, it is known that peaks have a FWHH (full width at half height, equals to twice the peak width) of  $12 \pm 5 \text{ cm}^{-1}$ . Therefore, an inverse gamma distribution is proposed whose mean and variance are equal to  $6 \text{ cm}^{-1}$  and  $2.5 \text{ cm}^{-1}$  respectively:

$$\mathbf{s}_k \sim I\mathcal{G}(\alpha_s, \beta_s) \quad (10)$$

where

$$\alpha_s = 16.4, \quad \beta_s = 82.$$

*Bernoulli parameter.* To avoid a degenerative behaviour when the peak number is too low, a conjugate prior (beta distribution) which penalizes high values is chosen:

$$\lambda \sim \mathcal{Be}(1, K_{\max} + 1). \quad (11)$$

*Peak Amplitude Variance.* Without any prior on  $r_w$ , its posterior distribution may not be defined when  $K \leq 2$ . To avoid this, we choose the following conjugate prior:

$$r_w \sim I\mathcal{G}(\alpha_w, \beta_w). \quad (12)$$

This prior is chosen to avoid numerical errors, but not to influence the estimation result; so, it has to be the less informative as possible. In the sequel, we suppose that the signal is multiplied by a constant in order to fix roughly the peak amplitude peak to a value close to 1. Then, we propose the following values for  $\alpha_w$  and  $\beta_w$ :

$$\alpha_w = 2 + \varepsilon, \quad \beta_w = 1 + \varepsilon, \quad \text{and} \quad \varepsilon \ll 1$$

*Noise variance.* Because of the lack of information about  $r_e$ , we opt for the traditional Jeffreys prior:

$$r_e \sim 1/r_e. \quad (13)$$

### 3.3. Conditional Posterior distributions

Then, the conditional posterior distributions are computed:

$$\mathbf{n}_k | \dots \sim \exp\left(-\left\|\mathbf{y} - \sum_{l=1}^{K_{\max}} \mathbf{f}(\mathbf{n}_l, \mathbf{w}_l, \mathbf{s}_l)\right\|^2 / 2r_e\right) \mathbb{1}_{[1, N]}(\mathbf{n}_k), \quad (14)$$

$$\mathbf{q}_k | \dots \sim \text{Ber}(\lambda_k), \quad (15)$$

$$\mathbf{w}_k | \dots \sim \begin{cases} \delta_0(\mathbf{w}_k) & \text{if } \mathbf{q}_k = 0, \\ \mathcal{N}^+(\mu_k, \rho_k) & \text{if } \mathbf{q}_k = 1, \end{cases} \quad (16)$$

$$\mathbf{s}_k | \dots \sim \exp\left(-\frac{1}{2r_e} \left\|\mathbf{y} - \sum_{l=1}^{K_{\max}} \mathbf{f}(\mathbf{n}_l, \mathbf{w}_l, \mathbf{s}_l)\right\|^2 - \beta_s / \mathbf{s}_k\right) \frac{1}{\mathbf{s}_k^{\alpha_s+1}} \mathbb{1}_{\mathbb{R}^+}(\mathbf{s}_k), \quad (17)$$

$$\lambda | \dots \sim \text{Be}(K+1, 2K_{\max} - K + 1), \quad (18)$$

$$r_w | \dots \sim \text{IG}(K/2 + \alpha_w, \mathbf{w}^T \mathbf{w} / 2 + \beta_w), \quad (19)$$

$$r_e | \dots \sim \text{IG}\left(N/2, \left\|\mathbf{y} - \sum_{l=1}^{K_{\max}} \mathbf{f}(\mathbf{n}_l, \mathbf{w}_l, \mathbf{s}_l)\right\|^2 / 2\right), \quad (20)$$

where  $K$  denotes the number of present peaks (*i.e.*  $K = \sum_{k=1}^K \mathbf{q}_k$ ) and:

$$\lambda_k = \left[1 + \frac{1-\lambda}{\lambda} \sqrt{\frac{r_w}{\rho_k}} \exp\left(-\frac{\mu_k^2}{2\rho_k}\right)\right]^{-1}, \quad \mu_k = \frac{\rho_k}{r_e} \mathbf{z}_{-k}^T \mathbf{f}(\mathbf{n}_k, 1, \mathbf{s}_k),$$

$$\rho_k = \frac{r_w r_e}{r_e + r_w \mathbf{f}(\mathbf{n}_k, 1, \mathbf{s}_k)^T \mathbf{f}(\mathbf{n}_k, 1, \mathbf{s}_k)}, \quad \mathbf{z}_{-k} = \mathbf{y} - \sum_{l=1, l \neq k}^{K_{\max}} \mathbf{f}(\mathbf{n}_l, \mathbf{w}_l, \mathbf{s}_l).$$

### 3.4. Variable Simulation

The simulation of  $\mathbf{n}_k$  is provided using a Metropolis-Hastings algorithm. To improve the convergence, we separate the two following cases:

- if  $\mathbf{q}_k = 1$ , the proposal distribution is a Gaussian (with bounded support since  $\mathbf{n}_k \in [1, N]$ ) with mean  $\mathbf{n}_k^{(i-1)}$  and variance  $r_n$  chosen by the user, so that the algorithm performs a random-walk algorithm. This is motivated by the fact that, as the peak exists, we aim at defining precisely its location: small perturbations around the current value allow to refine it;
- otherwise, if  $\mathbf{q}_k = 0$ , we do not dispose of any information to determine the location of a peak, then the proposal distribution is uniform:  $\mathcal{U}_{[1, N]}$ . The interest is to update the absent peak location and explore the entire space. This explain why an absent peak location is not zero.

So, the proposal distribution for the algorithm is:

$$q(\tilde{\mathbf{n}}_k) = \delta_0(\mathbf{q}_k) \mathcal{U}_{[1,N]} + \delta_1(\mathbf{q}_k) \mathcal{N}^{[1,N]}(\mathbf{n}_k^{(i-1)}, r_{\mathbf{n}}).$$

$\mathbf{q}_k$  is distributed according to a Bernoulli distribution while  $\mathbf{w}_k$  is distributed according to either a Dirac ( $\mathbf{q}_k = 0$ ) or a positive Gaussian distribution ( $\mathbf{q}_k = 1$ ). In the last case, several methods for positive normal variables simulation have been proposed [12, 13, 14]<sup>1</sup>. Again, a random-walk Metropolis Hastings algorithm is proposed to sample  $\mathbf{s}_k$ . The proposal distribution is a positive Gaussian whose variance is chosen by the user. At last, classical methods allow to simulate the hyperparameter posteriors (see *e.g.* [12]).

## 4. LABEL SWITCHING

The expected a posteriori estimator is natural to implement from a Markov chain, but its calculation is sometimes less straightforward that might be expected: this is due to the so-called ‘‘label-switching problem’’ [9, 15]. It is a common but difficult problem due to two phenomena: on the one hand, if there is not enough information to distinguish the variables, then the posterior distribution is the same for each permutation of the variable index  $k$ . Thus, it is impossible to distinguish two peaks, and then to avoid them to permute. On the other hand, the MCMC method does not avoid this, since it is able to explore the  $k!$  permutation possibilities (to some extent, the label-switching is evidence of good mixing).

It is now accepted that imposing an *identifiability constraint* [6] on parameters (for example:  $\mathbf{n}_1 < \dots < \mathbf{n}_{K_{\max}}$ ) can be inefficient to suppress the symmetry of the distribution. *Relabelling algorithms* [9, 16] consist in an iterative algorithm which choose, at each iteration, the permutation which minimizes a loss function. At last, Celeux *et al.* [17] propose to minimize a *label invariant cost function*, using simulated annealing.

As in [9, 16], we consider a relabelling algorithm by minimizing the following cost function:

$$\mathcal{L}_0(\mathbf{n}, \mathbf{w}, \mathbf{s}, \mu_{\mathbf{n}}, \rho_{\mathbf{n}}, \mu_{\mathbf{w}}, \rho_{\mathbf{w}}, \mu_{\mathbf{s}}, \rho_{\mathbf{s}}) = - \ln \left[ \prod_{k=1}^{K_{\max}} \mathcal{N}(\mathbf{n}_k | \mu_{\mathbf{n}k}, \rho_{\mathbf{n}k}) \mathcal{N}(\mathbf{w}_k | \mu_{\mathbf{w}k}, \rho_{\mathbf{w}k}) \mathcal{N}(\mathbf{s}_k | \mu_{\mathbf{s}k}, \rho_{\mathbf{s}k}) \right]. \quad (21)$$

The goal is to choose the hyperparameters  $(\mu_{\mathbf{n}k}, \mu_{\mathbf{w}k}, \mu_{\mathbf{s}k})$  and  $(\rho_{\mathbf{n}k}, \rho_{\mathbf{w}k}, \rho_{\mathbf{s}k})$  to minimise the function. At initialisation, the peak number is estimated in the sense of the MMAP (marginalized maximum a posteriori). The following is an iterative procedure on  $k = 1, \dots, \hat{K}^{\text{MMAP}}$ . We propose an alternative to general relabelling algorithms; the differences are threefold:

- first, a previous estimation is obtained by selecting the maxima in the histogram of  $(\mu_{\mathbf{n}}, \mu_{\mathbf{w}}, \mu_{\mathbf{s}})$  (each dimension representing an interest parameter; *e.g.*, in our case,

---

<sup>1</sup> The algorithm proposed in [14] is available for free at [http://www.iris.cran.uhp-nancy.fr/francais/si/Personnes/Personne\\_Mazet/rpnorm-en.htm](http://www.iris.cran.uhp-nancy.fr/francais/si/Personnes/Personne_Mazet/rpnorm-en.htm).

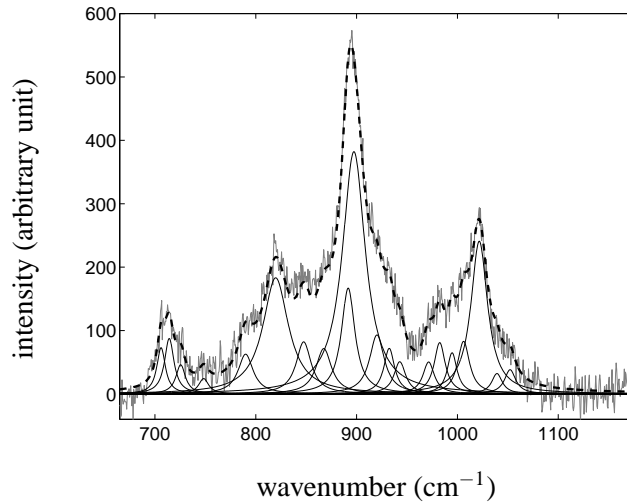
we have a 3D histogram). This first estimation provides an initialisation closer to the global optimum than a simple identity permutation;

- second, we do not try to choose the permutation to minimize  $\mathcal{L}_0$ . Instead, we prefer to relabel the sampled values  $(\mathbf{n}_l, \mathbf{w}_l, \mathbf{s}_l)$  at each iteration one after the other, by seeking, for each iteration, which one is the closest to the histogram maximum. This procedure comes down to minimize the cost-function  $\mathcal{L}_0$  too. Then, parameters  $(\mu_{\mathbf{n}_k}, \mu_{\mathbf{w}_k}, \mu_{\mathbf{s}_k})$  and  $(\rho_{\mathbf{n}_k}, \rho_{\mathbf{w}_k}, \rho_{\mathbf{s}_k})$  are updated by minimizing equation (21);
- at last, the proposed approach take into account the fact that the peak number is expected to change since the search is made only on peaks such that  $\mathbf{q}_l^{(i)} = 1$ .

Then the estimated peak has parameters  $(\mu_{\mathbf{n}_k}, \mu_{\mathbf{w}_k}, \mu_{\mathbf{s}_k})$ . The algorithm start again till  $k = \widehat{K}^{\text{MMAP}}$ , after fixing to zero the occurrences of the selected peaks, avoiding to select them again. There is not yet a mathematical background for this method, but all the experiences show that the estimation is satisfactory.

## 5. APPLICATION

In this section, we present an application of the proposed method on a part of a Raman spectrum of gibbsite  $\text{Al}(\text{OH})_3$ . We have performed 10,000 iterations with a burn-in period of 5,000 iterations, an initial spectrum with no peak, and the following initial values:  $\lambda = 0.5$ ,  $r_w = 10$ ,  $r_e = 0.1$ . The estimation is shown on figure 1. The results were validated by chemists. They are really satisfactory for two reasons: first, the peaks are mostly reproduced at each simulation (only one is presented here), second, the estimated peaks are physically significant. However, some real peaks are sometimes estimated by several ones, *e.g.* around  $710 \text{ cm}^{-1}$  or  $900 \text{ cm}^{-1}$ . In the last case, this must be due to the natural asymmetry of the real peaks.



**FIGURE 1.** Decomposition of a real Raman spectrum and the reconstructed signal.

## 6. CONCLUSION

In conclusion, we have proposed a method of signal decomposition into elementary pattern which appears as an alternative to blind deconvolution. A marked point process reveals outstanding results since it is able to set the peaks on a continuous space and to estimate them with different shapes. This approach performs better than a classical deconvolution approach where the peaks have inevitably the same width and is also faster than usual methods with a BG model. An alternative to RJMCMC is also proposed by considering a constant order model and adding a new variable coding the existence of the peaks. At last, we propose a new method of label switching.

## REFERENCES

1. A. Mohammad-Djafari, "A Bayesian estimation method for detection, localisation and estimation of superposed sources in remote sensing," in *SPIE 97 Annual Meeting*, San Diego, CA, USA, 1997, vol. 3163.
2. R. Fischer, and V. Dose, "Analysis of mixtures in physical spectra," in *ISBA 2000*, Heraklion, Crete, Greece, 2000.
3. S. Gulam Razul, W. Fitzgerald, and C. Andrieu, *Nucl. Instrum. Meth. A* **497**, 492–510 (2003).
4. N. Haan, *Statistical Models and Algorithms for DNA Sequencing*, Ph.D. thesis, University of Cambridge, UK (2001).
5. P. Green, *Biometrika* **82**, 711–732 (1995).
6. S. Richardson, and P. Green, *J. Roy. Stat. Soc. B* **59**, 731–792 (1997).
7. J. Kormylo, and J. Mendel, *IEEE Trans. Inform. Theory* **28**, 482–488 (1982).
8. Q. Cheng, R. Chen, and T.-H. Li, *IEEE Trans. Geosci. Remote* **34**, 377–384 (1996).
9. M. Stephens, *Bayesian methods for mixtures of normal distribution*, Ph.D. thesis, Magdalen College, University of Oxford (1997).
10. B. Carlin, and S. Chib, *J. Roy. Stat. Soc. B* **57**, 473–484 (1995).
11. M. Stephens, *Annals of Statistics* **28**, 40–74 (2000).
12. L. Devroye, *Non-uniform random variate generation*, Springer-Verlag, New York, 1986, ISBN 0-387-96305-7, URL <http://jeff.cs.mcgill.ca/~luc/rnbookindex.html>.
13. C. Robert, *Stat. Comput.* **5**, 121–125 (1995).
14. V. Mazet, D. Brie, and J. Idier, "Simulation of Positive Normal Variables using several Proposal Distributions," in *IEEE Workshop Statistical Signal Processing*, #190, Bordeaux, France, 2005.
15. C. Holmes, A. Jasra, and D. Stephens, *Statistical Science* **20**, 50–67 (2005).
16. G. Celeux, "Bayesian inference for mixtures: the label switching problem," in *COMPSTAT 98*, edited by R. Payne, and P. Green, Physica-Verlag, 1998, pp. 227–232.
17. G. Celeux, M. Hurn, and C. Robert, *J. Am. Stat. Assoc.* **95**, 957–970 (2000).