Signal Processing 109 (2015) 193–205

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Unsupervised joint decomposition of a spectroscopic signal sequence ☆

Vincent Mazet [a,*,1], Sylvain Faisan [a,1], Slim Awali [b,c], Marc-André Gaveau [b], Lionel Poisson [b]

[a] ICube, University of Strasbourg, CNRS, 300 boulevard Sébastien Brant, BP 10413, 67412 Illkirch, France
[b] Laboratoire Francis Perrin, CEA, CNRS, IRAMIS/Service des Photons Atomes et Molécules, 91191 Gif-sur-Yvette, France
[c] Laboratoire d'Étude des Milieux Ionisés et Réactifs, Institut Préparatoire aux Études d'Ingénieurs de Monastir, Tunisia

ARTICLE INFO

Article history:
Received 16 March 2014
Received in revised form
14 October 2014
Accepted 26 October 2014
Available online 4 November 2014

Keywords:
Signal decomposition
Sequence of spectroscopic signals
Hierarchical Bayesian model
Markovian prior
RJMCMC

ABSTRACT

This paper addresses the problem of decomposing a sequence of spectroscopic signals. Data are a series of signals modeled as a noisy sum of parametric peaks. We aim to estimate the peak parameters given that they change slowly between two contiguous signals. The key idea is to decompose the whole sequence rather than each signal independently. The problem is set within a Bayesian framework. The peaks with similar evolution are gathered into groups and a Markovian prior on the peak parameters of a same group is used to favor a smooth evolution of the peaks. In addition, the peak number and the group number are unknown and have to be estimated (the number of peaks in two contiguous signals change if peaks vanish). Therefore, the posterior distribution is sampled with a reversible jump Markov chain Monte Carlo algorithm. Simulations conducted on synthetic and real photoelectron data illustrate the performance of the method.

1. Introduction

1.1. Problem statement

A spectroscopic signal (we use in the sequel the shorter term "spectrum") represents the repartition of particles or electromagnetic waves with respect to their energy, wavelength, etc. It typically gathers several peaks whose positions and areas give information about the chemical composition of the analyzed sample. Some experiments now involve several spectra of the same sample acquired at different instants [1], so data are a temporal sequence of different spectra. The following assumptions are made on the data:

- each spectrum can be modeled as a noisy sum of parametric peaks. This is a widespread assumption for spectroscopic signals in which the peaks correspond to emission lines;
- the peaks exhibit a slow evolution with time, that is, their parameters do not vary too much between two contiguous spectra (*i.e.* a spectrum and the one acquired at the next or previous instant). This implies that two acquisitions have to be sufficiently close in time so as to correspond to almost the same reactions. If this assumption is not verified, then the sampling in time is too coarse, so the experiment cannot be well observed and analyzed;
- the number of peaks in each spectrum may vary in time: in the considered application (photoelectron spectroscopy), the peaks may appear and/or disappear.

☆ This work was initiated by the PEPS SpectroDec funded by the CNRS.
* Corresponding author.
E-mail addresses: vincent.mazet@unistra.fr (V. Mazet), faisan@unistra.fr (S. Faisan), slim.awali@cea.fr (S. Awali), marc-andre.gaveau@cea.fr (M.-A. Gaveau), lionel.poisson@cea.fr (L. Poisson).
[1] The two first authors contributed equally to this work.

http://dx.doi.org/10.1016/j.sigpro.2014.10.032
0165-1684/

Fig. 4 shows a simulated sequence with the three aforementioned properties. The aim of this work is to decompose each spectrum of the sequence — that is to estimate the peak number and their parameters : centers, amplitudes and widths (it amounts to the positions and areas) — and to follow the peaks through the sequence so as to estimate their evolution.

1.2. State of the art

As far as we know, nobody has addressed the problem, as it was previously defined, of the decomposition of a sequence of spectra, but many works provide solutions to similar problems.

Approaches related to source separation (convolutive or not) [2] or spectral unmixing [3] cannot be directly used because the peaks exhibit various locations and shapes through the sequence.

In spectroscopy, Gobinet et al. [4] propose several preprocessing algorithms (as, for example, peak alignment and width homogenization) so that usual methods of source separation can be used. However, the pre-processing algorithms require no overlap between peaks and strong knowledge of the peak locations.

Alternatives like dynamic time warping (DTW) or correlation optimized warping (COW) [5] do not require priors, but they cannot consider a varying peak number.

In time–frequency, some works decompose audio signals in time–frequency atoms using sparse approximation algorithms (*e.g.* Matching Pursuit [6] and Basis Pursuit [7]) or stochastic methods (reversible jump Monte Carlo Markov chain algorithm [8]). The spectrum sequence can be seen as a time–frequency signal whose atoms are the peaks. However, there is a relationship between the peaks (since they evolve with time) which cannot be managed by the aforementioned approaches.

Similar problems arise in the context of wavefront tracking [9,10], where seismic signals are acquired at different positions. In [9], a hidden Markov model is used to model delay profiles that are observed in these signals. Such a modeling cannot be used in our problem since peaks of a same spectrum are supposed to be independent. In [10], a Markov random field models the continuity between the decomposition of two neighbor signals, as it is expected in our problem. This requires the spikes (equivalent to the peak centers in spectrum decomposition) to lie on a discrete grid. We also use a Markovian prior, but it directly models the unknowns so that peak centers are not imposed to lie on a grid. Besides, the convolution model used in [9,10] cannot easily take into account the evolution of the peak widths.

Decomposing a sequence of spectroscopic signals could also be seen as a multi-target tracking problem since the peaks in spectra could be considered as targets. In most cases, multi-target tracking algorithms need point measurements which are obtained by thresholding the sensor output [11,12] or by using feature descriptors [13]. In our context, the detection step could consist in decomposing each spectrum separately using standard approaches, thus providing the data used for tracking. However, such methods are not adapted to the decomposition of sequence of spectra. Indeed, this leads to a sequential approach, in which the spectra are decomposed independently from the others, which is known to be unsuitable [10,14]: the decomposition of two contiguous spectra may lead to two very different decompositions even if the spectra are similar. On the other hand, the so-called track-before-detect algorithms consider the whole sensor output as a measurement [15,16]. Among them, the observation model of the Histogram Probabilistic Multi-Hypothesis Tracker (H-PMHT) [17,18] fits our data (data are the superposition of target and noise components). Moreover, a track management module can be added to handle a time varying number of targets (see *e.g.* [19] in the context of tracking groups of people in video). However, the track management module depends on the application and should be completely revisited in our context.

The works mentioned above are not directly adaptable to our problem. However, the decomposition of a *unique* spectrum has been intensively studied in the past. The first works propose simple approaches, such as using least squares [20] or gradient methods [21,22]; yet these methods are ineffective in the case of noisy data with overlapping peaks. Recently, sparse approximation methods [23,24] and Bayesian approaches using MCMC (Monte Carlo Markov Chain) algorithms [25–28] have been proposed. The latter framework yields very good results, thus we used it in [1] and also in this paper. However, Refs. [25–28] decompose a unique spectrum while we consider the decomposition of a sequence of spectra. In [1], we have also considered a sequence of spectra, but the spectra were processed independently and we did not aim at grouping the peaks to follow their evolution. Therefore, the problem in [1] came down to the decomposition of a unique spectrum. In this paper, the spectra are processed jointly, thus allowing to follow their evolution through the sequence. The smooth evolution of peaks is made possible with the introduction of groups: a group gathers the different apparition of a peak through the sequence. A formal definition is given in Section 2.1. The gathering of peaks into groups is one of the contribution of the paper. Because the peak number and the group number are unknown, inference is performed using a trans-dimensional MCMC algorithm.

1.3. Outline of the proposed approach

As said before, a sequential approach may lead to two very different decompositions of two similar contiguous spectra, which is physically impossible because the peaks evolve smoothly. Moreover, it does not allow to follow a peak through the sequence, so a post-processing is needed.

On the contrary, a joint decomposition approach, where spectra are decomposed simultaneously, may favor a smooth change of the peak parameters by regularizing their evolution, thus providing coherent and consistent results. In addition, a joint approach is also able to classify the peaks, thus giving the possibility to follow them through the sequence. The Bayesian methodology provides a rich framework for modeling inverse problems and enables us to define a joint decomposition approach.

We recently proposed in [14] a joint decomposition approach. However, the peak number is supposed to be known and constant through the sequence, which is obviously wrong on real data. In addition, real data contain more spectra that the synthetic sequence used in [14]: this leads to a significant increase in the size of the search space and the algorithm fails. So we extend the model of [14] by considering the peak number as a new random variable (Section 2), and we use the reversible jump MCMC (RJMCMC) algorithm [29] which handles problems with unknown dimension (Section 3). The simulated annealing scheme used in [14] appears now unnecessary. Also, the method proposed in [14] is supervised and requires to set several hyperparameters. The variability that may exist between different data sets makes this task difficult, that is why the method proposed in this paper is unsupervised. We also introduce some tricks (such as the use of mixed proposals or an overestimation of the hyperparameters in the burn-in time) to improve the convergence of the algorithm. In Section 4 the proposed approach is compared with the one in [14] and its performance in the presence of noise is discussed. Finally, we present some results on real photoelectron data in Section 5.

2. Bayesian model

Bold variables correspond to either matrices or vectors, and other variables to scalar.

2.1. Signal modeling

Data are a sequence of S spectra with N samples each. A spectrum $\boldsymbol{y}_s (s \in \{1,\dots,S\})$ is modeled as the sum of peaks $\boldsymbol{x}_s$ and an additive noise $\boldsymbol{v}_s$ (physical justifications of this model can be found in [1,26]) :

$$\boldsymbol{y}_s = \boldsymbol{x}_s + \boldsymbol{v}_s. \tag{1}$$

The peaks are gathered into groups so as to follow and regularize their evolution: a group is the association of $l \leq S$ similar peaks through the sequence; it is composed of one peak per spectrum, with the constraint that all peaks of a group lie on contiguous spectra. The group number is denoted K; each group $k \in \{1,\dots,K\}$ begins in spectrum b_k and has a length l_k (defined as the number of peaks in the group). We note $M = \sum_{k=1}^{K} l_k$ the total peak number.

In the sequel, we consider that peaks are Gaussian; however, any other parametric function (such as a Lorentzian or Voigt functions [1,26,30]) can be used to model the peaks. Thus, the sum of peaks $\boldsymbol{x}_s$ in (1) reads:

$$x_{s,n} = \sum_{k=1}^{K} \sum_{m=1}^{l_k} a_{k,m} \exp\left(-\frac{(n-c_{k,m})^2}{2w_{k,m}^2}\right) \delta_{b_k+m-1,s} \tag{2}$$

where the subscript n denotes the element $n \in \{1,\dots,N\}$ in the associated vector. The Kronecker delta $\delta_{b_k+m-1,s}$ equals 1 if $b_k+m-1=s$, 0 otherwise; it codes the presence or the absence of the mth peak of group k in spectrum s. The mth peak of group k is parametrized by its center $c_{k,m}$, amplitude $a_{k,m}$ and width $w_{k,m}$.

The model is set in a Bayesian framework: the priors are presented below and the directed acyclic graph is shown in Fig. 1.

2.2. Noise prior

The noise is supposed to be white, zero-mean Gaussian with a constant variance $r_{\boldsymbol{v}}$:

$$\forall s,n, \quad v_{s,n} | r_{\boldsymbol{v}} \sim \mathcal{N}(0, r_{\boldsymbol{v}}). \tag{3}$$

This assumption is reasonable since the number of photons collected onto the sensor is sufficiently high, which is often the case in usual spectroscopic applications.

2.3. Prior for K, $\boldsymbol{l}$ and $\boldsymbol{b}$

Since the data has to be described with the smallest number of peaks and groups, the prior $p(K,\boldsymbol{l},\boldsymbol{b})$ has to be chosen to favor solutions with small values of K and $M = \sum_{k=1}^{K} l_k$. A straightforward way is to define a joint prior using M and K:

$$p(K,\boldsymbol{l},\boldsymbol{b}) \propto \xi^{M+SK} \mathbb{1}_{\mathcal{X}}(K,\boldsymbol{l},\boldsymbol{b}) \tag{4}$$

where $\xi \in]0,1[$ so as to favor solutions with a small number of peaks and groups (the parameter ξ is no estimated, it allows the user to give a prior about the peak and group number), $\mathbb{1}$ is the indicator function, and $\mathcal{X}$ the set of allowed values for the triplet $(K,\boldsymbol{l},\boldsymbol{b})$:

- the group number K is bounded by a maximal value $K_{\max}$ so that $p(K,\boldsymbol{l},\boldsymbol{b})$ is integrable (because the number of configurations is bounded);
- the elements of $\boldsymbol{l}$ lies in $\{1,\dots,S\}$ (thus $M \in \{K,\dots,SK\}$);
- each beginning b_k takes values in $\{1,\dots,S-l_k+1\}$.

The proposed joint prior can be easily interpreted: it means that models with the same value for $M+SK$ have the same probability. Moreover it is sufficiently versatile to add another available information. On the opposite, a hierarchical model defines separately priors for K, l_k and b_k ($\forall k$), it makes the interpretation of the model in terms of M and K difficult (note that the probability of b_k depends on l_k and S). We recall that our goal is to fit the data with a low number of peaks and groups. So, it is natural to

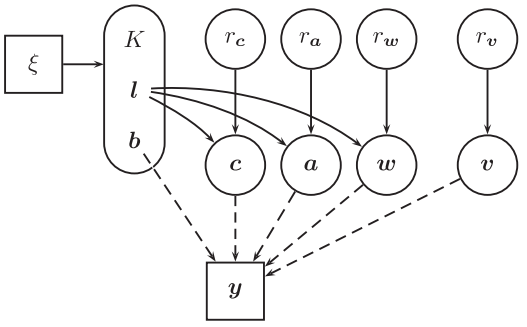

Fig. 1. Directed acyclic graph of the proposed model. Squares represent known or fixed variables, circles represent unknown random variables. Continuous line arrows indicate dependencies between variables while dashed arrows indicate deterministic dependencies.

express the prior in terms of the number of peaks and groups. Note finally that the constraints on K and M define the set of constant dimension spaces; they are schematized in Fig. 2.

2.4. Prior for the peak parameters

The peak centers $\boldsymbol{c}_k=(c_{k,1},\dots,c_{k,l_k})$ of group k are supposed to evolve slowly through the sequence. Therefore, the peak centers in group k are modeled as a one-dimensional Gaussian Markov random field, and the center of the first peak is distributed according to a uniform distribution on $\mathcal{C}=[1,N]$. This yields, for all $k\in\{1,\dots,K\}$:

$$p(\boldsymbol{c}_k|r_{\boldsymbol{c}},l_k)=p(c_{k,1})\times\prod_{m=2}^{l_k}p(c_{k,m}|c_{k,m-1},r_{\boldsymbol{c}},l_k)\approx\frac{1}{N-1}\frac{1}{(2\pi r_{\boldsymbol{c}})^{(l_k-1)/2}}\exp\left(-\frac{\|\boldsymbol{D}\boldsymbol{c}_k\|^2}{2r_{\boldsymbol{c}}}\right)\mathbb{1}_{\mathcal{C}^{l_k}}(\boldsymbol{c}_k)\quad(5)$$

where $\|\cdot\|^2$ is the ℓ_2 norm. The $(l_k-1)\times l_k$ matrix $\boldsymbol{D}$ is a first-order discrete derivative which promotes a straight evolution of the peak centers:

$$\boldsymbol{D}=\begin{pmatrix}1&-1&&0\\&\ddots&\ddots&\\0&&1&-1\end{pmatrix}\quad(6)$$

Higher order derivative can be used to promote other kinds of evolution of the peak centers. Whereas $\boldsymbol{D}$ is of rank l_k-1, the prior $p(\boldsymbol{c}_k|r_{\boldsymbol{c}},l_k)$ is proper thanks to the definition of $p(c_{k,1})$. The hyperparameter $r_{\boldsymbol{c}}>0$ allows to set the prior strength. The peaks are constrained to be in $\mathcal{C}$: this is a realistic assumption since the data are carefully acquired so that no peak is close to the boundaries. The hyperparameter $r_{\boldsymbol{c}}$ being the mean expected deviation of the peak centers, it is reasonably lower that N. So, the truncation induced by the constraint occurs far in the tails of the Gaussian $\exp(-\|\boldsymbol{D}\boldsymbol{c}_k\|^2/2r_{\boldsymbol{c}})$ and can be neglected so that $p(\boldsymbol{c}_k|r_{\boldsymbol{c}},l_k)$ is approximated by Eq. (5).

The centers $\boldsymbol{c}_1,\dots,\boldsymbol{c}_K$ are considered mutually and conditionally independent given $r_{\boldsymbol{c}}$ and $\boldsymbol{l}$. Moreover, as $p(\boldsymbol{c}_k|r_{\boldsymbol{c}},\boldsymbol{l})=p(\boldsymbol{c}_k|r_{\boldsymbol{c}},l_k)$, we obtain

$$p(\boldsymbol{c}|r_{\boldsymbol{c}},\boldsymbol{l})=\prod_{k=1}^{K}p(\boldsymbol{c}_k|r_{\boldsymbol{c}},l_k).\quad(7)$$

For the same reasons, the peak amplitudes $\boldsymbol{a}_k$ and widths $\boldsymbol{w}_k(\forall k)$ are distributed according to similar priors:

$$p(\boldsymbol{a}_k|r_{\boldsymbol{a}},l_k)\approx\frac{1}{a_{\max}-a_{\min}}\frac{1}{(2\pi r_{\boldsymbol{a}})^{(l_k-1)/2}}\times\exp\left(-\frac{\|\boldsymbol{D}\boldsymbol{a}_k\|^2}{2r_{\boldsymbol{a}}}\right)\mathbb{1}_{\mathcal{A}^{l_k}}(\boldsymbol{a}_k),\quad(8)$$

$$p(\boldsymbol{w}_k|r_{\boldsymbol{w}},l_k)\approx\frac{1}{w_{\max}-w_{\min}}\frac{1}{(2\pi r_{\boldsymbol{w}})^{(l_k-1)/2}}\times\exp\left(-\frac{\|\boldsymbol{D}\boldsymbol{w}_k\|^2}{2r_{\boldsymbol{w}}}\right)\mathbb{1}_{\mathcal{W}^{l_k}}(\boldsymbol{w}_k),\quad(9)$$

where $\mathcal{A}=[a_{\min},a_{\max}]$ and $\mathcal{W}=[w_{\min},w_{\max}]$ are the allowed values for the parameters, and $r_{\boldsymbol{a}}$ and $r_{\boldsymbol{w}}$ are hyperparameters controlling the prior strength.

This model for the peak parameters can be extended to the case of spectra lying on a multidimensional space. For example, a hyperspectral image can be seen as a set of spectra spread on a bidimensional spatial grid. In this case, the 1-D Markov random field has to be replaced by a 2-D Markov random field in (5), (8) and (9) and the matrix $\boldsymbol{D}$ has to be redefined to take into account the new neighborhoods. Also, we can consider the case of a non-uniform sampling on the grid; one then has to modify the coefficients in matrix $\boldsymbol{D}$.

2.5. Hyperparameter priors

Uniform distributions over $\mathbb{R}^+$ are chosen for the hyperparameters $r_{\boldsymbol{c}},r_{\boldsymbol{a}},r_{\boldsymbol{w}}$ and $r_{\boldsymbol{v}}$. Although these are improper distributions, the posterior will be integrable. Note that an inverse gamma prior is often preferred for variances since it is a conjugate distribution. We tried this prior with very small values for its parameters so as to tend to the Jeffreys prior [31,32] and to favor smooth evolutions of the peak parameters. However, this choice leads sometimes to an aberrant solution composed of a high number of very smooth groups.

2.6. Posterior distribution

The posterior distribution is sampled using the RJMCMC algorithm (see Section 3). In particular, the peak parameters and the hyperparameters are sampled by a Gibbs sampler which consists in sampling each variable according to its conditional posterior distribution [33]: these are detailed in Appendix A.

3. Algorithm

The non-linearity of the model with respect to the parameters leads to a difficult optimization problem admitting a high number of local minima. In addition, the dimension of the solution space is not fixed (the number of unknowns is itself an unknown). This has motivated the use of a MCMC algorithm with model uncertainty to achieve the estimation. In that way, several methods, known as trans-dimensional MCMC methods, have been developed over the last decades, see [29,34–36] and references therein for reviews and discussions. Among these methods, the reversible jump MCMC (RJMCMC) algorithm [29,37] appears to be popular, fast and flexible [38,34]. The next sub-sections describe the proposed moves as well as tricks to accelerate the convergence of the algorithm.

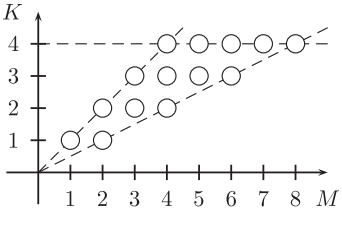

Fig. 2. Admissible values for M and K with $K_{\max}=4$ and $S=2$. Each circle stands for the solution space for the corresponding couple (M,K).

3.1. Moves

The dimension of the solution space actually depends on two variables: M and K. These two variables of dimension are highly correlated since $M = \sum_{k=1}^{K} l_k$ (see Fig. 2). In other words, the elementary objects of our model (the M peaks) are gathered into bigger objects (the K groups). This is a particularity of our problem which, as far as we know, has never been studied in the context of the RJMCMC algorithm. For example, Gulam Razul et al. [26] work with two variables of dimension, namely the peak number and the spline number (the splines model the continuum), but these two variables are not correlated. In a usual RJMCMC algorithm, a couple of dual moves (often called birth and death) is sufficient to change of dimension since it allows to explore the space in two opposite directions. To accelerate convergence, one often adds other moves which are only shortcuts for combinations of the two basic moves. When dealing with two dimension variables, two couples of moves are needed (these moves might be tricky to implement as the modification of one variable may involve to modify the other one). Again, to accelerate convergence, combinations of these basic moves are proposed. We also need a move to sample the posterior at constant dimension, so it yields the eight following moves (see also Fig. 3). Only the first six moves (birth, death, split, merge, increase, and reduce) are trans-dimensional, *i.e.* they perform dimension changes (when accepted).

- Birth and death consist respectively in adding or removing a group composed of only one peak: this is a jump from dimension (K, M) to dimension $(K+1, M+1)$ or $(K-1, M-1)$, respectively.
- A split consists in replacing a group by two contiguous groups (two groups i and j are contiguous if and only if $b_i + l_i = b_j$ or if $b_j + l_j = b_i$) while the peak number remains constant. This corresponds to a jump from (K, M) to $(K+1, M)$. On the opposite, a merge consists in merging two contiguous groups into a single one (jump from (K, M) to $(K-1, M)$).
- An increase consists in adding a peak at an extremity of a group: this corresponds to a jump from (K, M) to $(K, M+1)$. This move is equivalent to a combination of a birth and a merge. On the opposite, a reduce deletes a peak at the extremity of a group: this is a jump from (K, M) to $(K, M-1)$ and a combination of a split and a death.
- In the labeling move, peaks of group k_1 are exchanged with those of k_2 for all spectra whose index is greater than s; k_1, k_2, and s being randomly chosen. This move is equivalent to a combination of merge and split moves, but does not change the dimension since M and K are not altered.
- An update samples the peak parameters $\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{w}$ and the hyperparameters $r_{\boldsymbol{c}}, r_{\boldsymbol{a}}, r_{\boldsymbol{w}}, r_{\boldsymbol{v}}$. This is achieved within a random sweep Gibbs sampling [33] which samples each variable according to its conditional posterior. On the one hand, the peak centers $c_{k,m}$ and widths $w_{k,m}$ are sampled using a random-walk Metropolis-Hastings algorithm [38,33]. On the other hand, the amplitudes $a_{k,m}$ and hyperparameters $r_{\boldsymbol{c}}, r_{\boldsymbol{a}}, r_{\boldsymbol{w}}, r_{\boldsymbol{v}}$ are distributed according to usual distributions and can be sampled using direct methods, such as [39,40] for truncated Gaussian or [41,33] for inverse gamma.

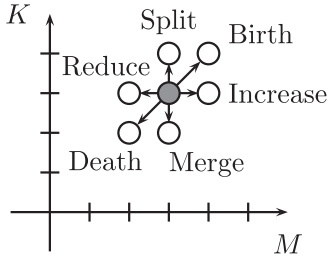

Fig. 3. Illustration of the trans-dimensional moves in space (M, K) (from the gray disk).

The implementation of the trans-dimensional moves is detailed in Appendix B.

3.2. Mixed proposals

The acceptance ratio α should be high enough to obtain a solution in an acceptable computation time. Suggestions have been made in [42–44]. In this work, we use mixed proposal distributions: the candidate values are randomly generated from a uniform distribution or from a distribution computed from the model and/or the data. A model/data-driven proposal is able to propose candidates that may have good chances to be accepted. But a model/data-driven proposal may lead sometimes to unsatisfactory behaviors. As an example, consider the case of a death: the acceptance ratio α for removing the group k requires to estimate the probability to propose the birth of this group k (from the model where it has been removed). If the group k does not explain the signal, then a model/data-driven proposal is very small and, consequently, the death may be rejected although the group does not explain the data. On the opposite, a uniform proposal is able to explore the whole space. The use of a uniform proposal in this context enables to circumvent this limitation since it prevents the proposal to become too small. However, the solution space being so huge, it would be inefficient to generate good candidates.

3.3. Implementation details

We compute an estimation of the maximum a posteriori (MAP). In the context of the RJMCMC algorithm, the usual approach for estimating the MAP consists in two steps: first, the variables of dimension are estimated, then the unknowns are estimated using the samples with the estimated group and peak numbers. Instead, we simply pick up the most probable generated sample. We have observed that these two estimation approaches provide almost the same results in our context. The benefit of our approach is to avoid storing the whole set of generated samples. The estimation being computed with only one sample, we do not face with the label switching problem

[45,46] (which is even more difficult in the context of a variable dimension problem).

To reduce the computation time, we adopt a multi-resolution framework (see Algorithm 1): only a small number of spectra (spread on the whole sequence) is first decomposed; then the other spectra are progressively added to these data (the new model is estimated from the old one using linear interpolation of the parameters). The RJMCMC algorithm is performed at each resolution: this is what we call a "run" in the sequel. This scheme is feasible since the assumption of smooth evolution also holds for a small number of spectra.

Moreover, we observed that the algorithm tends to underestimate the hyperparameters during the burn-in where only few small groups are detected. Small values of $r_{\boldsymbol{c}}, r_{\boldsymbol{a}}$ and $r_{\boldsymbol{w}}$ do favor very straight groups, thus the groups are split into many straight groups. To avoid this drawback, we force the hyperparameters to be overestimated during the burn-in: the estimated values of $r_{\boldsymbol{c}}, r_{\boldsymbol{a}}$ and $r_{\boldsymbol{w}}$ are systematically multiplied by 2. Then, overestimating the hyperparameters is a way to be less dependent of the initialization and to move more freely in the huge solution space.

Concerning the RJMCMC algorithm (Algorithm 2), it is essential to perform updates very often after a trans-dimensional move in order to achieve good convergence properties. For this reason, the probability of update is increased just after that a trans-dimensional move has been accepted. Note that this approach has similarities with the one proposed by [43] in which a secondary Markov chain is sampled after a trans-dimensional move. Besides, it is essential to perform less births and deaths than other trans-dimensional moves because there are more peaks than groups. But the probability of births and deaths is increased when no birth and death has been performed for a long time.

Finally, several methods for assessing convergence of RJMCMC algorithm have been proposed (see *e.g.* [47–49] for a discussion). However, the proposed methods often give different results of when convergence has been achieved, which does not give a rigorous answer to the convergence assessment. Besides, our solution space is huge (its dimension is typically higher than 1000), making the search for convergence even more difficult. That is why we adopt a conservative approach and ensure that the convergence is reached with the chosen iteration number by visually inspecting the chain behaviors.

4. Performances of the proposed approach

Otherwise stated, we set in the sequel the following values to the hyperparameters: $K_{\max}=100$, $c_{\min}=1$, $c_{\max}=N$, $a_{\min}=\max(\boldsymbol{y})/50$, $a_{\max}=\max(\boldsymbol{y})$, $w_{\min}=1$ and $w_{\max}=N/10$.

4.1. Benefits of the proposed approach compared to a constant dimension model

In this section, we compare the proposed approach with the former one published by the authors [14]. The method in [14] also performs a joint decomposition, but its model differs from the proposed one on two major points: first, the hyperparameters $r_{\boldsymbol{c}}, r_{\boldsymbol{a}}, r_{\boldsymbol{w}}$ and $r_{\boldsymbol{v}}$ and the group number K are set by the user whereas the proposed approach needs only to set ξ; second, the model is supposed to be of known and constant dimension (the peak number and the group number are not estimated: every group is of length S). Besides, the optimization is performed using a Gibbs sampler coupled with a simulated annealing algorithm [33,50,51].

The two methods are evaluated on a synthetic sequence of $S=60$ spectra with $N=256$ samples and $K=6$ groups. Fig. 5 represents the centers, amplitudes, and widths of the peaks. White Gaussian noise of variance $r_{\boldsymbol{v}}$ is added to each signal so that the peak signal-to-noise ratio (PSNR), defined in (10), equals 30 dB.

$$\mathrm{PSNR}=10\log_{10}\left[\left(r_{\boldsymbol{v}}\frac{1}{S}\sum_{s=1}^{S}\frac{1}{\max(\boldsymbol{y}_s)^2}\right)^{-1}\right]. \tag{10}$$

Fig. 4 shows some of the spectra. The sequence shows up four noticeable behaviors:

- three groups (2, 4, and 5) have a length smaller than S;
- two groups are very close (3, 6);
- two groups cross each other (1, 2);
- at the junction of these two groups ($s=14$), the peaks have exactly the same center and width so that a unique Gaussian peak is observed.

The simulations were performed in Matlab R2010a running on Ubuntu 12.04 with a 3.47 GHz Intel Xeon CPU; the code and data are available on-line at miv.u-strasbg.fr/mazet/jointdec. The method of [14] has been run for 150,000 iterations (approximately 4 h 16), the hyperparameters are set as indicated in [14], the noise variance has been set to the value used for generating the data, and the group number has been fixed to the maximal number of peaks in a spectrum, *i.e.* 5. The proposed method has been run for 30,000 iterations on 8 spectra

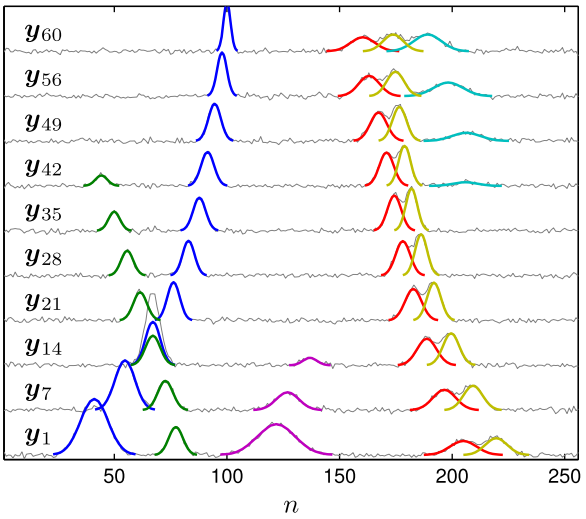

Fig. 4. Spectra $\boldsymbol{y}_s$ for $s\in\{1,7,14,21,28,35,42,49,56,60\}$ (in gray). The colored peaks stand for the actual peaks. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

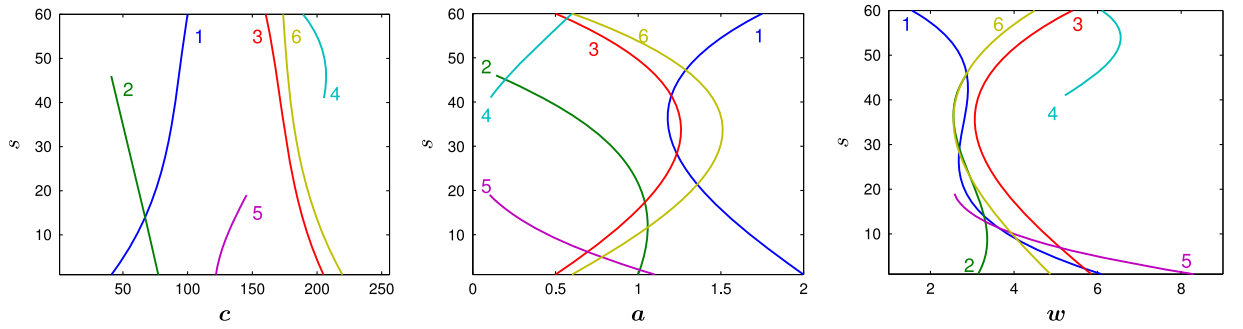

Fig. 5. Centers (left), amplitudes (middle) and widths (right) of the peaks of the synthetic sequence (ground truth). Each color and number stands for a group. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

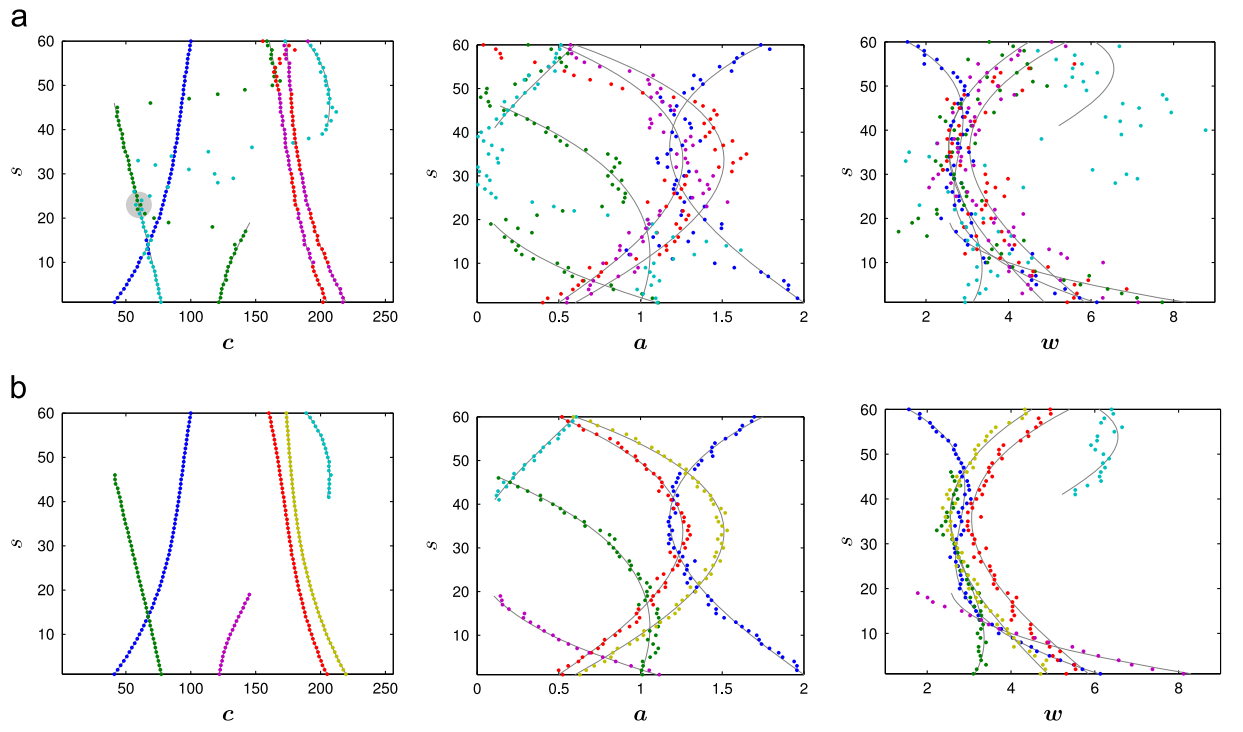

Fig. 6. Centers (left), amplitudes (middle) and widths (right) of the estimated peaks. Gray lines represent the ground truth (see Fig. 5) and colored dots the estimations (each color corresponds to a group). (a) Estimation with the method of [14]. (b) Estimation with the proposed approach. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

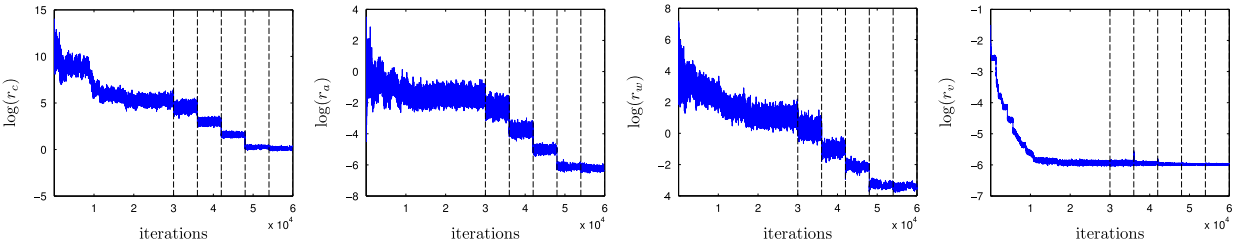

Fig. 7. Traces of the hyperparameters (in logarithmic scale) with respect to the iterations. The vertical dashed lines limit the different RJMCMC runs.

with an overestimation of the hyperparameters, then on 6000 iterations in the following runs of the algorithm (this takes approximately 30 min). An illustrative example of the result that can be obtained is shown in Fig. 6a and b.

In the method of [14], each estimated group has a length of S. Even if the group number has been fixed to the maximal number of peaks in a spectrum (namely 5), the number of peaks to estimate is greater than the real

number of peaks (so, setting the group number to its real value, *i.e.* 6, leads to worse results). Then, it is not surprising that some real peaks are estimated with two peaks (see *e.g.* the gray zone in Fig. 6a, left), favoring swaps between groups. Besides, the labeling move of [14] is different from the one proposed in this paper. While it works well for a small number of spectra, it appears to be inefficient for realistic spectrum number. We also observe that the crossing in spectrum 14, where the peaks 1 and 2 have same center and width, is not well estimated. Indeed, the estimated peaks zigzag over about 5 spectra rather than crossing properly. Nevertheless, the estimation of the peak parameters is relatively satisfactory (except when one real peak is estimated by two), and the mean square error between the real $\boldsymbol{x}$ and estimated $\widehat{\boldsymbol{x}}$ signals, defined in (11), equals 10.062×10^{-4}:

$$\mathrm{MSE} = \frac{1}{SN}\sum_{s=1}^{S}\sum_{n=1}^{N}(x_{s,n}-\widehat{x}_{s,n})^2. \tag{11}$$

Note that the Markovian priors as well as the misclassification of the groups lead to peaks that do not fit the spectra and result in an increase of the MSE.

In the proposed approach, ξ is set to 10^{-3}, as discussed in Section 4.2. The estimation of the peak parameters is also satisfactory (Fig. 6b), and since the peaks are well classified, we obtained a better MSE ($15{,}843 \times 10^{-4}$). Besides, the proposed approach estimates very well the group and peak number: the six groups are obtained, with correct beginnings and lengths. Furthermore, the proposed

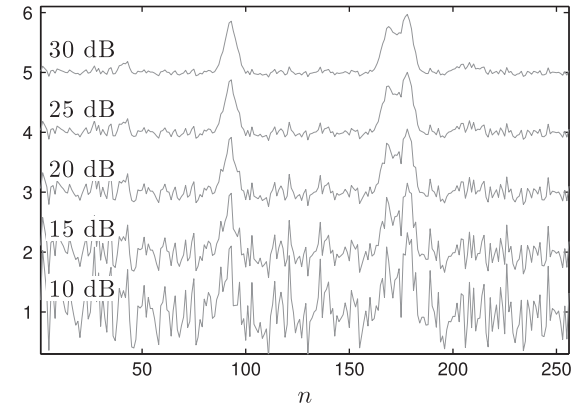

Fig. 8. Spectrum at $s=45$ for different values of PSNR.

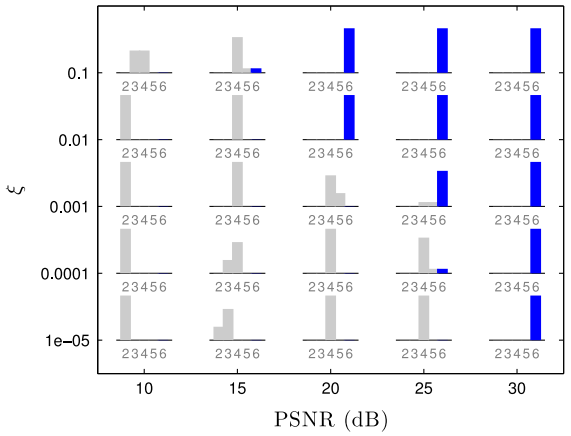

Fig. 9. Estimated number of groups with respect to the PSNR and the value of ξ. The blue bins correspond to the real value of K. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

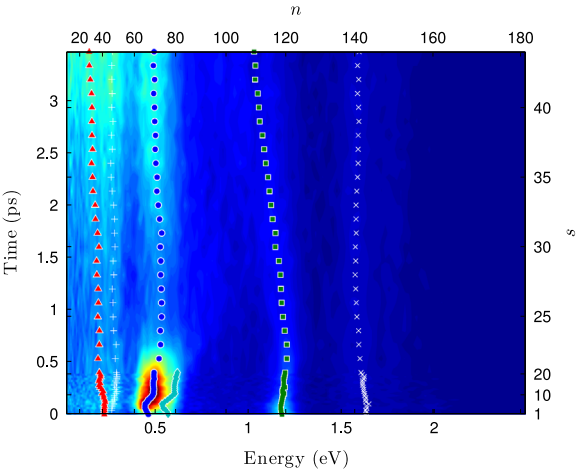

Fig. 11. First photoelectron data plotted as a function of the electron energy (n, horizontal axis) and acquisition time (s, vertical axis). The points are the peak centers estimated with the proposed approach.

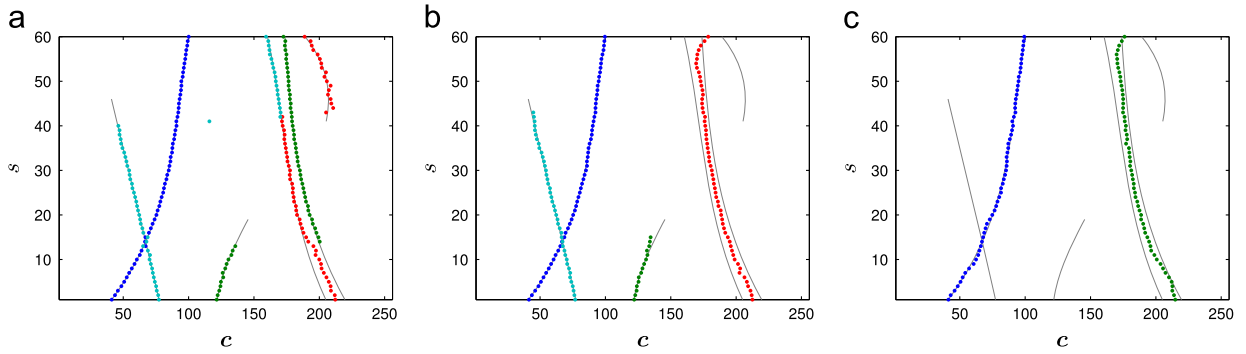

Fig. 10. Some examples of estimation obtained with the proposed approach for different PSNR and values of ξ. Only the peak centers are plotted (each color corresponds to a group). (a) PSNR=25 dB, $\xi=10^{-3}$. (b) PSNR=15 dB, $\xi=10^{-3}$. (c) PSNR=10 dB, $\xi=10^{-3}$. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

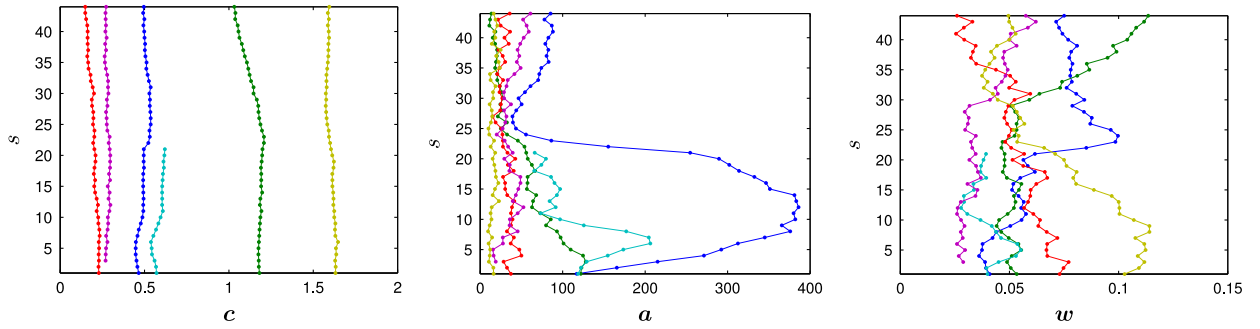

Fig. 12. Estimated parameters on the first photoelectron data. Each color and number stands for a group. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

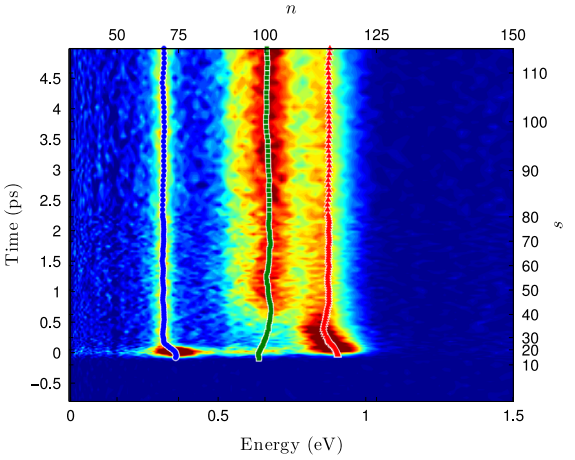

Fig. 13. Second photoelectron data (zoom on the region of interest) plotted as a function of the electron energy (n, horizontal axis) and acquisition time (s, vertical axis). The points are the peak centers estimated with the proposed approach. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

approach manages correctly the crossing in spectrum 14: although the two real peaks can be estimated perfectly by only one peak, the proposed approach finds their parameters with a very good accuracy.

Fig. 7 shows the evolution of the hyperparameters r_c, r_a, r_w and r_v. The convergence of these hyperparameters is reached at each run of the RJMCMC. Besides, one can observe that the hyperparameter values decrease at each new run. The first decrease (iteration 30,000) is due to the end of overestimation. The next decreases are more difficult to explain. Indeed, look at the conditional posteriors (see Eqs. (A.6)–(A.9)): the first parameter of the inverse gamma increases when adding new spectra (since M increases), but one cannot conclude on the evolution of the second parameter: depending on the evolution of the peaks, the discrete derivative can increase or decrease when adding new spectra.

4.2. Influence of the noise and of the parameter ξ

We now discuss the behavior of the proposed approach with respect to the noise. Five different values of PSNR are derived by adding white Gaussian noise to the same non-noisy data. The PSNR values lie between 10 dB and 30 dB; Fig. 8 shows the spectrum 45 for the different data sets. The proposed approach requires the setting of the parameter ξ: five values have been tested: 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. For each PSNR and each value of ξ, the algorithm has been performed ten times. Fig. 9 plots the histograms of the estimated number of groups with respect to the PSNR and the value of ξ. Figs. 6b, 10a–c show illustrative examples of the estimation for different PSNR and ξ.

It appears that for high PSNR, the estimation is very good (see *e.g.* Fig. 6b): the groups have been found in all cases and their beginnings and lengths are correct in most cases. Specifically, for $\xi = 10^{-3}$, the beginnings and lengths of each group are correct in every simulation. The lengths of groups 2, 4, and 5 may be slightly too small for smaller values of ξ, and slightly too large for larger values of ξ. The length of these groups is rather difficult to estimate because it is smaller than S; so, the amplitude of the groups vanishes.

A similar observation also holds for smaller PSNRs. The length of a group of length S is always well-estimated, but the groups with vanishing peaks are generally too small (higher value of ξ leads naturally to larger groups than smaller values of ξ). In addition, ξ has an impact on the number of estimated groups. As we can observe in Fig. 9, reducing the PSNR and ξ leads to reduce the number of estimated groups. Indeed, a low value of ξ favors a priori a small number of groups and decreasing the PSNR reduces the interest of creating a new group. Therefore, for small PSNR, one should set a high value of ξ so as to not penalize the group number. See *e.g.* the case of a PSNR of 20 dB: excellent results are only achieved with $\xi = 10^{-1}$ or 10^{-2} (the 6 groups are detected). Note also that for small values of ξ, the number of estimated groups is never overestimated as it may happen in real data. This can be explained by the fact that the synthetic data are simulated from the model.

Note also that the labeling may not be correct for small values of $\xi (\xi \leq 10^{-3})$ and intermediate PSNR (20 or 25 dB). The model favors a small number of groups (as ξ is small) while explaining well the data. For instance in Fig. 10a, 4 groups are enough to explain the data, but this implies a bad labeling and, consequently, non-smooth groups: the model prefers groups to be fewer than smoother.

Anyway, this problem arises only with small values of ξ which is seldom used in practice.

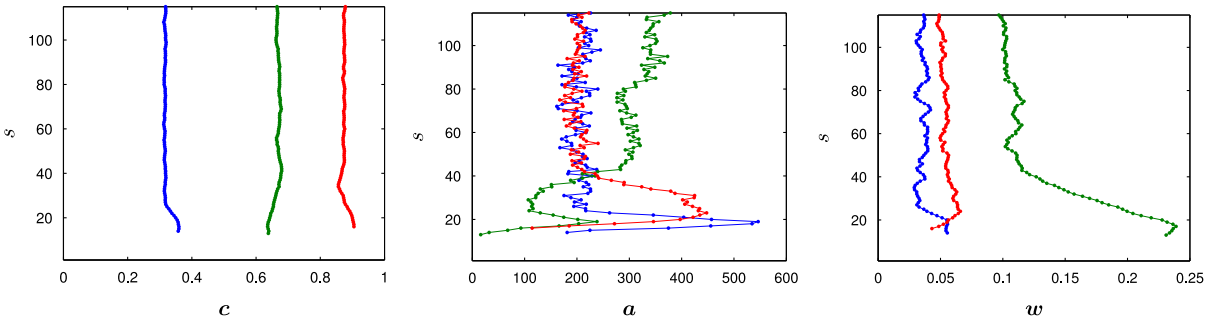

Fig. 14. Estimated parameters on the second photoelectron data. Each color and number stands for a group. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Finally, for low PSNR (less than 15 dB), we observe the following kinds of result:

- if 2 groups are detected, these are group 1 and a unique group fitting the real groups 3 and 6. Indeed, the benefit of creating a group to better explain the data is smaller than the "cost" induced by its creation (see Fig. 10c);
- if 3 groups are detected: they are the two previous ones plus the group 5;
- if 4 groups are detected: they are the three previous ones plus the group 2 and the crossing between groups 1 and 2 is always well estimated (see Fig. 10b);
- finally if 5 groups are detected: they are the four previous ones plus the group 4.

The group 4 is difficult to detect because its peaks have a small amplitude, so a birth on this group is seldom accepted. On the opposite, some small peaks with similar amplitudes in groups 2 and 5 are more likely to be found. This is due to the fact that at least one peak with a sufficient amplitude is estimated in these groups, then, thanks to the increase move, the group can be extended.

5. Results on real photoelectron data

Time-resolved photoelectron spectroscopy [52] studies the energy relaxation occurring after absorption of a photon by an isolated molecule, atom or a blend of both. The energy relaxation is probed by ionization of the studied system with another delayed photon, thereby ejecting a so-called photoelectron. The distribution of the photoelectrons according to their energy is measured at different times to get a temporal sequence of photoelectron spectra [53]. The experiments presented in this paper study the relaxation of an atom of barium and of a molecule called DABCO [54] on an argon cluster (droplet isolated from several hundred of argon atoms): it is question of understanding the mechanism that leads the excited barium atom or molecule to lose its energy through an interaction with an inert and cold medium (34 K). The proposed approach was applied on a sequence of photoelectron spectra: the goal is to determine how the energy, intensity and shape (*i.e.* centers, amplitudes and widths) of the peaks evolve through the sequence, indicating the changes undergone by the studied system. More information on the technique as well as the experimental setup can be found in [1].

A photoelectron spectrum may show a continuum, usually modeled as an exponential of the form $\alpha_s \exp(-n/\beta_s)$ where α_s and β_s are unknowns and depend of the spectrum s. Thus, we consider the model presented in Section 2 and add priors on $\boldsymbol{\alpha}=\{\alpha_1,\ldots,\alpha_S\}$ and $\boldsymbol{\beta}=\{\beta_1,\ldots,\beta_S\}$. To get a smooth evolution of the continuum, the priors are Markovian such as $\boldsymbol{c}_k$, $\boldsymbol{a}_k$ or $\boldsymbol{w}_k$ (this also implies to estimate the hyperparameters $r_{\boldsymbol{\alpha}}$ and $r_{\boldsymbol{\beta}}$ of these priors). In addition, the sampling in both energy (n) and time (s) are irregular. The sampling in energy is modeled by replacing in (2) the variable n by the energy λ_n at sample n. This implies to choose $c_{\min}=\lambda_1$, $c_{\max}=\lambda_N$, $w_{\min}$ as the maximal step between two samples and $w_{\max}=(\lambda_N-\lambda_1)/10$. However, we do not consider the irregular sampling in the model. Indeed, the assumption of smooth evolution between contiguous spectra remains valid since the sampling is fine only at times where the evolution is significant. In conclusion, the model for real data is the same than the one presented in Section 2, except for the aforementioned differences.

We first process the data considered in [1] (Fig. 11). The sequence gathers $S=44$ spectra (covering a duration of 3,47 ps), each of $N=181$ samples (from 0.02 eV to 2.52 eV). The results obtained with the proposed approach are shown in Figs. 11 and 12 (we do not find any relevance to show the results obtained with the method of [14] since they are totally unexploitable). The peaks are properly classified (there is no permutation between the peaks), and it appears that the peak evolution is smoother than the one obtained with a sequential approach (see [1, Fig. 4]). In particular, some groups are close to each other leading to overlapped peaks (see the two groups at 0.25 eV, or those around 0.5 eV at short times). These results have been validated by the experts and are not straightforward when looking at Fig. 11. Also, the energy of a peak varies in time: its energy increases slightly from 1.18 to 1.20 eV between 0 and 0.4 ps, then decreases and reaches 1.05 eV at 2.5 ps. The quantitative results confirm the qualitative observations of [1] namely that the energy of an electronic level of the studied system varies in time and opens the door to a complete analysis of the time evolution (energy *vs.* time and intensity *vs.* time) to understand the energy relaxation. The dynamics observed

suggests that the energy level is very sensitive to the surrounding atoms of argon, and that the excited barium atom is colliding the argon cluster prior to the electronic relaxation.

In the second experiment we are interested in the energy relaxation of the excited DABCO molecule (see [54]) solvated upon argon cluster (Fig. 13). The analysis presented here focuses on the second order polarized part of the photoelectron spectra. The sequence gathers $S=115$ spectra (from -0.80 ps to 4.99 ps), each of $N=191$ samples (from 10^{-5} eV to 2.44 eV). The results displayed in Fig. 13 and 14 point out three different behaviors. The group around 0.3 eV (blue trajectory), already reported in [54] deals with the isolated DABCO molecule (not deposited). Basically, it shows that the isolated DABCO molecule poorly relaxes its internal electronic energy. The two other groups are linked to the dynamics of the solvated molecule. Roughly, two solvated structures are identified, corresponding to the groups around 0.7 eV and 0.85 eV (respectively green and red trajectories). At time zero, only one is excited (the one at 0.85 eV), but after 1 ps part of the molecule is observed with the other one. The quantitative values provided by this analysis denote accurately the different steps of the relaxation dynamics.

6. Conclusion

The problem addressed in this paper concerns the decomposition of a spectroscopic signal sequence: data are a series of spectroscopic signals (aka spectra) that can be modeled as a noisy sum of peaks. The goal is to decompose the spectra and to classify the estimated peaks into groups. The key idea is to achieve a joint decomposition by processing the whole sequence instead of each spectrum separately. We propose a Bayesian method using the RJMCMC algorithm. The main difficulties concern the size of the solution space (it is huge and unknown) and the high correlation between the unknowns. To tackle these difficulties, we bring the following contributions:

1. we have proposed an original Bayesian generative model. In particular, the model considers two variables of dimension (K and M) which are not independent, and the smooth evolution of the peaks is modeled using Markovian priors on the peak parameters;
2. a joint prior on K, $\boldsymbol{l}$ and $\boldsymbol{b}$ has been used;
3. we have proposed novel RJMCMC moves adapted to the model;
4. lastly, some techniques such as mixed proposals, overestimation of the hyperparameters or a multiresolution approach in the burn-in time are also implemented to accelerate the convergence of the algorithm.

Simulations and results on real data show the relevance of the proposed approach compared to a constant dimension approach [14] (more results on real photoelectron data are shown in [55]). Future works will be dedicated to the decomposition of bidimensional sequences where spectra lie on a bidimensional grid. A possible application concerns the decomposition of spectra in radio-astronomical hyperspectral images for estimating the kinematic of galaxies.

Algorithm 1. Main algorithm.

1: select $\tilde{S}$ (say 8) spectra almost uniformly distributed on the sequence
2: initialize with a unique group with a small number of peaks
3: decompose (see Algorithm 2) these $\tilde{S}$ spectra with an overestimation of the hyperparameters
4: use the MAP estimation as initialization for the next step
5: **while** $\tilde{S} \leq S$ **do**
6: decompose (see Algorithm 2) the $\tilde{S}$ spectra without overestimation
7: increase $\tilde{S}$ (generally by a factor around 2)
8: interpolate the MAP estimation on the $\tilde{S}$ spectra so as to get an initialization for the next step
9: **end while**

Algorithm 2. RJMCMC algorithm (I is the iteration number of the algorithm).

1: initialize the move probabilities
2: **for** $i=0$ **to** I **do**
3: normalize the probabilities so that their sum equal 1
4: choose randomly a move according to its probability
5: perform the chosen move
6: **if** a birth or death has been accepted **then**
7: set the probability of birth and death to a low value
8: **end if**
9: **if** no birth nor death has been performed for many iterations **then**
10: set the probability of birth and death to a high value
11: **end if**
12: **if** a trans-dimensional move has been accepted **then**
13: set the probability of update to a high value
14: **end if**
15: **if** no trans-dimensional move has been accepted for many iterations **and** p_U not too small **then**
16: reduce the probability of update p_U
17: **end if**
18: compute the posterior
19: **if** the posterior is greater than the best posterior **then**
20: save the current parameters as the MAP estimation
21: best posterior $\leftarrow$ posterior
22: **end if**
23: **end for**

Appendix A. Conditional posterior distributions

For each $k \in \{1,\ldots,K\}$ and $m \in \{1,\ldots,l_k\}$, the peak center $c_{k,m}$ and peak width $w_{k,m}$ are distributed according to

$$p(c_{k,m}|\ldots) \propto \exp\left(-\frac{\|\boldsymbol{y}_s-\boldsymbol{x}_s\|^2}{2r_{\boldsymbol{v}}}-\frac{\|D\boldsymbol{c}_k\|^2}{2r_{\boldsymbol{c}}}\right)\mathbb{1}_{\mathcal{C}}(c_{k,m}), \tag{A.1}$$

$$p(w_{k,m}|\ldots) \propto \exp\left(-\frac{\|\boldsymbol{y}_s-\boldsymbol{x}_s\|^2}{2r_{\boldsymbol{v}}}-\frac{\|D\boldsymbol{w}_k\|^2}{2r_{\boldsymbol{w}}}\right)\mathbb{1}_{\mathcal{W}}(w_{k,m}) \tag{A.2}$$

where $\boldsymbol{x}_s$ is defined in (2).

For each k, m, the peak amplitude $a_{k,m}$ is distributed according to a normal distribution truncated on $\mathcal{A}$:

$$p(a_{k,m}|\ldots) \propto \exp\left(-\frac{(a_{k,m}-\mu_{k,m})^2}{2\rho_{k,m}}\right)\mathbb{1}_{\mathcal{A}}(a_{k,m}) \tag{A.3}$$

where:

$$\mu_{k,m}=\frac{r_{\boldsymbol{a}}\boldsymbol{z}_{s,k,m}^T\boldsymbol{g}_{k,m}+r_{\boldsymbol{v}}B_{k,m}}{r_{\boldsymbol{a}}\boldsymbol{g}_{k,m}^T\boldsymbol{g}_{k,m}+r_{\boldsymbol{v}}A_{k,m}} \tag{A.4}$$

Table 1
Coefficients $A_{k,m}$ and $B_{k,m}$ in (A.4) and (A.5) (D being a first-order discrete derivative).

m	$A_{k,m}$	$B_{k,m}$
1	1	$a_{k,2}$
$2,\dots,l_k-1$	2	$a_{k,m-1}+a_{k,m+1}$
l_k	1	a_{k,l_k-1}

$$\rho_{k,m}=\frac{r_{\boldsymbol{a}}r_{\boldsymbol{v}}}{r_{\boldsymbol{a}}\boldsymbol{g}_{k,m}^T\boldsymbol{g}_{k,m}+r_{\boldsymbol{v}}A_{k,m}} \tag{A.5}$$

and $\boldsymbol{g}_{k,m}$ is a Gaussian peak of amplitude 1, center $c_{k,m}$, and width $w_{k,m}$; vector $\boldsymbol{z}_{s,k,m}$ is the residual of spectrum $s=b_k+m-1$ from which the contribution of peak (k, m) is not considered; and coefficients $A_{k,m}$ and $B_{k,m}$ are given in Table 1.

The conditional posteriors of the hyperparameters are inverse gamma distributions:

$$r_{\boldsymbol{v}}|\cdots\sim\mathcal{IG}\left(NS/2-1,\ \sum_{s=1}^{S}\|\boldsymbol{y}_s-\boldsymbol{x}_s\|^2/2\right), \tag{A.6}$$

$$r_{\boldsymbol{c}}|\cdots\sim\mathcal{IG}\left((M-K)/2-1,\ \sum_{k=1}^{K}\|D\boldsymbol{c}_k\|^2/2\right), \tag{A.7}$$

$$r_{\boldsymbol{a}}|\cdots\sim\mathcal{IG}\left((M-K)/2-1,\ \sum_{k=1}^{K}\|D\boldsymbol{a}_k\|^2/2\right), \tag{A.8}$$

$$r_{\boldsymbol{w}}|\cdots\sim\mathcal{IG}\left((M-K)/2-1,\ \sum_{k=1}^{K}\|D\boldsymbol{w}_k\|^2/2\right). \tag{A.9}$$

Appendix B. Detailed move description

B.1. Birth move

A birth adds into the model a group of one peak (so the group length is 1). A mixed proposal distribution is used (see Section 3.2). With a uniform proposal, the group beginning and the peak parameters (center, amplitude, and width) are respectively chosen uniformly on $\{1,\dots,S\}$, $\mathcal{C}$, $\mathcal{A}$, and $\mathcal{W}$. Otherwise, the following steps are performed:

1. the group beginning and an index $\tilde{i}$ are randomly chosen proportionally to the square of the residuals computed at each sample. The peak center $\tilde{c}$ is generated according to $\mathcal{U}_{[\tilde{i}-(1/2),\tilde{i}+(1/2)]}$ so that $\tilde{c}$ is defined on the continuous interval $[1, N]$. The signal amplitude at $\tilde{i}$ is denoted $\tilde{a}$;
2. the minimal distance between $\tilde{c}$ and the first sample lower than $\tilde{a}/2$ is computed and stored in $\tilde{d}$, so $\tilde{d}$ can be considered as half the FWHH (full width at half height) of the peak to generate. The peak width is generated according to a Gaussian $\mathcal{N}(\mu_w,\sigma_w)$ truncated on $\mathcal{W}$, where $\mu_w=\tilde{d}/\sqrt{2\log(2)}$ and $\sigma_w=\mu_w/10$;
3. the peak amplitude is generated according to the posterior A.3 considering the peak parameters detailed in the previous items.

B.2. Death move

A death simply chooses uniformly a group of length 1 and removes it.

B.3. Split move

A split consists in "cutting" a group into two contiguous groups. A mixed proposal is used: the group to split and the location of cut are either chosen uniformly, or with a probability proportional to the drifts. A drift is computed for each peak (k, s) and equals

$$\begin{aligned}d_{k,s}=&\frac{(c_{k,s}-c_{k,s-1})^2}{r_{\boldsymbol{c}}}+\frac{(c_{k,s}-c_{k,s+1})^2}{r_{\boldsymbol{c}}}\\&+\frac{(a_{k,s}-a_{k,s-1})^2}{r_{\boldsymbol{a}}}+\frac{(a_{k,s}-a_{k,s+1})^2}{r_{\boldsymbol{a}}}\\&+\frac{(w_{k,s}-w_{k,s-1})^2}{r_{\boldsymbol{w}}}+\frac{(w_{k,s}-w_{k,s+1})^2}{r_{\boldsymbol{w}}}.\end{aligned} \tag{B.1}$$

B.4. Merge move

A merge consists in assembling two contiguous groups into a single one. The two groups to merge are uniformly chosen among the couples of contiguous groups.

B.5. Increase move

An increase adds a peak at an extremity of a group. The group to increase as well as the choice of its extremity are uniformly chosen. Then a mixed proposal is used. With a uniform proposal, the parameters of the new peak are chosen uniformly on $\mathcal{C}$, $\mathcal{A}$, and $\mathcal{W}$. With a model/data-driven proposal, the peak center is generated according to a Gaussian whose means corresponds to the actual last peak center, whose variance equals $r_{\boldsymbol{c}}$ and truncated on $\mathcal{C}$. We proceed similarly for the amplitude and the width.

B.6. Reduce move

A reduce deletes a peak at the extremity of a group. The group and its extremity to reduce are uniformly chosen.

B.7. Labeling move

The labeling switches the peaks in $s'\geq s$ between two groups using a mixed proposal (s being the split location). With a uniform proposal, the two groups and the switch location are uniformly chosen. Otherwise, they are chosen according to the posteriors computed for each possible switch.

References

[1] A. Masson, L. Poisson, M.-A. Gaveau, B. Soep, J.-M. Mestdagh, V. Mazet, F. Spiegelman, Dynamics of highly excited barium atoms deposited on large argon clusters. I. General trends, J. Chem. Phys. 133 (054307) (2010) 1–14.

[2] H.-L. Nguyen Thi, C. Jutten, Blind source separation for convolutive mixtures, Signal Process. 45 (2) (1995) 209–229.
[3] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (2) (2012) 354–379.
[4] C. Gobinet, V. Vrabie, M. Manfait, O. Piot, Preprocessing methods of Raman spectra for source extraction on biomedical samples: application on paraffin-embedded skin biopsies, IEEE Trans. Biomed. Eng. 56 (5) (2009) 1371–1382.
[5] G. Tomasi, F. van den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241.
[6] S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415, http://dx.doi.org/10.1109/78.258082.
[7] S. Chen, D. Donoho, M. Saunders, Atomic Decomposition by Basis Pursuit, Technical Report, Standford University, 1995.
[8] M. Davy, S. Godsill, J. Idier, Bayesian analysis of polyphonic western tonal music, J. Acoust. Soc. Am. 119 (4) (2006) 2498–2517.
[9] M. Nicoli, V. Rampa, U. Spagnolini, Hidden Markov model for multidimensional wavefront tracking, IEEE Trans. Geosci. Remote Sens. 40 (3) (2002) 651–662.
[10] J. Idier, Y. Goussard, Markov modeling for Bayesian restoration of two-dimensional layered structures, IEEE Trans. Inf. Theory 39 (4) (1993) 1356–1373.
[11] Y. Bar-Shalom, T. Fortmann, Tracking and Data Association, vol. 179, Academic Press Professional, Inc., San Diego, CA, 1987.
[12] S. Blackman, Multiple-Target Tracking with Radar Applications, Mass. Artech House, Norwood, 1986.
[13] S. Gauglitz, T. Höllerer, M. Turk, Evaluation of interest point detectors and feature descriptors for visual tracking, Int. J. Comput. Vis. 94 (3) (2011) 335–360.
[14] V. Mazet, Joint Bayesian decomposition of a spectroscopic signal sequence, IEEE Signal Process. Lett. 18 (3) (2011) 181–184.
[15] J. Czyz, B. Ristic, B. Macq, A particle filter for joint detection and tracking of color objets, Image Vis. Comput. 25 (8) (2007) 1271–1281.
[16] B.-N. Vo, B.-T. Vo, N.-T. Pham, D. Suter, Joint detection and estimation of multiple objects from image observations, IEEE Trans. Signal Process. 58 (10) (2010) 5129–5141.
[17] R.-L. Streit, Tracking on Intensity-Modulated Data Streams, Technical Report 11221, Naval Undersea Warfare Center Division, Newport, Rhode Island, May 2000.
[18] R.-L. Streit, M.-L. Graham, M.-J. Walsh, Multitarget tracking of distributed targets using histogram-PMHT, Digit. Signal Process. 12 (2) (2002).
[19] S.-J. Davey, M. Wieneke, Tracking groups of people in video with histogram-PMHT, in: Workshop on Defense Applications of Signal Processing, 2011.
[20] R. Caruana, R. Searle, T. Heller, S. Shupack, Fast algorithm for the resolution of spectra, Anal. Chem. 58 (1986) 1162–1167.
[21] G. Allen, R. McMeeking, Deconvolution of spectra by least-squares fitting, Anal. Chim. Acta 103 (1) (1978) 73–108.
[22] R. Fraser, E. Suzuki, Resolution of overlapping absorption bands by least squares procedures, Anal. Chem. 38 (12) (1966) 1770–1773.
[23] D. Donoho, Y. Tsaig, Fast Solution of l1-Norm Minimization Problems When the Solution may be Sparse, Technical Report, Standford University, 2006.
[24] S. Sahnoun, E.-H. Djermoune, C. Soussen, D. Brie, Sparse multi-resolution modal estimation, in: IEEE Statistical Signal Processing Workshop, Nice, France, 2011.
[25] R. Fischer, V. Dose, Analysis of mixtures in physical spectra, Bayesian Methods (2001) 145–154.
[26] S. Gulam Razul, W. Fitzgerald, C. Andrieu, Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC, Nucl. Instrum. Methods Phys. Res. A 497 (2003) 492–510.
[27] N. Haan, S. Godsill, Bayesian models for DNA sequencing, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002, pp. IV-4020–4023.
[28] V. Mazet, Développement de méthodes de traitement de signaux spectroscopiques: estimation de la ligne de base et du spectre de raies (Ph.D. thesis), Université Henri Poincaré, Nancy 1, December 2005.
[29] P. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika 82 (4) (1995) 711–732. http://dx.doi.org/10.1093/biomet/82.4.711.
[30] T. Ida, M. Ando, H. Toraya, Extended pseudo-Voigt function for approximating the Voigt profile, J. Appl. Spectrosc. 33 (2000) 1311–1316.
[31] H. Jeffreys, An invariant form for the prior probability in estimation problems, Proc. R. Soc. Lond. Ser. A 186 (1007) (1946) 453–461.
[32] R. Kass, L. Wasserman, The selection of prior distributions by formal rules, J. Am. Stat. Assoc. 91 (1996) 1343–1370.
[33] C. Robert, G. Casella, Monte Carlo Statistical Methods, 2nd edition, Springer, New York, NY, 2004.
[34] O. Cappé, C. Robert, T. Rydén, Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers, J. R. Stat. Soc.: Ser. B 65 (3) (2003) 679–700.
[35] S. Godsill, On the relationship between Markov chain Monte Carlo methods for model uncertainty, J. Comput. Gr. Stat. 10 (2) (2001) 230–248.
[36] P. Dellaportas, J. Forster, I. Ntzoufras, On Bayesian Model and Variable Selection Using MCMC, Stat. Com. 12 (2002) 27–36. http://dx.doi.org/1023/A:1013164120801.
[37] S. Richardson, P. Green, On Bayesian analysis of mixtures with an unknown number of components, J. R. Stat. Soc. B 59 (4) (1997) 731–792.
[38] C. Andrieu, N. De Freitas, A. Doucet, M. Jordan, An introduction to MCMC for machine learning, Mach. Learn. 50 (2003) 5–43.
[39] N. Chopin, Fast simulation of truncated Gaussian distributions, Stat. Comput. 21 (2011) 275–288.
[40] V. Mazet, Simulation of Truncated Gaussian Distribution, URL ⟨http://miv.u-strasbg.fr/mazet/rtnorm/⟩, 2012.
[41] L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, New York, NY, 1986 URL: ⟨http://luc.devroye.org/rnbookindex.html⟩.
[42] S. Brooks, P. Giudici, G. Roberts, Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions, J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 65 (1) (2003) 3–39.
[43] F. Al-Awadhi, M. Hurn, C. Jennison, Improving the acceptance rate of reversible jump MCMC proposals, Stat. Probab. Lett. 69 (2) (2004) 189–198.
[44] P. Green, A. Mira, Delayed rejection in reversible jump Metropolis-Hastings, Biometrika 88 (4) (2001) 1035–1053.
[45] M. Stephens, Dealing with label switching in mixture models, J. R. Statist. Soc. B 62 (4) (2000) 795–809.
[46] A. Jasra, C. Holmes, D. Stephens, Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling, Stat. Sci. 20 (1) (2005) 50–57.
[47] S. Brooks, P. Giudici, A. Philippe, Nonparametric convergence assessment for mcmc model selection, J. Comput. Gr. Stat. 12 (1) (2003) 1–22.
[48] J. Castelloe, D. Zimmerman, Convergence Assessment for Reversible Jump MCMC Samplers, Technical Report, University of Iowa, 2002.
[49] S. Brooks, A. Gelman, G. Jones, X.-L. Meng (Eds.), Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC, Boca Raton, FL, 2011.
[50] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984) 721–741.
[51] S. Kirkpatrick, C. Gelatt Jr., M. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[52] A. Stolow, A. Bragg, D. Neumark, Femtosecond time-resolved photoelectron spectroscopy, Chem. Rev. 104 (4) (2004) 1719–1757.
[53] E. Gloaguen, J.-M. Mestdagh, L. Poisson, F. Lepetit, J.-P. Visticot, B.S.M. Coroiu, A. Eppink, D.-H. Parker, Experimental evidence for ultrafast electronic relaxation in molecules, mediated by diffuse states, J. Am. Chem. Soc. 127 (47) (2005) 16529–16534.
[54] L. Poisson, R. Maksimenska, B. Soep, J.-M. Mestdagh, D. Parker, M. Nsangou, M. Hochlaf, Unusual quantum interference in the S-1 state of DABCO and observation of intramolecular vibrational redistribution, J. Phys. Chem. A 114 (2010) 3313–3319.
[55] S. Awali, L. Poisson, B. Soep, M.-A. Gaveau, M. Briant, C. Pothier, J.-M. Mestdagh, M.B.E.H. Rhouma, M. Hochlaf, V. Mazet, S. Faisan, Time resolved observation of the solvation dynamics of a Rydberg excited molecule deposited on an Argon cluster – I: DABCO∗ at short times, Phys. Chem. Chem. Phys. 16 (2014) 516–526.