JOINT BAYESIAN DECOMPOSITION OF A SPECTROSCOPIC SIGNAL SEQUENCE WITH RJMCMC

V. Mazet, S. Faisan

LSIIT UMR 7005, UDS, CNRS,
Bd Sébastien Brant, BP 10413,
67412 Illkirch Cedex, France
{vincent.mazet,faisan}@unistra.fr

A. Masson†, M.-A. Gaveau, L. Poisson, J.-M. Mestdagh

CNRS, CEA, IRAMIS, SPAM,
Laboratoire Francis Perrin, URA 2453,
91191 Gif-sur-Yvette, France
lionel.poisson@cea.fr

ABSTRACT

This article presents a method for decomposing a sequence of spectroscopic signals into a sum of peaks whose centers, amplitudes and widths are estimated. Since the peaks exhibit a slow evolution through the sequence, the decomposition is performed jointly on every spectra. To this end, we have developed a Bayesian model where a Markov random field favors a smooth evolution of the peaks through the sequence. The main contribution concerns the estimation of the peak number using the reversible jump MCMC algorithm. We show the accuracy of this approach on synthetic and real data.

***Index Terms*—** Decomposition of a sequence of spectroscopic signals; photoelectron spectrum; hierarchical Bayesian model; RJMCMC.

1. INTRODUCTION

We face with the problem of decomposing a sequence of spectra (*i.e.* spectroscopic signals that are typically acquired at different times in an experiment) into peaks. This leads to estimate the centers, amplitudes and widths of the peaks in each spectrum. Because the peaks exhibit various locations and shapes through the sequence, the problem cannot be seen as a source separation nor a spectral unmixing problem.

To our knowledge, nobody has addressed the problem of the decomposition of a sequence of spectra. Similar problems appears in time-frequency [1, 2] or seismic deconvolution [3] but the solutions proposed in these works cannot be directly applied. Yet, several methods have been proposed for decomposing a unique spectrum using deterministic methods [4], sparse approximation [5] or Bayesian approaches [6, 7]. Since the latters give very satisfactory results, we have developed a hierarchical Bayesian model whose posterior is simulated with a Monte Carlo Markov chain (MCMC) method [8].

*This work was supported by the CNRS (PEPS Rupture "SpectroDec").

†Now with EPFL, ISIC, LCPM, 1015 Lausanne, Switzerland.

We introduced in [9, 10] a hierarchical Bayesian model and showed that a sequential approach, in which each spectrum is decomposed independently from the others, is not appropriate. Indeed, since the peaks evolve slowly through the sequence, two contiguous spectra (a spectrum and the one acquired at the next or previous instant) are similar but a sequential approach may lead to two very different decompositions. On the contrary, a joint approach takes into account the slow evolution of the peaks, leading to coherent decompositions. Moreover, a decomposition is aided by the contiguous ones since they are favored to be similar, and a joint approach furnishes a classification of the peaks.

However, the approach proposed in [9, 10] considers that the peak number is known and constant through the sequence, which is obviously wrong on real data. So we extend this model by considering the peak number as a new random variable (section 2), and we use the reversible jump MCMC (RJMCMC) algorithm [8, 11] which handles problems with unknown dimension (section 3). The simulated annealing scheme used in [9, 10] appears now unnecessary. The performance of the method is assessed on synthetic and real photoelectron data (section 4).

2. HIERARCHICAL BAYESIAN MODEL

2.1. Model for the Data

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 Gaussian peaks $\boldsymbol{x}_s$ lying on an exponential baseline $\boldsymbol{z}_s$ and an additive noise $\boldsymbol{v}_s$ (physical justifications of this model can be found in [12]):

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

Since the peaks evolve slowly, we introduce the notion of "track" which corresponds to the association of similar peaks through the sequence. By definition, a track is composed of one peak per spectrum, with the constraint that all peaks of a track lie on contiguous spectra. The length of a track is defined as the number of peaks in the track. Thus, data contain

K tracks of length l_k each, whose first and last peaks are respectively located in spectra b_k and $b_k + l_k - 1$: the tracks are considered as the unknown "objects" whose number and length are also unknown. This makes the problem to be of unknown dimension since the number of peaks $M = \sum_k^K l_k$ has to be estimated.

Finally, the terms in equation 1 reads:

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

$$(\boldsymbol{z}_s)_n = \alpha_s \exp\left(-\frac{n}{\beta_s}\right)$$

where the subscript n denotes the element $n \in \{1, \ldots, N\}$ 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 track k in spectrum s. The mth peak of track k is parametrized by its center $c_{k,m}$, amplitude $a_{k,m}$ and width $w_{k,m}$; the baseline parameters are α_s and β_s.

2.2. Prior Distributions

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

$$p(\boldsymbol{v}|r_{\boldsymbol{v}}) = \frac{1}{(2\pi r_{\boldsymbol{v}})^{\frac{NS}{2}}} \exp\left(-\frac{\sum_{s=1}^{S} \|\boldsymbol{v}_s\|^2}{2r_{\boldsymbol{v}}}\right).$$

The track number K is drawn from a Poisson distribution as in [11] ($\mathbb{1}$ is the indicator function, $\xi \ll 1$):

$$p(K) = e^{-\xi} \frac{\xi^K}{K!} \mathbb{1}_{\mathbb{R}^+}(K).$$

The track length l_k is distributed according to a binomial distribution ($\zeta \ll 1$):

$$\forall k, \quad p(l_k) = \binom{S}{l_k} \zeta^{l_k} (1 - \zeta)^{S - l_k} \mathbb{1}_{\{0,\ldots,S\}}(l_k).$$

The track beginning b_k is drawn from a uniform distribution:

$$\forall k, \quad p(b_k|l_k) = \frac{1}{S - l_k + 1} \mathbb{1}_{\{1,\ldots,S-l_k+1\}}(b_k).$$

The peak centers $\boldsymbol{c}_k$ in track k are supposed to evolve slowly through the sequence, so they are modeled using a Gaussian Markov chain with hyperparameter $r_{\boldsymbol{c}}$. The center of the first peak in a track is distributed according to a uniform distribution $\mathcal{U}_{[1,N]}$ and all the centers are bounded to the interval $[1, N]$. Note that $[1, N]$ is large enough compared to the small values of $r_{\boldsymbol{c}}$, thus the effect of the truncation is negligible and the prior can be approximated by:

$$\forall k, \quad p(\boldsymbol{c}_k|r_{\boldsymbol{c}}) = p(c_{k,1}) \prod_{m=2}^{l_k} p(c_{k,m}|c_{k,m-1}, r_{\boldsymbol{c}}),$$

$$= \frac{1}{N-1} \frac{1}{(2\pi r_{\boldsymbol{c}})^{\frac{l_k-1}{2}}} \exp\left(-\frac{\|D\boldsymbol{c}_k\|^2}{2r_{\boldsymbol{c}}}\right) \mathbb{1}_{[1,N]}(\boldsymbol{c}_k).$$

where $\|\cdot\|$ denotes the L^2 norm and D is a first-order discrete derivative which promotes no evolution of the parameters.

For the same reasons, amplitudes $\boldsymbol{a}_k$, widths $\boldsymbol{w}_k$ ($\forall k$) and parameter $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are distributed from

$$p(\boldsymbol{a}_k|r_{\boldsymbol{a}}) = \frac{1}{a_{\max}} \frac{\exp\left(-\|D\boldsymbol{a}_k\|^2/2r_{\boldsymbol{a}}\right)}{(2\pi r_{\boldsymbol{a}})^{(l_k-1)/2}} \mathbb{1}_{[0,a_{\max}]}(\boldsymbol{a}_k),$$

$$p(\boldsymbol{w}_k|r_{\boldsymbol{w}}) = \frac{1}{w_{\max}} \frac{\exp\left(-\|D\boldsymbol{w}_k\|^2/2r_{\boldsymbol{w}}\right)}{(2\pi r_{\boldsymbol{w}})^{(l_k-1)/2}} \mathbb{1}_{[0,w_{\max}]}(\boldsymbol{w}_k),$$

$$p(\boldsymbol{\alpha}|r_{\boldsymbol{\alpha}}) = \frac{1}{A_{\max}} \frac{\exp\left(-\|D\boldsymbol{\alpha}\|^2/2r_{\boldsymbol{\alpha}}\right)}{(2\pi r_{\boldsymbol{\alpha}})^{(S-1)/2}} \mathbb{1}_{[0,A_{\max}]}(\boldsymbol{\alpha}),$$

$$p(\boldsymbol{\beta}|r_{\boldsymbol{\beta}}) = \frac{1}{B_{\max}} \frac{\exp\left(-\|D\boldsymbol{\beta}\|^2/2r_{\boldsymbol{\beta}}\right)}{(2\pi r_{\boldsymbol{\beta}})^{(S-1)/2}} \mathbb{1}_{[0,B_{\max}]}(\boldsymbol{\beta}).$$

where $a_{\max}$, $w_{\max}$, $A_{\max}$, $B_{\max}$ are the maximal allowed values for the parameters and $r_{\boldsymbol{a}}$, $r_{\boldsymbol{w}}$, $r_{\boldsymbol{\alpha}}$, $r_{\boldsymbol{\beta}}$ are hyperparameters controlling the prior strength.

Finally, a conjugate prior (inverse gamma) is chosen for the hyperparameters ($r_\bullet$ denotes $r_{\boldsymbol{v}}$, $r_{\boldsymbol{c}}$, $r_{\boldsymbol{a}}$, $r_{\boldsymbol{w}}$, $r_{\boldsymbol{\alpha}}$ or $r_{\boldsymbol{\beta}}$):

$$p(r_\bullet|\varepsilon) = \frac{\varepsilon^\varepsilon}{\Gamma(\varepsilon)} \frac{\exp\left(-\varepsilon/r_\bullet\right)}{r_\bullet^{\varepsilon+1}} \mathbb{1}_{\mathbb{R}^+}(r_\bullet).$$

whose parameter ε is small (typically lesser than 1) so as to favor smooth evolutions of the parameters.

2.3. Conditional Posterior Distributions

A Gibbs sampler is used in a move of the RJMCMC algorithm (section 3) for updating the peak and baseline parameters as well as the hyperparameters. It samples each variable according to its conditional posterior. The conditional posterior for peak centers $\boldsymbol{c}_k$ reads $\forall k$:

$$p(\boldsymbol{c}_k|...) \propto \exp\left(-\frac{\|D\boldsymbol{c}_k\|^2}{2r_{\boldsymbol{c}}} - \frac{\|\boldsymbol{y}_s - \boldsymbol{x}_s - \boldsymbol{z}_s\|^2}{2r_{\boldsymbol{v}}}\right) \mathbb{1}_{[1,N]}(\boldsymbol{c}_k). \tag{2}$$

The conditional posteriors for $\boldsymbol{a}_k$, $\boldsymbol{w}_k$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are very similar (they turn out to be normal distributions with positive support for $\boldsymbol{a}_k$ and $\boldsymbol{\alpha}$), and the ones for the hyperparameters are inverse Gamma distributions. The Metropolis-Hastings algorithm is used to sample $\boldsymbol{c}_k$, $\boldsymbol{w}_k$ and $\boldsymbol{\beta}$ and direct methods simulate the other posteriors.

3. ALGORITHM

As the posterior has not a closed form expression and the model dimension is not fixed, the RJMCMC algorithm [8, 11] has been chosen to explore the posterior over the different spaces. This is a Metropolis–Hastings algorithm designed to sample a distribution on a union of spaces of different dimensions. At each iteration, a move from the current state to a proposed state is randomly chosen. The dimension of the two states may be different. The proposed state is then accepted

according to an acceptance ratio ensuring invariance of the Markov chain with respect to the posterior distribution. We proposed the eight following moves:

- **B: Birth of a new track.** A new track with only one peak is proposed. The proposal distribution for the beginning of the proposed track and the center of the peak is proportional to the residual, that is, we propose a new peak with a high probability where the data is not well explained by the estimated peaks. The proposal for the width of the peak is an inverse Gamma distribution and the amplitude of the peak is drawn from a Gaussian whose parameters are computed from the residual and the noise variance.
- **D: Death of a track.** This is the dual move of the birth: a track with only one peak is proposed to be deleted.
- **M: Merge of two tracks.** Two contiguous tracks i and j (that is, if $b_i < b_j$, $b_i + l_i = b_j$) are proposed to be replaced by a unique track from b_i to $b_i + l_i + l_j - 1$.
- **S: Split of a track.** This is the dual move of the merge: a track (with several peaks) is divided in two contiguous tracks. The location of the division is uniformly chosen.
- **I: Increasing a track.** A peak is added at one of the extremity of a randomly chosen track. The proposal distribution for its center is proportional to the residual (of the spectrum where the new peak is proposed) weightened by a Gaussian whose mean equals the actual center of the extreme peak. The proposal distribution for its width is a Gaussian whose mean equals the actual width of the extreme peak. The amplitude of the peak is drawn in the same way as in move B. This move is equivalent to a birth of peak at the extremity of a track and then to a merge.
- **R: Reducing a track.** This is the dual move of the increasing: a peak at an extremity of a randomly chosen track is proposed to be deleted. This comes down to do a split followed by a death.
- **L: Relabelling the peaks.** This move consists in inverting some peaks of two tracks: the peaks of tracks i and j located between $s = s_1$ and $s = s_2$ are permuted; i, j, s_1 and s_2 being uniformly randomly chosen. This move is equivalent to a combination of split and merge moves.
- **U: Updating of the parameters.** Parameters $\boldsymbol{c}$, $\boldsymbol{a}$, $\boldsymbol{w}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and hyperparameters $r_{\boldsymbol{c}}$, $r_{\boldsymbol{a}}$, $r_{\boldsymbol{w}}$, $r_{\boldsymbol{\alpha}}$, $r_{\boldsymbol{\beta}}$, $r_{\boldsymbol{v}}$ are simulated according to their conditional posterior distributions. This is performed with a Gibbs sampling (see section 2.3).

The choice of these moves comes from heuristical facts, although moves B, D, M and S are usual in the literature [7, 11]. Even if these moves are sufficient to explore all the solution space, the introduction of moves I, R and L improves the convergence rate. Also, it is essential to perform moves L and U (which do not change the problem dimension) very often. We choose to set the probability of moves B and D to 0.01, the probability of moves S, M, I and R to 0.03 and the probability of moves L and U to 0.43. Finally, note that only one track with one peak is considered at initialization and that the maximum a posteriori (MAP) is estimated as the most probable generated sample.

4. RESULTS

4.1. Results on Synthetic Data

To quantify the benefits of the reversible jump scheme, the proposed approach is compared on a synthetic sequence with the approach presented in [9]. The two approaches have very similar modelling, but in [9] the track number is supposed to be known, each track gathers S peaks and a Gibbs sampler embedded in a simulated annealing scheme is used to get a MAP estimation. The two methods were run on a synthetic sequence with $S = 60$ spectra of length $N = 256$ and $K = 5$ tracks with approximately the same computation time ($\approx$ 25 min). The signal-to-noise ratio is about 24 dB. The Matlab code and data are available online at lsiit-miv.u-strasbg.fr/mazet/jointdec. The estimated parameters are shown in figure 1.

In the constant dimension approach, the track number has been set to its real value, namely 5. But some actual tracks have not a length of $S = 60$ so there are not always 5 real peaks per spectrum. Consequently, some estimated peaks do not correspond to real peaks and merely fit noise (they typically have non-relevant centers, small amplitudes and large widths, see left column in figure 1); some real peaks can be fitted with two estimated peaks (upper left corner of figure 1); and different tracks are fitted with peaks with the same label. In conclusion, the results obtained with the constant dimension approach can lead to misleading interpretations. On the opposite, results obtained with the proposed approach are very satisfactory: the number of track, their beginning and ending, and the parameters of each peak are well estimated. Note also that the proposed approach has shown better convergence properties, which can be explained by two reasons: (i) less parameters have often to be estimated (since tracks can be shorter than S) thus reducing the number of unknowns, and (ii) starting from a low dimension space and increasing slowly the number of tracks and peaks avoid being trapped in local minima that are hard to escape.

4.2. Results on Real Photoelectron Data

Photoelectrons refer to electrons emitted from matter as a consequence of the absorption of an electromagnetic radiation. Their distribution with respect to their energy is acquired at different times, providing a temporal sequence of spectra. The one shown in figure 2 gathers $S = 44$ spectra with $N = 181$ samples each (see [12] for details about the

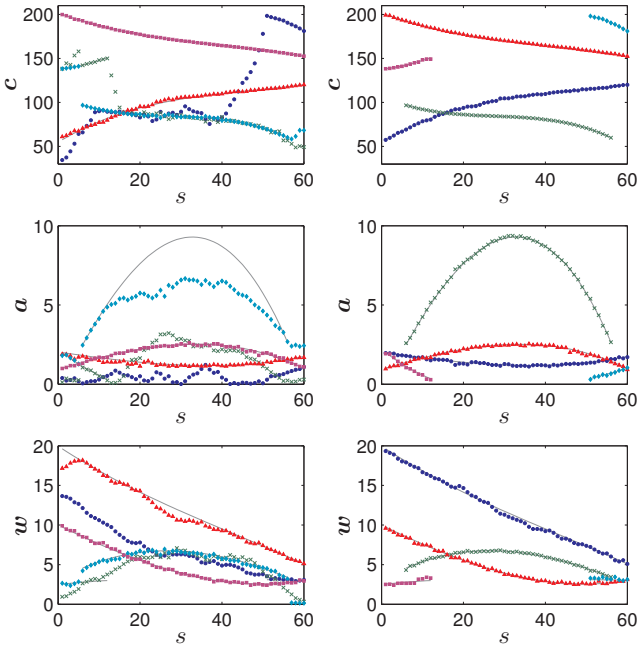

Fig. 1. Estimated parameters for a synthetic sequence. Left column: constant dimension approach of [9], right column: proposed approach. The gray and colored lines correspond respectively to the actual and estimated parameters.

experiment). The proposed approach ran with $I = 5 \cdot 10^5$ iterations (≈ 7 h). The major difference with the constant dimension model of [9] is that the proposed approach finds 5 tracks at short time (before 0.5 ps) whereas only 4 are needed at longer time. This is an important addition since it was known by the experimentalists that the short time dynamics is very complex with several competing processes, one of which disappears at longer time. In this framework, it is very stimulating that the model successfully detects such vanishing track. The estimation confirms quantitatively the qualitative observation of [12] that the energy (*i.e.* center) of a peak varies in time. Finally, the present analysis points out the contribution of a peak around 1.6 eV which could not be analyzed before.

5. CONCLUSION

We presented a new approach to estimate the peak parameters and their number in a sequence of spectroscopic signals. The whole sequence is decomposed jointly so as to deal with the peak evolution. The evolution has been modeled in a hierarchical Bayesian model using a Markov random field. An RJMCMC algorithm is implemented to jump between spaces with different dimension so as to estimate the track numbers and lengths. Simulations show the relevance of this scheme and the result obtained on real photoelectron data confirms the expert view.

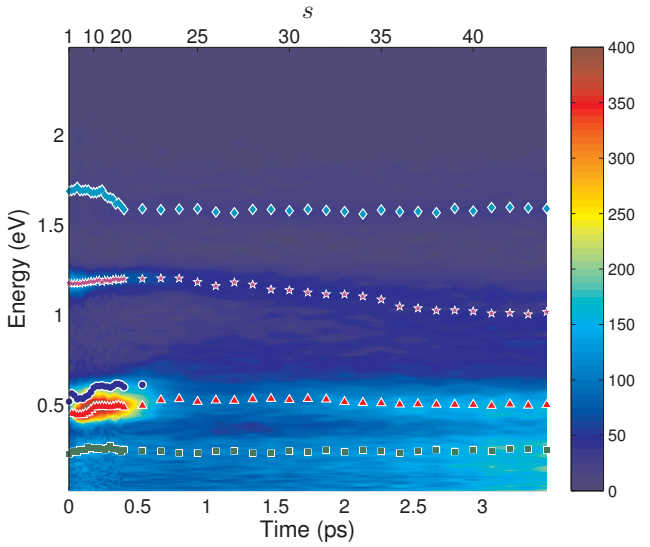

Fig. 2. Photoelectron data plotted as a function of the electron energy (n, vertical axis) and acquisition time (s, horizontal axis). The points are the peak centers estimated with the proposed approach.

6. REFERENCES

[1] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE T. Signal Proces.*, vol. 41, 1993.

[2] S.S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," Tech. Rep., Stanford University, 1995.

[3] J. Idier and Y. Goussard, "Markov modeling for Bayesian restoration of two-dimensional layered structures," *IEEE T. Inform. Theory*, vol. 39, 1993.

[4] R.A. Caruana, R.B. Searle, T. Heller, and S.I. Shupack, "Fast algorithm for the resolution of spectra," *Anal. Chem.*, 1986.

[5] D.L. Donoho and Y. Tsaig, "Fast solution of l1-norm minimization problems when the solution may be sparse," Tech. Rep., Stanford University, 2006.

[6] R. Fischer and V. Dose, "Analysis of mixtures in physical spectra," *Bayesian methods*, 2001.

[7] S. Gulam Razul, W.J. Fitzgerald, and C. Andrieu, "Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC," *Nucl. Instrum. Meth. A*, vol. 497, 2003.

[8] C.P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer, 2nd edition, 2004.

[9] V. Mazet, S. Faisan, A. Masson, M.-A. Gaveau, and L. Poisson, "Unsupervised joint bayesian decomposition of a sequence of photoelectron spectra," in *WHISPERS*, 2011.

[10] V. Mazet, "Joint Bayesian decomposition of a spectroscopic signal sequence," *IEEE Signal Proc. Let.*, vol. 18, 2011.

[11] P.J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, 1995.

[12] A. Masson, L. Poisson, M.-A. Gaveau, B. Soep, J.-M. Mestdagh, V. Mazet, and F. Spiegelman, "Dynamics of highly excited barium atoms deposited on large argon clusters. I. General trends," *J. Chem. Phys.*, vol. 133, 2010.