UNSUPERVISED JOINT BAYESIAN DECOMPOSITION OF A SEQUENCE OF PHOTOELECTRON SPECTRA

V. Mazet, S. Faisan

LSIIT
(University of Strasbourg, CNRS–UMR 7005),
Boulevard Sébastien Brant, BP 10413,
67412 Illkirch Cedex, France.
{vincent.mazet,faisan}@unistra.fr

A. Masson, M.-A. Gaveau, L. Poisson

Laboratoire Francis Perrin
(CNRS–URA 2453, CEA, IRAMIS),
Service des Photons Atomes et Molécules,
91191 Gif-sur-Yvette Cedex, France.
{firstname.lastname}@cea.fr

ABSTRACT

This article presents a method for decomposing a temporal sequence of photoelectron spectra into a parameter set reflecting the positions, amplitudes, and widths of the peaks. Since the peaks exhibit a slow evolution with time, we propose to take into account this temporal information by jointly decomposing the whole sequence. To this end, we have developed a Bayesian model where a Gaussian Markov random field favors a smooth evolution of the peaks. The approach is made completely unsupervised and a Gibbs sampler with simulated annealing algorithm enables to estimate the maximum a posteriori. We show the accuracy of this approach compared to a method in which the spectra are decomposed separately and present an application on real photoelectron data.

***Index Terms*—** Spectroscopic signal sequence decomposition, Bayesian inference, Markov chain Monte Carlo (MCMC) method, simulated annealing, photoelectron spectroscopy.

1. INTRODUCTION

Photoelectrons refer to electrons emitted from a sample of matter as a consequence of the absorption of an electromagnetic radiation. The distribution of the electrons with respect to their energy is the so-called photoelectron spectrum; it can be modeled as a sum of peaks (which informs on the electronic energy distribution of the system under consideration) superimposed on a continuum [1]. The system considered in section 5 is a single very excited barium atom (Ba) bound to a cluster of several hundreds of argon atoms (Ar). The electronic energy distribution of the Ba-Ar$_n$ system is followed as a function of time providing a temporal sequence of photoelectron spectra (i.e. a multispectral signal). An issue in this work is to determine whether the energy of the peaks varies as a function of the time, indicating that progressive solvation of barium in the cluster is at play.

We aim to decompose each spectrum of the sequence, that is to estimate the centers, amplitudes and widths of the peaks as well as the continuum. We emphasize that this problem is not a source separation or a spectral unmixing problem because the data cannot be modeled as a sum of source signals. This is clearly a decomposition problem since each spectrum is a sum of peaks whose locations and shape vary.

Several methods have been proposed for decomposing spectra [2, 3, 4] but, to our knowledge, nobody has addressed the problem of the decomposition of a sequence of spectra. We proposed recently in [5] such a method using a Bayesian framework and showed that a sequential approach, in which each spectrum is decomposed independently from the others, is not appropriate. Indeed, the decomposition of two contiguous spectra (i.e. recorded at two adjacent times) may lead to two very different decompositions whereas the spectra are quite similar. We also showed the interest of a joint decomposition, which takes into the fact that peaks evolve slowly with time: a decomposition is aided by the contiguous ones since they are favored to be similar; the decompositions are coherent because they smoothly evolve; and a joint decomposition furnishes a classification of the peaks, thus providing the possibility to follow the peaks through the sequence. However, the approach proposed in [5] has only been validated on synthetic data and several limitations appear when processing real photoelectron data.

Indeed, the real data gather more spectra than the synthetic data used in [5]: it leads to a significant increase of the search space size and the algorithm proposed in [5] fails. A solution is presented in section 3. Moreover, the continuum has not been modeled in [5], so we propose to model it with an exponential function (Eq. 1). Last but not least, the method proposed in [5] requires the tuning of the hyperparameters. Even if some arguments may help the user to set them, the variability of real signals makes the choice difficult. That is why the method proposed in this article is fully unsupervised.

The model and the priors are detailed in section 2. The Maximum A Posteriori (MAP) estimator is obtained using a Markov chain Monte Carlo (MCMC) simulated annealing method (section 3). Finally, the performance of the method is assessed on synthetic (section 4) and real data (section 5).

2. BAYESIAN MODEL

Data are a sequence of S spectra with N samples each. A spectrum $\boldsymbol{y}_s$ ($s \in \{1,\dots,S\}$) is modeled as a noisy sum of K Gaussian peaks lying on an exponential baseline [1]:

$$(\boldsymbol{y}_s)_n = \sum_{k=1}^{K} a_{s,k} \exp\left(-\frac{(n-c_{s,k})^2}{2w_{s,k}^2}\right) + \alpha_s e^{-n/\beta_s} + (\boldsymbol{b}_s)_n \tag{1}$$

where $(\boldsymbol{y}_s)_n$ denotes the element $n \in \{1,\dots,N\}$ in vector $\boldsymbol{y}_s$. The peak parameters are its center $c_{s,k}$, amplitude $a_{s,k}$ and width $w_{s,k}$; the baseline parameters are α_s and β_s; $\boldsymbol{b}_s$ is an additive noise modeling measurement errors and model uncertainties. Note also that the peak number K is supposed to be known and the same for each spectrum. Besides, we introduce a label $z_{s,k} \in \{1,\dots,K\}$ for each peak so as to proceed to the classification of the peaks and to regularize their evolution.

For convenience, we will note in the sequel $\boldsymbol{\theta}$ one of the $3K+2$ following vectors:

$$\boldsymbol{c}^1,\dots,\boldsymbol{c}^K,\boldsymbol{a}^1,\dots,\boldsymbol{a}^K,\boldsymbol{w}^1,\dots,\boldsymbol{w}^K,\boldsymbol{\alpha},\text{ or }\boldsymbol{\beta},$$

where the superscripts denote the parameters with the same label, e.g. $\boldsymbol{c}^l = \{c_{s,k} | z_{s,k} = l, \forall s, \forall k\}$.

2.1. Priors

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

$$\forall s, n, \quad (\boldsymbol{b}_s)_n | r_{\boldsymbol{b}} \sim \mathcal{N}(0, r_{\boldsymbol{b}}). \tag{2}$$

Because each component of the parameter $\boldsymbol{\theta}$ is supposed to evolve slowly through the sequence, a smoothness prior helps to regularize the solution: this comes down to use a (constrained) Gaussian Markov random field (GMRF) model [6]:

$$p(\boldsymbol{\theta}|\boldsymbol{z}) \propto \frac{1}{r_{\boldsymbol{\theta}}^{S/2}} \exp\left(-\frac{1}{2r_{\boldsymbol{\theta}}}\|D\boldsymbol{\theta}\|^2\right) \mathbb{1}_{\Theta}(\boldsymbol{\theta}). \tag{3}$$

The hyperparameter $r_{\boldsymbol{\theta}}$ controls the prior strength, it stands for one of the 5 hyperparameters related to the vectors represented by $\boldsymbol{\theta}$: $r_{\boldsymbol{c}}$, $r_{\boldsymbol{a}}$, $r_{\boldsymbol{w}}$, $r_{\boldsymbol{\alpha}}$ or $r_{\boldsymbol{\beta}}$. The L^2 norm is denoted by $\|\cdot\|$. D is a difference operator (discrete derivative), a first-order derivative promotes no evolution:

$$D\boldsymbol{\theta} = (\theta_{s-1} - \theta_s)_{s\in\{2,\dots,S\}} \tag{4}$$

while a second-order derivative promotes a linear evolution:

$$D\boldsymbol{\theta} = (\theta_{s-1} - 2\theta_s + \theta_{s+1})_{s\in\{2,\dots,S-1\}}. \tag{5}$$

One has to choose D with respect to the expected peak evolution in the data, in the same way as the regularization matrix in a Tikhonov criterion. Finally, $\mathbb{1}_{\Theta}(\boldsymbol{\theta})$ is the indicator function: $\mathbb{1}_{\Theta}(\boldsymbol{\theta}) = 1$ if $\boldsymbol{\theta} \in \Theta^S$, 0 otherwise. It is necessary to impose bound limits on the parameters. The solution space Θ for the peak centers is $[1, N]$ and it is $\mathbb{R}^+$ for the other parameters. 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 real probability density function $p(\boldsymbol{c}^l|\boldsymbol{z})$ ($l \in \{1,\dots,K\}$) can be approximated by the expression in Eq. (3). Lastly, each parameter $\boldsymbol{\theta}$ is considered to be independent of the others (for example, $\boldsymbol{c}^1$ and $\boldsymbol{c}^2$ are independent).

A priori, there is no information available on the labels. So they are distributed according to a uniform distribution:

$$\forall s, \qquad \boldsymbol{z}_s \sim \mathcal{U}_{\mathcal{S}_K} \tag{6}$$

where $\mathcal{S}_K$ is the permutation group of the set $\{1,\dots,K\}$.

A common choice for the noise variance ($r_{\boldsymbol{b}}$) prior is the Jeffreys distribution which is unfortunately non-proper. Therefore, we adopt an inverse gamma prior for $r_{\boldsymbol{b}}$:

$$r_{\boldsymbol{b}} \sim \mathcal{IG}(\epsilon, \epsilon) \tag{7}$$

whose parameter ϵ is close to zero which is in fact the limit case corresponding to the Jeffreys prior.

A conjugate inverse-gamma prior is chosen for the 5 last hyperparameters ($r_{\boldsymbol{\theta}}$). Note that three of these hyperparameters ($r_{\boldsymbol{c}}$, $r_{\boldsymbol{a}}$ and $r_{\boldsymbol{w}}$) were set manually in [5] and two others ($r_{\boldsymbol{\alpha}}$ and $r_{\boldsymbol{\beta}}$) were not used at all. Since they have to be as small as possible so as to get a smooth evolution of the parameters, a prior similar to the noise variance is used:

$$r_{\boldsymbol{\theta}} \sim \mathcal{IG}(\epsilon, \epsilon). \tag{8}$$

2.2. Posteriors

The global posterior is sampled using a Gibbs sampler (section 3) which consists in simulating each variable according to its conditional posterior distribution [7]. The latter are briefly presented below (see [5] for a detailed presentation).

The centers, widths and β_s are distributed according to

$$c_{s,k}|\dots \sim \exp\left(-\frac{\|\boldsymbol{e}_s\|^2}{2r_{\boldsymbol{b}}} - \frac{\|D\boldsymbol{c}^l\|^2}{2r_{\boldsymbol{c}}}\right) \mathbb{1}_{[1,N]}(c_{s,k}), \tag{9}$$

$$w_{s,k}|\dots \sim \exp\left(-\frac{\|\boldsymbol{e}_s\|^2}{2r_{\boldsymbol{b}}} - \frac{\|D\boldsymbol{w}^l\|^2}{2r_{\boldsymbol{w}}}\right) \mathbb{1}_{\mathbb{R}^+}(w_{s,k}), \tag{10}$$

$$\beta_s|\dots \sim \exp\left(-\frac{\|\boldsymbol{e}_s\|^2}{2r_{\boldsymbol{b}}} - \frac{\|D\boldsymbol{\beta}\|^2}{2r_{\boldsymbol{\beta}}}\right) \mathbb{1}_{\mathbb{R}^+}(\beta_s) \tag{11}$$

with $l = z_{s,k}$ and

$$(\boldsymbol{e}_s)_n = (\boldsymbol{y}_s)_n - \sum_{k=1}^{K} a_{s,k} \exp\left(-\frac{(n-c_{s,k})^2}{2w_{s,k}^2}\right) - \alpha_s e^{-n/\beta_s}. \tag{12}$$

It can be shown that the amplitude posterior is a Gaussian distribution with positive support:

$$a_{s,k}|\cdots \sim \exp\left(-\frac{(a_{s,k}-\mu_{s,k})^2}{2\rho_{s,k}}\right)\mathbb{1}_{\mathbb{R}^+(a_{s,k})} \quad (13)$$

where $\mu_{s,k}$ and $\rho_{s,k}$ depend on $c_{s,k}$, $w_{s,k}$, $r_{\boldsymbol{b}}$, $r_{\boldsymbol{a}}$, $\boldsymbol{e}_s$ and on some other amplitudes [5].

Similarly, the conditional posterior of α_s is a Gaussian distribution with positive support:

$$\alpha_s|\cdots \sim \exp\left(-\frac{(\alpha_s-\lambda_s)^2}{2\nu_s}\right)\mathbb{1}_{\mathbb{R}^+}(\alpha_s). \quad (14)$$

The labels are distributed according to

$$\boldsymbol{z}|\cdots \sim \prod_s \exp\left[-\sum_l\left[\frac{\|D\boldsymbol{c}^l\|^2}{2r_{\boldsymbol{c}}}+\frac{\|D\boldsymbol{a}^l\|^2}{2r_{\boldsymbol{a}}}+\frac{\|D\boldsymbol{w}^l\|^2}{2r_{\boldsymbol{w}}}\right]\right]\mathbb{1}_{\mathcal{S}_K}(\boldsymbol{z}_s). \quad (15)$$

Finally, the hyperparameter posteriors are:

$$r_{\boldsymbol{b}}|\ldots \sim \mathcal{IG}\left(NS/2+\epsilon, \textstyle\sum_l\|\boldsymbol{e}_s\|^2/2+\epsilon\right), \quad (16)$$

$$r_{\boldsymbol{c}}|\ldots \sim \mathcal{IG}\left(KS/2+\epsilon, \textstyle\sum_l\|D\boldsymbol{c}^l\|^2/2+\epsilon\right), \quad (17)$$

$$r_{\boldsymbol{a}}|\ldots \sim \mathcal{IG}\left(KS/2+\epsilon, \textstyle\sum_l\|D\boldsymbol{a}^l\|^2/2+\epsilon\right), \quad (18)$$

$$r_{\boldsymbol{w}}|\ldots \sim \mathcal{IG}\left(KS/2+\epsilon, \textstyle\sum_l\|D\boldsymbol{w}^l\|^2/2+\epsilon\right), \quad (19)$$

$$r_{\boldsymbol{\alpha}}|\ldots \sim \mathcal{IG}\left(S/2+\epsilon, \|D\alpha\|^2/2+\epsilon\right), \quad (20)$$

$$r_{\boldsymbol{\beta}}|\ldots \sim \mathcal{IG}\left(S/2+\epsilon, \|D\beta\|^2/2+\epsilon\right). \quad (21)$$

3. MCMC SIMULATED ANNEALING ALGORITHM

The posterior is sampled using a Gibbs sampler because spectrum decomposition is usually processed with MCMC algorithms [2, 3, 4] and the Gibbs sampler is a common solution when dealing with GMRF. The Gibbs sampler requires to simulate each variable according to its conditional posterior. Parameters $c_{s,k}$, $w_{s,k}$ and β_s are sampled using a random walk Metropolis-Hastings algorithm [7]. Parameters $a_{s,k}$, α_s and the hyperparameters are sampled directly. Lastly the labels are sampled using the following symmetric Metropolis-Hastings algorithm: the proposed labels are identical to the actual ones, except that the labels k_1 and k_2 between $s=s_1$ and $s=s_2$ are permuted; k_1, k_2, s_1 and s_2 being uniformly randomly chosen and 100 candidates are proposed at each iteration of the Gibbs sampler.

In practice, the Gibbs sampler can be trapped in a local minimum [7], so it is included within a simulated annealing scheme [8, 9]. In this way, the MAP estimate is straightforward and corresponds to the sample of the Markov chain which maximizes the posterior. A geometric cooling schedule is used where the initial and final temperatures are set respectively to 10 and 0.1. Note that the simulated annealing is not implemented on the variables $a_{s,k}$ and α_s as well as the hyperparameters. The main reason is that simulated annealing on these variables introduces numerical problems and divergences. Also, this approach can be seen as a simulated annealing method with different temperatures as in [10].

4. SIMULATION RESULTS ON SYNTHETIC DATA

To quantify the proposed method benefits, we compare the proposed method (UJD: unsupervised joint decomposition) with the sequential decomposition (SD) and the supervised joint decomposition (SJD) both presented in [5]. These two methods have been enriched to model the continuum but they are still supervised (the hyperparameters have been set according to [5]). Even if the three decompositions handle the same model (Eq.1), the SD cannot handle the label estimation nor the smooth evolution of the peaks and the labels in the SJD are sampled as in [5]. For each method, a MCMC simulated annealing algorithm with 5000 iterations is used to estimate the MAP solution. We observed that using more iterations did not lead to a significant improvement. The three methods were run on 30 synthetic sequences with $S=10$, $K=4$ and D is the second-order derivate. The Matlab code and data are available on-line at lsiit-miv.u-strasbg.fr/mazet/jointdec. The performances are compared by computing the mean square error (MSE) between the observed signals and the reconstructed ones.

The SD and the UJD achieve respectively a MSE equal to $11.54\cdot10^{-2}$ and $3.78\cdot10^{-2}$. One can explain these results by the fact that the SD may miss some peaks while the smoothness prior of the UJD favors the search for peaks on certain zones in the solution space: in other terms, the estimation obtained for a spectrum is used to help the estimation of a contiguous spectrum.

The UDJ also outperforms the SJD (whose MSE equals $7.40\cdot10^{-2}$). This is due to the fact that setting the hyperparameters is not easy. Note also that the hyperparameters control the relative importance of the regularization terms against the data-dependent terms. This means that they have to be modified according to the processed sequence depending on its signal-to-noise ratio and on the peak evolutions. This clearly highlights the benefit of the unsupervised proposed method.

5. RESULTS ON PHOTOELECTRON SPECTRA

The photoelectron data gathers $S=44$ spectra (covering a delay time from 0 ps to 3.47 ps) with $N=182$ samples each (from 0.02 eV to 2.52 eV). The sampling in both time (s) and energy (n) is irregular. The proposed approach was run with $K=4$ peaks, $I=10^4$ iterations, and D is the first-order derivate (promoting no evolution). The estimated parameters are shown Fig. 1. Their evolution is quite smooth and the peaks are well classified, that is, there is no switch between peaks. This validates the presented label sampling because due to the high value of S, the method proposed in [5] does not provide good results. Also, it appears that the evolutions of the peaks are smoother than the one obtained after a sequential decomposition as it was implemented in [1] (not shown). The estimation points out that one peak is shifting

in energy: its energy increases slightly from 1.18 to 1.20 eV between 0 and 0.5 ps, then decreases, and reaches 1.11 eV at 2 ps. After 2 ps, this peak has a very low amplitude and a very high width: this is typical of a peak that disappears by vanishing into the background. The estimation confirms quantitatively the qualitative observation of [1] that the energy of one of the electronic levels of the Ba-Ar$_n$ system varies in time. This suggests that the corresponding energy level is very sensitive to the argon environment about barium and that barium solvates progressively into the argon cluster.

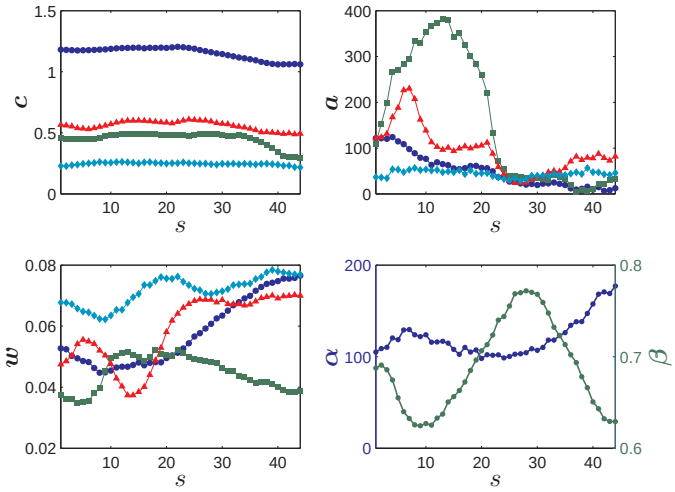

Fig. 1. Estimated parameters: centers (top-left), amplitudes (top-right) and widths (bottom-left) of the peaks; parameters α and β of the baseline (bottom-right).

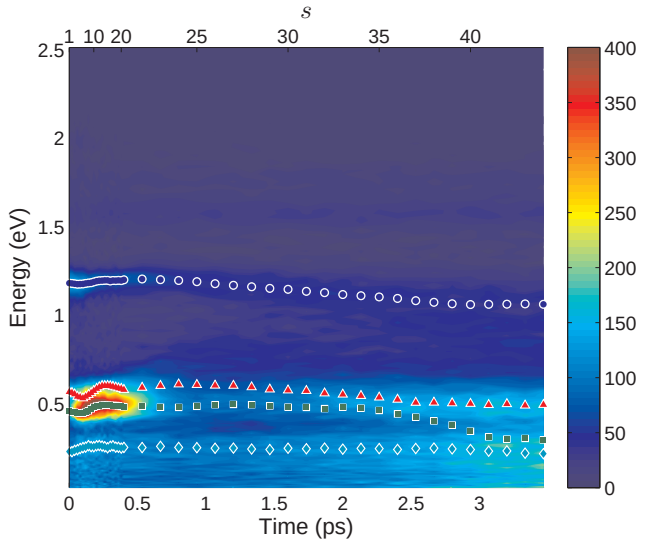

Fig. 2. Photoelectron data: the electron signal (false colors) is plotted as a function of the electron energy (n, vertical axis) and delay time (s, horizontal axis). The points are the estimated centers of the peaks.

6. CONCLUSION

We presented a new approach to estimate the peak parameters in a sequence of spectroscopic signals. The key idea is to achieve a joint decomposition by processing the whole sequence instead of each spectrum separately. This is done by using a Bayesian model, with a GMRF favoring the smoothness of the solution. The approach has been made completely unsupervised by using conjugate priors on the hyperparameters, and the solution is estimated using a MCMC simulated annealing algorithm. Simulations show the relevance of this scheme and the result obtained on real photoelectron data confirms the expert view. Future works will be mainly dedicated to the case where the peak number varies.

7. REFERENCES

[1] 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," *The Journal of Chemical Physics*, vol. 133, no. 5 (054307), 2010.

[2] R. Fischer and V. Dose, "Analysis of mixtures in physical spectra," in *Bayesian methods*, 2001, pp. 145–154.

[3] S. Gulam Razul, W.J. Fitzgerald, and C. Andrieu, "Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC," *Nuclear Instruments and Methods in Physics Research A*, vol. 497, pp. 492–510, 2003.

[4] N.M. Haan and S.J. Godsill, "Bayesian models for DNA sequencing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[5] V. Mazet, "Joint Bayesian decomposition of a spectroscopic signal sequence," *IEEE Signal Processing Letters*, vol. 18, no. 3, pp. 191–184, 2011.

[6] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

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

[8] S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[10] B. Perret, V. Mazet, C. Collet, and É. Slezak, "Hierarchical multispectral galaxy decomposition using a MCMC algorithm with multiple temperature simulated annealing," *Pattern Recognition*, vol. 44, no. 6, pp. 1328–1342, 2011.