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.