DECOMPOSITION OF A CHEMICAL SPECTRUM USING A MARKED POINT PROCESS AND A CONSTANT DIMENSION MODEL

Vincent Mazet[1], David Brie[1], Jérôme Idier[2]
(1) CRAN (UHP, CNRS UMR 7039), BP 239,
54506 Vandœuvre-lès-Nancy Cedex, France
(e-mail: firstname.lastname@cran.uhp-nancy.fr)
(2) IRCCyN (ÉCN, CNRS UMR 6597),
1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3, France
(e-mail: jerome.idier@irccyn.ec-nantes.fr)

Abstract

We consider the problem of estimating the peak parameters in a chemical spectrum, *i.e.* their locations, amplitudes and widths. The spectrum $\mathbf{y}$ is modelled as a noisy sum of K positive Lorentzian peaks $\boldsymbol{f}$: $\mathbf{y} = \sum_{k=1}^{K} \boldsymbol{f}(\mathbf{t}_k, \mathbf{a}_k, \mathbf{s}_k) + \mathbf{e}$ where $\mathbf{t}_k$, $\mathbf{a}_k$ and $\mathbf{s}_k$ stand respectively for the location, amplitude and width of the peak k, and $\mathbf{e}$ denotes the noise and model errors. A non-supervised Bayesian approach coupled with MCMC methods is retained to solve the problem.

A marked point process provides a suitable representation for this phenomenon: it is a a finite set of objects (*i.e.* a configuration of points with some marks) lying in a bounded space, corresponding in our application to the observation space while the objects model the peaks, characterized by their locations and marks (amplitude and widths). A stochastic model for these quantities is then proposed.

But the peak number is also unknown. Numerous MCMC methods for model uncertainty have been proposed, such as the RJMCMC algorithm. Nevertheless, we propose in this paper an approach in which the dimension model is constant. Thus, the use of a Gibbs sampler is possible and natural due to the hierarchical structure of the model. The idea consists in considering an upper bound for peak number and modelling the peak occurrence by a Bernoulli distribution. However, the estimation is not straightforward because of the label switching phenomenon; we then propose a label switching method adapted to the proposed approach.

In conclusion, this approach performs better than a classical deconvolution approach where the peaks have inevitably the same width. Moreover, the input is generally modelled as a Bernoulli-Gaussian process of N points (N being the length of $\mathbf{y}$) though we consider only a signal with length K, and, obviously, $K < N$ (the peak number is less than the signal length). Therefore, there is less variables to estimate, so the estimation is better and the method faster.