Bayesian orthogonal component analysis for sparse representation. Extension to non-homogeneous sparsity level over times

Nicolas Dobigeon and Jean-Yves Tourneret
E-mail : {Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr

TECHNICAL REPORT – 2009, November

University of Toulouse, IRIT/INP-ENSEEIHT/TéSA
2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

Abstract

In this technical note, an extension of the work detailed in [1] is presented. More precisely, one allows for different sparsity levels over times, i.e., the probability of having an active component source to be dependent on t.

I. Problem formulation

Let $\boldsymbol{x}(t) = [x_1(t), \ldots, x_M(t)]^T$ denote measurement vectors of $\mathbb{R}^M$ observed at time instants $t = 1, \ldots, T$ by M sensors. These observations are assumed to be related to $N < M$ unobserved sources denoted $\boldsymbol{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ via the matrix $\boldsymbol{\Psi}$ in the following noisy linear model

$$\boldsymbol{x}(t) = \boldsymbol{\Psi}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{1}$$

where $\boldsymbol{n}(t)$ stands for an additive measurement noise. Standard matrix notations yield

$$\boldsymbol{X} = \boldsymbol{\Psi}\boldsymbol{S} + \boldsymbol{N} \tag{2}$$

with $\boldsymbol{X} = [\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)]$, $\boldsymbol{S} = [\boldsymbol{s}(1), \ldots, \boldsymbol{s}(T)]$ and $\boldsymbol{N} = [\boldsymbol{n}(1), \ldots, \boldsymbol{n}(T)]$. The $M \times 1$ noise vectors $\boldsymbol{n}(t)$ ($t = 1, \ldots, T$) are assumed to be independent and distributed according to a centered multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}_M, \sigma^2 \mathbf{I}_M)$.

In this work, the $M \times N$ matrix $\boldsymbol{\Psi}$ is assumed to be an unknown orthogonal matrix

$$\boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{3}$$

where the sources to be recovered can be sparsely represented. Consequently, since only a few sources are assumed to be active at time index t, the unobserved vector of N sources $\boldsymbol{s}(t)$ is sparse and contains only a few components that are non-zero.

II. Bayesian model

The prior distribution of the source component is the following Bernoulli-Gaussian prior

$$f\left(s_n(t)|\lambda(t), a_n^2\right) = (1-\lambda(t))\,\delta\left(s_n(t)\right) + \lambda(t) g_{a_n^2}\left(s_n(t)\right). \tag{4}$$

By assigning independent uniform distributions over $(0,1)$ as priors for the probabilities $\lambda(t)$ ($t = 1, \ldots, T$), the conditional posterior distribution of $\lambda(t)$ becomes

$$f\left(\lambda(t)|\boldsymbol{s}(t)\right) \propto (1-\lambda(t))^{n_0(t)}\,\lambda(t)^{n_1(t)} \tag{5}$$

with $n_1(t) = \sum_{n=1}^{N} q_n(t)$ and $n_0(t) = N - n_1(t)$. Consequently, in the Gibbs algorithm, sampling according to $f\left(\lambda(t)|\boldsymbol{s}(t), \boldsymbol{\Psi}, \sigma^2, \boldsymbol{X}\right)$ consists of drawing samples according to the following Beta distribution

$$\lambda(t)|\boldsymbol{s}(t), \boldsymbol{\Psi}, \boldsymbol{X} \sim \mathcal{B}e\left(n_0(t)+1, n_1(t)+1\right). \tag{6}$$

Of course, with such model, the number of sources N should be sufficient to obtain accurate estimates of the probabilities $\lambda(t)$. Indeed, we can notice that

$$\mathrm{E}\left[\lambda(t)|\boldsymbol{s}(t)\right] = \frac{n_1(t)+1}{n_1(t)+n_0(t)+2} = \frac{\sum_{n=1}^{N} q_n(t)+1}{N+2}. \tag{7}$$

Finally, the prior model defined by (4) and (5) provides the following the conditional

$$\begin{aligned} f\left(\boldsymbol{S}, \boldsymbol{\Psi}|\boldsymbol{X}\right) \propto & \frac{\prod_{t=1}^{T} B\left(1+n_1(t), 1+n_0(t)\right)}{\left[\sum_{t=1}^{T} \|\boldsymbol{x}(t) - \boldsymbol{\Psi}\boldsymbol{s}(t)\|^2\right]^{\frac{TM}{2}}} \\ & \times \frac{\Gamma\left(\frac{\alpha_0}{2} + \frac{1}{2}\sum_{t=1}^{T} n_1(t)\right)}{\left[\frac{\alpha_1}{2} + \frac{1}{2}\sum_{t=1}^{T}\|\boldsymbol{s}(t)\|^2\right]^{\frac{\alpha_0}{2}+\frac{1}{2}\sum_{t=1}^{T} n_1(t)}}. \end{aligned} \tag{8}$$

III. Simulation results

Some simulations have been conducted based on this model. $N = 8$ sources have been generated of length $T = 100$ according to the prior (4) with the active source variance $a^2 = 100$ and the following active source probabilities: $\lambda(t) = 0.85$ for $t = 1, \ldots, 64$ and $\lambda(t) = 0.01$ for $t = 65, \ldots, 100$. The MMSE estimates of the probabilities $\lambda(1), \ldots, \lambda(100)$ provided by the modified algorithm are depicted in Fig. 1. These results are in good agreement with the actual values of the probabilities.

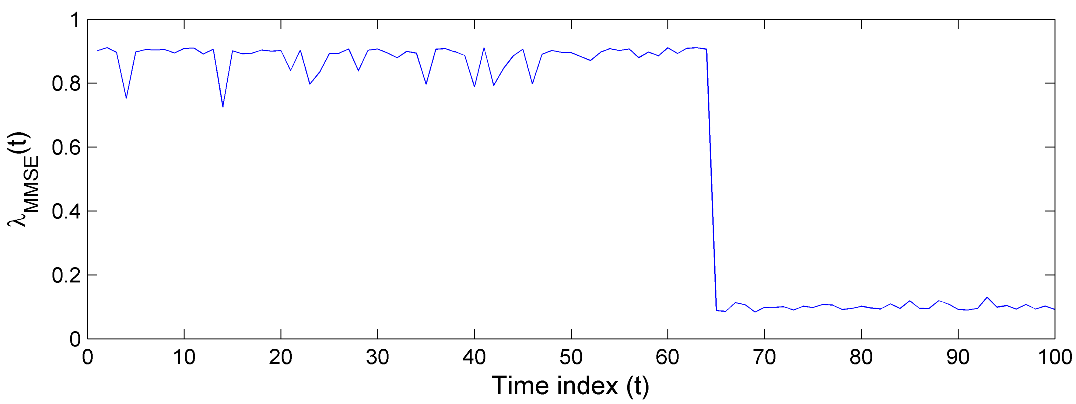

Fig. 1. MMSE estimates of the probabilities $\lambda(1), \ldots, \lambda(100)$.

References

[1] N. Dobigeon and J.-Y. Tourneret, “Bayesian orthogonal component analysis for sparse representation,” *IEEE Trans. Signal Process.*, 2009, submitted. [Online]. Available: http://arxiv.org/abs/0908.4489