

Bayesian sparse sources separation

Ali Mohammad-Djafari

Laboratoire des signaux et systmes (L2S)

UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD

SUPELEC, Plateau de Moulon, 91192 Gif-sur-Yvette, France

Email: djafari@lss.supelec.fr http://djafari.free.fr

Abstract—In this paper, first a Bayesian estimation approach is proposed for sources separation which is viewed as an inference problem where we want to estimate the mixing matrix, the sources and all the hyperparameters associated to modeling (likelihood and priors). Then, the sources separation problem is considered in three steps: i) Estimation of the sources when the mixing matrix is known; ii) Estimation of the mixing matrix when the sources are known; and iii) Joint estimation of sources and the mixing matrix. In each of these cases, we consider also the cases where we do not know the hyperparameters and we want to estimate them too. In all cases, one of the main steps is modeling of sources and the mixing matrix prior laws. We propose to use sparsity enforcing probability laws (such as Generalized Gaussians, Student-t, Elastic net and mixture models) both for the sources and the mixing matrix. For algorithmic and computational aspects, we consider either Joint MAP, MCMC Gibbs sampling and Variational Bayesian Approximation tools.

I. INTRODUCTION

The general sources separation problem can be viewed as an inference problem where first we provide a model linking the observed data (mixed signals) $\mathbf{g}(t)$ to unknown sources $\mathbf{f}(t)$ through a forward model. In this paper, we only consider the instantaneous mixing model:

$$\mathbf{g}(t) = \mathbf{A}\mathbf{f}(t) + \boldsymbol{\epsilon}(t), \quad t \in [1, \dots, T] \quad (1)$$

where $\boldsymbol{\epsilon}(t)$ represents the errors of modelling and measurement. \mathbf{A} is called mixing matrix and when it is invertible, its inverse $\mathbf{B} = \mathbf{A}^{-1}$ is called the separating matrix. The second step is to write down the expression of the joint posterior law:

$$p(\mathbf{f}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}, \mathbf{A}, \boldsymbol{\theta}_1) p(\mathbf{f}|\boldsymbol{\theta}_2) p(\mathbf{A}|\boldsymbol{\theta}_3) p(\boldsymbol{\theta})}{p(\mathbf{g})} \quad (2)$$

where $p(\mathbf{g}|\mathbf{f}, \mathbf{A}, \boldsymbol{\theta}_1)$ is the likelihood and $p(\mathbf{f}|\boldsymbol{\theta}_2)$ and $p(\mathbf{A}|\boldsymbol{\theta}_3)$ are the priors on sources and the mixing matrix, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ represent the hyperparameters of the problem and $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_1) p(\boldsymbol{\theta}_2) p(\boldsymbol{\theta}_3)$ their associated prior laws.

In this paper, we will consider different prior modelling for sources $p(\mathbf{f}|\boldsymbol{\theta}_2)$ and different priors for the mixing matrix $p(\mathbf{A}|\boldsymbol{\theta}_3)$ and use conjugate priors for $p(\boldsymbol{\theta})$. In particular, we consider the Generalized Gaussian (GG), Student-t (St), Elastic net (EN) and Mixture of Gaussians (MoG) models. Some of these models are well-known [?], [?], [?], [?], [?], [?], [?], some others less. In general, we can classify them in two categories: i) Simple Non Gaussian models with heavy tails and ii) Mixture models with hidden variables \mathbf{z} which result to hierarchical models.

The second main step in the Bayesian approach is to do the computations. The Bayesian computations in general can be:

- Joint optimization of $p(\mathbf{f}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{g})$ which needs optimisation algorithms;
- MCMC Gibbs sampling methods which need generation of samples from the conditionals $p(\mathbf{f}|\mathbf{A}, \boldsymbol{\theta}, \mathbf{g})$, $p(\mathbf{A}|\mathbf{f}, \boldsymbol{\theta}, \mathbf{g})$ and $p(\boldsymbol{\theta}|\mathbf{f}, \mathbf{A}, \mathbf{g})$;
- Bayesian Variational Approximation (BVA) methods which approximate $p(\mathbf{f}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{g})$ by a separable

$$q(\mathbf{f}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{g}) = q_1(\mathbf{f}|\tilde{\mathbf{A}}, \tilde{\boldsymbol{\theta}}, \mathbf{g}) q_2(\mathbf{A}|\tilde{\mathbf{f}}, \tilde{\boldsymbol{\theta}}, \mathbf{g}) q_3(\boldsymbol{\theta}|\tilde{\mathbf{f}}, \tilde{\mathbf{A}}, \mathbf{g})$$

one and then using them for the estimation.

The rest of the final paper will be organized as follows: In section II, we review a few prior models which are frequently used in particular when sparsity has to be enforced [?] and select a few most important ones such as the Generalized Gaussian (GG) with two particular cases of Gaussian (G) and Double Exponential (DE) or Laplace, the Student-t model which can be interpreted as an infinite mixture with a variance hidden variable, Elastic net and the mixture model. In Section III, first we examine in details the estimation of the sources \mathbf{f} when the mixing matrix \mathbf{A} is known, then the estimation of the mixing matrix \mathbf{A} when the sources \mathbf{f} are known, then the joint estimation of the mixing matrix \mathbf{A} and the sources \mathbf{f} , and finally, the more realistic case of joint estimation of the mixing matrix \mathbf{A} , the sources \mathbf{f} , their hidden variables \mathbf{z} and the hyperparameters $\boldsymbol{\theta}$. In section IV, we give principal practical algorithms which can be used in real applications, and finally, in section V we show some results and real applications.