

Bayesian Source Separation: Beyond PCA and ICA

Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes,
Unité mixte de recherche 8506 (CNRS-Supélec-UPS)
Supélec, Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France.

Abstract. Blind source separation (BSS) has become one of the major signal and image processing area in many applications. Principal component analysis (PCA) and Independent component analysis (ICA) have become two main classical approaches for this problem. However, these two approaches have their limits which are mainly, the assumptions that the data are temporally iid and that the model is exact (no noise). In this paper, we first show that the Bayesian inference framework gives the possibility to go beyond these limits while obtaining PCA and ICA algorithms as particular cases. Then, we propose different a priori models for sources which progressively account for different properties of the sources. Finally, we illustrate the application of these different models in spectrometry, in astrophysical imaging, in satellite imaging and in hyperspectral imaging.

1 Introduction

Even if PCA and ICA have their origin in data analysis [1], they have become two main classical approaches for blind source separation problem which can be presented as :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\epsilon}(t), \quad (1)$$

where $\mathbf{x}(t) = [x_1(t), \dots, x_m(t)]'$ is the observed vector of data, $\mathbf{A}$ the mixing matrix, $\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]'$ the sources and $\boldsymbol{\epsilon}(t) = [\epsilon_1(t), \dots, \epsilon_m(t)]'$ represents all the uncertainties (modeling uncertainty and measurement noise). t represents either time, frequency, pixel position or index number of a coefficient in a transform domain such as Fourier, spline or wavelet domain. The main problem is the estimation of the mixing matrix $\mathbf{A}$ and the sources $\mathbf{s}(t)$.

1.1 PCA and ICA

One of the main hypothesis in PCA and ICA approaches is that the data are temporally iid, in such a way that, the time index t can be omitted. The second main hypothesis is that the observation model is exact (no noise). With these two hypotheses, we have

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \quad (2)$$

All probabilistic methods are based on trying to assign or to model the probability law of the sources $p(\mathbf{s})$ and consequently $p(\mathbf{x})$. PCA based methods use only the second order statistics. The main hypothesis is $\mathbf{R}_s = \text{cov}[\mathbf{s}]$ is diagonal. Then, assuming $\mathbf{x} = \mathbf{A}\mathbf{s}$, we have $\mathbf{R}_x = \text{cov}[\mathbf{x}] = \mathbf{A}\mathbf{R}_s\mathbf{A}^t$. Thus, the PCA

algorithm, starts by estimating $\text{cov}[\mathbf{x}]$ from the data and then using the SVD, the mixing matrix $\mathbf{A}$ can be determined up to a rotation. It is also easy to see that the PCA solution is the maximum likelihood solution :

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} \{p(\mathbf{x}|\mathbf{A})\} = \arg \min_{\mathbf{A}} \{-\ln p(\mathbf{x}|\mathbf{A})\} \quad (3)$$

assuming the sources are uncorrelated Gaussian :

$$p(\mathbf{s}) = \mathcal{N}(\mathbf{0}, \mathbf{R}_s) \longrightarrow p(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{A}\mathbf{R}_s\mathbf{A}^t).$$

This can easily be seen if we note that

$$-\ln p(\mathbf{x}|\mathbf{A}) = \frac{n}{2}(\ln(2\pi) + \ln |\mathbf{A}|) + \frac{1}{2}\mathbf{x}^t(\mathbf{A}\mathbf{R}_s\mathbf{A}^t)^{-1}\mathbf{x}$$

In ICA, the main idea is to form an auxiliary vector $\mathbf{y} = \mathbf{B}\mathbf{x}$ and then determine the vector $\mathbf{y}$ which has the most independent components, for example by

$$\hat{\mathbf{B}} = \arg \max_{\mathbf{B}} \left\{ \text{KL}(p(\mathbf{y}) : \prod_i p_{y_i}(y_i)) \right\} \quad (4)$$

where $\text{KL}(p, q)$ is the Kullback-Leibler distance or the mutual information between two probability laws p and q .

The main point is then to choose $\mathbf{y} = \hat{\mathbf{B}}\mathbf{x}$ as an estimate of $\mathbf{s}$. It is not too difficult to show that, if $\mathbf{A}$ is invertible and if $\hat{\mathbf{B}} = \mathbf{A}^{-1}$, then $\mathbf{y} = \hat{\mathbf{B}}\mathbf{x} = \hat{\mathbf{B}}\mathbf{A}\mathbf{s} = \mathbf{s}$. But, can we say anything about the relation between $\mathbf{s}$ and $\mathbf{y}$ if the matrix is singular or if we now that the data are noisy ?

It is also interesting to note that, when $\mathbf{A}$ is invertible, the solution obtained by the relation (4) is, mathematically, the same as the ML solution (3) when replacing $\mathbf{A}$ by $\mathbf{B}^{-1}$, i.e.

$$\hat{\mathbf{B}} = \arg \max_{\mathbf{B}} \{p(\mathbf{x}|\mathbf{B})\} = \arg \min_{\mathbf{B}} \{-\ln p(\mathbf{x}|\mathbf{B})\}$$

This interpretation gives also the interesting point which is that ICA extends the PCA to the case of non Gaussian signals.

One more remark is that to obtain an explicit expression for the mutual information criterion, we need to make more assumptions about $p(\mathbf{y})$ and consequently about $p(\mathbf{x})$ and thus about $p(\mathbf{s})$. For example, if we limit our analysis to the second order statistics, we obtain PCA and if we go farther, we obtain all the HOS algorithms, and if we do any approximation for the expression of $p(\mathbf{s})$ by troncation of its development in Taylor's series, we obtain the ICA algorithms based on the contrast functions [2, 3, 4, 5, 6, 7].

2 Bayesian approach

In the Bayesian approach, we start by accounting explicitly for the errors in modeling and the noise. The simplest model is then assuming the errors to be

additive $\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon}$. Then, we assign a probability law to $\boldsymbol{\epsilon}$ to translate our prior knowledge about it. The simplest choice is to assume $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\epsilon)$ with $\mathbf{R}_\epsilon = \text{diag}[\sigma_{\epsilon_i}^2, i = 1, \dots, M]$. Then:

$$p(\mathbf{x}|\mathbf{A}, \mathbf{s}) = \mathcal{N}(\mathbf{A}\mathbf{s}, \mathbf{R}_\epsilon) \quad (5)$$

The next step is to assign $p(\mathbf{s})$ which gives the possibility to derive the joint pdf $p(\mathbf{x}, \mathbf{s}|\mathbf{A}) = p(\mathbf{x}|\mathbf{A}, \mathbf{s})p(\mathbf{s})$ and the marginal pdf

$$p(\mathbf{x}|\mathbf{A}) = \int p(\mathbf{x}|\mathbf{A}, \mathbf{s})p(\mathbf{s})d\mathbf{s}. \quad (6)$$

We can then stop here and look for the Maximum likelihood (ML) solution

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} \{p(\mathbf{x}|\mathbf{A})\} = \arg \min_{\mathbf{A}} \{-\ln p(\mathbf{x}|\mathbf{A})\} \quad (7)$$

either directly when it is possible to obtain an analytical expression for $p(\mathbf{x}|\mathbf{A})$ as is the case of Gaussian hypothesis for $p(\mathbf{s})$, or using the EM algorithm to obtain the solution.

The interesting point is that, by this approach, we obtain the PCA and ICA methods as particular case in the limiting case of no noise ($\boldsymbol{\epsilon} = \mathbf{0}$) and invertible mixing matrix $\mathbf{A}$.

But, now, we can go farther. For example, if $p(\mathbf{s})$ depends on a set of unknown parameters $\boldsymbol{\theta}$ and if $p(\mathbf{x}|\mathbf{A}, \mathbf{s})$ depends on a set of unknown parameters (for example the $\mathbf{R}_\epsilon$), then we can write the joint pdf $p(\mathbf{x}, \mathbf{s}|\mathbf{A}, \boldsymbol{\theta}, \mathbf{R}_\epsilon) = p(\mathbf{x}|\mathbf{A}, \mathbf{s}, \mathbf{R}_\epsilon)p(\mathbf{s}|\boldsymbol{\theta})$ and the marginal pdf $p(\mathbf{x}|\mathbf{A}, \boldsymbol{\theta}, \mathbf{R}_\epsilon)$ and then estimate them jointly by

$$(\hat{\mathbf{A}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{R}}_\epsilon) = \arg \max_{\mathbf{A}, \boldsymbol{\theta}, \mathbf{R}_\epsilon} \{p(\mathbf{x}|\mathbf{A}, \boldsymbol{\theta}, \mathbf{R}_\epsilon)\} \quad (8)$$

But, we can still do better, by assigning appropriate priors to obtain an expression for the joint posterior of all the unknowns

$$p(\mathbf{s}, \mathbf{A}, \boldsymbol{\theta}, \mathbf{R}_\epsilon|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{A}, \mathbf{s}, \mathbf{R}_\epsilon)p(\mathbf{s}|\boldsymbol{\theta})p(\mathbf{A})p(\boldsymbol{\theta})p(\mathbf{R}_\epsilon) \quad (9)$$

from which we can infer on all the unknowns at the same time. However, this needs to assign the appropriate priors $p(\mathbf{A})$, $p(\boldsymbol{\theta})$ and $p(\mathbf{R}_\epsilon)$. There are many works on this subjects. Sometimes they can really represent our prior knowledge based on the physics of the data measurement system [8, 9], but sometimes they can only represent a mathematical tool to eliminate the degeneracy of the likelihood based criteria [10, 11, 12, 13, 14, 15].

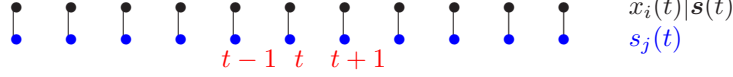
3 Proposed a priori sources models

3.1 Classical models

The main step in any classical Maximum likelihood (ML) or Bayesian approach is the prior modeling of the sources. Two main classes of prior models are the Generalized Gaussians (GG) and Mixture of Gaussians (MoG). The GG model

$$p(s_j) \propto \exp -\beta_j |s_j|^{\gamma_j} \quad (10)$$

has been used mainly to model the sparse sources. Many other non Gaussian prior laws have also been used. All these simple and direct models can be represented in the following scheme:



Model 1: The iid model for the sources.

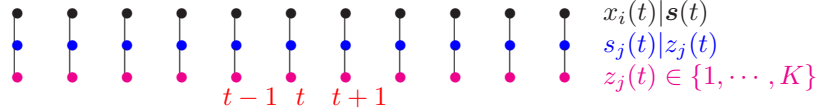
Another interesting model used frequently is the MoG:

$$p(s_j) = \sum_{k=1}^K \alpha_{j_k} \mathcal{N}(m_{j_k}, \sigma_{j_k}^2) \quad (11)$$

The interesting point in this model is in the fact that it can be described via a hidden discrete valued variable z_j in a hierarchical way

$$p(s_j|z_j = k) = \mathcal{N}(m_{j_k}, \sigma_{j_k}^2) \text{ with } P(z_j = k) = \alpha_{j_k} \quad (12)$$

This interpretation of the MoG model can be illustrated by the following scheme:

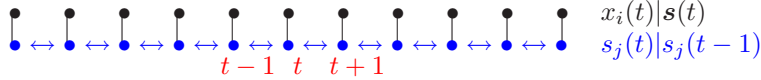


Model 2: A three level iid model for sources with an iid hidden variable $z_j(t)$. This model is equivalent to **Model 1** with a mixture of Gaussians (MoG) prior law for sources.

3.2 Accounting for time evolution

We remember that, we omitted the time index t from the equation (1), because, in all the previous cases, we assumed the sources and thus the data to be temporally iid. Now, we are going to relax this unrealistic hypothesis. Indeed, we are going to explicitly account for this time evolution.

An easy and explicit way is to model directly the sources by using a Markovian model which accounts for local dependency of the sources at time t and $t - 1$.



Model 3: A simple Markovian model for the sources.

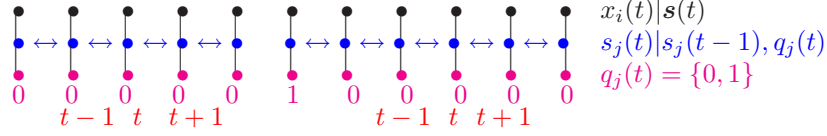
In this case, if the expression of $p(s_j(t)|s_j(t-1))$ has the parameters θ_j which are independent of time, we will have a stationary process. There are also models in which these parameters are allowed to vary during the time allowing

to account for non stationarity of the sources. In particular, we may mention the Gauss-Markov models where

$$p(s_j(t)|s_j(t-1)) = \mathcal{N}(s_j(t-1), \theta_j(t)) \quad (13)$$

where $\theta_j(t)$ is linked to the power spectra of the sources [16, 17, 18, 19, 20, 21, 22].

Another easy and explicit way which also has the possibility to account for non stationarity of the time evolution and in particular the discontinuities in the sources is the following:



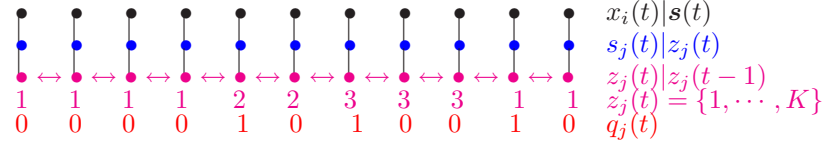
Model 4: A composite Markovian model with hidden contours variables $q_j(t)$.

In this model the binary valued hidden variable $q_j(t)$ represents the discontinuities (contours). The expression of $p(s_j(t)|s_j(t-1))$ depends on $q_j(t)$, for example:

$$\begin{aligned} p(s_j(t)|s_j(t-1), q_j(t) = 0) &= \mathcal{N}(m_{j_k}, v_{j_k}) \\ p(s_j(t)|s_j(t-1), q_j(t) = 1) &= \mathcal{N}(s_j(t-1), v_{j_k}) \end{aligned}$$

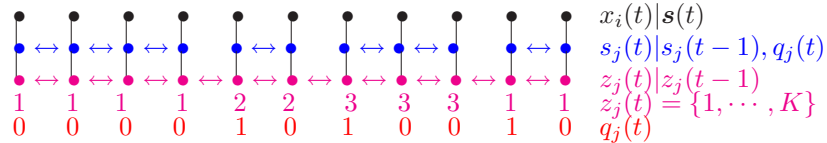
which results to $p(s_j(t)|s_j(t-1)) = (1 - \lambda_j)\mathcal{N}(m_{j_k}, v_{j_k}) + \lambda_j\mathcal{N}(s_j(t-1), v_{j_k})$ with $\lambda_j = P(q_j = 1)$.

Still another easy and explicit way which also has the possibility to account for non stationarity of the time evolution and in particular the notion of piecewise homogeneity in the sources is the following:



Model 5: A composite Markovian model with hidden classification variables $z_j(t)$. Note that $s_j(t)|z_j(t)$ are iid, but $z_j(t)$ are markovian (Potts). Note also that the contour variables $q_j(t)$ are obtained in a deterministic way from $z_j(t)$, i.e. $q_j(t) = 1 - \delta(z_j(t) - z_j(t-1))$. These contours are closed by construction.

Finally, the following model combines both **Model 3:** and **Model 5:** which gives the possibility to have a local dependency inside a region.



Model 6: A doubly Markovian model with hidden Potts variables.

4 Bayesian estimation computation

In all these models, we have to assign the prior laws:

— $p(\mathbf{x}(t)|\mathbf{s}(t), \mathbf{A}, \mathbf{R}_\epsilon)$: For this, we use the observation model (1) and assume that the noises $\epsilon_i(t)$ are centered, white, Gaussians and mutually independent. This then leads to:

$$p(\mathbf{x}(t)|\mathbf{s}(t), \mathbf{A}, \mathbf{R}_\epsilon) = \mathcal{N}(\mathbf{A}\mathbf{s}(t), \mathbf{R}_\epsilon)$$

with $\mathbf{R}_\epsilon = \text{diag}[\sigma_{\epsilon_1}^2, \dots, \sigma_{\epsilon_m}^2]$.

— $p(\mathbf{s}(t))$ or $p(\mathbf{s}(t)|\mathbf{z}(t))$: For this, first, in general, we assume that the sources are *a priori* independents, i.e., $p(\mathbf{s}(t)) = \prod_j p(s_j(t))$ or only conditionally independent, i.e. $p(\mathbf{s}(t)|\mathbf{z}(t)) = \prod_j p(s_j(t)|z_j(t))$. Then, we have either to choose expressions for $p(s_j(t))$ such as GG or for $p(s_j(t)|z_j(t))$, for example in the case of MoG.

— For the Markovian models, in general, we choose a first order Markov chain in 1D case and a Markov field with first order neighborhood in 2D case (images). For example, in 1D signals

$$p(s_j(0)) = N(0, \sigma_j^2), \quad p(s_j(t)|s_j(t-1)) = N(s_j(t-1), \sigma_j^2)$$

and for images where $t \in \mathcal{T}$ represents a pixel position and $t' \in \mathcal{V}(t)$ the neighboring pixels positions

$$p(s_j(t)|s_j(t'), t' \in \mathcal{T}_{-t}) \propto \exp \left[-\alpha_j \sum_{t' \in \mathcal{V}(t)} \phi(s_j(t), s_j(t')) \right]$$

where ϕ is the potential associated to the cliques of the Markov field neighborhood system.

— For the Markovian models for the discrete valued hidden variables $q_j(t)$ and $z_j(t)$, we use the Ising and Potts models respectively. For example

$$p(z_j(0) = k) = \pi_{jk}, \quad p(z_j(t) = k | z_j(t-1) = l) = \pi_{jkl}$$

and for images

$$p(z_j(t)|z_j(t'), t' \in \mathcal{T}_{-t}) \propto \exp \left[-\alpha_j \sum_{t' \in \mathcal{V}(t)} \delta(z_j(t) - z_j(t')) \right]$$

where $\delta(t) = 1$ for $t = 0$ and $\delta(t) = 0$ for $t \neq 0$.

— Finally, we also have to assign prior laws to the elements of the matrix $\mathbf{A}$ and to the hyperparameters Θ such as the noise covariance $\mathbf{R}_\epsilon$ or only the noise variances $\theta_\epsilon = \{\sigma_{\epsilon_i}^2\}$, the parameters of the source distributions such as $\theta_s = \{(m_{jk}, \sigma_{jk}^2)\}$. In general, we choose conjugate priors for the hyperparameters.

When all these priors are appropriately assigned, we can obtain an expression for the posterior law

$$p(\mathbf{S}, \mathbf{Z}, \mathbf{A}, \Theta | \mathbf{X}) \propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \theta_\epsilon) p(\mathbf{S} | \mathbf{Z}, \theta_s) p(\Theta) \quad (14)$$

where $\mathbf{S} = \{\mathbf{s}(t), t \in \mathcal{T}\}$, $\mathbf{Z} = \{\mathbf{z}(t), t \in \mathcal{T}\}$, $\mathbf{X} = \{\mathbf{x}(t), t \in \mathcal{T}\}$ and $\Theta = \{\mathbf{A}, \theta_\epsilon, \theta_s\}$. We can then use this posterior law to define an estimator such as Joint Maximum A Posteriori (JMAP) or the Posterior Means (PM). The first needs optimization algorithms and the second integration methods. Both are computationally demanding. Alternate optimization is generally used for the first while the MCMC techniques are used for the second. We present here three such algorithms:

Algorithm A:	Algorithm B:	Algorithm C:
$\hat{\mathbf{S}} \sim p(\mathbf{S} \hat{\mathbf{Z}}, \hat{\mathbf{A}}, \hat{\Theta}, \mathbf{X})$	$\hat{\mathbf{S}} \sim p(\mathbf{S} \hat{\mathbf{Z}}, \hat{\mathbf{A}}, \hat{\Theta}, \mathbf{X})$	$\hat{\mathbf{S}} \sim p(\mathbf{S} \hat{\mathbf{Z}}, \hat{\mathbf{A}}, \hat{\Theta}, \mathbf{X})$
$\hat{\mathbf{Z}} \sim p(\mathbf{Z} \hat{\mathbf{S}}, \hat{\mathbf{A}}, \hat{\Theta}, \mathbf{X})$	$\hat{\mathbf{Z}} \sim p(\mathbf{Z} \hat{\mathbf{A}}, \hat{\Theta}, \mathbf{X})$	$\hat{\mathbf{Z}} \sim p(\mathbf{Z} \hat{\mathbf{A}}, \hat{\Theta}, \mathbf{X})$
$\hat{\mathbf{A}} \sim p(\mathbf{A} \hat{\mathbf{Z}}, \hat{\mathbf{S}}, \hat{\Theta}, \mathbf{X})$	$\hat{\mathbf{A}} \sim p(\mathbf{A} \hat{\mathbf{Z}}, \hat{\Theta}, \mathbf{X})$	$\hat{\mathbf{A}} \sim p(\mathbf{A} \hat{\Theta}, \mathbf{X})$
$\hat{\Theta} \sim p(\Theta \hat{\mathbf{A}}, \hat{\mathbf{S}}, \hat{\mathbf{Z}}, \mathbf{X})$	$\hat{\Theta} \sim p(\Theta \hat{\mathbf{A}}, \hat{\mathbf{Z}}, \mathbf{X})$	$\hat{\Theta} \sim p(\Theta \hat{\mathbf{A}}, \mathbf{X})$

In these algorithms, $\sim$ represents either *argmax* or *generate sample using*. The first is used for alternate optimization and the second for MCMC Gibbs sampling. Implementing the **Algorithm A:** is easy, because, in general we have all the expressions of all the conditional laws used. The **Algorithm B:** needs integration with respect to the sources and **Algorithm C:** integration with respect to the sources and the hidden variables.

5 Examples of applications

To illustrate the application of these different models we give here a few examples of results obtained with the proposed prior models.

5.1 Spectrometry

The following is an example of simulation result in near infra red spectroscopy where 10 mixtures of 3 sources (A, B and C) have been observed. It is assumed that there is no interaction between these 3 components. Thus the hypothesis of a linear mixture is valid. Here, we have to account for the positivity of the spectra as well as the positivity of the mixing matrix elements. The model used here is **Model 1** with *a priori* Gamma prior laws for both the sources and the mixing matrix elements. The estimator is the joint MAP and the implemented algorithm is alternate maximization with respect to sources, the mixing matrix elements and the hyperparameters (S. Moussaoui et al. [23, 24, 25]).

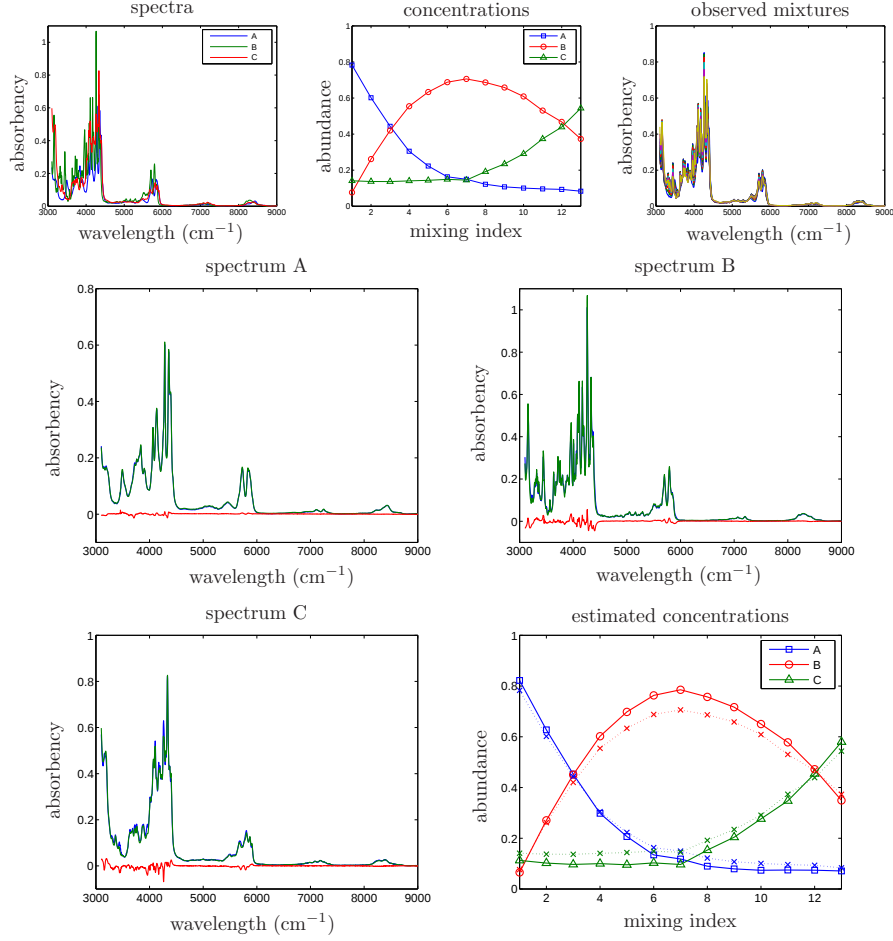


Fig. 1: Source separation in spectrometry: a) The three sources, the mixing matrix elements and the ten mixture data, b) Estimated sources and concentrations.

5.2 Source separation in astrophysics

One of the problems in astrophysics imaging is to separate different cosmic activities and in particular the Cosmological Microwave Background (CMB). The observation model, when written in Fourier domain is exactly as in (1), but here t represents the frequency. The example provided here is a simulation in the case where there are 3 sources and 6 observations. The appropriate model here **Model 3** and the implemented algorithm is an EM like [26, 27, 28, 29].

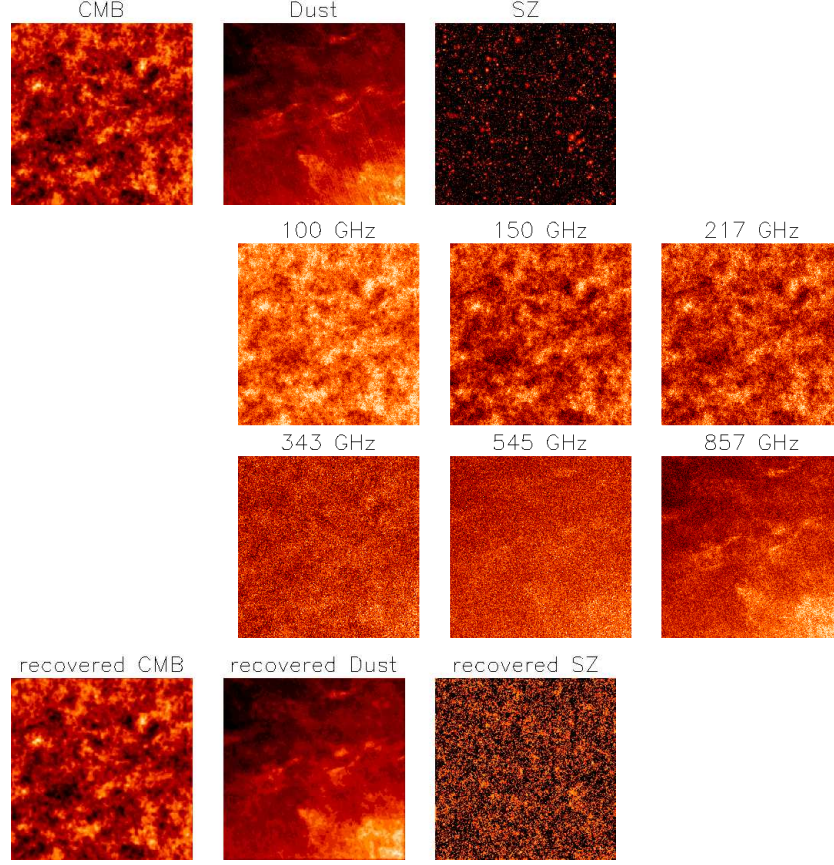


Fig. 2: Source separation in astrophysics : a) the 3 sources $\mathbf{S}$, b) the 6 observations $\mathbf{X}$, c) the estimated sources $\hat{\mathbf{S}}$.

5.3 Source separation in satellite imaging

Here, we give two results in an application related to eliminating clouds in satellite imaging. In the first example, the model used is **Model 4** where the pixels of each source image are assumed to be in one of the two classes [28].

The second example concerns the same problem but in wavelet domain. Indeed, if we apply an unitary transform such as the orthogonal wavelet transform on both sides of the equation (1), we obtain exactly the same type of equation but on the wavelet coefficients of sources and observations. Then, we can do the separation in the wavelet domain. The main advantage being that simpler *a priori* models such as **Model 1** or **Model 2** can be used more efficiently. The following figure shows an example of results obtained using either a Generalized Gaussian (GG) or a Mixture of two Gaussians (MoG) models [30, 31].

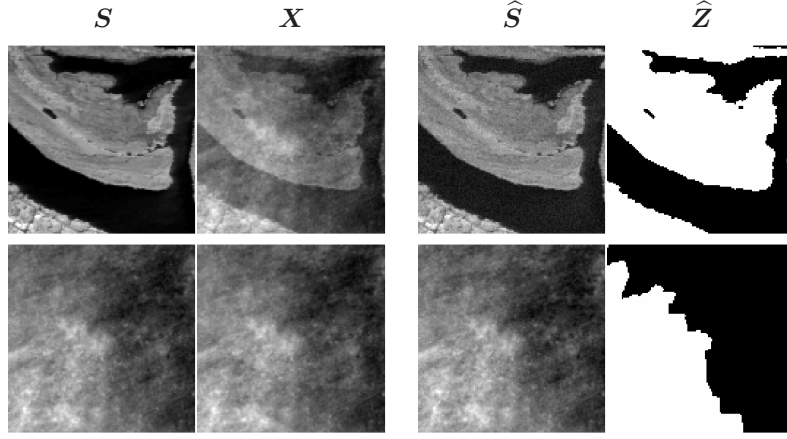


Fig. 3: Source separation in satellite imaging : a) sources $\mathbf{S}$, b) observations $\mathbf{X}$, c) estimated sources $\hat{\mathbf{S}}$ and d) estimated segmentations $\hat{\mathbf{Z}}$.

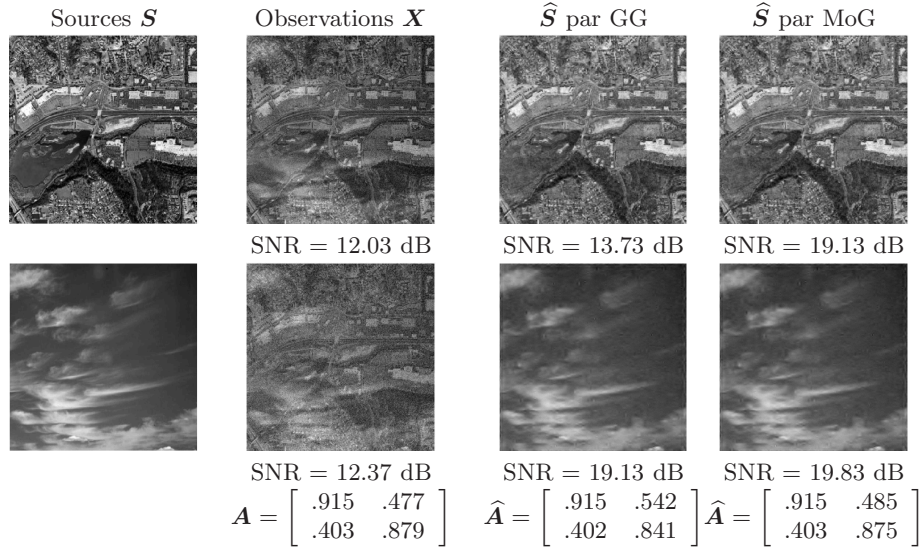


Fig. 4: Source separation using Models 1 et 2: a) sources $\mathbf{S}$, b) observations $\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E}$, c) estimated sources using GG and d) estimated sources using MoG.

5.4 Data reduction, classification and segmentation of hyper-spectral images

Hyper-spectral images data are often represented as a set of images $x_\omega(\mathbf{r})$. However, a specificity of these images is that the pixels at a position $\mathbf{r}$ along the wavelength bands ω form an spectral line [32, 33, 34, 35]. This is the reason for considering also these data as a set of spectra $x_{\mathbf{r}}(\omega)$. In both representations, the data are dependent in both spatial positions and in spectral bands. One of the image processing problems is then data reduction.

If we consider the data as a set of spectra, then we want to write:

$$x_{\mathbf{r}}(\omega) = \sum_{k=1}^K A_{\mathbf{r},k} s_k(\omega) + \epsilon_{\mathbf{r}}(\omega)$$

where the $s_k(\omega)$ are the K spectral sources and each column of the mixing matrix $\mathbf{A}$ is in fact an images $A_k(\mathbf{r})$. The ideal case here would be to obtain an estimate for $\mathbf{A}$ such that each column $A_k(\mathbf{r})$ represents an image where only non-zero values for the pixels in the regions which are associated to the spectrum $s_k(\omega)$.

If we consider the data as a set of images $x_\omega(\mathbf{r})$, then we have:

$$x_\omega(\mathbf{r}) = \sum_{j=1}^N A_{\omega,j} s_j(\mathbf{r}) + \epsilon_\omega(\mathbf{r})$$

where the sources $s_j(\mathbf{r})$ are the N source images and each column of the mixing matrix $\mathbf{A}$ in this case correspond to the spectrum $A_j(\omega)$. The ideal case here would be to obtain an estimate for the sources such that each the pixels of each image $s_j(\mathbf{r})$ be non-zero only for the pixels in the regions which are associated to the spectrum $A_j(\omega)$.

A priori, one may consider these two problems independently, but we may also account for these specificities. Indeed, in this last case, we may want to assign an *a priori* law to the elements of the matrix $\mathbf{A}$ to insure their positivity. We may also note that the columns $A_k(\mathbf{r})$ of the mixing matrix in the first model is related to the sources $s_j(\mathbf{r})$ in the second model and the columns $A_j(\omega)$ of the mixing matrix in the second model is related to the sources $s_k(\omega)$ in the first model. This becomes still more explicit if we choose $K = N$

In the following examples, we considered the first model where each source represents an image which is composed of homogeneous regions. As the objective here is data reduction, it is natural to assume that *a priori* the sources $s_j(\mathbf{r})$ are mutually independent and are all composed of a finite set of connected regions and a finite set of K characteristic components. Then, we can easily see that appropriate models to use are either **Model 5** or **Model 6**.

In the following results, we used the **Model 5** with additional assumption that all sources share the same hidden variable $z(\mathbf{r})$. More details on this method can be obtained in [35] in the same proceeding.

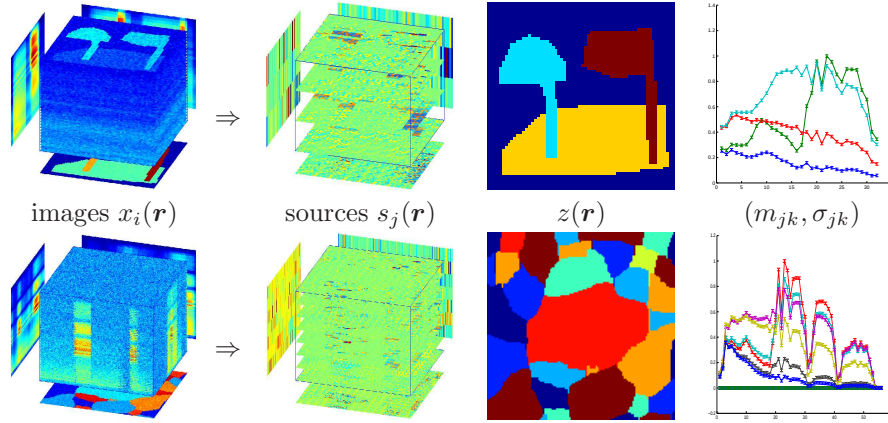


Fig. 5: Data reduction, classification and segmentation of hyper-spectral images: a) 32 observed images $x_i(\mathbf{r})$, b) 8 estimated sources $s_j(\mathbf{r})$, c) estimated common segmentation $z(\mathbf{r})$, d) estimated means m_{jk} and variances σ_{jk}^2 associated to each class. The means m_{jk} represent the columns of $\mathbf{A}$ and σ_{jk} their standard deviation.

6 Conclusion

The Bayesian estimation approach pushes farther the limits of classical PCA and ICA based methods for blind sources separation. In this work, we proposed six different models for the prior laws of the sources which can be used in different applications. The two first models are classical, the third accounts for temporal evolution of stationnary sources. The fourth model which account for contours in images is appropriate for piecewise stationnary sources, and finally, the fifth and sixth models are particularly appropriate for piecewise homogeneous sources. We then showed a few examples of simulation results which show different application areas of such models in spectrometry, in astrophysical imaging, in satellite and in hyperspectral imaging.

References

- [1] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, 2000.
- [2] C. Jutten and J. Herault, "Blind separation of sources .1. an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.
- [3] P. Comon, C. Jutten, and J. Héroult, "Blind separation of sources, Part II: Problems statement," *Signal Processsing*, vol. 24, pp. 11–20, 1991.
- [4] P. Comon, "Independent component analysis – a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [5] D.-T. Pham, "Blind separation of instantaneous mixture sources via independent component analysis," *IEEE Trans. Signal Processing*, vol. 44, 1996.

- [6] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, 1999.
- [7] C. Jutten, "Source separation: from dusk till dawn," in *Proc. of 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'2000)*, (Helsinki, Finland), pp. 15–26, 2000.
- [8] K. Knuth, "A Bayesian approach to source separation," in *Proceedings of International Workshop on Independent Component Analysis and Signal Separation (ICA'99)* (J.-F. Cardoso, C. Jutten, and P. Loubaton, eds.), (Aussios, France), pp. 283–288, 1999.
- [9] K. Knuth, "Source separation as an exercise in logical induction," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (A. Mohammad-Djafari, ed.), (Paris, France), pp. 340–349, American Institute of Physics, 2001.
- [10] H. Snoussi and A. Mohammad-Djafari, "Penalized maximum likelihood for multivariate gaussian mixture," in *Bayesian Inference and Maximum Entropy Methods* (R. L. Fry, ed.), pp. 36–46, MaxEnt Workshops, Amer. Inst. Physics, Aug. 2001.
- [11] H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of gaussians prior," *To appear in Int. Journal of VLSI Signal Processing Systems*, 2002.
- [12] H. Snoussi and A. Mohammad-Djafari, "Unsupervised learning for source separation with mixture of Gaussians prior for sources and Gaussian prior for mixture coefficients," in *Proc. Neural Networks for Signal Processing XI*, pp. 293–302, D.J. Miller, Ed, September 2001.
- [13] S. Roberts, "Independent component analysis: Source assessment and separation, a Bayesian approach," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 145, no. 3, pp. 149–154, 1998.
- [14] S. Roberts and R. Choudrey, "Data decomposition using independent component analysis with prior constraints," *Pattern Recognition*, vol. 36, pp. 1813–1825, 2003.
- [15] D. Rowe, *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Boca Raton, Florida, USA: CRC Press, 2003.
- [16] A. Souloumiac, "Blind source detection and separation using second order non-stationarity," in *Proc. ICASSP*, pp. 1912–1915, 1995.
- [17] A. Belouchrani and J.-F. Cardoso, "Maximum likelihood source separation by the expectation-maximization technique: deterministic and stochastic implementation," in *proceeding of International Symposium on Nonlinear Theory and Application (NOLTA'95)*, pp. 49–53, 1995.
- [18] A. Hyvärinen, "Independent component analysis for time-dependent stochastic processes," in *Proceedings of International Conference Neural Networks*, pp. 541–546, 1998.
- [19] H. Attias, "Independent factor analysis with temporally structured sources," in *Advances in Neural Information Processing Systems* (S. Solla, T. Leen, and K.-R. Müller, eds.), vol. 12, pp. 386–392, MIT Press, 2000.
- [20] D.-T. Pham and J. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Processing*, vol. 49, 9, no. 11, pp. 1837–1848, 2001.

- [21] D.-T. Pham, "Blind Separation of Non Stationary Non Gaussian Sources," in *Proceeding of EUSIPCO'02*, (Toulouse, France), Sep. 2002.
- [22] S. Hosseini, C. Jutten, and D. T. Pham, "Markovian source separation," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3009–3019, 2003.
- [23] S. Moussaoui, C. Carteret, A. Mohammad-Djafari, O. Caspary, D. Brie, and B. Humbert, "Approche bayésienne pour l'analyse de mélanges en spectroscopies.," in *Chimimétrie'2003*, (Paris, France), Décembre 2003.
- [24] S. Moussaoui, A. Mohammad-Djafari, D. Brie, and O. Caspary, "A Bayesian method for positive source separation," in *proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, (Montreal, Canada), pp. 485–488, May 2004.
- [25] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling." Accepté pour publication dans *IEEE Transactions on Signal Processing*, Octobre 2005.
- [26] H. Snoussi and A. Mohammad-Djafari, "Bayesian source separation with mixture of Gaussians prior for sources and Gaussian prior for mixture coefficients," in *Bayesian Inference and Maximum Entropy Methods* (A. Mohammad-Djafari, ed.), (Gif-sur-Yvette, France), pp. 388–406, Proc. of MaxEnt, Amer. Inst. Physics, July 2000.
- [27] H. Snoussi and A. Mohammad-Djafari, "Unsupervised learning for source separation with mixture of Gaussians prior for sources and Gaussian prior for mixture coefficients.," in *Neural Networks for Signal Processing XI* (D. J. Miller, ed.), pp. 293–302, IEEE workshop, Sep. 2001.
- [28] H. Snoussi, *Approche bayésienne en séparation de sources. Applications en imagerie*. Phd thesis, Université de Paris-Sud, Orsay, France, Sep. 2003.
- [29] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images," *Journal of Electronic Imaging*, vol. 13, pp. 349–361, Apr. 2004.
- [30] A. Mohammad-Djafari and M. Ichir, "Wavelet domain image separation," in *proceedings of the 22th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt'2002)*, (Moscow, Idaho, USA), August 3–9 2002.
- [31] A. Mohammad-Djafari and M. Ichir, "Wavelet domain blind image separation," in *SPIE, Mathematical Modeling, Wavelets X*, August 2003.
- [32] K. Sasaki, S. Kawata, and S. Minami, "Component analysis of spatial and spectral patterns in multispectral images. I. basics," *Journal of the Optical Society of America. A*, vol. 4, no. 11, pp. 2101–2106, 1987.
- [33] L. Parra, C. Spence, A. Ziehe, K.-R. Mueller, and P. Sajda, "Unmixing hyperspectral data," in *Advances in Neural Information Processing Systems 13*, (*NIPS'2000*), pp. 848–854, MIT Press, 2000.
- [34] N. Bali and A. Mohammad-Djafari, "Mean Field Approximation for BSS of images with compound hierarchical Gauss-Markov-Potts model," in *MaxEnt05, San José CA, US*, American Institute of Physics (AIP), Aug. 2005.
- [35] N. Bali and A. Mohammad-Djafari, "Approxiamtion en champ moyen pour la séparation de sources appliqué aux images hyperspectrales," in *Actes 20^e coll. GRETSI*, (Louvains Belgium), Sep. 2005.