Wavelet Domain Blind Image Separation

Mahieddine M. Ichir[1] and Ali Mohammad-Djafari[2]

Laboratoire des Signaux et Systèmes
3, rue Joliôt Curie, Supélec, Plateau de Moulon
91192 Gif sur Yvette, France

ABSTRACT

In this work, we consider the problem of blind source separation in the wavelet domain via a Bayesian estimation framework. We use the sparsity and multiresolution properties of the wavelet coefficients to model their distribution by heavy tailed *prior* probability laws: the generalized exponential family and the Gaussian mixture family. Appropriate MCMC algorithms are developped in each case for the estimation purposes and simulation results are presented for comparaison.

Keywords: Blind source separation, Bayesian estimation, wavelet transform, MCMC algorithm

1. INTRODUCTION

Blind source separation (BSS) is an important field of research in signal processing and data analysis. Independent compenent analysis (ICA)[2] is one solution to the problem. However, in some applications, ICA fails to work particularly when the observations are too noisy and/or when the instantaneous mixture model is not totally verfied.

Bayesian estimation has been applied with success to solve the BSS problem.[3, 8, 11] It allows to account for any *prior* information we may have about the observational process, hence to model any independance or correlation (temporal and/or spacial) of the sources parameters and mixing matrix.

The BSS problem has been considered either directly in the original domain of observations (time 1D-signal or pixel 2D-image) or in a transform domain: Fourrier[7] or wavelet domain.[6, 12] The idea behind transform domains is that usually an invertible linear transform restructures the signal/image leaving the transform coefficients a structure easier to model.

Wavelet transform is a particularly interesting representation of (non-stationnary) signals/images. This property makes it a powerful tool in many signal processing domains: encoding, compression and signal denoising. But its application in blind source separation is new[6, 12] and it still remains to be more explored.

2. BAYESIAN APPROACH AND BSS

We consider linear and instantaneous mixing model, with noisy obseravtions given by:

$$x_m(t) = \sum_{n=1}^{N} A_{mn} s_n(t) + \epsilon_m(t) \quad \text{for} \quad m = 1, \ldots, M \tag{1}$$

for $t = 1, \ldots, T$. Or in a vector form

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t) \tag{2}$$

where $\boldsymbol{x}(t)$ represents the noisy observed data vector, $\boldsymbol{A}$ represents the the unknown mixing matrix, $\boldsymbol{s}(t)$ represents the source vector and $\boldsymbol{\epsilon}(t)$ the noise vector. The index t may be a single index for example the time

Further author information:
[1]E-mail : ichir@lss.supelec.fr, Tel.: +33 (0)169 851 747, web: http://www.lss.supelec.fr/perso/ichir/
[2]E-mail: djafari@lss.supelec.fr, Tel.: +33 (0)169 851 741, web: http://www.lss.supelec.fr/perso/djafari/

index for time series signals or a composite index for example pixel index for images. Noise models both the measurement noise and any uncertainty on the obsrvation model (2).

Since the wavelet transform Ψ is, in general, an orthonormal transform ($\Psi^*\Psi = \mathbb{I}$), the model (2) is still valid and can be written in the transform domain as:

$$x_m^j(k) = \sum_{n=1}^{N} A_{mn} s_n^j(k) + \epsilon_m^j(k) \quad \text{for} \quad m = 1, \ldots, M, \quad j = 1, \ldots, J \quad \text{and} \quad k = 1, \ldots, T/2^j \tag{3}$$

or equivalently

$$\boldsymbol{x}^j(k) = \boldsymbol{A}\boldsymbol{s}^j(k) + \boldsymbol{\epsilon}^j(k) \tag{4}$$

where $\{x_m^j(k), s_n^j(k)$ and $\epsilon_m^j(k)\}$ represent the k^{th} wavelet coefficients of $\{x_m(t), s_n(t)$ and $\epsilon_m(t)\}$ respectively at the resolution j (k beeing the dual index of t in the transform domain).

In a Bayesian estimation framework, the joint *posterior* distribution of the parameters of interest is given by:

$$p(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{\theta}, \boldsymbol{S})\ \pi(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{\theta})\ \pi(\boldsymbol{\theta}) \tag{5}$$

where $\boldsymbol{S} = \{\boldsymbol{s}^j(k)\}$, $\boldsymbol{X} = \{\boldsymbol{x}^j(k)\}$, $p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta})$ is the likelihood function of the model (4) and $\pi(\boldsymbol{s}, \boldsymbol{A}|\boldsymbol{\theta})$ is the *prior* distribution reflecting (encoding) any *prior* information we may have about these parameters. $\pi(\boldsymbol{\theta})$ is the hyperparameters *prior* distribution. It may reflect some behaviour of these parameters (positivity of the noise variance for example). In this work, we assume that the noise $\boldsymbol{\epsilon}(t)$ is centered, temporarly and spacially white and Gaussian with a covariance matrix $\boldsymbol{R}_{\boldsymbol{\epsilon}} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_M^2\right)$. Then the likelihood is given by:

$$p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\theta}) = \prod_{j,k} p\big(\boldsymbol{x}^j(k)|\boldsymbol{s}^j(k), \boldsymbol{A}, \boldsymbol{\theta}\big) \tag{6}$$

with

$$p\big(\boldsymbol{x}^j(k)|\boldsymbol{s}^j(k), \boldsymbol{A}, \boldsymbol{\theta}\big) = \mathcal{N}\left(\boldsymbol{x}^j(k)|\boldsymbol{A}\boldsymbol{s}^j(k), \boldsymbol{R}_{\boldsymbol{\epsilon}}\right) \tag{7}$$

$$\propto |\boldsymbol{R}_{\boldsymbol{\epsilon}}|^{-1/2} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}^j(k) - \boldsymbol{A}\boldsymbol{s}^j(k)\right)^* \boldsymbol{R}_{\boldsymbol{\epsilon}}{}^{-1}\left(\boldsymbol{x}^j(k) - \boldsymbol{A}\boldsymbol{s}^j(k)\right)\right) \tag{8}$$

The main issue in the Bayesian framework is the appropriate choice of the *prior* laws $\pi(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ which is developed in the following sections.

3. WAVELET COEFFICIENTS STATISTICAL MODEL

The wavelet transform is an interesting representation of signals, it has some properties that makes it rich for modeling.[5] The wavelet tranform of signals/images is sparse: the wavelet transform of a signal/image (Fig. 1) results in a *large number of small coefficients and a small number of large coefficients.* This property makes the wavelet transform a suitable choice for compression, encoding and signal denoising. We can statistically model this property by some convenient probability distributions.[1,4]

3.1. Heavy tailed distributions

Mallat[4] has porposed to assign to the wavelet coefficients a Generalized Exponential (GE) like distribution given by:

$$\pi(x|\gamma, \alpha) = \mathcal{E}\text{xp}\left(x|\gamma, \alpha\right) = K \exp\left(-\frac{1}{2\gamma}|x|^{\alpha}\right) \tag{9}$$

where K is a normalisation constant, $\gamma > 0$ and $1 \leq \alpha \leq 2$.

Figure 1. Lena image (left) and its wavelet tansform (right).

Another family of laws that describes well this sparsity are the Gaussian mixture distributions (for example a two component Gaussian mixture), as adopted by Crouse et al.[1]:

$$\pi(x|p,\tau_1,\tau_2) = p\,\mathcal{N}\left(x|0,\tau_1\right) + (1-p)\,\mathcal{N}\left(x|0,\tau_2\right) \tag{10}$$

where $\tau_1 >> \tau_2$ and $0 \leq p \leq 1$.

3.2. Independance

The wavelet transform is known to have a decorrelation porperty, we say that the wavelet transform *nearly decorrelates the signal*, resulting in uncorrelated coefficients. So we can model the wavelet coefficients distribution by a separable probablity distrbution:

$$p(\boldsymbol{S}) = \prod_{j,k} \pi\left(\boldsymbol{s}^j(k)\right) \tag{11}$$

where $\boldsymbol{S}$ is joint set of the wavelet coefficients at all resolutions and $\pi\left(\boldsymbol{s}^j(k)\right) = \prod_{n} \pi\left(s_n^j(k)\right)$, with $\pi\left(s_n^j(k)\right)$ given by Equation (9) or Equation (10).

3.3. Inter-scale correlation

The decorrelation property of the wavelet transform is not totally ensured, and in addition to that, decorrelation is not independance, thus the validity of independant models (11) is not really verified. We can enhance the statistical description of the wavelet coefficients by taking into account some of their additional properties:

- *persistence* large/small values of wavelet coefficients tend to propagate across scales.
- *locality* each wavelet atom is localized simultaneously in time and frequency (scale).

We have presented in Fig. 2, the continuous wavelet transform (in abolute values) of a one dimensional signal where we observe that if a wavelet coefficient is present at a given reolution, then it tends to propagate through the coarser resolutions. However, we aleviate the model by assuming that the wavelet coefficients are independant inside each scale.

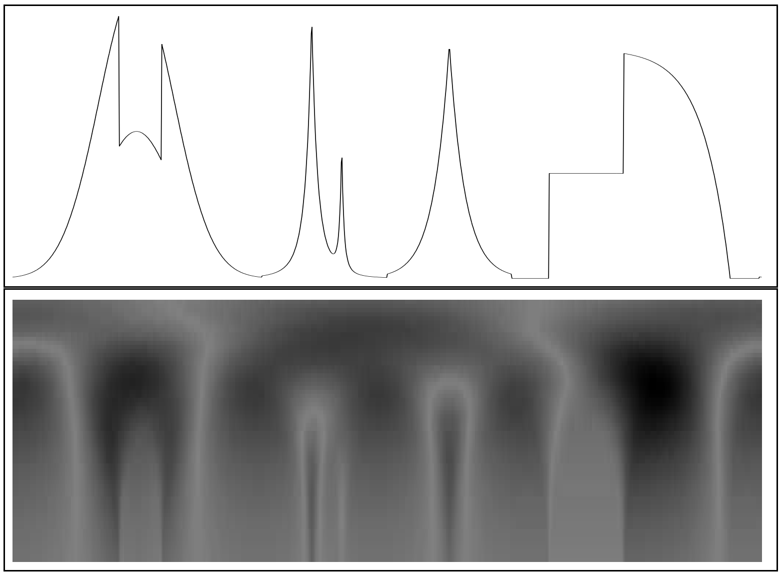

Figure 2. A one dimensionnal signal (top) and its continous wavlet transform (in absolute values) (bottom).

The *prior* probability distribution of the sources coefficients is then given by:

$$p\left(\boldsymbol{S}\right) = \pi\left(\boldsymbol{S}^1\right) \prod_{j=2}^{J} \pi\left(\boldsymbol{S}^j | \boldsymbol{S}^{P(j)}\right), \quad \text{with} \quad \pi\left(\boldsymbol{S}^j|.\right) = \prod_{k}^{T_j} \pi\left(\boldsymbol{s}^j(k)|.\right) \tag{12}$$

with $\boldsymbol{S}^{P(j)} = \{S_n^{P(j)}\}$ represents the set of the direct ancestors of the coefficients $\boldsymbol{S}^j = \{S_n^j\}$ (Fig. 3).

Equation (12) and Fig. 3 describes a first order Makov model, where each wavelet coefficient at a given resolution is independant of the other coefficients at the same resolution, but depends on those at the higher resolution given by the set of its direct ancestors $\boldsymbol{S}^{P(j)}$.

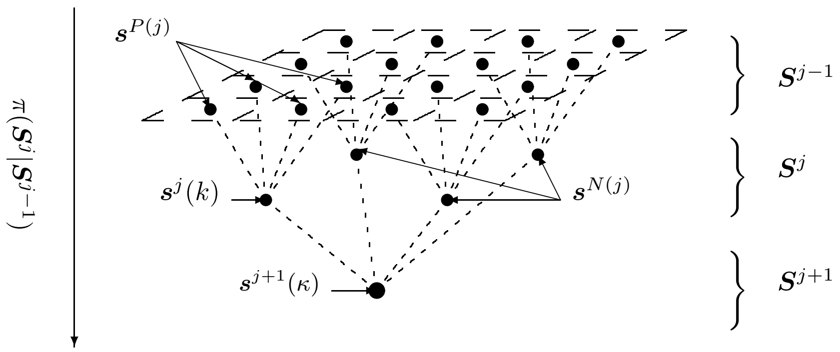

Figure 3. Graphical model describing the inter-scale correlation.

In this work, the correlation property is introduced and taken into account only for Generalized Exponential (GE) models. Correlation property in the Gaussian mixture models needs the definition of what is known as Hidden Markov Models (HMM) which has not yet been treated.

The *prior* probability law (12) in the case of GE models, is given more explicitely by:

$$\pi\left(\boldsymbol{s}^j(k)|\boldsymbol{S}^{P(j)}\right) = \mathcal{E}\text{xp}\left(\boldsymbol{s}^j(k)|\boldsymbol{S}^{P(j)}, \boldsymbol{R}_{\gamma_j}, \alpha_j\right) = \prod_n \mathcal{E}\text{xp}\left(s_n^j(k)|S_n^{P(j)}, \gamma_n^j, \alpha_j\right) \tag{13}$$

where

$$\mathcal{E}\text{xp}\left(s_n^j(k)|S_n^{P(j)}, \gamma_n^j, \alpha_j\right) = K \exp\left(-\frac{1}{2\gamma_n^j}\left|s_n^j(k) - \phi_n^j(s_n^{P(j)})\right|^{\alpha_j}\right) \tag{14}$$

where $\phi_n^j(s_n^{P(j)})$ is some *function* of the set of the direct ancestors of $s_n^j(k)$ $\left(\phi_n^j(.)\right.$ can be defined as beeing some wheighted sum for example$\left.\right)$, and $\boldsymbol{R}_{\gamma_j}=\text{diag}\left(\gamma_1^j, \ldots, \gamma_N^j\right)$.

4. MCMC ALGORITHM

As estimates of the coefficients $\boldsymbol{s}^j(k)$, the mixing matrix $\boldsymbol{A}$ and the hyperparameters $\boldsymbol{\theta} = [\boldsymbol{R}_{\boldsymbol{\epsilon}}, \boldsymbol{R}_{\gamma_j}]$, we take their *posterior* means where:

1. The *prior* law of the sources coefficients is independant of the mixing matrix and is given either by Equation (11) (for independant coefficients) or by Equation (12) (for inter-scale correlated coefficients).
2. The elements of the mixing matrix are supposed Gaussian, of mean $\boldsymbol{\mu}_A$, and of covariance matrix $\boldsymbol{R}_A$:

$$\pi(\boldsymbol{A}|\boldsymbol{\mu}_A, \boldsymbol{R}_A) = \mathcal{N}(\boldsymbol{A}|\boldsymbol{\mu}_A, \boldsymbol{R}_A) \triangleq \prod_{i,j} \mathcal{N}\left(a_{ij}|\mu_{a_{ij}}, \sigma^2_{a_{ij}}\right) \tag{15}$$

3. The parameters $\left(\sigma_i^2, \{\gamma_j^j\}\right)$ are assigned an inverse gamma *prior* distribution (to encode their positif character):

$$\pi(x|\nu,\beta) = \mathcal{IG}(x|\nu,\beta) \propto \frac{e^{-\beta x}}{x^{\nu+1}} \tag{16}$$

   Indeed, this choice corresponds to the conjugate *prior*[10] and eliminates the degeneracy of the likelihood function for the Gaussian mixture model.[9]

The *posterior* distribution (Equation (5)) is then given by

$$p\left(\boldsymbol{s}^j(k), \boldsymbol{A}, \boldsymbol{\theta}|\boldsymbol{x}^j(k), \boldsymbol{s}^{P(j)}\right) \propto \mathcal{N}\left(\boldsymbol{x}^j(k)|\boldsymbol{A}\boldsymbol{s}^j(k), \boldsymbol{R}_{\boldsymbol{\epsilon}}\right)\mathcal{N}(\boldsymbol{A}|\boldsymbol{\mu}_A, \boldsymbol{R}_A)\,\pi\left(\boldsymbol{s}^j(k)\right)\mathcal{IG}(\boldsymbol{\theta}|\nu,\beta) \tag{17}$$

for $j = 1, \ldots, J$ and $k = 1, \ldots, T/2^j$.

We make use of an MCMC (Monte Carlo Markov Chain) algorithm to generate samples from the *posterior* distribution (Equation (17)). In what follows, we present the details of the developped algorithms for the estimation purposes. However, we essentially classify them into two algorithms, a Gibbs/Gibbs algorithm corresponding to an Independant Gaussian Mixture (IGM) *prior* model, and a hybrid Hastings-Metroplis/Gibbs algorithm corresponding to a Generalized Exponential (GE) *prior* model (independant or correlated).

4.1. Gibbs sampling

The sampling step of the sources coefficients $\boldsymbol{s}^j(k)$, the mixing matrix $\boldsymbol{A}$ and the parameters $\boldsymbol{\theta}$ is done in an alternate manner according to their conditionnal laws.

At iteration (i)

1. $\boldsymbol{s}^j(k)^{(i)} | \{\boldsymbol{A}^{(i-1)}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{x}^j(k)\} \sim \mathcal{N}\left(\boldsymbol{x}^j(k) | \boldsymbol{A}\boldsymbol{s}^j(k), \boldsymbol{R}_{\boldsymbol{\epsilon}}\right) \pi(\boldsymbol{S})$
Ref. to the source coefficients sampling step (section 4.2).

2. $\boldsymbol{A}^{(i)} | \{\boldsymbol{S}^{(i)}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{X}\} \sim \mathcal{N}(\boldsymbol{A} | \boldsymbol{\mu}, \boldsymbol{R})$
where

$$\boldsymbol{\mu} = \boldsymbol{R}\Big(\big(\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1} \otimes \mathbb{I}_n\big) \textstyle\sum_{j,k} C_{xs}(j,k) + \boldsymbol{\mu}_A\Big), \quad \boldsymbol{R} = \Big(\textstyle\sum_{j,k} \boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1} \otimes C_{ss}(j,k) + \boldsymbol{R}_A^{-1}\Big)^{-1},$$

$C_{ss}(j,k) = \boldsymbol{s}^j(k)\boldsymbol{s}^{j*}(k)$ and $C_{xs}(j,k) = \boldsymbol{x}^j(k) \otimes \boldsymbol{s}^j(k)$.

3. $\{\sigma_m^2\}^{(i)} | \{\boldsymbol{S}^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{X}\} \sim \mathcal{IG}\big(\nu', \beta'(m)\big)$
where

$$\nu' = T/2 + \nu \quad \text{and} \quad \beta'(m) = \Big(\tfrac{1}{2}\textstyle\sum_t \left(x_m(t) - [\boldsymbol{A}\boldsymbol{s}(t)]_m\right)^2 + \beta\Big)$$

For the sampling step of the parameters $\{\gamma^j\}_n$, we define two slightly different steps, one corresponding two an independant model, and two the other two the multi-resolution correlation model.

Independant model:

4. $\{\gamma_n^j\}^{(i)} | \{\boldsymbol{S}^{(i)}, \boldsymbol{A}^{(i)}\} \sim \mathcal{IG}\big(\nu'(j), \beta'(n,j)\big)$
where

$$\nu'(j) = \tfrac{T/2^j}{\alpha_j} + \nu \quad \text{and} \quad \beta'(j,n) = \big(\tfrac{1}{2}\textstyle\sum_k \left|s_n^j(k)\right|^{\alpha_j} + \beta\big)$$

Inter-scale correlation model:

4. $\{\gamma_n^j\}^{(i)} | \{\boldsymbol{S}^{(i)}, \boldsymbol{A}^{(i)}\} \sim \mathcal{IG}\big(\nu'(j), \beta'(n,j)\big)$
where

$$\nu'(j) = \tfrac{T/2^j}{\alpha_j} + \nu \quad \text{and} \quad \beta'(n,j) = \Big(\tfrac{1}{2}\textstyle\sum_k \left|s_n^j(k) - \delta_n^j(s^{P(j)})\right|^{\alpha_j} + \beta\Big)$$

4.2. Sources coefficients sampling step

a. Independant Gaussian Mixture model (IGM)

When the coefficients are modeled by an independant Gaussian mixture model (Equation (10)), a Gibbs samling algorithm is used:

To each coefficient $s_n^j(k)$, we associate a discrete hidden variable $z_n^j(k) \in \{1,2\}$, such that the *prior* model is now a conditional model given by:

$$\pi\left(s_n^j(k) \middle| z_n^j(k) = l_n\right) = \mathcal{N}\left(s_n^j(k) | 0, \tau_{l,n}\right), \quad l = 1,2 \tag{18}$$

At iteration (i)

1.1 $z_n^j(k)^{(i)} \sim \mathcal{M}_2(1; p, 1-p)$

1.2 $\boldsymbol{s}^j(k)^{(i)} \big| \boldsymbol{z}^j(k)^{(i)} = \boldsymbol{l} \sim \mathcal{N}\left(\boldsymbol{s}^j(k) | \boldsymbol{\mu}_z, \boldsymbol{R}_z\right)$
where

$$\boldsymbol{\mu}_z = \boldsymbol{R}_z \boldsymbol{A}^* \boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{x}^j(k), \quad \boldsymbol{R}_z = \left(\boldsymbol{A}^* \boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{A} + \boldsymbol{R}_l^{-1}\right)^{-1}$$

$\boldsymbol{l} = \operatorname{diag}(l_1, \ldots, l_N)$ and $\boldsymbol{R}_l = \operatorname{diag}\left(\tau_{l,1}^2, \ldots, \tau_{l,N}^2\right)$.

b. Genralized Exponential model (GE)

When the coefficients are modeled by generalized exponential (GE) *prior* distibutions, their sampling process is not straight forward since the conditionnal *posterior* law of the coefficients is a product of a Gaussian distribution (Equation (8)) with the GE *prior* law (Equation (9)). We use then a Hastings-Metropolis step:

First, we approximate the *prior* generalized exponential law by a Gaussian law:

$$\pi\left(\boldsymbol{s}^j(k)\right) \sim \tilde{\pi}\left(\boldsymbol{s}^j(k)\right) = \mathcal{N}\left(\boldsymbol{s}^j(k)|\boldsymbol{\mu}, \boldsymbol{R}_{\gamma_j}\right) \tag{19}$$

where $\boldsymbol{\mu} = \boldsymbol{\Phi}^j(\boldsymbol{s}^{P(j)})$ in the correlated case and $\boldsymbol{\mu} = 0$ in the independant case and $\boldsymbol{\Phi}^j = \mathrm{diag}\left(\phi_1^j, \ldots, \phi_N^j\right)$.

The *posterior* approximate law is then given by:

$$\tilde{p}(\boldsymbol{s}_j(k)) \propto \mathcal{N}\left(\boldsymbol{s}^j(k)|\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{R}}\right) \tag{20}$$

where, in the correlated case:

$$\tilde{\boldsymbol{R}} = \left(\boldsymbol{A}^*\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A} + \boldsymbol{R}_{\gamma_j}^{-1} + \boldsymbol{\Phi}^{j+1^*}\boldsymbol{R}_{\gamma_{j+1}}^{-1}\boldsymbol{\Phi}^{j+1}\right)^{-1},$$

$$\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{R}}\left(\boldsymbol{A}^*\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{x}^j(k) + \boldsymbol{R}_{\gamma_j}^{-1}\boldsymbol{\Phi}^j\boldsymbol{s}^{P(j)} + \boldsymbol{\Phi}^{j+1^*}\boldsymbol{R}_{\gamma_{j+1}}^{-1}\left(\boldsymbol{s}^{j+1}(\kappa) - \boldsymbol{\Phi}^{j+1}\boldsymbol{s}^{N(j)}\right)\right)$$

The expressions of $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{R}}$ simplifies, in the independant case, to:

$$\tilde{\boldsymbol{R}} = \left(\boldsymbol{A}^*\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A} + \boldsymbol{R}_{\gamma_j}{}^{-1}\right)^{-1} \quad \text{and} \quad \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{R}}\boldsymbol{A}^*\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{x}^j(k)$$

The Hastings-Metropolis sampling step is given by:

`At iteration (i)`

`1.1` $\boldsymbol{y}|\{\boldsymbol{A}^{(i-1)}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{x}^j(k)\} = \boldsymbol{U}\boldsymbol{z} + \tilde{\boldsymbol{\mu}}$

`where`

$$\boldsymbol{z} \sim \prod_n \exp\left(-\frac{1}{2d_n}|z_n|\right)$$

$\tilde{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*, \boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_N)$

`1.2`

$$\boldsymbol{s}^j(k)^{(i)} = \begin{cases} \boldsymbol{y} & \text{with prob. } \rho, \\ \boldsymbol{s}^j(k)^{(i-1)} & \text{with prob. } 1-\rho \end{cases}$$

`with`

$$\rho = \left\{1 \wedge \left(\frac{p(\boldsymbol{y})}{g(\boldsymbol{y})} \Big/ \frac{p\left(\boldsymbol{s}^j(k)^{(i-1)}\right)}{g\left(\boldsymbol{s}^j(k)^{(i-1)}\right)}\right)\right\}$$

$$p\left(\boldsymbol{s}^j(k)|\boldsymbol{A}^{(i-1)}, \boldsymbol{\theta}^{(i-1)}\right) \propto \mathcal{N}\left(\boldsymbol{x}^j(k)|\boldsymbol{A}\boldsymbol{s}^j(k), \boldsymbol{R}_{\boldsymbol{\epsilon}}\right)\mathcal{E}\mathrm{xp}\left(\boldsymbol{s}^j(k)|\gamma_j^j, \alpha_j\right)$$

$$g\left(\boldsymbol{s}^j(k)\right) \propto \prod_n \exp\left(-\frac{1}{2d_n}\left|\boldsymbol{U}^*\left(\boldsymbol{s}^j(k) - \tilde{\boldsymbol{\mu}}\right)\right|_n\right)$$

5. SIMULATIONS

We have tested the algorithms detailed in the previous section to simulated data. The obtained results are presented in Fig. 5. Two images (Fig. 4.a) are mixed with a mixing matrix given by:

$$\boldsymbol{A} = \begin{bmatrix} 0.875 & 0.508 \\ 0.484 & 0.861 \end{bmatrix}$$

and a noise of 20dB is added to each image to obtain the images in Fig. 4.b. The estimated sources obtained according to the independant GE model are presented in Fig. 5.a, those obtained by taking into account an inter-scale correlation with a GE model are presented in Fig. 5.b, and finally, those obtained when the coefficients are modeled by IGM models are presented in Fig. 5.c.

To quantify the obtained results, we have chosen as a quality measure, a measure of the normed erreur given by:

$$P_\beta(\tilde{\boldsymbol{S}}) = \frac{\|\boldsymbol{S} - \tilde{\boldsymbol{S}}\|_\beta}{\|\boldsymbol{S}\|_\beta}, \quad 1 \le \beta \le 2 \tag{21}$$

We have presented the numerical results in Table 1. We notice that concerning the sources, it is not easy to say which model is better than the othe, however concerning the estimation of the mixing matrix, we can say that the IGM model gives a better result. Even visually (observing the sources), we have tendancy to say the IGM gives better estimates than the two other models.

a. b.

Figure 4. a. Source images, b. Mixed images

a. b. c.

Figure 5. a. Estimated images in the independant GE case, b. Estimated sources in the correlated GE case, c. Estimated sources in the IGM case

Table 1. Numrical simulation results

	$P_1(\tilde{\boldsymbol{S}})$		$P_2(\tilde{\boldsymbol{S}})$		$P_2(\tilde{\boldsymbol{A}})$
	Source 1	Source 2	Source 1	Source 2	
Independant GE model	0.129	0.146	0.148	0.167	0.037
Correlated GE model	0.134	0.147	0.155	0.166	0.021
IGM model	0.139	0.142	0.160	0.157	0.015

6. CONCLUDING REMARKS

In this work, we have proposed a Bayesian approach to BSS by assigning to the wavelet coefficients of the sources to estimate (signals/images) *prior* laws that try to encode the sparsity of the latter. In the GE models, we have even tried to encode some inter-correlation information of the multi-resolution representation of signals. We have proposed MCMC algorithms adapted to each case and presented the obtained results.

We think that the Gaussian mixture models encode better the sparsity property of the wavelet coefficients, and even from an algorithmic point of view, algorithms based on such models are more tractable than the generalized exponential models.

For futur works, we will be interested on Hidden Markov Models (HMM), which are extensions of the Independant Gaussian Mixture models used in this work. The HMM models have the ability to account for the inter-scale correlation more easily than the generalized exponential models and they have already proven their performances on treating complex signal problems.

REFERENCES

1. Matthiew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4), April 1998.
2. A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley, New York, 2001.
3. K. Knuth. A Bayesian approach to source separation. In *Proceedings of Independent Component Analysis Workshop*, pages 283–288, 1999.
4. Stephane G. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), july 1989.
5. Stephane G. Mallat. *a Wavelet Tour of Signal Processing*. Academic Press, 1999.
6. A. Mohammad-Djafari and Mahieddine M. Ichir. Wavelet domain image separation. In *Proceedings of American Institute of Physics*, MaxEnt2002, pages 208–223, Moscow, Idaho, USA, Aug. 2002. International Workshop on Bayesian and Maximum Entropy methods.
7. H. Snoussi, G. Patanchon, J.F. Macías-Pérez, A. Mohammad-Djafari, and J. Delabrouille. Bayesian blind component separation for cosmic microwave background observations. In Robert L. Fry, editor, *Bayesian Inference and Maximum Entropy Methods*, pages 125–140. MaxEnt Workshops, American Institute of Physics, August 2001.
8. Hichem Snoussi and Ali Mohammad-Djafari. Bayesian separation of HMM sources. In Robert L. Fry, editor, *Bayesian Inference and Maximum Entropy Methods*, pages 77–88. MaxEnt Workshops, American Institute of Physics, August 2001.
9. Hichem Snoussi and Ali Mohammad-Djafari. Dégénérescences des estimateurs MV en séparation de sources. Technical report ri-s0010, GPI–L2S, 2001.
10. Hichem Snoussi and Ali Mohammad-Djafari. Penalized maximum likelihood for multivariate gaussian mixture. In Robert L. Fry, editor, *Bayesian Inference and Maximum Entropy Methods*, pages 36–46. MaxEnt Workshops, American Institute of Physics, August 2001.
11. Hichem Snoussi and Ali Mohammad-Djafari. Bayesian unsupervised learning for source separation with mixture of gaussians prior. *To appear in Int. Journal of VLSI Signal Processing Systems*, 2003.
12. Michael Zibulevsky and Barak A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionnary. *MIT Letters on Neural Computation*, 13:863–882, 2001.