

# Fast joint separation and segmentation of mixed images

Hichem Snoussi\* and Ali Mohammad-Djafari\*

*\*Laboratoire des Signaux et Systèmes (L2S),  
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France*

**Abstract.** We consider the problem of the blind separation of noisy instantaneously mixed images. The images are modeled by hidden Markov fields with unknown parameters. Given the observed images, we give a Bayesian formulation and we propose a fast version of the MCMC algorithm based on the Bartlett decomposition for the resulting data augmentation problem. We separate the unknown variables into two categories:

1. The parameters of interest which are the mixing matrix, the noise covariance and the parameters of the sources distributions.
2. The hidden variables which are the unobserved sources and the unobserved pixels segmentation labels.

The proposed algorithm provides, in the stationary regime, samples drawn from the posterior distributions of all the variables involved in the problem leading to great flexibility in the cost function choice.

Finally, we show the results for both synthetic and real data to illustrate the feasibility of the proposed solution.

## I. INTRODUCTION AND MODEL ASSUMPTIONS

The observations are  $m$  images  $(\mathbf{X}^i)_{i=1..m}$ , each image  $\mathbf{X}^i$  is defined on a finite set of sites  $\mathcal{S}$ , corresponding to the pixels of the image:  $\mathbf{X}^i = (x_r^i)_{r \in \mathcal{S}}$ . The observations are noisy linear instantaneous mixture of  $n$  source images  $(\mathbf{S}^j)_{j=1..n}$  defined on the same set  $\mathcal{S}$ :

$$x_r^i = \sum_{j=1}^n a_{ij} s_r^j + n_r^i, \quad r \in \mathcal{S}, i = 1..m$$

where  $\mathbf{A} = (a_{ij})$  is the unknown mixing matrix,  $\mathbf{N}^i = (n_r^i)_{r \in \mathcal{S}}$  is a zero-mean white Gaussian noise with variance  $\sigma_{\epsilon_i}^2$ . At each site  $r \in \mathcal{S}$ , the matrix notation is:

$$\mathbf{x} = \mathbf{A} \mathbf{s} + \mathbf{n} \tag{1}$$

The noise and source components  $(\mathbf{N}^i)_{i=1..m}$  and  $(\mathbf{S}^j)_{j=1..n}$  are supposed to be independent. However, the noise can be correlated across detectors that is the noise covariance matrix  $\mathbf{R}_\epsilon = \mathbb{E}[\mathbf{n} \mathbf{n}^*]$  is not necessarily diagonal.

Many techniques have been proposed for the source separation problem based on entropy and information theoretic approach [1, 2, 3, 4, 5] and the maximum likelihood principle [6, 7, 8, 9, 10, 11, 12] leading to contrast functions [13, 14, 15, 16] and estimating functions [17, 18, 19, 20]. Among the limitations of these methods, we

can mention: **(i)** the lack of possibility to account for some prior information about the mixing coefficients or other parameter involved in the problem, **(ii)** the lack of information about the degree of uncertainty of the mixing matrix estimate particularly in the noisy mixture, **(iii)** the objective functions are intractable or difficult to optimize when the source model is more elaborate ...

Recently, a few works using the Bayesian approach have been presented to push further the limits of these methods [21, 22, 23, 24, 2, 25, 26, 27, 28, 29]. For example, in the Bayesian framework, we can introduce some *a priori* information on the sources and on the mixing elements as well as on the hyperparameters by assigning appropriate prior laws for them. Also, thanks to the posterior laws, we can quantify the uncertainty of any estimated parameter. Finally, thanks to sampling schemes, we can propose tractable estimation algorithms.

In previous works, we have assumed Gaussian mixture models for sources where the labels are white [30, 31]. However, this model does not take into account the temporal correlation of the sources. An extension to the hidden Markov models HMM is considered in the one dimensional case [29] and its formulation in the two dimensional case (hidden Markov fields HMF) seems to be appropriate in image separation applications. The main objective of this paper is to study the image separation problem using the HMF model. Each source is modeled by a double stochastic process  $(\mathbf{S}^j, \mathbf{Z}^j)$ .  $\mathbf{S}^j$  is a field of values in a continuous set  $\mathcal{R}$  and represents the real observed image in the absence of noise and mixing deformation.  $\mathbf{Z}^j$  is the hidden Markov field representing the unobserved pixel classification whose components are in a discrete set,  $\mathbf{Z}_r^j \in \{1..K^j\}$ . The joint probability distribution of  $\mathbf{Z}^j$  satisfies the following properties,

$$\begin{cases} \forall \mathbf{Z}^j, & P_M(z_r^j | \mathbf{Z}_{\mathcal{S} \setminus \{r\}}^j) = P_M(z_r^j | \mathbf{Z}_{N(r)}^j) \\ \forall \mathbf{Z}^j, & P_M(\mathbf{Z}^j) > 0 \end{cases}$$

where  $\mathbf{Z}_{\mathcal{S} \setminus \{r\}}^j$  denotes the field restricted to  $\mathcal{S} \setminus \{r\} = \{l \in \mathcal{S}, l \neq r\}$  and  $N(r)$  denotes the set of neighbors of  $r$ , according to the neighborhood system defined on  $\mathcal{S}$  for each source component. According to the Hammersley-Clifford theorem, there is an equivalence between a Markov random field and a Gibbs distribution,

$$P_M(\mathbf{Z}^j) = [W(\alpha_j)]^{-1} \exp\{-H_{\alpha_j}(\mathbf{Z}^j)\}$$

where  $H_{\alpha_j}$  is the energy function and  $\alpha_j$  is a parameter weighting the spatial dependencies supposed to be known.

Conditionally to the hidden discrete field  $\mathbf{Z}^j$ , the source pixels  $S_r^j, r \in \mathcal{S}$  are supposed to be independent and have the following conditional distribution:

$$p(\mathbf{S}^j | \mathbf{Z}^j, \boldsymbol{\eta}^j) = \prod_{r \in \mathcal{S}} p_r(s_r^j | z_r^j, \boldsymbol{\eta}^j)$$

where the positive conditional distributions depend on the parameter  $\boldsymbol{\eta}^j \in \mathcal{R}^d$ . We assume in the following that  $p_r(\cdot | z)$  is a Gaussian distribution with parameters  $\boldsymbol{\eta}^j = (\mu_{jk}, \sigma_{jk}^2)_{k=1..K}$ .

We note that we have a two-level inverse problem:

1. The problem described by (1) when the mixing matrix  $\mathbf{A}$  is unknown which is the source separation problem [32, 33, 27].
2. Given the source component  $\mathbf{S}^j$ , the estimation of the parameter  $\boldsymbol{\eta}^j$  and the recovering of the hidden classification labels  $\mathbf{Z}^j$  which is known as the unsupervised segmentation [34].

In this contribution, given the observations  $\mathbf{X}^i (i = 1..m)$  we propose a solution to jointly separate the  $n$  unknown sources and perform their unsupervised segmentations. In section II, we give a Bayesian formulation of the problem. In section III, we propose an original construction of the parameter *a priori*. In section IV, a fast implementation of an MCMC algorithm based on the data augmentation modelization is proposed. In section V, numerical simulations on synthetic and real data are shown to illustrate the feasibility of the solution.

## II. BAYESIAN FORMULATION

Given the observed data  $\mathbf{X} = (\mathbf{X}^1, \dots, \mathbf{X}^m)$ , our objective is the estimation of the mixing matrix  $\mathbf{A}$ , the noise covariance  $\mathbf{R}_\epsilon$ , the means and variances  $(\mu_{jk}, \sigma_{jk}^2)_{j=1..n, k=1..K}$  of the conditional Gaussians of the prior distribution of the sources. The *a posteriori* distribution of the whole parameter  $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{R}_\epsilon, \mu_{jk}, \sigma_{jk}^2)$  contains all the information that we can extract from the data. According to the Bayesian rule, we have

$$p(\boldsymbol{\theta} | \mathbf{X}) \propto p(\mathbf{X} | \boldsymbol{\theta})p(\boldsymbol{\theta})$$

In the section III, we will discuss the attribution of appropriate prior distribution  $p(\boldsymbol{\theta})$ . Concerning the likelihood, it has the following expression,

$$\begin{aligned} p(\mathbf{X} | \boldsymbol{\theta}) &= \sum_{\mathbf{Z}} \int_{\mathbf{S}} p(\mathbf{X}, \mathbf{S}, \mathbf{Z} | \boldsymbol{\theta}) d\mathbf{S} \\ &= \sum_{\mathbf{Z}} \left\{ \prod_{r \in \mathcal{S}} \mathcal{N}(\mathbf{x}_r; \mathbf{A}\boldsymbol{\mu}_{z_r}, \mathbf{A}\mathbf{R}_{z_r}\mathbf{A}^* + \mathbf{R}_\epsilon) \right\} P_M(\mathbf{Z}) \end{aligned} \quad (2)$$

where  $\mathcal{N}$  denotes the Gaussian distribution,  $\mathbf{x}_r$  the  $(m \times 1)$  vector of observations on the site  $r$ ,  $\mathbf{z}_r$  is the vector label,  $\boldsymbol{\mu}_{z_r} = [\mu_{1z_1}, \dots, \mu_{nz_n}]^t$  and  $\mathbf{R}_{z_r}$  the diagonal matrix  $\text{diag}[\sigma_{1z_1}^2, \dots, \sigma_{nz_n}^2]$ . We note that the expression (2) has not a tractable form with respect to the parameter  $\boldsymbol{\theta}$  because of the integration of the hidden variables  $\mathbf{S}$  and  $\mathbf{Z}$ . This remark leads us to consider the data augmentation algorithm [35] where we complete the observations  $\mathbf{X}$  by the hidden variables  $(\mathbf{Z}, \mathbf{S})$ , the complete data are then  $(\mathbf{X}, \mathbf{S}, \mathbf{Z})$ . In a previous work [36], we implemented restoration-maximization algorithms in the one dimensional case to estimate the maximum *a posteriori* estimate of  $\boldsymbol{\theta}$ . We extend this work in two directions: **(i)** first, the sources are two-dimensional signals, **(ii)** second, we implement an MCMC algorithm to obtain samples of  $\boldsymbol{\theta}$  drawn from their *a posteriori*

distribution. This gives the possibility of not being restricted to estimate the parameter by its maximum *a posteriori*, we can consider another cost function and compute the corresponding estimate.

### III. A PRIORI SELECTION

In the following, our parameter of interest is  $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{R}_\epsilon, \boldsymbol{\eta})$ .  $\mathbf{A}$  is the mixing matrix,  $\mathbf{R}_\epsilon$  the noise covariance and  $\boldsymbol{\eta}$  contains all the parameters of the sources model:

$$\begin{cases} \boldsymbol{\eta}^j = (\boldsymbol{\eta}_k^j)_{k=1..K_j} \\ \boldsymbol{\eta}_k^j = (\mu_k^j, v_k^j = (\sigma_k^j)^2) \end{cases}$$

where the index  $j$  indicates the source  $j$  and  $k$  indicates the Gaussian component  $k$  of the distribution of the source  $j$ . Our objective is the computation of the  $\delta$  priors [37]. We have an incomplete data problem with two hierarchies of hidden variables, the sources  $\mathbf{s}_{1..T}$  and the labels  $\mathbf{z}_{1..T}$  so that the complete data are  $(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T})$ . The  $\delta$ -prior has the following form:

$$\Pi(\boldsymbol{\theta}) \propto e^{-\frac{\gamma_e}{\gamma_u} D_\delta(p_\theta, p_0)} \sqrt{\|g(\boldsymbol{\theta})\|} \quad (3)$$

where  $\frac{\gamma_e}{\gamma_u}$  is the tradeoff between the confidence degree  $\gamma_e$  in the reference distribution  $p_0$  and the uniformity degree  $\gamma_u$ .  $g(\boldsymbol{\theta})$  is the Fisher information matrix and  $D_\delta$  is the  $\delta$ -divergence [38]:

$$D_\delta(p, q) = \frac{\int p}{1-\delta} + \frac{\int q}{\delta} - \frac{\int p^\delta q^{1-\delta}}{\delta(1-\delta)}$$

We assume that the reference distribution  $p_0$  belongs to the parametric family  $\{p_\theta\}$  so that it is defined by the reference parameters  $\boldsymbol{\theta} = (\mathbf{A}^0, \mathbf{R}_\epsilon^0, \boldsymbol{\eta}^0)$ . The expressions of the Fisher matrix and the  $\delta$ -divergence are intractable for the incomplete model. Consequently, they are approximated in the following by their expression in the complete model case.

We begin by the computation of the Fisher information matrix.

#### a/ Fisher information matrix

The Fisher matrix  $g(\boldsymbol{\theta})$  is defined as:

$$g_{ij}(\boldsymbol{\theta}) = - \underset{\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T}}{E} \left[ \frac{\partial^2}{\partial_i \partial_j} \log p(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T} | \boldsymbol{\theta}) \right]$$

The factorization of the joint distribution  $p(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T} | \boldsymbol{\theta})$  as:

$$p(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T} | \boldsymbol{\theta}) = p(\mathbf{x}_{1..T} | \mathbf{s}_{1..T}, \mathbf{z}_{1..T}, \boldsymbol{\theta}) p(\mathbf{s}_{1..T} | \mathbf{z}_{1..T}, \boldsymbol{\theta}) p(\mathbf{z}_{1..T} | \boldsymbol{\theta})$$

and the corresponding expectations as

$$\underset{\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T}}{E} [\cdot] = \underset{\mathbf{z}_{1..T}}{E} \left[ \underset{\mathbf{s}_{1..T} | \mathbf{z}_{1..T}}{E} \left[ \underset{\mathbf{x}_{1..T} | \mathbf{s}_{1..T}, \mathbf{z}_{1..T}}{E} [\cdot] \right] \right]$$

and taking into account the conditional independences  $((\mathbf{x}_{1..T} | \mathbf{s}_{1..T}, \mathbf{z}_{1..T}) \Leftrightarrow (\mathbf{x}_{1..T} | \mathbf{s}_{1..T})$  and  $(\mathbf{s}_{1..T} | \mathbf{z}_{1..T}) \Leftrightarrow \prod s_{1..T}^j | \mathbf{z}_{1..T}^j$ , the Fisher information matrix will have a block diagonal structure as follows:

$$g(\boldsymbol{\theta}) = \begin{bmatrix} g(\mathbf{A}, \mathbf{R}_\epsilon) & \dots & [0] \\ \vdots & g(\boldsymbol{\eta}^1) & \\ & & \ddots \\ [0] & & \dots & g(\boldsymbol{\eta}^n) \end{bmatrix}$$

### a.1/ $(\mathbf{A}, \mathbf{R}_\epsilon)$ -block

The Fisher information matrix of  $(\mathbf{A}, \mathbf{R}_\epsilon)$  is:

$$\mathcal{F}_{ij}(\mathbf{A}, \mathbf{R}_\epsilon) = -E_{\mathbf{s}} E_{\mathbf{x} | \mathbf{s}} \left[ \frac{\partial^2}{\partial_i \partial_j} \log p(\mathbf{x}_{1..T} | \mathbf{s}_{1..T}, \mathbf{A}, \mathbf{R}_\epsilon) \right]$$

which is very similar to the Fisher information matrix of the mean and covariance of a multivariate Gaussian distribution. The obtained expression is

$$g(\mathbf{A}, \mathbf{R}_\epsilon) = \begin{bmatrix} \left( E_{\mathbf{s}_{1..T}} \mathbf{R}_{ss} \right) \otimes \mathbf{R}_\epsilon^{-1} & [0] \\ [0] & -\frac{1}{2} \frac{\partial \mathbf{R}_\epsilon^{-1}}{\partial \mathbf{R}_\epsilon} \end{bmatrix}$$

where  $\mathbf{R}_{ss} = \frac{1}{T} \sum \mathbf{s}_t \mathbf{s}_t^*$  and  $\otimes$  is the Kronecker product.

We note the block diagonality of the  $(\mathbf{A}, \mathbf{R}_\epsilon)$ -Fisher matrix. The term corresponding to the mixing matrix  $\mathbf{A}$  is the signal to noise ratio as can be expected. Thus, the amount of information about the mixing matrix is proportional to the signal to noise ratio. The induced volume of  $(\mathbf{A}, \mathbf{R}_\epsilon)$  is then:

$$|g(\mathbf{A}, \mathbf{R}_\epsilon)|^{1/2} d\mathbf{A} d\mathbf{R}_\epsilon = \frac{|\mathbf{R}_{ss}|^{m/2}}{|\mathbf{R}_\epsilon|^{\frac{m+n+1}{2}}} \eta d\mathbf{A} d\mathbf{R}_\epsilon$$

### a.2/ $(\boldsymbol{\eta}^j)$ -block

Each  $g(\boldsymbol{\eta}^j)$  is the Fisher information of a one-dimensional Gaussian distribution (see [29] for details).

$$|g(\boldsymbol{\eta}^j)|^{1/2} d\boldsymbol{\eta}^j = \prod_{k=1}^{K_j} \frac{1}{v_k^{3/2}} d\boldsymbol{\eta}^j$$

### b/ $\delta$ -Divergence ( $\delta = 0$ )

In this paper, we fix the value of  $\delta$  to 0. The 0-divergence between two parameters  $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{R}_\epsilon, \boldsymbol{\eta})$  and  $\boldsymbol{\theta}^0 = (\mathbf{A}^0, \mathbf{R}_\epsilon^0, \boldsymbol{\eta}^0)$  for the complete data likelihood  $p(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T} | \boldsymbol{\theta})$

is:

$$D_0(\boldsymbol{\theta} : \boldsymbol{\theta}^0) = E_{x,s,z|\boldsymbol{\theta}^0} \log \frac{p(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T} | \boldsymbol{\theta}^0)}{p(\mathbf{x}_{1..T}, \mathbf{s}_{1..T}, \mathbf{z}_{1..T} | \boldsymbol{\theta})}$$

Similar developments of the above equation as in the computation of the Fisher matrix based on the conditional independences, lead to an affine form of the divergence, which is a sum of the expected divergence between the  $(\mathbf{A}, \mathbf{R}_\epsilon)$  parameters and the divergence between the sources parameters  $\boldsymbol{\eta}$ :

$$D_0(\boldsymbol{\theta} : \boldsymbol{\theta}^0) = E_{s|\boldsymbol{\eta}^0} D_0(\mathbf{A}, \mathbf{R}_\epsilon : \mathbf{A}^0, \mathbf{R}_\epsilon^0) + D_0(\boldsymbol{\eta} : \boldsymbol{\eta}^0)$$

where  $D_0$  means the divergence between the distributions  $p(\mathbf{x}_{1..T} | \mathbf{A}, \mathbf{R}_\epsilon, \mathbf{s}_{1..T})$  and  $p(\mathbf{x}_{1..T} | \mathbf{A}^0, \mathbf{R}_\epsilon^0, \mathbf{s}_{1..T})$  keeping the sources  $\mathbf{s}_{1..T}$  fixed.

The 0-divergence between  $\boldsymbol{\eta}$  and  $\boldsymbol{\eta}^0$  is the sum of the 0-divergences between each source parameter  $\boldsymbol{\eta}^j$  and  $\boldsymbol{\eta}^{0j}$  due to the *a priori* independence between the sources. In the following, we omit the superscript  $j$  referring to the source  $j$  to have clear expressions. The divergence between  $\boldsymbol{\eta}$  and  $\boldsymbol{\eta}^0$  is obtained as a particular case ( $n = 1$ ) of the general expression derived in the multivariate case in [29] leading to a Normal Inverse Gamma prior:

$$\Pi_0(\boldsymbol{\eta}) = \prod_{k=1}^K \Pi_0(\boldsymbol{\eta}_k) = \prod_{k=1}^K \mathcal{N}(\mu_k; \mu^0, \frac{v_k}{\nu^0}) \mathcal{G}(v_k^{-1}; \frac{\nu^0}{2}, \frac{\nu^0}{2} v^0) \quad (4)$$

with  $\nu^0 = \alpha w_k^0$ ,  $\alpha = \frac{\gamma_\epsilon}{\gamma_u}$ ,  $w_k^0$  is the marginal probability of the label  $k$  and  $\mathcal{G}(\cdot)$  the Gamma distribution:

$$\mathcal{G}(x | d, \beta) \propto x^{d-1} \exp[-\beta x]$$

The expressions of the averaged divergences between the  $(\mathbf{A}, \mathbf{R}_\epsilon)$  parameters are:

$$\begin{aligned} E_{s|\boldsymbol{\eta}^0} D_0(\mathbf{A}, \mathbf{R}_\epsilon : \mathbf{A}^0, \mathbf{R}_\epsilon^0) &= \frac{1}{2} \left( \log |\mathbf{R}_\epsilon \mathbf{R}_\epsilon^{0-1}| + \text{Tr}(\mathbf{R}_\epsilon^{-1} \mathbf{R}_\epsilon^0) \right. \\ &\quad \left. + \text{Tr} \left( \mathbf{R}_\epsilon^{-1} (\mathbf{A} - \mathbf{A}^0) E_{s|\boldsymbol{\eta}^0} [\mathbf{R}_{ss}] (\mathbf{A} - \mathbf{A}^0)^* \right) \right) \end{aligned}$$

leading to the following 0-priors on  $(\mathbf{A}, \mathbf{R}_\epsilon)$ :

$$\Pi_0(\mathbf{A}, \mathbf{R}_\epsilon^{-1}) = \mathcal{N} \left( \mathbf{A}; \mathbf{A}^0, \frac{1}{\alpha} \mathbf{R}_{ss}^0^{-1} \otimes \mathbf{R}_\epsilon \right) \mathcal{W}_{im} \left( \mathbf{R}_\epsilon^{-1}; \alpha, \mathbf{R}_\epsilon^{0-1} \right) | E_{s|\boldsymbol{\eta}} [\mathbf{R}_{ss}] |^{\frac{m}{2}}$$

where  $\mathbf{R}_{ss}^0 = E_{s|\boldsymbol{\eta}^0} \mathbf{R}_{ss}$  and  $\mathcal{W}_n$  is the wishart distribution of an  $n \times n$  matrix:

$$\mathcal{W}_n(\mathbf{R}; \nu, \boldsymbol{\Sigma}) \propto |\mathbf{R}|^{\frac{\nu-(n+1)}{2}} \exp \left[ -\frac{\nu}{2} \text{Tr}(\mathbf{R} \boldsymbol{\Sigma}^{-1}) \right]$$

Therefore, the 0-prior is a normal inverse Wishart prior (conjugate prior). The mixing matrix and the noise covariance are not *a priori* independent. In fact, the covariance

matrix of  $\mathbf{A}$  is the noise to signal ratio  $\frac{1}{\alpha} \mathbf{R}_{ss}^0{}^{-1} \otimes \mathbf{R}_\epsilon$ . We note a multiplicative term which is a power of the determinant of the *a priori* expectation of the source covariance  $E[\mathbf{R}_{ss}]$ . This term can be injected in the prior  $p(\boldsymbol{\eta})$  and thus the  $(\mathbf{A}, \mathbf{R}_\epsilon)$  parameters and the  $\boldsymbol{\eta}$  parameters are *a priori* independent.

We note that the precision matrix for the mixing matrix  $\mathbf{A}$  ( $\alpha \mathbf{R}_{ss}^0 \otimes \mathbf{R}_\epsilon^{-1}$ ) for  $\Pi_0$  is the product of the confidence term  $\alpha = \frac{\gamma_e}{\gamma_u}$  in the reference parameters and the signal to noise ratio. Therefore, the resulting precision of the reference matrix  $\mathbf{A}^0$  is not only our *a priori* coefficient  $\gamma_e$  but the product of this coefficient and the signal to noise ratio.

## IV. MCMC IMPLEMENTATION

We divide the vector of unknown variables into two sub-vectors: The hidden variables  $(\mathbf{Z}, \mathbf{S})$  and the parameter  $\boldsymbol{\theta}$  and we consider a Gibbs sampler:

repeat until convergence,

1. draw  $(\tilde{\mathbf{Z}}^{(h)}, \tilde{\mathbf{S}}^{(h)}) \sim p(\mathbf{Z}, \mathbf{S} | \mathbf{X}, \tilde{\boldsymbol{\theta}}^{(h-1)})$
2. draw  $\tilde{\boldsymbol{\theta}}^{(h)} \sim p(\boldsymbol{\theta} | \mathbf{X}, \tilde{\mathbf{Z}}^{(h)}, \tilde{\mathbf{S}}^{(h)})$

This Bayesian sampling [39] produces a Markov chain  $(\tilde{\boldsymbol{\theta}}^{(h)})$ , ergodic with stationary distribution  $p(\boldsymbol{\theta} | \mathbf{X})$ . After  $h_0$  iterations (warming up), the samples  $(\tilde{\boldsymbol{\theta}}^{(h_0+h)})$  can be considered to be drawn approximately from their *a posteriori* distribution  $p(\boldsymbol{\theta} | \mathbf{X})$ . Then, by the ergodic theorem, we can approximate *a posteriori* expectations by empirical expectations:

$$E[f(\boldsymbol{\theta}) | \mathbf{X}] \approx \frac{1}{H} \sum_{h=1}^H f(\tilde{\boldsymbol{\theta}}^{(h_0+h)}) \quad (5)$$

**Sampling  $(\mathbf{Z}, \mathbf{S})$ :** The sampling of the hidden fields  $(\mathbf{Z}, \mathbf{S})$  from  $p(\mathbf{Z}, \mathbf{S} | \mathbf{X}, \boldsymbol{\theta})$  is obtained by,

1. draw  $\tilde{\mathbf{Z}}$  from

$$p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\theta}) \propto p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\theta}) P_M(\mathbf{Z})$$

In this expression, we have two kinds of dependencies: (i)  $\mathbf{Z}$  are independent across components,  $p(\mathbf{Z}) = \prod_{j=1}^n p(\mathbf{Z}^j)$  but each discrete image  $\mathbf{Z}^j \sim P_M(\mathbf{Z}^j)$  has a Markovian structure. (ii) Given  $\mathbf{Z}$ , the fields  $\mathbf{X}$  are independent through the set  $\mathcal{S}$ ,  $p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\theta}) = \prod_{r \in \mathcal{S}} p(\mathbf{x}_r | \mathbf{z}_r, \boldsymbol{\theta})$  but dependent through the components because of the mixing operation  $p(\mathbf{x}_r | \mathbf{z}_r, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_r; \mathbf{A}\boldsymbol{\mu}_{\mathbf{z}_r}, \mathbf{A}\mathbf{R}_{\mathbf{z}_r}\mathbf{A}^* + \mathbf{R}_\epsilon)$  where  $\mathbf{z}_r$  is the vector label on the site  $r$ ,  $\boldsymbol{\mu}_{\mathbf{z}_r} = [\mu_{1z_1}, \dots, \mu_{nz_n}]^t$  and  $\mathbf{R}_{\mathbf{z}_r}$  the diagonal matrix  $\text{diag}[\sigma_{1z_1}^2, \dots, \sigma_{nz_n}^2]$ .

2. draw  $\tilde{\mathbf{S}} | \tilde{\mathbf{Z}}$  from

$$p(\mathbf{S} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) = \prod_{r \in \mathcal{S}} \mathcal{N}(\mathbf{s}_r; \mathbf{m}_r^{\text{apost}}, \mathbf{V}_r^{\text{apost}})$$

where the *a posteriori* mean and covariance are easily computed [40],

$$\begin{aligned} \mathbf{V}_r^{apost} &= [\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{R}_{z_r}^{-1}]^{-1} \\ \mathbf{m}_r^{apost} &= \mathbf{V}_r^{apost} (\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{x}_r + \mathbf{R}_{z_r}^{-1} \boldsymbol{\mu}_{z_r}) \end{aligned}$$

**Sampling  $\boldsymbol{\theta}$ :** Given the observations  $\mathbf{X}$  and the samples  $(\mathbf{Z}, \mathbf{S})$ , the sampling of the parameter  $\boldsymbol{\theta}$  becomes an easy task (this represents the principal reason of introducing the hidden sources). The conditional distribution  $p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{Z}, \mathbf{S})$  is factorized into two conditional distributions,

$$p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{Z}, \mathbf{S}) \propto p(\mathbf{A}, \mathbf{R}_\epsilon | \mathbf{X}, \mathbf{S}) p(\boldsymbol{\mu}, \boldsymbol{\sigma} | \mathbf{S}, \mathbf{Z})$$

leading to a separate sampling of  $(\mathbf{A}, \mathbf{R}_\epsilon)$  and  $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ . Choosing the 0-priors developed in the previous section, the *a posteriori* distributions are:

- Inverse Wishart for the noise covariance and Inverse Gamma for sources variances.
- Normal for the mixing matrix and for the sources means.

The expressions of these distributions are developed in the appendix A. We give below the expressions for  $(\mathbf{A}, \mathbf{R}_\epsilon)$  in the particular case when  $\alpha = 0$  (Jeffreys prior):

$$\begin{cases} \mathbf{R}_\epsilon^{-1} \sim \text{Wishart}(\nu_p, \boldsymbol{\Sigma}_p), \nu_p = \frac{|\mathcal{S}|-n}{2}, \boldsymbol{\Sigma}_p = \frac{|\mathcal{S}|}{2} (\mathbf{R}_{xx} - \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1} \mathbf{R}_{xs}^*) \\ p(\mathbf{A} | \mathbf{R}_\epsilon) \sim \mathcal{N}(\mathbf{A}_p, \boldsymbol{\Gamma}_p), \mathbf{A}_p = \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1}, \boldsymbol{\Gamma}_p = \frac{1}{|\mathcal{S}|} \mathbf{R}_{ss}^{-1} \otimes \mathbf{R}_\epsilon \end{cases} \quad (6)$$

where we define the empirical statistics  $\mathbf{R}_{xx} = \frac{1}{|\mathcal{S}|} \sum_r \mathbf{x}_r \mathbf{x}_r^*$ ,  $\mathbf{R}_{xs} = \frac{1}{|\mathcal{S}|} \sum_r \mathbf{x}_r \mathbf{s}_r^*$  and  $\mathbf{R}_{ss} = \frac{1}{|\mathcal{S}|} \sum_r \mathbf{s}_r \mathbf{s}_r^*$  (the sources  $\mathbf{S}$  are generated in the first step of the Gibbs sampling). We note that the covariance matrix of  $\mathbf{A}$  is proportional to the noise to signal ratio. This explains the fact noted in [41] concerning the slow convergence of the EM algorithm.

## FAST MCMC IMPLEMENTATION

A critical aspect of the above implementation is the computational cost of the sampling steps. Indeed, the convergence of the MCMC sampling may require a great number of iterations to ensure the convergence. Therefore, we need fast steps in the proposed algorithm to obtain a great number of iterations with a reasonable computational cost.

We investigated this direction by avoiding the sources sampling. In fact, the sources  $\mathbf{S}$  are sampled in the MCMC algorithm but only the statistics  $\mathbf{R}_{xs}$  and  $\mathbf{R}_{ss}$  are used in the generation of the parameters  $(\mathbf{A}, \mathbf{R}_\epsilon)$  (see equation (6)). Therefore we avoid the sampling of the sources  $\mathbf{S}$  and we sample directly the statistic matrices  $\mathbf{R}_{xs}$  and  $\mathbf{R}_{ss}$ . We show in the following how these simulations are easily performed in our problem formulation.

After the drawing of the labels  $\mathbf{Z}$ , the multidimensional source images  $\mathbf{S}$  are classified into  $K = K_1 \times \dots \times K_n$  regions  $(\mathcal{S}_z)_{z=1..K}$  defined by:

$$\mathcal{S}_z = \{r \in \mathcal{S} | Z(r) = z\}$$



In each region  $\mathcal{S}_z$ , the sources are Gaussians with mean and covariance:

$$\begin{aligned} \mathbf{V}_z &= [\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{R}_z^{-1}]^{-1} \\ \mathbf{m}_z &= \mathbf{V}_z (\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{x}_r + \mathbf{R}_z^{-1} \boldsymbol{\mu}_z) \end{aligned} \quad (7)$$

We then define the statistic matrices  $\mathbf{R}_{xs}^{(z)}$  and  $\mathbf{R}_{ss}^{(z)}$  on the region  $\mathcal{S}_z$  as:

$$\begin{aligned} \mathbf{R}_{xs}^{(z)} &= \frac{1}{|\mathcal{S}_z|} \sum_{r \in \mathcal{S}_z} \mathbf{x}_r \mathbf{s}_r^* \\ \mathbf{R}_{ss}^{(z)} &= \frac{1}{|\mathcal{S}_z|} \sum_{r \in \mathcal{S}_z} \mathbf{s}_r \mathbf{s}_r^* \\ \mathbf{R}_{xx}^{(z)} &= \frac{1}{|\mathcal{S}_z|} \sum_{r \in \mathcal{S}_z} \mathbf{x}_r \mathbf{x}_r^* \end{aligned} \quad (8)$$

From the expressions (7) and (8) and some algebraic manipulations, the statistics  $\mathbf{R}_{xs}^{(z)}$  and  $\mathbf{R}_{ss}^{(z)}$  can be decomposed as follows:

$$\begin{aligned} \mathbf{R}_{xs}^{(z)} &= \mathbf{R}_1 + \mathbf{U}_{n,1}^* \mathbf{C}_z^* \\ \mathbf{R}_{ss}^{(z)} &= \mathbf{R}_2 + \mathbf{V}_z (\mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{U}_{n,1}^* + \mathbf{U}_{n,2}^*) \mathbf{C}_z^* \\ &\quad \mathbf{C}_z (\mathbf{U}_{n,1} \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{U}_{n,2}) + \mathbf{C}_z \mathbf{U}_w \mathbf{C}_z^* \end{aligned}$$

where  $\mathbf{V}_z = \mathbf{C}_z \mathbf{C}_z^*$ . The matrices  $\mathbf{R}_1$  and  $\mathbf{R}_2$  are not random matrices and are updated at each iteration. The matrices  $\mathbf{U}_{n,1}$ ,  $\mathbf{U}_{n,2}$  and  $\mathbf{U}_w$  are random matrices and have the following distributions:

$$\begin{aligned} \mathbf{U}_{n,1} &\sim \mathcal{N}(0, \frac{1}{|\mathcal{S}_z|} \mathbf{R}_{xx}^{(z)} \otimes \mathbf{I}_n) \\ \mathbf{U}_{n,2} &\sim \mathcal{N}(0, \frac{1}{|\mathcal{S}_z|} \mathbf{R}_z^{-1} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^* \mathbf{R}_z^{-1} \otimes \mathbf{I}_n) \\ \mathbf{U}_w &\sim \mathcal{W}i_n(|\mathcal{S}_z|, \mathbf{I}_n) \end{aligned}$$

$\mathcal{W}i_n(\nu, \boldsymbol{\Sigma})$  denotes the Wishart distribution with degree of freedom  $\nu$  and parameter matrix  $\boldsymbol{\Sigma}$ . We have thus avoided the sampling of the sources and, instead, we generate directly the random statistic matrices in each class  $z$  from Normal and Wishart distributions, then we compute the total statistics  $\mathbf{R}_{xs}$  and  $\mathbf{R}_{ss}$  by linear combination of the matrices  $\mathbf{R}_{xs}^{(z)}$  and  $\mathbf{R}_{ss}^{(z)}$  as follows:

$$\begin{aligned} \mathbf{R}_{xs} &= \frac{1}{|\mathcal{S}|} \sum_{z=1}^K |\mathcal{S}_z| \mathbf{R}_{xs}^{(z)} \\ \mathbf{R}_{ss} &= \frac{1}{|\mathcal{S}|} \sum_{z=1}^K |\mathcal{S}_z| \mathbf{R}_{ss}^{(z)} \end{aligned}$$

The key point of this acceleration in the use of the Bartlett decomposition to sample from a Wishart distribution [42]. This procedure is summarized in the appendix B.

## V. SIMULATION RESULTS

To illustrate the feasibility of the algorithm, we generate two discrete fields of  $64 \times 64$  pixels from the Potts model,

$$P_M(\mathbf{Z}^j) = [W(\alpha_j)]^{-1} \exp\{\alpha_j \sum_{r \sim s} I_{z_r = z_s}\}, \alpha_j = 2,$$

$\alpha_j = 2$  implies an homogeneous structure (see figure 1). The first source has 3 colors (3 Gaussians) whereas the second has 2 colors (Ising model).

Conditionally to  $\mathbf{Z}$ , the continuous sources are generated from Gaussian distributions of means  $\mu_1 = \begin{bmatrix} -3 & 0 & 3 \end{bmatrix}$  and variances  $\sigma_1 = \begin{bmatrix} 1 & 0.3 & 0.5 \end{bmatrix}$  for the first source and  $\mu_2 = \begin{bmatrix} -3 & 3 \end{bmatrix}$ ,  $\sigma_2 = \begin{bmatrix} 0.1 & 2 \end{bmatrix}$  for the second source.

The sources are then mixed with the matrix  $\mathbf{A} = \begin{bmatrix} 0.85 & 0.44 \\ 0.50 & 0.89 \end{bmatrix}$  and a white Gaussian noise with covariance  $\mathbf{R}_\epsilon = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}$  is added. The signal to noise ratio is 1 to 3 dB. Figure-1 shows the true discrete labels, the true sources and the mixed images obtained on the detectors.

We apply the MCMC algorithm described in section IV to obtain the Markov chains  $\mathbf{A}^{(h)}$ ,  $\mathbf{R}_\epsilon^{(h)}$ ,  $\mu_{jk}^{(h)}$  and  $\sigma_{jk}^{2(h)}$ . Figures 2 shows the histograms of the element samples of  $\mathbf{A}$  and their empirical expectations (5). We note the concentration of the histograms representing approximately the marginal distributions around the true values and the convergence of the empirical expectations after about 2000 iterations. Figures 3, 4 and 5 show the convergence of the empirical expectations of the sources parameters and the noise covariance. We note that the convergence of the variances is slower than the mixing elements and the means. Finally, the figure 6 shows a sample from source and labels marginal distributions compared to the original images, illustrating the ability of the algorithm to recover the true signals.

We test our algorithm on real data. The first source is a satellite image of an earth region and the second source represents the clouds (First row of figure 7). The mixed images are shown in the second row of the figure. We choose an Ising model for the labels (two colors). The results of the algorithm are illustrated in the third row of the figure where the sources are successfully separated. The last row illustrates the joint segmentation of the sources. We note that the results of the two segmentations are the same as the results which can be obtained if we apply directly the segmentation on the original sources.

## VI. CONCLUSION

In this contribution, we propose an MCMC algorithm to jointly estimate the mixing matrix and the parameters of the hidden Markov fields. The problem has an interesting

natural hidden variable structure leading to a two-level data augmentation procedure. The observed images are embedded in a wider space composed of the observed images, the original unknown images and hidden discrete fields modelizing a second attribute of the images and allowing to take into account a Markovian structure. In this work the number of sources and the number of the discrete values of the hidden Markov field are assumed to be known. However, the implementation of the algorithm could be extended to involve the reversible jump procedure on which we are working.

## APPENDIX A: A *POSTERIORI* DISTRIBUTIONS

### ( $\mathbf{A}, \mathbf{R}_\epsilon$ ) - *a posteriori*

According to the Bayes rule, the *a posteriori* distribution of the ( $\mathbf{A}, \mathbf{R}_\epsilon$ ) parameters is given by,

$$\begin{aligned} p(\mathbf{A}, \mathbf{R}_\epsilon | \mathbf{X}, \mathbf{S}, \mathbf{Z}) &\propto p(\mathbf{X}, \mathbf{S}, \mathbf{Z} | \mathbf{A}, \mathbf{R}_\epsilon) \Pi_0(\mathbf{A}, \mathbf{R}_\epsilon) \\ &\propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \mathbf{R}_\epsilon) \Pi_0(\mathbf{A}, \mathbf{R}_\epsilon) \end{aligned}$$

The prior  $\Pi_0$  has the same advantage of the conjugate prior in that the posterior distribution remains in the same family of the prior distribution. In the case of the ( $\mathbf{A}, \mathbf{R}_\epsilon$ ) parameters, the *a posteriori* distribution is Normal Inverse Wishart:

$$p(\mathbf{A}, \mathbf{R}_\epsilon | \mathbf{X}, \mathbf{S}, \mathbf{Z}) = \mathcal{N}(\mathbf{A}; \mathbf{A}_p, \mathbf{\Gamma}_p) \mathcal{W}_{im}(\mathbf{R}_\epsilon^{-1}; \nu_p, \mathbf{\Sigma}_p)$$

The parameters of these distributions are updated according to the following equations:

$$\left\{ \begin{array}{l} \nu_p = K + \alpha, \quad (K = |\mathcal{S}|) \\ \text{Vec}(\mathbf{A}_p) = [\mathbf{R}_v^{-1} + \mathbf{R}_a^{-1}]^{-1} [\mathbf{R}_v^{-1} \text{Vec}(\mathbf{A}_v) + \mathbf{R}_a^{-1} \text{Vec}(\mathbf{A}_0)] \\ \mathbf{\Gamma}_p^{-1} = \mathbf{R}_v^{-1} + \mathbf{R}_a^{-1} \\ \mathbf{R}_v = K^{-1} \mathbf{R}_{ss}^{-1} \otimes \mathbf{R}_\epsilon \\ \mathbf{R}_a = \alpha^{-1} \mathbf{R}_{ss}^{0^{-1}} \otimes \mathbf{R}_\epsilon \\ \mathbf{A}_v = \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1} \\ \mathbf{\Sigma}_p^{-1} = \frac{1}{K+\alpha} \left[ k \hat{\mathbf{R}}_\epsilon + \alpha \mathbf{R}_0 + (\mathbf{A}_0 - \mathbf{A}_v) (\mathbf{R}_v^{-1} + \alpha^{-1} \mathbf{R}_{ss}^{0^{-1}})^{-1} (\mathbf{A}_0 - \mathbf{A}_v)^T \right] \\ \hat{\mathbf{R}}_\epsilon = \mathbf{R}_{xx} - \mathbf{R}_{xs} \mathbf{R}_{ss}^{-1} \mathbf{R}_{sx} \end{array} \right.$$

The statistics  $\mathbf{R}_{xs}$  and  $\mathbf{R}_{ss}$  are computed from the sampled sources  $\tilde{\mathbf{S}}$  or directly sampled according to their *a posteriori* distributions in the fast version of the MCMC implementation.  $\mathbf{R}_{ss}^0$  is the *a priori* expectation of the matrix  $\mathbf{R}_{ss}$ :

$$\mathbf{R}_{ss}^0 = E_{s|\eta^0} [\mathbf{R}_{ss}]$$

### ( $\mu_k, v = \sigma_k^2$ ) - *a posteriori*

The same computations as in the previous section lead to a Normal Inverse Gamma for the means and variances of the univariate Gaussians:

$$p(\mu_k, v_k^{-1} | \mathbf{X}, \mathbf{S}, \mathbf{Z}) = \mathcal{N}(\mu_k; \mu_p, v_p) \mathcal{G}(v_k^{-1}; \eta_p, \beta_p)$$

The parameters of these distributions are updated according to the following equations:

$$\left\{ \begin{array}{l} \mu_p = \frac{N_k \bar{s} + \alpha w_k^0 \mu_0}{N_k^\dagger \alpha w_k^0} \\ v_p = \frac{v_k}{N_k^\dagger \alpha w_k^0} \\ \eta_p = \frac{N_k^\dagger \alpha w_k^0}{2} \\ \beta_p = \frac{\alpha w_k^0 v_0}{2} + \frac{s^2}{2} + \frac{1}{2} \frac{N_k \alpha w_k^0}{N_k^\dagger \alpha w_k^0} (\bar{s} - \mu_0)^2 \\ \bar{s} = \frac{\sum_{r \in \mathcal{S}_k} s(r)}{N_k} \\ s^2 = \sum_{r \in \mathcal{S}_k} s(r)^2 - N_k \bar{s}^2 \end{array} \right.$$

where  $\mathcal{S}_k$  is the region of the image  $j$  such that the sampled label is equal to  $k$ :

$$\left\{ \begin{array}{l} \mathcal{S}_k = \{r \in \mathcal{S} \mid Z(r) = k\} \\ N_k = |\mathcal{S}_k| \end{array} \right.$$

## APPENDIX B: BARTLETT DECOMPOSITION

Let  $\mathbf{W}$  an  $m \times m$  matrix distributed from a Wishart distribution  $\mathcal{W}_{im}(\nu, \Sigma)$ . A direct simulation from this distribution consists in sampling  $\nu$   $m$ -variate normal vectors  $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_m)$  and then computing:

$$\mathbf{W} = \mathbf{B} \frac{\sum_1^\nu \mathbf{v}_k \mathbf{v}_k^T}{\nu} \mathbf{B}^T$$

where  $\Sigma = \mathbf{B} \mathbf{B}^T$ . This method involves  $\nu m$  simulations from univariate Normal distribution leading to a high computational cost when  $\nu$  increases.

An alternative is to use the Bartlett decomposition which can be summarized in the following theorem:

**Theorem** Let  $\mathbf{W}$  be  $\mathcal{W}_{im}(\nu, \Sigma)$  and  $\Sigma = \mathbf{B} \mathbf{B}^T$ . Put  $\mathbf{W} = \frac{1}{\nu} \mathbf{B} \mathbf{V} \mathbf{V}^T \mathbf{B}^T$ , where  $\mathbf{V}$  is a lower-triangular  $m \times m$  matrix with positive diagonal elements. Then the elements  $v_{ij}$  ( $1 \leq j \leq i \leq m$ ) are independent, and each  $v_{ii}^2$  is  $\chi_{\nu-i+1}^2$  ( $i = 1, \dots, m$ ) while each  $v_{ij}$  is  $\mathcal{N}(0, 1)$  ( $j < i$ ).

The pseudo code of this algorithm is:

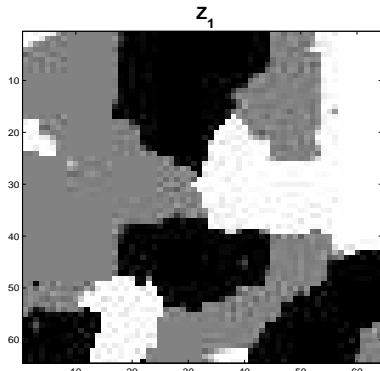
```
// --- Sampling Wishart distribution --- //
.tdf = nu;
.W = zeros(m, m);
.for i = 1 : 1 : m - 1,
    .W(i, i) = sqrt(2 * gamrnd(tdf/2, 1));
    .tdf = tdf - 1;
    .W(i + 1 : m, i) = randn(m - i, 1);
.end
.W(m, m) = sqrt(2 * gamrnd(tdf/2, 1));
.B = (chol(Sigma))';
.W = B * W;
.W = W * W' / nu;
// --- End of sampling --- //
```

## REFERENCES

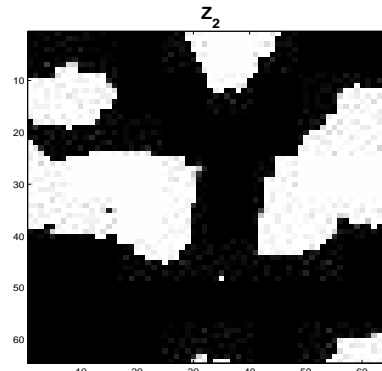
1. A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources", *IEEE Trans. on Circuits and Systems*, vol. 43, no. 11, pp. 894–906, 1996.
2. S. J. Roberts, "Independent component analysis: Source assessment, and separation, a Bayesian approach", *IEE Proceedings - Vision, Image, and Signal Processing*, vol. 145, no. 3, 1998.
3. T. Lee, M. Lewicki, and T. Sejnowski, "Unsupervised classification with non-Gaussian mixture models using ICA", *Advances in Neural Information Processing Systems*, p. in press, 1999.
4. T. Lee, M. Lewicki, and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources", *Neural Computation*, vol. 11, pp. 409–433, 2 1999.
5. T. Lee, M. Girolami, A. Bell, and T. Sejnowski, "A unifying information-theoretic framework for independent component analysis", *Int. Journal of computers and mathematics with applications Computation*, p. in press, 1999.
6. I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-36, no. 10, pp. 1553–1560, October 1988.
7. M. Wax, "Detection and localization of multiple sources via the stochastic signals model", *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2450–2456, November 1991.
8. J.-F. Cardoso, "Infomax and maximum likelihood for source separation", *IEEE Letters on Signal Processing*, vol. 4, no. 4, pp. 112–114, April 1997.
9. J.-L. Lacoume, "A survey of source separation", in *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, Aussois, France, January 11–15, 1999, pp. 1–6.
10. E. Oja, "Nonlinear PCA criterion and maximum likelihood in independent component analysis", in *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, Aussois, France, January 11–15, 1999, pp. 143–148.
11. R. B. MacLeod and D. W. Tufts, "Fast maximum likelihood estimation for independent component analysis", in *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, Aussois, France, January 11–15, 1999, pp. 319–324.
12. O. Bermond and J.-F. Cardoso, "Approximate likelihood for noisy mixtures", in *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, Aussois, France, January 11–15, 1999, pp. 325–330.
13. P. Comon, C. Jutten, and J. Herault, "Blind separation of sources .2. problems statement", *Signal Processing*, vol. 24, no. 1, pp. 11–20, 1991.
14. 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.
15. E. Moreau and B. Stoll, "An iterative block procedure for the optimization of constrained contrast functions", in *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, Aussois, France, January 11–15, 1999, pp. 59–64.
16. P. Comon and O. Grellier, "Non-linear inversion of underdetermined mixtures", in *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, Aussois, France, January 11–15, 1999, pp. 461–465.
17. J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation", *IEEE Trans. on Sig. Proc.*, vol. 44, no. 12, pp. 3017–3030, December 1996.
18. A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and Éric Moulines, "A blind source separation technique based on second order statistics", *IEEE Trans. on Sig. Proc.*, vol. 45, no. 2, pp. 434–44, February 1997.
19. S.-I. Amari and J.-F. Cardoso, "Blind source separation — semiparametric statistical approach", *IEEE Trans. on Sig. Proc.*, vol. 45, no. 11, pp. 2692–2700, November 1997.
20. J.-F. Cardoso, "Blind signal separation: statistical principles", *Proceedings of the IEEE. Special issue on blind identification and estimation*, , no. 10, pp. 2009–2025, October 1998.
21. J. J. Rajan and P. J. W. Rayner, "Decomposition and the discrete karhunen-loeve transformation using a bayesian approach", *IEE Proceedings - Vision, Image, and Signal Processing*, vol. 144, no. 2, pp. 116–123, 1997.

22. K. Knuth, "Bayesian source separation and localization", in *SPIE'98 Proceedings: Bayesian Inference for Inverse Problems*, San Diego, CA, A. Mohammad-Djafari, Ed., July 1998, pp. 147–158.
23. K. Knuth and H. Vaughan JR., "Convergent Bayesian formulation of blind source separation and electromagnetic source estimation", in *MaxEnt 98 Proceedings: Int. Workshop on Maximum Entropy and Bayesian methods*, Garching, Germany, F. R. von der Linden W., Dose W. and P. R., Eds., 1998, p. in press.
24. S. E. Lee and S. J. Press, "Robustness of Bayesian factor analysis estimates", *Communications in Statistics – Theory And Methods*, vol. 27, no. 8, 1998.
25. K. Knuth, "A Bayesian approach to source separation", in *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99*, Aussios, France, C. J. J.-F. Cardoso and P. Loubaton, Eds., 1999, pp. 283–288.
26. T. Lee, M. Lewicki, M. Girolami, and T. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representation", *IEEE Signal Processing Letters*, p. in press, 1999.
27. A. Mohammad-Djafari, "A Bayesian approach to source separation", in *Bayesian Inference and Maximum Entropy Methods*, J. R. G. Erikson and C. Smith, Eds., Boise, IH, July 1999, MaxEnt Workshops, Amer. Inst. Physics.
28. P. Senecal, S. Amblard, "Mcmc methods for discrete source separation", in *Bayesian Inference and Maximum Entropy Methods*, Gif-sur-Yvette, France, July 2000, Proc. of MaxEnt, pp. 350–360, Amer. Inst. Physics.
29. 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*, 2003.
30. H. Attias, "Blind separation of noisy mixture: An EM algorithm for independent factor analysis", *Neural Computation*, vol. 11, pp. 803–851, 1999.
31. 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. IEEE workshop, September 2001, pp. 293–302.
32. J. F. Cardoso, "Infomax and maximum likelihood for source separation", *IEEE Letters on Signal Processing*, vol. 4, pp. 112–114, avril 1997.
33. K. Knuth, "A Bayesian approach to source separation", in *Proceedings of Independent Component Analysis Workshop*, 1999, pp. 283–288.
34. N. Peyrard, "Convergence of MCEM and related algorithms for hidden markov random field", Research Report 4146, INRIA, 2001.
35. M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation", *J. Amer. Statist. Assoc.*, vol. 82, no. 398, pp. 528–540, June 1987.
36. H. Snoussi and A. Mohammad-Djafari, "Bayesian separation of HMM sources", in *Bayesian Inference and Maximum Entropy Methods*, R. L. Fry, Ed. MaxEnt Workshops, August 2002, pp. 77–88, Amer. Inst. Physics.
37. H. Snoussi and A. Mohammad-Djafari, "Information Geometry and Prior Selection.", in *Bayesian Inference and Maximum Entropy Methods*. MaxEnt Workshops, August 2002, to appear in Amer. Inst. Physics.
38. S. Amari and H. Nagaoka, *Methods of Information Geometry*, vol. Volume 191 of Translations of Mathematical Monographs, AMS, OXFORD, University Press, 2000.
39. C. Robert, *Méthodes de Monte-Carlo par chaînes de Markov*, Economica, Paris, France, 1996.
40. 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, July 2000, Proc. of MaxEnt, pp. 388–406, Amer. Inst. Physics.
41. O. Bermond, *Méthodes statistiques pour la séparation de sources*, Phd thesis, Ecole Nationale Supérieure des Télécommunications, 2000.
42. B. D. Ripley, *Stochastic Simulation*, John Wiley & Sons, 1987.

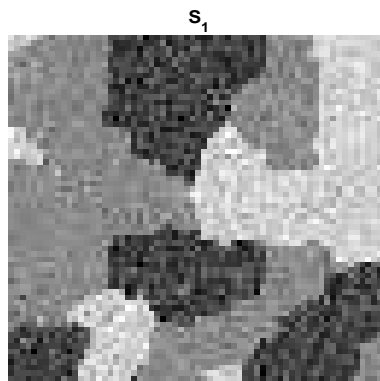




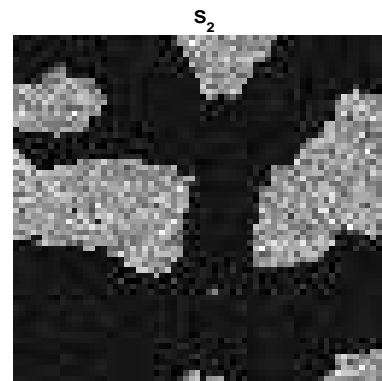
(a)- True classification for source 1



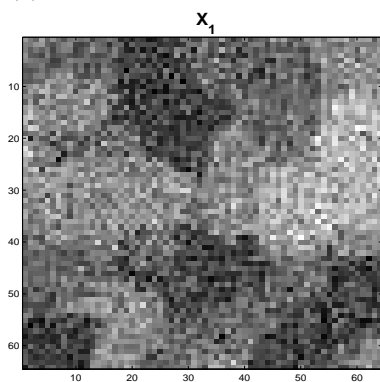
(a)- True classification for source 2



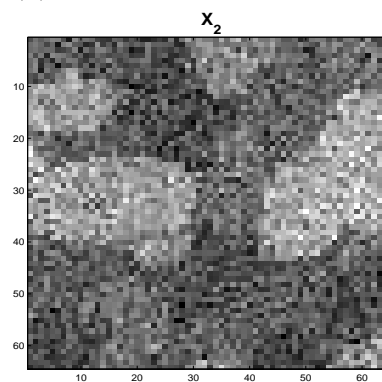
(c)- Source 1



(d)- Source 2



(e)- Observed image 1



(f)- Observed image 2

Figure -1- True and observed images

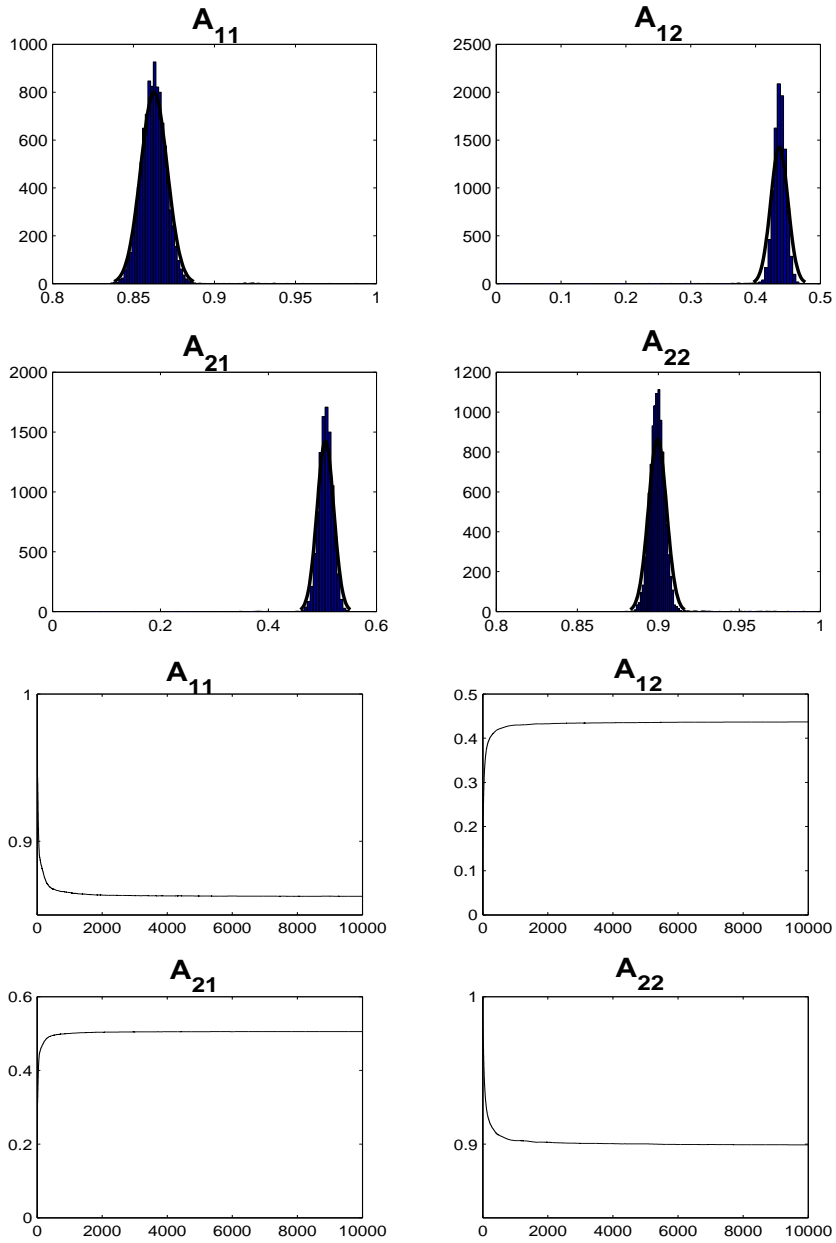
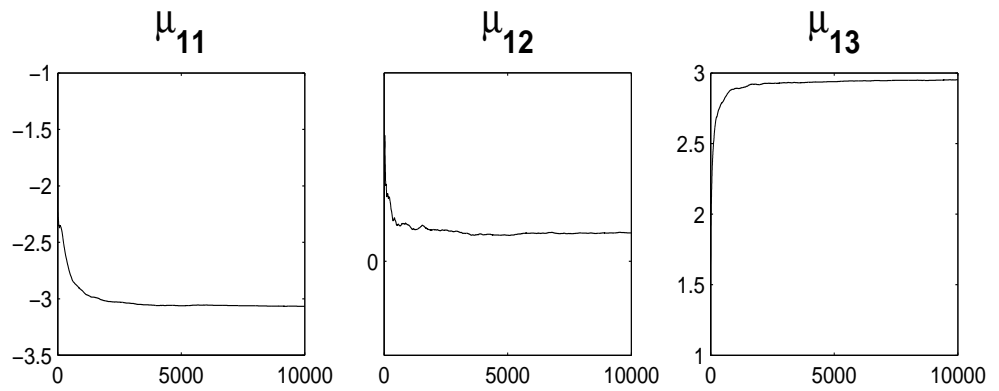
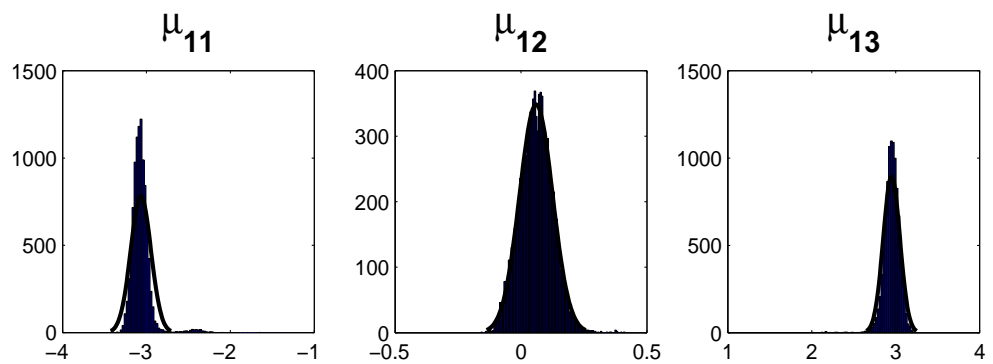


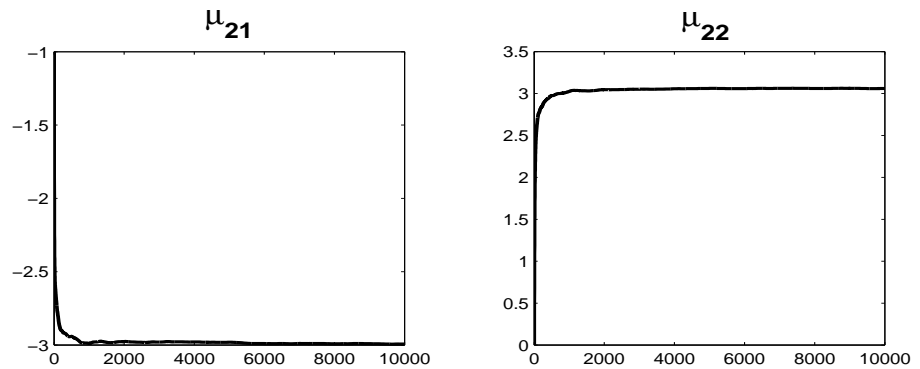
Figure-2 Convergence of the empirical expectations of  $a_{ij}$  after 2000 iterations and the corresponding histograms



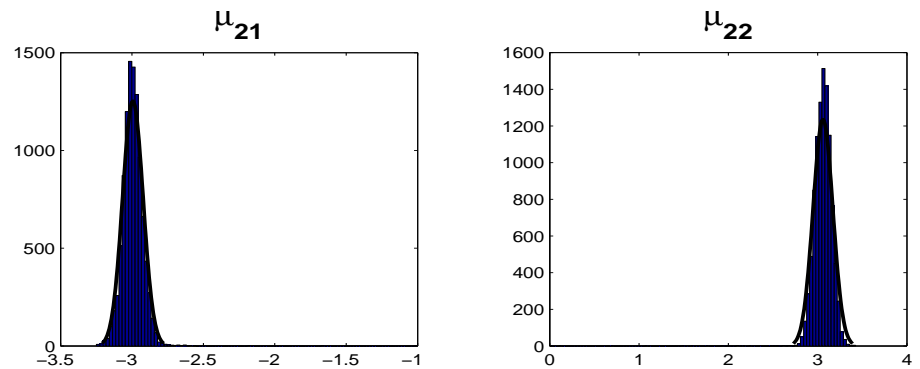
(a)



(b)

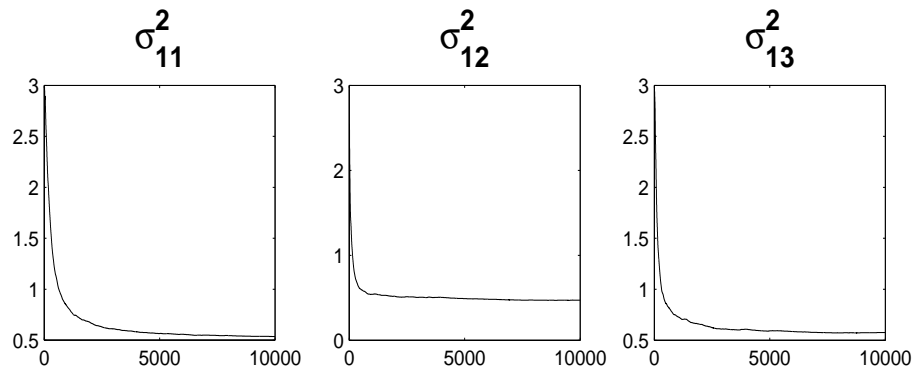


(c)

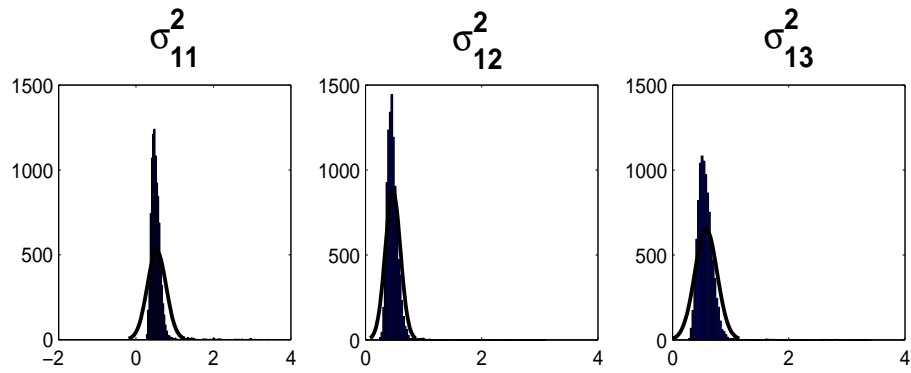


(d)

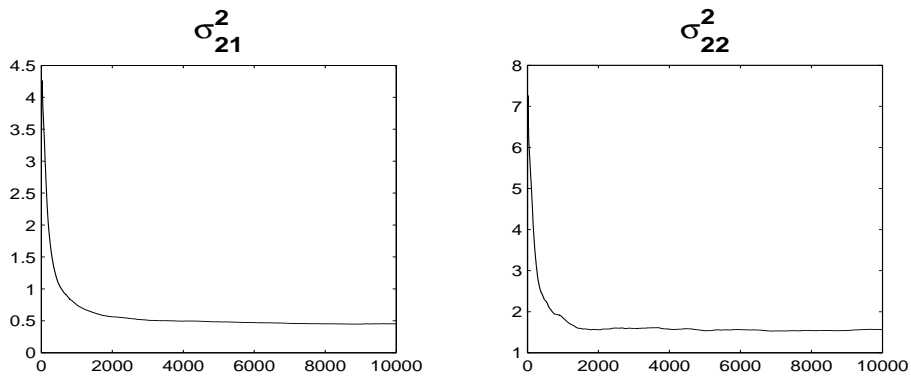
Figure-3 (a)- Convergence of the empirical expectations of the means  $m_{ij}$  of the source 1  
 (b)- Histograms of the means of the source 1  
 (c)- Convergence of the empirical expectations of the means  $m_{ij}$  of the source 2  
 (d)- Histograms of the means of the source 2



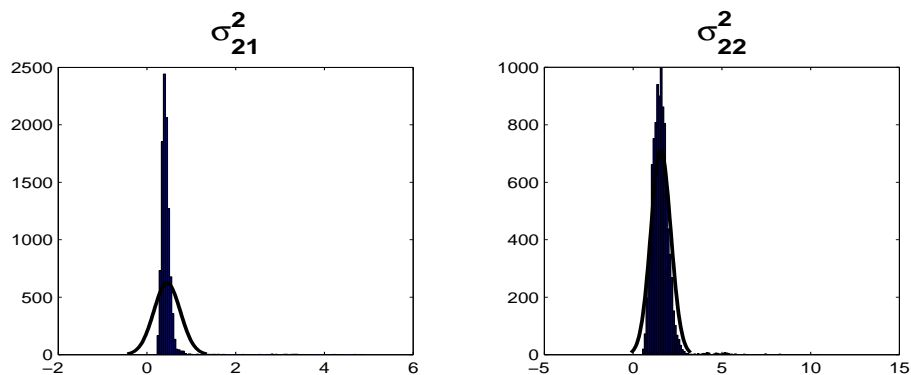
(a)



(b)

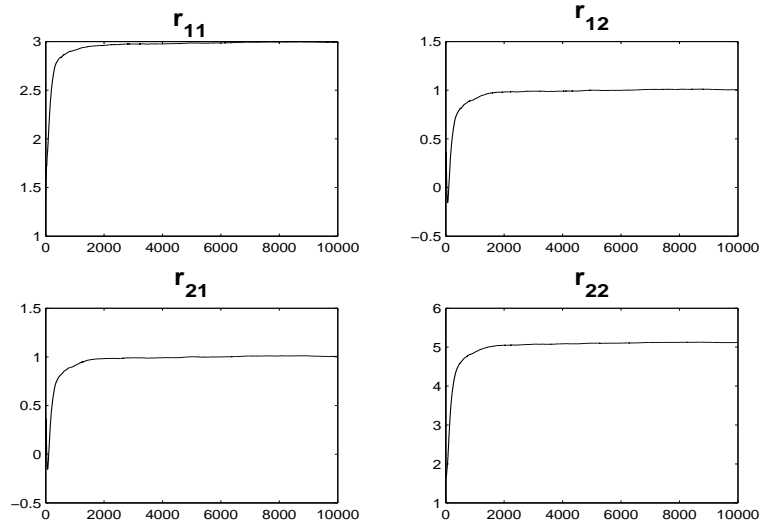


(c)

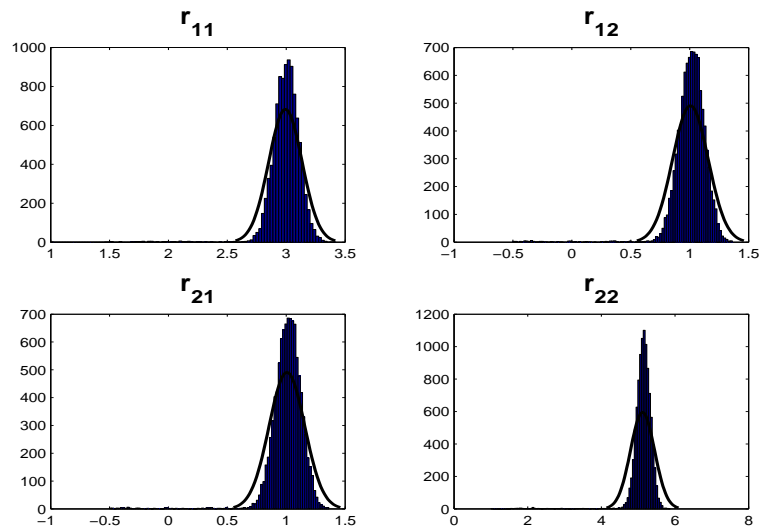


(d)

Figure-4 (a)- Convergence of the empirical expectations of the variances  $\sigma_{ij}$  of the source 1  
 (b)-Histograms of the variances of the source 1  
 (c)- Convergence of the empirical expectations of the variances  $\sigma_{ij}$  of the source 2  
 (d)- Histograms of the variances of the source 2

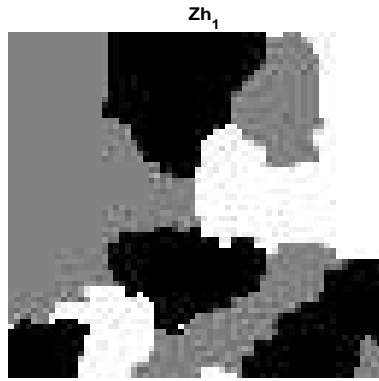


(a)

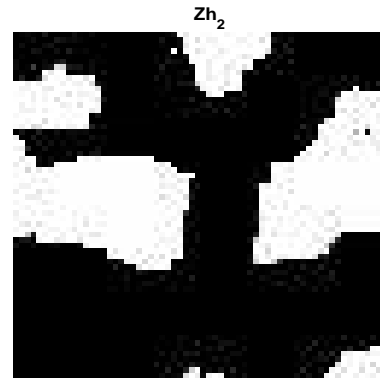


(b)

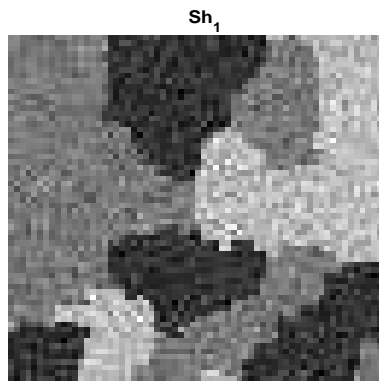
Figure-5 (a)- Convergence of the empirical expectations of the noise variances  
 (b) Histograms of the noise variances



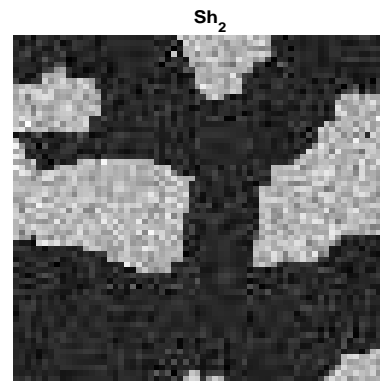
(a)- True classification for source 1



(a)- True classification for source 2



(c)- Source 1



(d)- Source 2

Figure 6

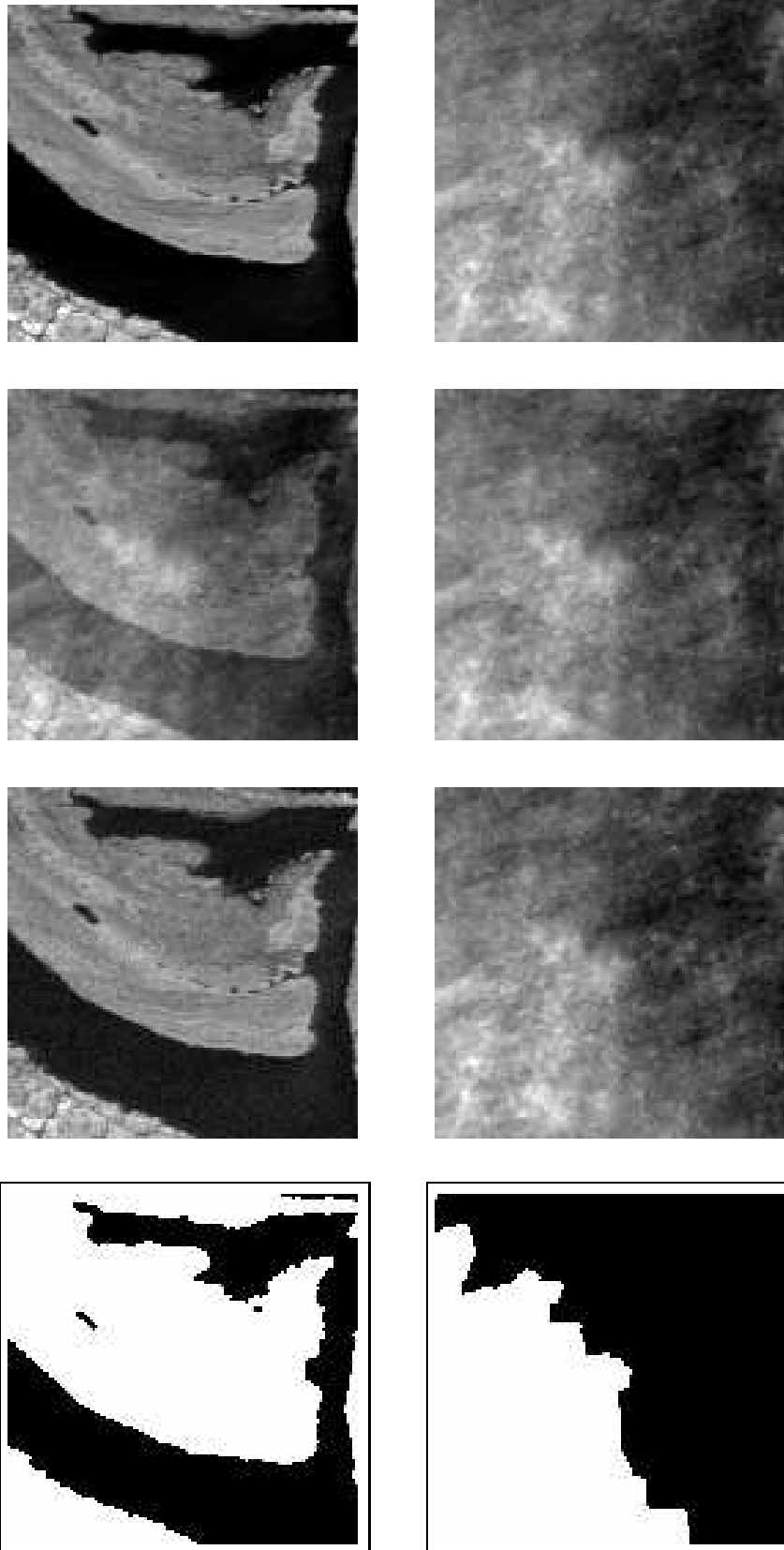


Figure.7: from top to bottom: original sources, mixed sources, estimated sources and segmented images