FAST MCMC JOINT IMAGE SEPARATION AND SEGMENTATION

Hichem SNOUSSI *and Ali* MOHAMMAD–DJAFARI

Laboratoire des Signaux et Systèmes (CNRS-SUPELEC-UPS).
SUPELEC, Plateau de Moulon, 91192 Gif-Sur-Yvette Cedex, France.
snoussi@lss.supelec.fr, djafari@lss.supelec.fr

ABSTRACT

In this contribution, we consider the problem of the blind separation of noisy instantaneously mixed images. The images are modelized by hidden Markov fields with unknown parameters. Given the observed images, we give a Bayesian formulation and we propose to solve the resulting data augmentation problem by implementing a Monte Carlo Markov Chain (MCMC) procedure. 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 classification 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 a flexibility in the cost function choice.

An acceleration of the MCMC implementation based on Wishart distributions is proposed.

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

1. I. INTRODUCTION AND MODEL ASSUMPTIONS

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

$$x^i_r = \sum_{j=1}^{n} a_{ij}\, s^j_r + n^i_r,\ r \in \mathcal{S}, i = 1..m$$

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

$$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} + \boldsymbol{n} \qquad (1)$$

The noise and source components $(\boldsymbol{N}^i)_{1..m}$ and $(\boldsymbol{S}^j)_{j=1..n}$ are supposed to be independent. Each source is modelized by a double stochastic process $(\boldsymbol{S}^j, \boldsymbol{Z}^j)$. $\boldsymbol{S}^j$ 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. $\boldsymbol{Z}^j$ is the hidden Markov field representing the unobserved pixels classification whose components are in a discrete set, $\boldsymbol{Z}^j_r \in \{1..K^j\}$. The joint probability distribution of $\boldsymbol{Z}^j$ satisfies the following properties,

$$\begin{cases} \forall\, \boldsymbol{Z}^j, \quad P_M(z^j_r \mid \boldsymbol{Z}^j_{\mathcal{S}\setminus\{r\}}) = P_M(z^j_r \mid \boldsymbol{Z}^j_{N(r)}) \\ \forall\, \boldsymbol{Z}^j, \quad P_M(\boldsymbol{Z}^j) > 0 \end{cases}$$

where $\boldsymbol{Z}^j_{\mathcal{S}\setminus\{r\}}$ 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,

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

where H_{α_j} is the energy function and α_j is a parameter weighting the spatial dependencies supposed to be known. Conditionally to the hidden discrete field $\boldsymbol{Z}^j$, the source pixels $S^j_r, r \in \mathcal{S}$ are supposed to be independent and have the following conditional distribution:

$$p(\boldsymbol{S}^j \mid \boldsymbol{Z}^j, \boldsymbol{\eta}^j) = \prod_{r\in\mathcal{S}} p_r(s^j_r \mid z^j_r, \boldsymbol{\eta}^j)$$

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

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

1. The problem described by (1) when the mixing matrix $\boldsymbol{A}$ is unkown is the source separation problem [1, 2, 3].
2. Given the source component $\boldsymbol{S}^j$, the estimation of the parameter $\boldsymbol{\eta}^j$ and the recovering of the hidden classification $\boldsymbol{Z}^j$ is known as the unsupervised segmentation [4].

In this contribution, given the observations $\boldsymbol{X}^i, i = 1..m$, 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, an MCMC algorithm based on the data augmentation modelization is proposed. In section IV, we focus on the problem of the non identifiability and the degeneracies occurring in the source separation problem and their effects on the MCMC implementation. In section V, numerical simulations are shown to illustrate the feasibility of the solution.

2. II. BAYESIAN FORMULATION

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

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

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

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

where $\mathcal{N}$ denotes the Gaussian distribution, $\boldsymbol{x}_r$ the $(m \times 1)$ vector of observations on the site r, $\boldsymbol{z}_r$ is the vector label, $\boldsymbol{\mu}_{\boldsymbol{z}_r} = [\mu_{1z_1}, ..., \mu_{nz_n}]^t$ and $\boldsymbol{R}_{\boldsymbol{z}_r}$ the diagonal matrix $diag[\sigma_{1z_1}^2, ..., \sigma_{nz_n}^2]$. We note that the expression (2) hasn't a tractable form with respect to the parameter $\boldsymbol{\theta}$ because of the integration of the hidden variables $\boldsymbol{S}$ and $\boldsymbol{Z}$. This remark leads us to consider the data augmentation algorithm [5] where we complete the observations $\boldsymbol{X}$ by the hidden variables $(\boldsymbol{Z}, \boldsymbol{S})$, the complete data are then $(\boldsymbol{X}, \boldsymbol{S}, \boldsymbol{Z})$. In a previous work [6], 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 its *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.

3. III. MCMC IMPLEMENTATION

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

repeat until convergence,

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

This Bayesian sampling [7] produces a Markov chain $(\tilde{\boldsymbol{\theta}}^{(k)})$, ergodic with stationary distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{X})$. After k_0 iterations (warming up), the samples $(\tilde{\boldsymbol{\theta}}^{(k_0+h)})$ can be considered to be drawn approximately from their *a posteriori* distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{X})$. Then, by the ergodic theorem, we can approximate *a posteriori* expectations by empirical expectations:

$$\mathrm{E}\left[h(\boldsymbol{\theta}) \,|\, \boldsymbol{X}\right] \approx \frac{1}{K} \sum_{k=1}^{K} h(\tilde{\boldsymbol{\theta}}^{(k)}) \tag{3}$$

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

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

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

   In this expression, we have two kinds of dependencies: (**i**) $\boldsymbol{Z}$ are independent across components, $p(\boldsymbol{Z}) = \prod_{j=1}^n p(\boldsymbol{Z}^j)$ but each discrete image $\boldsymbol{Z}^j \sim P_M(\boldsymbol{Z}^j)$ has a Markovian structure. (**ii**) Given $\boldsymbol{Z}$, the fields $\boldsymbol{X}$ are independent through the set $\mathcal{S}$, $p(\boldsymbol{X} \,|\, \boldsymbol{Z}, \boldsymbol{\theta}) = \prod_{r \in \mathcal{S}} p(\boldsymbol{x}_r \,|\, \boldsymbol{z}_r, \boldsymbol{\theta})$ but dependent through the components because of the mixing operation $p(\boldsymbol{x}_r \,|\, \boldsymbol{z}_r, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_r \,;\, \boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{z}_r}, \boldsymbol{A}\boldsymbol{R}_{\boldsymbol{z}_r}\boldsymbol{A}^* + \boldsymbol{R}_\epsilon)$.

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

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

   where the *a posteriori* mean and covariance are easily computed [8],

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

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

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

leading to a separate sampling of $(\boldsymbol{A}, \boldsymbol{R}_\epsilon)$ and $(\boldsymbol{\mu}, \boldsymbol{\sigma})$. The choice of the *a priori* distributions is not an easy task [9, 10]. The complete likelihood of $(\boldsymbol{A}, \boldsymbol{R}_\epsilon)$ belongs to the location scale family [11] and applying the Jeffrey's rule we have,

$$p(\boldsymbol{A}, \boldsymbol{R}_\epsilon) \propto |\mathcal{F}(\boldsymbol{R}_\epsilon)|^{\frac{1}{2}} = |\boldsymbol{R}_\epsilon|^{\frac{-(m+1)}{2}}$$

where $p(\boldsymbol{A})$ is locally uniform and $\mathcal{F}$ is the Fisher information matrix. We obtain an inverse Wishart distribution for the noise covariance and a Gaussian distribution for the mixing matrix,

$$\begin{cases} \boldsymbol{R}_\epsilon^{-1} \sim Wi_m(\alpha_\epsilon, \boldsymbol{\beta}_\epsilon), \\ \alpha_\epsilon = \frac{|\mathcal{S}|-n}{2},\ \boldsymbol{\beta}_\epsilon = \frac{|\mathcal{S}|}{2}(\boldsymbol{R}_{xx} - \boldsymbol{R}_{xs}\boldsymbol{R}_{ss}^{-1}\boldsymbol{R}_{xs}^*) \end{cases} \tag{4}$$

$$\begin{cases} \boldsymbol{A} \,|\, \boldsymbol{R}_\epsilon \sim \mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{R}_a), \\ \boldsymbol{\mu}_a = vec(\boldsymbol{R}_{xs}\boldsymbol{R}_{ss}^{-1}),\ \boldsymbol{R}_a = \frac{1}{|\mathcal{S}|}\boldsymbol{R}_{ss}^{-1} \otimes \boldsymbol{R}_\epsilon \end{cases} \tag{5}$$

where we define the empirical statistics $\boldsymbol{R}_{xx} = \frac{1}{|\mathcal{S}|}\sum_r \boldsymbol{x}_r\boldsymbol{x}_r^*$, $\boldsymbol{R}_{xs} = \frac{1}{|\mathcal{S}|}\sum_r \boldsymbol{x}_r\boldsymbol{s}_r^*$ and $\boldsymbol{R}_{ss} = \frac{1}{|\mathcal{S}|}\sum_r \boldsymbol{s}_r\boldsymbol{s}_r^*$. We note that the covariance matrix of $\boldsymbol{A}$ is proportional to the noise to signal ratio. This explains the fact noted in [12] concerning the slow convergence of the EM algorithm. For the parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma})$, we choose conjugate priors [7]. The reason of this choice is the elimination of degeneracies occurring when estimating the variances σ_{jz}. This point is elucidated in section IV. The *a posteriori* distribution remains in the same family as the likelihood function, Gaussian for the means and Inverse Gamma for the variances. The expressions are the same as in [7].

4. 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 $\boldsymbol{S}$ are sampled in the MCMC algorithm but only the statistics $\boldsymbol{R}_{xs}$ and $\boldsymbol{R}_{ss}$ are used in the generation of the parameters ($\boldsymbol{A}, \boldsymbol{R}_\epsilon$ (see equation (4) and (5)). Therefore we avoid the sampling of the sources $\boldsymbol{S}$ and we sample directly the statistic matrices $\boldsymbol{R}_{xs}$ and $\boldsymbol{R}_{ss}$. We show in the following how these simulations are easly performed in our problem formulations.

After the drawing of the labels $\boldsymbol{Z}$, the source images $\boldsymbol{S}$ are classified into K regions $(\mathcal{S}_z)_{z=1..K^n}$ defined by:

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

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

$$\begin{aligned} \boldsymbol{V}_z &= \left[\boldsymbol{A}^* \boldsymbol{R}_\epsilon^{-1} \boldsymbol{A} + \boldsymbol{R}_z^{-1}\right]^{-1} \\ \boldsymbol{m}_z &= \boldsymbol{V}_z \left(\boldsymbol{A}^* \boldsymbol{R}_\epsilon^{-1} \boldsymbol{x}_r + \boldsymbol{R}_z^{-1} \boldsymbol{\mu}_z\right) \end{aligned} \qquad (6)$$

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

$$\begin{aligned} \boldsymbol{R}_{xs}^{(z)} &= \frac{1}{|\mathcal{S}_z|} \sum_{r \in \mathcal{S}_z} \boldsymbol{x}_r \boldsymbol{s}_r^* \\ \boldsymbol{R}_{ss}^{(z)} &= \frac{1}{|\mathcal{S}_z|} \sum_{r \in \mathcal{S}_z} \boldsymbol{s}_r \boldsymbol{s}_r^* \\ \boldsymbol{R}_{xx}^{(z)} &= \frac{1}{|\mathcal{S}_z|} \sum_{r \in \mathcal{S}_z} \boldsymbol{x}_r \boldsymbol{x}_r^* \end{aligned} \qquad (7)$$

From the expressions (6) and (7) and some algebraic manipulations, the statistics $\boldsymbol{R}_{xs}^{(z)}$ and $\boldsymbol{R}_{xs}^{(z)}$ can be decomposed as follows:

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

where $\boldsymbol{C}_z$ is the Cholesky root of the matrix $\boldsymbol{V}_z$. The matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are not random matrices and are updated at each iteration. The matrices $\boldsymbol{U}_{n,1}$, $\boldsymbol{U}_{n,2}$ and $\boldsymbol{U}_w$ are random matrices and have the following distributions:

$$\boldsymbol{U}_{n,1} \sim \mathcal{N}(0, \tfrac{1}{|\mathcal{S}_z|} \boldsymbol{R}_{xx}^{(z)} \otimes \boldsymbol{I}_n)$$

$$\boldsymbol{U}_{n,2} \sim \mathcal{N}(0, \tfrac{1}{|\mathcal{S}_z|} \boldsymbol{R}_z^{-1} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^* \boldsymbol{R}_z^{-1} \otimes \boldsymbol{I}_n)$$

$$\boldsymbol{U}_w \sim \mathcal{W}i_n(|\mathcal{S}_z|, \boldsymbol{I}_n))$$

$\mathcal{W}i_n(\nu, \boldsymbol{\Sigma}))$ denotes the Wishart distribution with degree of freedom ν 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 $\boldsymbol{R}_{xs}$ and $\boldsymbol{R}_{xs}$ by linear combination of the $\boldsymbol{R}_{xs}^{(z)}$ and $\boldsymbol{R}_{ss}^{(z)}$ as follows:

$$\boldsymbol{R}_{xs} = \frac{1}{|\mathcal{S}|} \sum_{z=1}^{K^n} |\mathcal{S}_z| \, \boldsymbol{R}_{xs}^{(z)}$$

$$\boldsymbol{R}_{ss} = \frac{1}{|\mathcal{S}|} \sum_{z=1}^{K^n} |\mathcal{S}_z| \, \boldsymbol{R}_{ss}^{(z)}$$

5. V. SIMULATION RESULTS

To illustrate the feasibility of the algorithm, we generate two discrete fields of 64×64 pixels from the Ising model,

$$P_M(\boldsymbol{Z}^j) = [W(\alpha_j)]^{-1} \exp\{\alpha_j \sum_{r \sim s} I_{z_r = z_s}\}, \alpha_1 = 2, \ \alpha_2 = 0.8$$

$\alpha_1 > \alpha_2$ implies that the first source is more homogeneous than the second source. Conditionally to $\boldsymbol{Z}$, the continuous sources are generated from Gaussian distributions of means $\mu_{jz} = \begin{bmatrix} -2 & 2 \\ -3 & 3 \end{bmatrix}$ and variances $\sigma_{jz} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$.

The sources are then mixed with the matrix $\boldsymbol{A} = \begin{bmatrix} 0.85 & 0.44 \\ 0.51 & 0.89 \end{bmatrix}$ and a white Gaussian noise with covariance $\sigma_\epsilon^2 \boldsymbol{I}$ ($\sigma_\epsilon^2 = 5$) is added. The signal to noise ratio is 1 to 3 dB. Figure-1 shows the mixed signals obtained on the detectors.

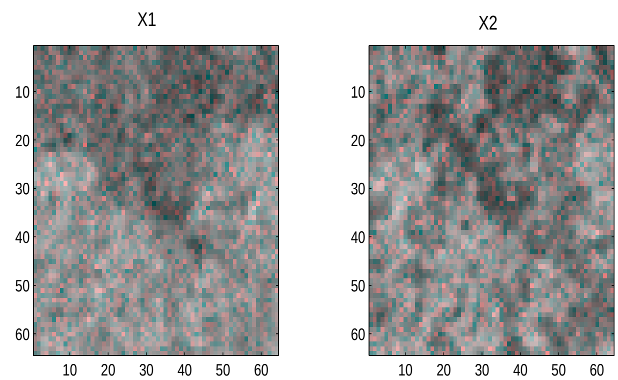


figure-1. The noisy mixed images $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$

We apply the MCMC algorithm described in section III to obtain the Markov chans $\boldsymbol{A}^{(k)}$, $\boldsymbol{R}_\epsilon^{(k)}$, $\mu_{jz}^{(k)}$ and $\sigma_{jz}^{2\,(k)}$. Figures 2 and 3 show the histograms of the element samples of $\boldsymbol{A}$ and their empirical expectations (3). 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 1000 iterations. Figures 4 and 5 show the convergence of the empirical expectations. We note that the convergence of the variances is slower that the mixing elements and the means.

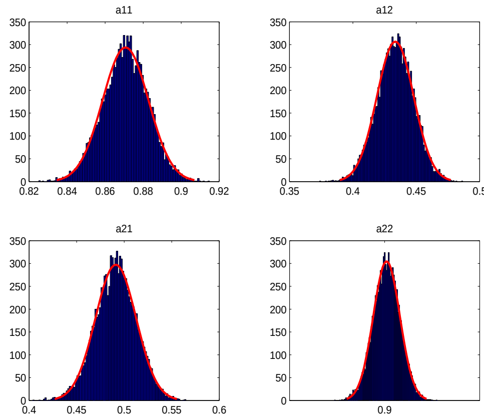


Figure-2 The histograms of the samples of mixing elements a_{ij}

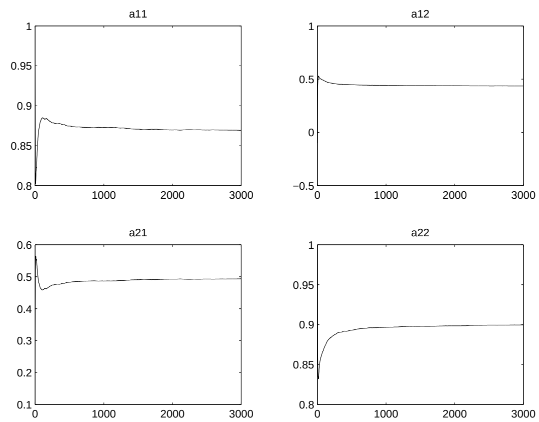


Figure-3 Convergence of the empirical expectations of a_{ij} after 1000 iterations

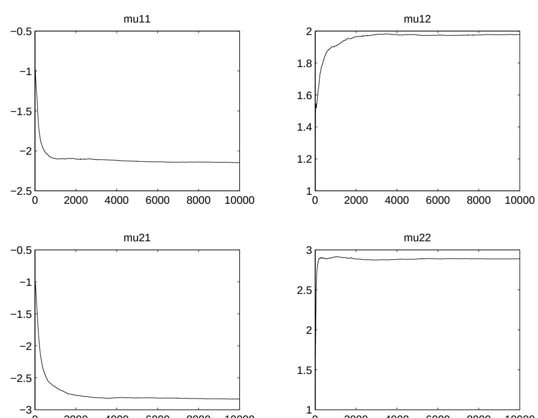


Figure-4 Convergence of the empirical expectations of the means m_{ij}

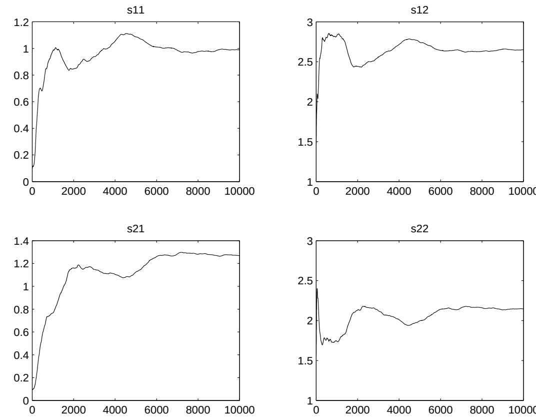
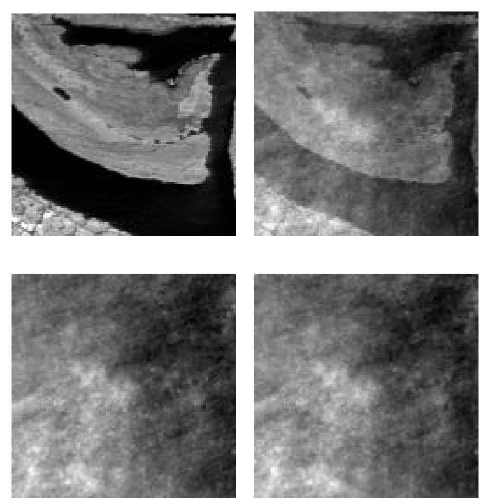


Figure-5 Convergence of the empirical expectations of the variance σ^2_{ij}

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 column of figure 6). The mixed images are shown in the second column of figure 6. The results of the algorithm are illustrated in figure 7 where the sources are successfully separated. The figure 8 illustrate 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 the segmentation on the original sources.



(a) (b)

Figure.6: (a) Original sources, (b) Mixed sources

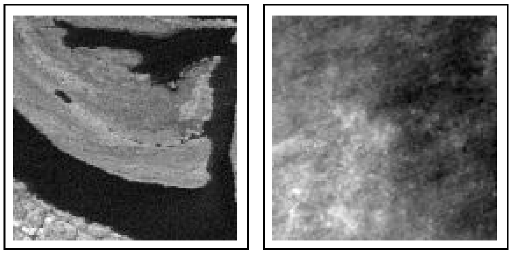


Figure.7: Estimated sources

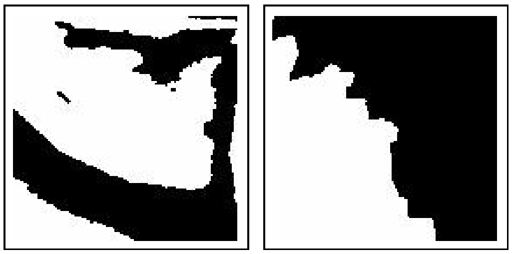


Figure 8: Segmented images

6. 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. A fast implementation of the MCMC algorithm is proposed. It is essentially based on sampling directly the statistic matrices from Normal and Wishart distributions.

7. REFERENCES

[1] J. F. Cardoso, "Infomax and maximum likelihood for source separation", *IEEE Letters on Signal Processing*, vol. 4, pp. 112–114, avril 1997.

[2] K. Knuth, "A Bayesian approach to source separation", in *Proceedings of Independent Component Analysis Workshop*, 1999, pp. 283–288.

[3] 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.

[4] N. Peyrard, "Convergence of MCEM and related algorithms for hidden markov random field", Research Report 4146, INRIA, 2001.

[5] 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.

[6] 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.

[7] C. Robert, *Méthodes de Monte-Carlo par chaînes de Markov*, Economica, Paris, France, 1996.

[8] 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.

[9] R. E. Kass and L. Wasserman, "Formal rules for selecting prior distributions: A review and annotated bibliography", Technical report no. 583, Department of Statistics, Carnegie Mellon University, 1994.

[10] 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.

[11] G. E. P. Box and G. C. Tiao, *Bayesian inference in statistical analysis*, Addison-Wesley publishing, 1972.

[12] O. Bermond, *Méthodes statistiques pour la séparation de sources*, Phd thesis, Ecole Nationale Supérieure des Télécommunications, 2000.