Journal of Electronic Imaging 13(2), 1–0 (April 2004).

Fast joint separation and segmentation of mixed images

Hichem Snoussi
Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes (L2S)
Supélec, Plateau de Moulon
91192 Gif-sur-Yvette Cedex, France
E-mail: snoussi@lss.supelec.fr

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 pixel 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.* © *2004 SPIE and IS&T.*
[DOI: 10.1117/1.1666873]

1 Introduction and Model Assumptions

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

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

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

$$x_r=As_r+n_r. \tag{1}$$

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

Paper 03021 received Feb. 10, 2003; revised manuscript received Aug. 4, 2003; accepted for publication Aug. 12, 2003.
1017-9909/2004/$15.00 © 2004 SPIE and IS&T.

Many techniques have been proposed for the source separation problem based on entropy and information theoretic approach[1–5] and the maximum likelihood principle[6–12] leading to contrast functions[13–16] and estimating functions.[17–20] 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 parameters 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.[2,21–32] 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, based on 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.[33,34] However, this model does not take into account the temporal correlation of the sources. An extension to the hidden Markov models 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 (S^j,Z^j). 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. Z^j is the hidden Markov field representing the unobserved pixel classification whose components are in a discrete set, $Z_r^j\in\{1...K^j\}$. The joint probability distribution of Z^j satisfies the following properties:

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

where $Z^j_{\mathcal{S}\setminus\{r\}}$ denotes the field restricted to $\mathcal{S}\setminus\{r\}=\{\ell$

Table 1

r	Pixel location
$\mathcal{S}$	The set of all pixel locations
Variables	
X^i	The image on the detector number i
S^j	The source image number j
Z^j	The image classification of the source number j
X	$(X^i)_{i=1...m}$ (known data)
S	$(S^j)_{j=1...n}$ (hidden sources)
Z	$(Z^j)_{j=1...n}$ (hidden labels)
x_r	The vector of observed data at the pixel r
s_r	The vector of sources at the pixel r
Parameters	
A	The mixing matrix
R_ϵ	Noise covariance
$\eta_k^j=(\mu_k^j, v_k^j=(\sigma_k^j)^2)$ $k=1...K_j$	The means and variances of the source j

$\in \mathcal{S}, \ell \neq r\}$ 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

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

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

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

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

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

1. The problem described by (1) when the mixing matrix A is unknown which is the source separation problem.[27,35,36]
2. Given the source component S^j, the estimation of the parameter η^j and the recovering of the hidden classification labels Z^j which is known as the unsupervised segmentation.[37]

Table 1 gives a summary of the notations for the variables, the parameters and the indices used in this paper and their meanings.

In this contribution, given the observations X^i $(i=1...m)$ we propose a solution to jointly separate the n unknown sources and perform their unsupervised segmentations. In Sec. 2, we give a Bayesian formulation of the problem. In Sec. 3, we propose an original construction of the prior law selection for the parameters. In Sec. 4, a fast implementation of an MCMC algorithm based on the data augmentation technique is proposed. In Sec. 5, numerical simulations on synthetic and real data are shown to illustrate the feasibility and the performances of the proposed solution.

2 Bayesian Formulation

Given the observed data $X=(X^1,\ldots,X^m)$, our objective is the estimation of the mixing matrix A, the noise covariance R_ϵ, 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 $\theta=(A,R_\epsilon,\mu_{jk},\sigma_{jk}^2)$ contains all the information that we can extract from the data. According to the Bayesian rule, we have

$$p(\theta|X)\propto p(X|\theta)p(\theta).$$

In Sec. 3, we will discuss the attribution of appropriate prior distribution $p(\theta)$. Concerning the likelihood, it has the following expression:

$$p(X|\theta)=\sum_Z \int_S p(X,S,Z|\theta)dS$$
$$=\sum_Z \left\{\prod_{r\in\mathcal{S}} \mathcal{N}(x_r;A\mu_{z_r},AR_{z_r}A^*+R_\epsilon)\right\} P_M(Z), \quad (2)$$

where $\mathcal{N}$ denotes the Gaussian distribution, x_r the $(m\times 1)$ vector of observations on the site r, z_r is the vector label, $\mu_{z_r}=[\mu_{1z_1},\ldots,\mu_{nz_n}]^t$ and R_{z_r} the diagonal matrix $\mathrm{diag}[\sigma_{1z_1}^2,\ldots,\sigma_{nz_n}^2]$. We note that the expression (2) does not have a tractable form with respect to the parameter θ because of the integration of the hidden variables S and Z. This remark leads us to consider the data augmentation algorithm[38] where we complete the observations X by the

hidden variables (Z,S); the complete data are then (X,S,Z). In a previous work,[39] we implemented restoration-maximization algorithms in the one-dimensional case to estimate the maximum *a posteriori* estimate of θ. We extend this work in two directions: (i) the sources are two-dimensional signals, (ii) we implement an MCMC algorithm to obtain samples of θ drawn from their *a posteriori* distribution. This gives the possibility of not being restricted to estimate the parameters by its maximum *a posteriori*. We can consider another cost function and compute the corresponding estimate.

3 *a priori* Selection

The Bayesian method is more and more attracting practitioners. The basic reason is its ability to combine, in a simple way, two sources of information: information from collected data and *a priori* information. This combination consists in multiplying the likelihood by the prior: $p(data|\theta)\Pi(\theta)$ to obtain the posterior $p(\theta|data)$. However, the problem that arises with this method is the choice of a prior distribution for the parameter θ. In a recent work,[40] the author proposed an original rule to construct a prior. The rule's principle consists in exploiting the prior knowledge without adding irrelevant information. The resulting prior distribution (called δ prior) is the minimizer of a cost function representing a trade-off between some desirable behavior (given by a distance to a reference prior) and uniformity (given by a distance to Jeffreys prior). The proposed cost function has the following expression:

$$\mathcal{J}(\Pi)=\gamma_e\int \Pi(\theta)D_\delta(p_\theta,p_0)d\theta + \gamma_u\int \Pi(\theta)\log\Pi(\theta)/\sqrt{\|g(\theta)\|}d\theta,$$

where $p_\theta=p(data|\theta)$ is the likelihood of θ, γ_e/γ_u is the trade-off between the confidence degree γ_e in the reference distribution p_0 and the uniformity degree γ_u, $g(\theta)$ is the Fisher information matrix and D_δ is the δ divergence[41]

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

By variational calculus, the δ prior has the following form (see Ref. 40 for details):

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

We note that the prior selection needs to be established in a specific geometry in that it depends on the measure of distinguishability (here the δ divergence) between probability distributions.

The rest of this section is the computation of the δ prior in our special case.* Our parameter of interest is $\theta=(A,R_\epsilon,\eta)$. A is the mixing matrix, R_ϵ the noise covariance and η contains all the parameters of the sources model

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

where the index j indicates the source j, k indicates the Gaussian component k of the distribution of the source j and (μ_k^j,v_k^j) are the corresponding mean and variance. Our objective is the computation of the δ priors (3). We have an incomplete data problem with two hierarchies of hidden variables, the sources $s_{1\ldots T}$ and the labels $z_{1\ldots T}$ so that the complete data are $(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T})$.

We assume that the reference distribution p_0 belongs to the parametric family $\{p_\theta\}$ so that it is defined by the reference parameters $\theta=(A^0,R_\epsilon^0,\eta^0)$. The expressions of the Fisher matrix and the δ 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.

3.1 *Fisher Information Matrix*

The Fisher matrix $g(\theta)$ is defined as

$$g_{ij}(\theta)=-\underset{x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}}{E}\left[\frac{\partial^2}{\partial_i\partial_j}\log p(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}|\theta)\right].$$

The factorization of the joint distribution $p(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}|\theta)$ as

$$p(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}|\theta) =p(x_{1\ldots T}|s_{1\ldots T},z_{1\ldots T},\theta)p(s_{1\ldots T}|z_{1\ldots T},\theta)p(z_{1\ldots T}|\theta)$$

and the corresponding expectations as

$$\underset{x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}}{E}[\cdot]=\underset{z_{1\ldots T}}{E}\left[\underset{s_{1\ldots T}|z_{1\ldots T}}{E}[\cdot]\right]\underset{x_{1\ldots T}|s_{1\ldots T},z_{1\ldots T}}{E}[\cdot]$$

and taking into account the conditional independences $[(x_{1\ldots T}|s_{1\ldots T},z_{1\ldots T})\Leftrightarrow(x_{1\ldots T}|s_{1\ldots T})$ and $(s_{1\ldots T}|z_{1\ldots T})\Leftrightarrow\Pi s_{1\ldots T}^j|z_{1\ldots T}^j)]$, the Fisher information matrix will have a block diagonal structure as follows:

$$g(\theta)=\begin{bmatrix} g(A,R_\epsilon) & \cdots & [0]\\ \vdots & g(\eta^1) & \\ & & \ddots \\ [0] & \cdots & g(\eta^n)\end{bmatrix}.$$

3.1.1 *(A,Rϵ)-block*

The Fisher information matrix of (A,R_ϵ) is

$$\mathcal{F}_{ij}(A,R_\epsilon)=-E\underset{s}{E}\left[\frac{\partial^2}{\partial_i\partial_j}\log p(x_{1\ldots T}|s_{1\ldots T},A,R_\epsilon)\right]$$

*The reader who is not interested in derivation details can directly consult the prior expressions obtained in the end of this section.

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

$$g(A,R_\epsilon)=\begin{bmatrix} \left(\underset{s_{1\ldots T}}{E}\ R_{ss}\right)\otimes R_\epsilon^{-1} & [0] \\ [0] & -\frac{1}{2}\frac{\partial R_\epsilon^{-1}}{\partial R_\epsilon} \end{bmatrix},$$

where $R_{ss}=1/T\Sigma\, s_t s_t^*$ and $\otimes$ is the Kronecker product (defined as in Ref. 42).

We note the block diagonality of the (A,R_ϵ)-Fisher matrix. The term corresponding to the mixing matrix 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 (A,R_ϵ) [the local volume of the differential manifold of the set of distributions $p(X|S,A,R_\epsilon)$] is then

$$|g(A,R_\epsilon)|^{1/2}dAdR_\epsilon=\frac{\left|\underset{\eta}{ER_{ss}}\right|^{m/2}}{|R_\epsilon|^{m+n+1/2}}dAdR_\epsilon.$$

3.1.2 (η^j) *block*

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

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

3.2 *δ Divergence (δ=0)*

In this paper, we fix the value of δ to 0. The 0 divergence between two parameters $\theta=(A,R_\epsilon,\eta)$ and $\theta^0=(A^0,R_\epsilon^0,\eta^0)$ for the complete data likelihood $p(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}|\theta)$ is

$$D_0(\theta:\theta^0)=\underset{x,s,z|\theta^0}{E}\log\frac{p(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}|\theta^0)}{p(x_{1\ldots T},s_{1\ldots T},z_{1\ldots T}|\theta)}.$$

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 (A,R_ϵ) parameters and the divergence between the sources' parameters η

$$D_0(\theta:\theta^0)=\underset{s|\eta^0}{E}\ D_0(A,R_\epsilon:A^0,R_\epsilon^0)+D_0(\eta:\eta^0),$$

where D_0 means the divergence between the distributions $p(x_{1\ldots T}|A,R_\epsilon,s_{1\ldots T})$ and $p(x_{1\ldots T}|A^0,R_\epsilon^0,s_{1\ldots T})$ keeping the sources $s_{1\ldots T}$ fixed.

The 0 divergence between η and η^0 is the sum of the 0 divergences between each source parameter η^j and η^{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 η and η^0 is obtained as a particular case $(n=1)$ of the general expression derived in the multivariate case in Ref. 29 leading to a normal inverse gamma prior

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

with $\nu^0=\alpha w_k^0$, $\alpha=\gamma_e/\gamma_u$, 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 (A,R_ϵ) parameters are

$$\underset{s|\eta^0}{E}\ D_0(A,R_\epsilon:A^0,R_\epsilon^0)=\frac{1}{2}\Bigg(\log|R_\epsilon R_\epsilon^{0\ -1}|+\mathrm{Tr}(R_\epsilon^{-1}R_\epsilon^0)+\mathrm{Tr}\Big[R_\epsilon^{-1}(A-A^0)\ \underset{s|\eta^0}{E}\ [R_{ss}]\times(A-A^0)^*\Big]\Bigg)$$

leading to the following 0 priors on (A,R_ϵ):

$$\Pi_0(A,R_\epsilon^{-1})=\mathcal{N}\left(A;A^0,\frac{1}{\alpha}R_{ss}^{0-1}\otimes R_\epsilon\right)\mathcal{W}_{im}(R_\epsilon^{-1};\alpha,R_\epsilon^{0\ -1})\times\left|\underset{s|\eta}{E}[R_{ss}]\right|^{m/2}$$

where $R_{ss}^0=E_{s|\eta^0}R_{ss}$ and $\mathcal{W}_n$ is the Wishart distribution of an $n\times n$ matrix

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

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 A is the noise to signal ratio $1/\alpha R_{ss}^{0-1}\otimes R_\epsilon$. We note a multiplicative term which is a power of the determinant of the *a priori* expectation of the source covariance $E_{s|\eta}[R_{ss}]$. This term can be injected in the prior $p(\eta)$ and thus the (A,R_ϵ) parameters and the η parameters are *a priori* independent.

We note that the precision matrix for the mixing matrix A $(\alpha R_{ss}^0\otimes R_\epsilon^{-1})$ for Π_0 is the product of the confidence term $\alpha=\gamma_e/\gamma_u$ in the reference parameters and the signal to noise ratio. Therefore, the resulting precision of the reference matrix A^0 is not only our *a priori* coefficient γ_e but the product of this coefficient and the signal to noise ratio.

4 MCMC Implementation

We divide the vector of unknown variables into two subvectors: the hidden variables (Z,S) and the parameter θ and we consider a Gibbs sampler

repeat until convergence,

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

This Bayesian sampling[43] produces a Markov chain $(\tilde{\theta}^{(h)})$, ergodic with stationary distribution $p(\theta|X)$. After h_0 iterations (warming up), the samples $(\tilde{\theta}^{(h_0+h)})$ can be considered to be drawn approximately from their *a posteriori* distribution $p(\theta|X)$. Then, by the ergodic theorem, we can approximate *a posteriori* expectations by empirical expectations

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

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

(1) draw $\tilde{Z}$ from

$$p(Z|X,\theta) \propto p(X|Z,\theta)P_M(Z).$$

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

(2) draw $\tilde{S}|\tilde{Z}$ from

$$p(S|X,Z,\theta)=\prod_{r\in\mathcal{S}} \mathcal{N}(s_r;m_r^{apost},V_r^{apost}),$$

where the *a posteriori* mean and covariance are easily computed[44]

$$V_r^{apost}=[A^*R_\epsilon^{-1}A+R_{z_r}^{-1}]^{-1}$$

$$m_r^{apost}=V_r^{apost}(A^*R_\epsilon^{-1}x_r+R_{z_r}^{-1}\mu_{z_r}).$$

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

$$p(\theta|X,Z,S) \propto p(A,R_\epsilon|X,S)p(\mu,\sigma|S,Z)$$

leading to a separate sampling of (A,R_ϵ) and (μ,σ). Choosing the 0 priors developed in the previous section, the *a posteriori* distributions are

1. Inverse Wishart for the noise covariance and inverse gamma for sources' variances.
2. 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 (A,R_ϵ) in the particular case when $\alpha=0$ (Jeffreys prior)

$$\begin{cases} R_\epsilon^{-1} \sim Wi_m(\nu_p,\Sigma_p), \quad \nu_p=\dfrac{|\mathcal{S}|-n}{2}, \\ \Sigma_p=\dfrac{|\mathcal{S}|}{2}(R_{xx}-R_{xs}R_{ss}^{-1}R_{xs}^*), \\ p(A|R_\epsilon) \sim \mathcal{N}(A_p,\Gamma_p), \quad A_p=R_{xs}R_{ss}^{-1}, \\ \Gamma_p=\dfrac{1}{|\mathcal{S}|}R_{ss}^{-1}\otimes R_\epsilon, \end{cases} \tag{6}$$

where we define the empirical statistics $R_{xx}=1/|\mathcal{S}|\Sigma_r x_r x_r^*$, $R_{xs}=1/|\mathcal{S}|\Sigma_r x_r s_r^*$ and $R_{ss}=1/|\mathcal{S}|\Sigma_r s_r s_r^*$ (the sources S are generated in the first step of the Gibbs sampling). We note that the covariance matrix of A is proportional to the noise to signal ratio. This explains the fact noted in Ref. 45 concerning the slow convergence of the Einstein–Maxwell algorithm.

4.1 *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 S are sampled in the MCMC algorithm but only the statistics R_{xs} and R_{ss} are used in the generation of the parameters $[A, R_\epsilon$ (see Eq. (6)]. Therefore we avoid the sampling of the sources S and we sample directly the statistic matrices R_{xs} and R_{ss}. We show in the following how these simulations are easily performed in our problem formulation.

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

$$V_z=[A^*R_\epsilon^{-1}A+R_z^{-1}]^{-1}, \tag{7}$$

$$m_z=V_z(A^*R_\epsilon^{-1}x_r+R_z^{-1}\mu_z).$$

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

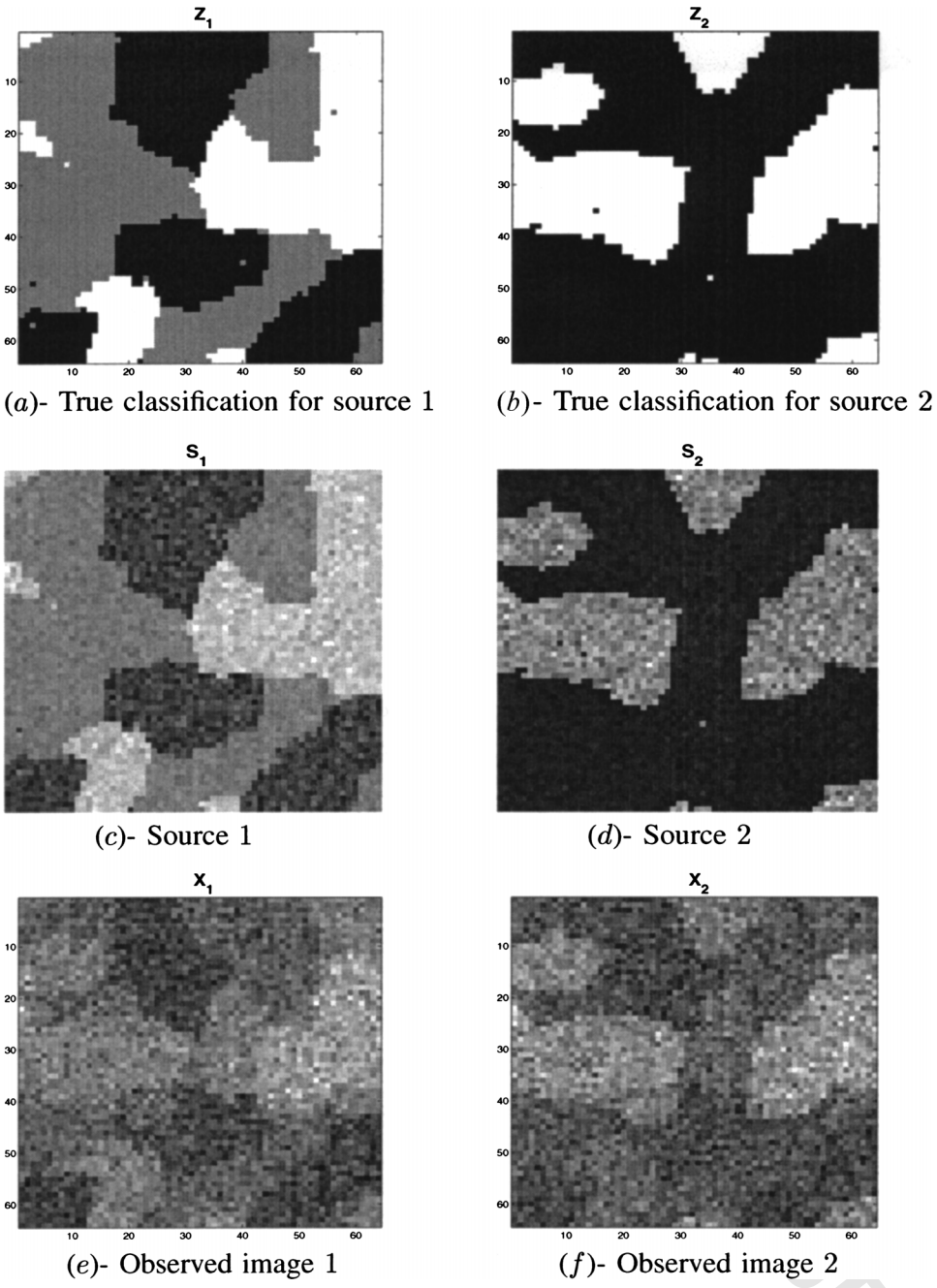

Fig. 1 True and observed images: the observations are a noisy mixture of true images.

$$R_{xs}^{(z)}=\frac{1}{|\mathcal{S}_z|}\sum_{r\in\mathcal{S}_z} x_r s_r^*, \quad R_{ss}^{(z)}=\frac{1}{|\mathcal{S}_z|}\sum_{r\in\mathcal{S}_z} s_r s_r^*,$$

$$R_{xx}^{(z)}=\frac{1}{|\mathcal{S}_z|}\sum_{r\in\mathcal{S}_z} x_r x_r^*. \tag{8}$$

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

$$R_{xs}^{(z)}=R_1+U_{n,1}^* C_z^*,$$

$$R_{ss}^{(z)}=R_2+V_z(A^* R_\epsilon^{-1} U_{n,1}^*+U_{n,2}^*)C_z^*,$$

$$C_z(U_{n,1}R_\epsilon^{-1}A+U_{n,2})+C_z U_w C_z^*,$$

where $V_z=C_z C_z^*$. The matrices R_1 and R_2 are not random matrices and are updated at each iteration. The matrices $U_{n,1}$, $U_{n,2}$, and U_w are random matrices and have the following distributions:

$$U_{n,1} \sim \mathcal{N}\left(0, \frac{1}{|\mathcal{S}_z|} R_{xx}^{(z)} \otimes I_n\right),$$

$$U_{n,2} \sim \mathcal{N}\left(0, \frac{1}{|\mathcal{S}_z|} R_z^{-1} \mu_z \mu_z^* R_z^{-1} \otimes I_n\right),$$

$$U_w \sim \mathcal{W}i_n[|\mathcal{S}_z|, I_n)].$$

$\mathcal{W}i_n(\nu, \Sigma)$ denotes the Wishart distribution with degree of freedom ν and parameter matrix Σ. 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 R_{xs} and R_{xs} by linear combination of the matrices $R_{xs}^{(z)}$ and $R_{ss}^{(z)}$ as follows:

$$R_{xs} = \frac{1}{|\mathcal{S}|} \sum_{z=1}^{K} |\mathcal{S}_z| R_{xs}^{(z)},$$

$$R_{ss} = \frac{1}{|\mathcal{S}|} \sum_{z=1}^{K} |\mathcal{S}_z| R_{ss}^{(z)}.$$

The key point of this acceleration is the use of the Bartlett decomposition to sample from a Wishart distribution.[46] This procedure is summarized in Appendix B.

5 Simulation Results

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

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

where $\alpha_j = 2$ implies a homogeneous structure (see Fig. 1). The first source has three colors (three Gaussians) whereas the second has two colors (Ising model).

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

The sources are then mixed with the matrix $A = \begin{bmatrix} 0.85 & 0.44 \\ 0.50 & 0.89 \end{bmatrix}$ and a white Gaussian noise with covariance $R_\epsilon = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}$ is added. The signal to noise ratio is 1–3 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 Sec. IV to obtain the Markov chans $A^{(h)}$, $R_\epsilon^{(h)}$, $\mu_{jk}^{(h)}$, and $\sigma_{jk}^{2(h)}$. Figure 3 shows the histograms of the element samples of 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 4, 5 and 6 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 that the mixing elements and the means. Finally, Fig. 2 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 and their classifications.

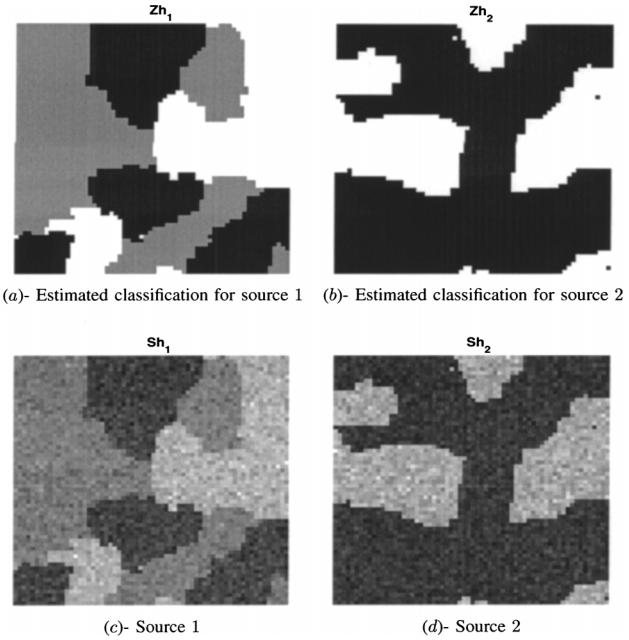

Fig. 2 Source reconstruction and segmentation.

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 Fig. 7). We have artificially mixed these two images and added a Gaussian noise. 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 obtained from the noisy mixed images are the same as the results which can be obtained if we directly apply the segmentation on the original sources.

6 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 modeling 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.

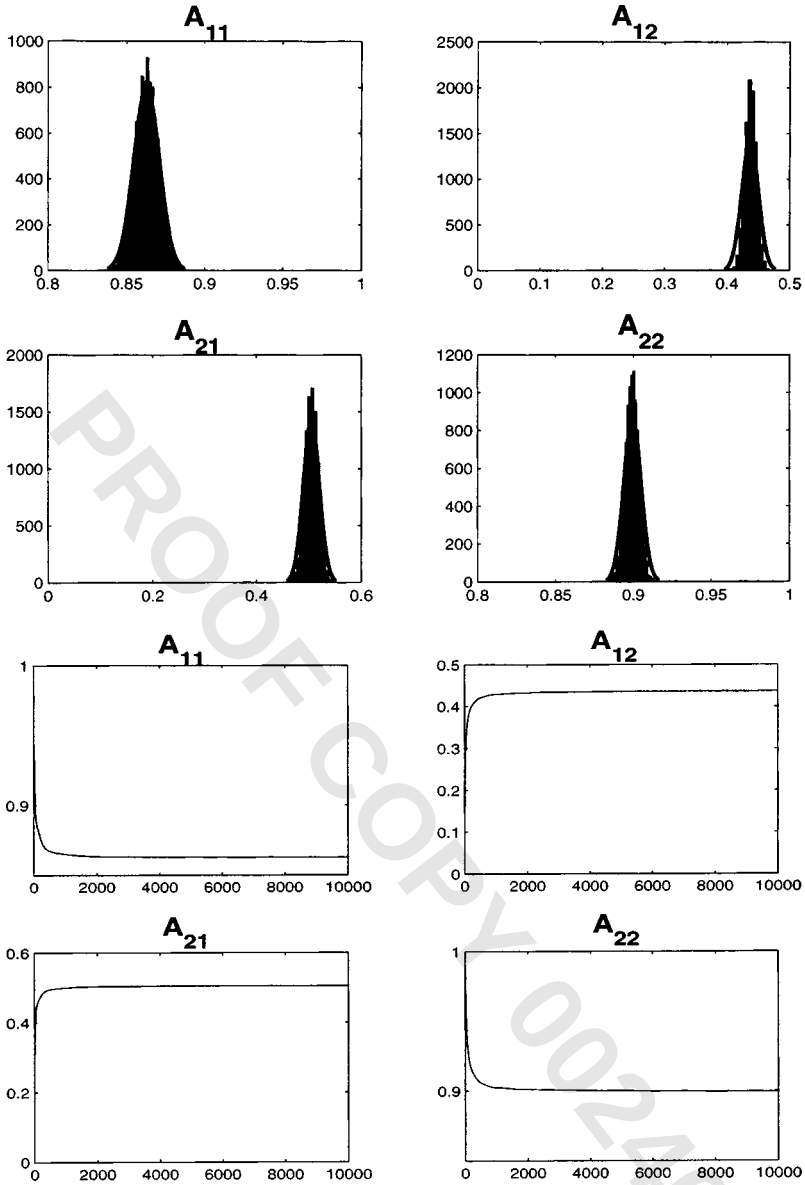

Fig. 3 Convergence of the empirical expectations of a_{ij} after 2000 iterations and the corresponding histograms.

7 Appendix A: *a posteriori* Distributions

7.1 (A,R_ϵ)—*a posteriori*

According to the Bayes rule, the *a posteriori* distribution of the (A,R_ϵ) parameters is given by

$$p(A,R_\epsilon|X,S,Z)\propto p(X,S,Z|A,R_\epsilon)\Pi_0(A,R_\epsilon),$$

$$\propto p(X|S,A,R_\epsilon)\Pi_0(A,R_\epsilon).$$

The prior Π_0 has the same advantage as the conjugate prior in that the posterior distribution remains in the same family of the prior distribution. In the case of the (A,R_ϵ) parameters, the *a posteriori* distribution is normal inverse Wishart

$$p(A,R_\epsilon|X,S,Z)=\mathcal{N}(A;A_p,\Gamma_p)\mathcal{W}_{im}(R_\epsilon^{-1};\nu_p,\Sigma_p).$$

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

$$\begin{cases}
\nu_p=K+\alpha,(K=|\mathcal{S}|),\\
Vec(A_p)=[R_v^{-1}+R_a^{-1}]^{-1}[R_v^{-1}Vec(A_v)+R_a^{-1}Vec(A_0)],\\
\Gamma_p^{-1}=R_v^{-1}+R_a^{-1},\\
R_v=K^{-1}R_{ss}^{-1}\otimes R_\epsilon,\\
R_a=\alpha^{-1}R_{ss}^{0\ -1}\otimes R_\epsilon,\\
A_v=R_{xs}R_{ss}^{-1},\\
\Sigma_p^{-1}=\dfrac{1}{K+\alpha}[k\hat{R}_\epsilon+\alpha R_0+(A_0-A_v)\\
\qquad\times(K^{-1}R_{ss}^{-1}+\alpha^{-1}R_{ss}^{0\ -1})^{-1}(A_0-A_v)^T,\\
\hat{R}_\epsilon=R_{xx}-R_{xs}R_{ss}^{-1}R_{sx}.
\end{cases}$$

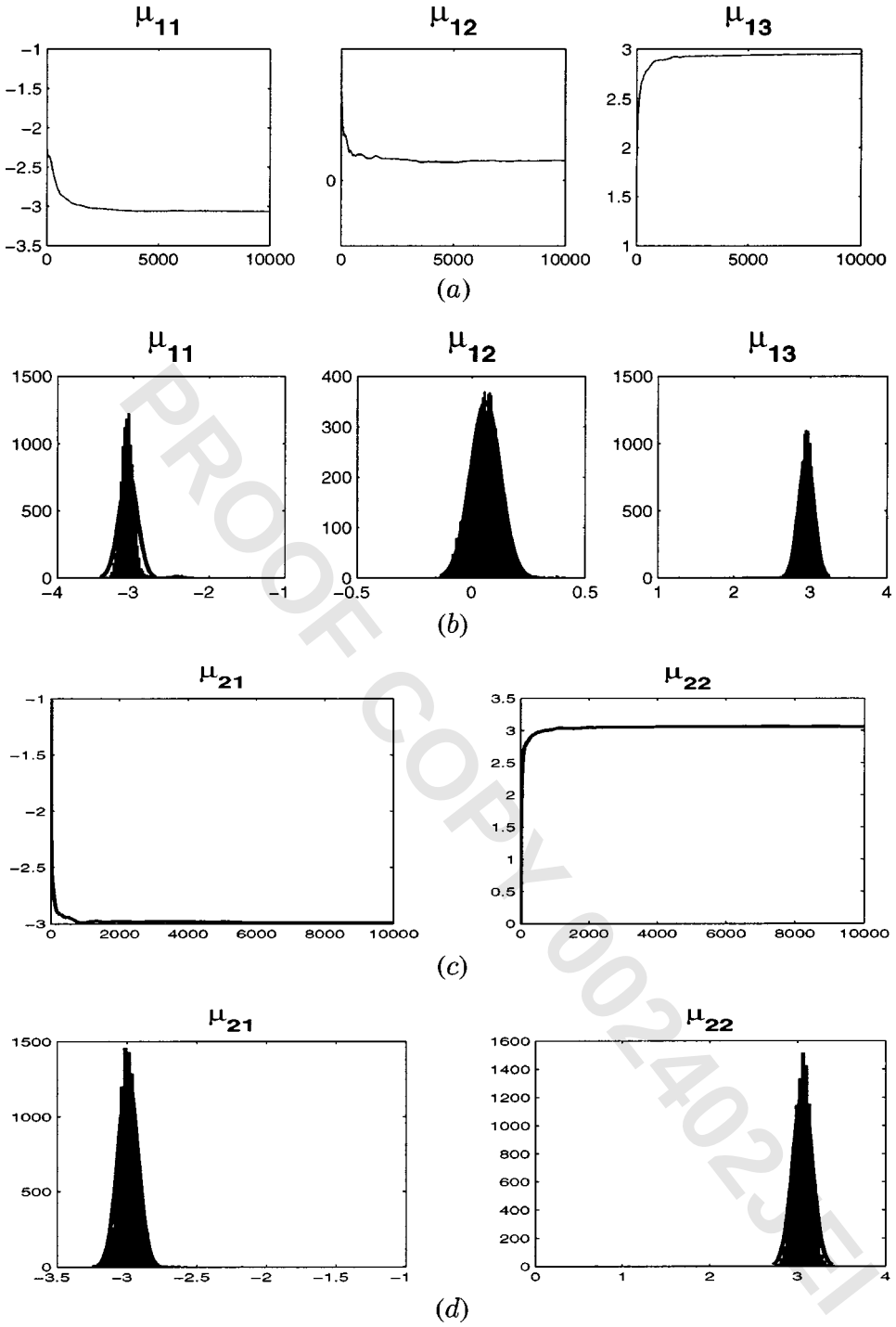

Fig. 4 (a) Convergence of the empirical expectations of the means m_{ij} of the source 1, (b) histograms of the means of the source 1, (c) convergence of the empirical expectations of the means m_{ij} of the source 2, (d) histograms of the means of the source 2.

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

$$R_{ss}^0 = \underset{s|\eta^0}{E}[R_{ss}].$$

7.2 $(\mu_k, v=\sigma_k^2)$—*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

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

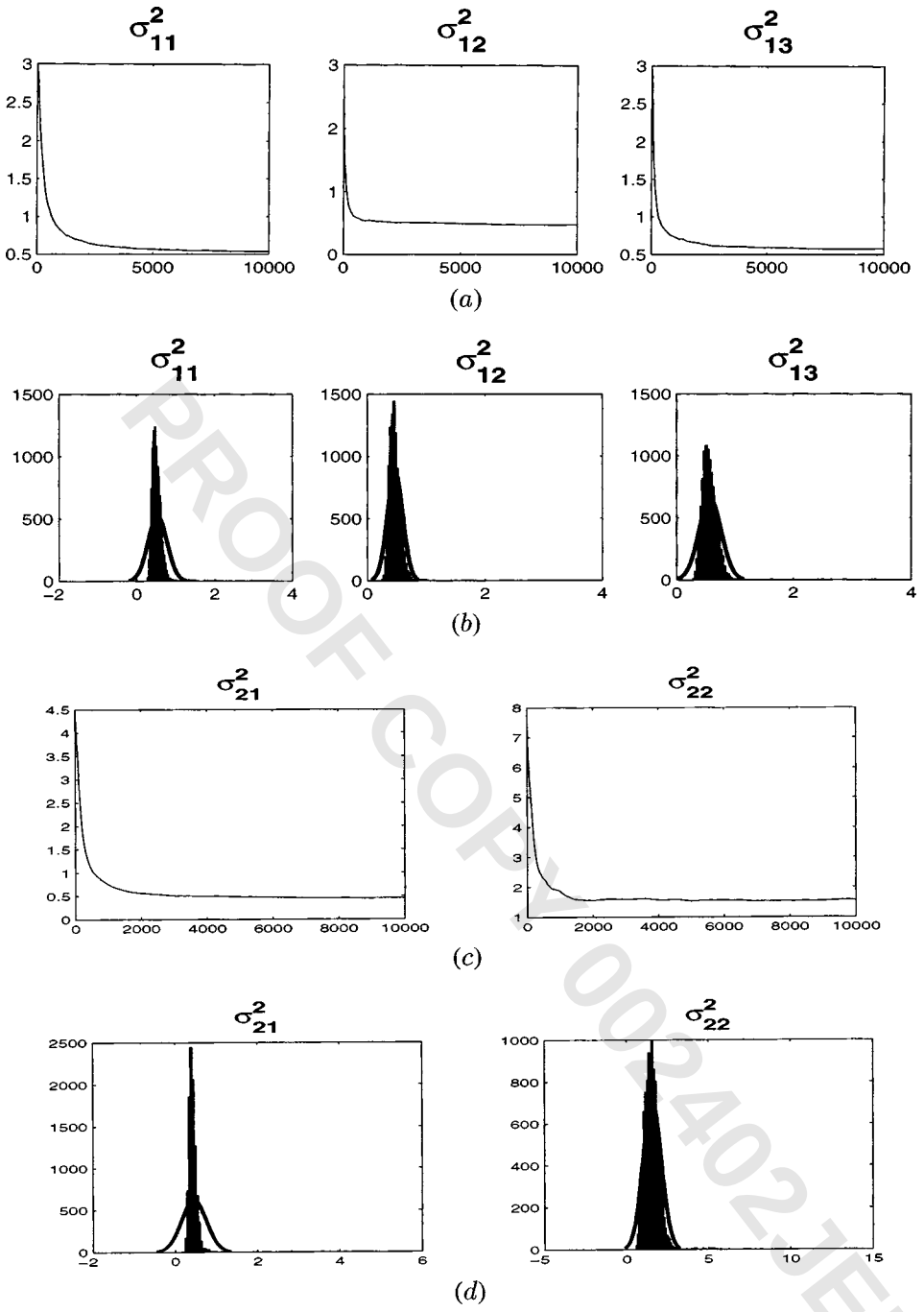

Fig. 5 (a) Convergence of the empirical expectations of the variances σ_{ij} of the source 1, (b) histograms of the variances of the source 1, (c) convergence of the empirical expectations of the variances σ_{ij} of the source 2, (d) histograms of the variances of the source 2.

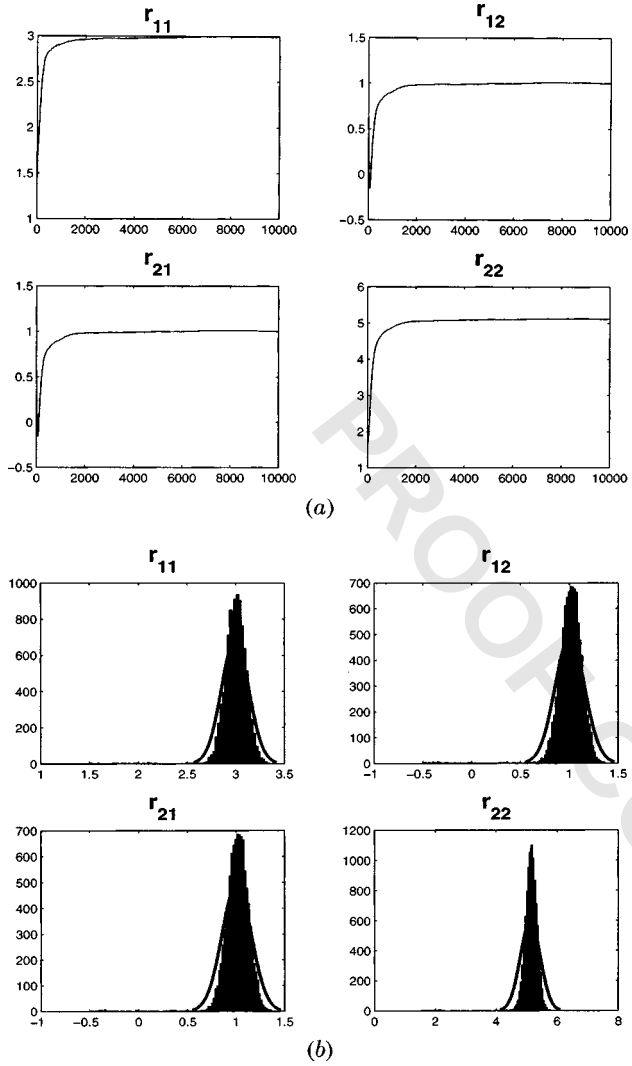

Fig. 6 (a) Convergence of the empirical expectations of the noise variances, (b) histograms of the noise variances.

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

$$
\begin{cases}
\mu_p = \dfrac{N_k \overline{s} + \alpha w_k^0 \mu_0}{N_k^{\dagger} \alpha w_k^0}, \\
v_p = \dfrac{v_k}{N_k^{\dagger} \alpha w_k^0}, \\
\eta_p = \dfrac{N_k^{\dagger} \alpha w_k^0}{2}, \\
\beta_p = \dfrac{\alpha w_k^0 v_0}{2} + \dfrac{s^2}{2} + \dfrac{1}{2} \dfrac{N_k \alpha w_k^0}{N_k^{\dagger} \alpha w_k^0} (\overline{s} - \mu_0)^2, \\
\overline{s} = \dfrac{\sum_{r \in s_k} s(r)}{N_k}, \\
s^2 = \sum_{r \in \mathcal{S}_k} s(r)^2 - N_k \overline{s}^2,
\end{cases}
$$

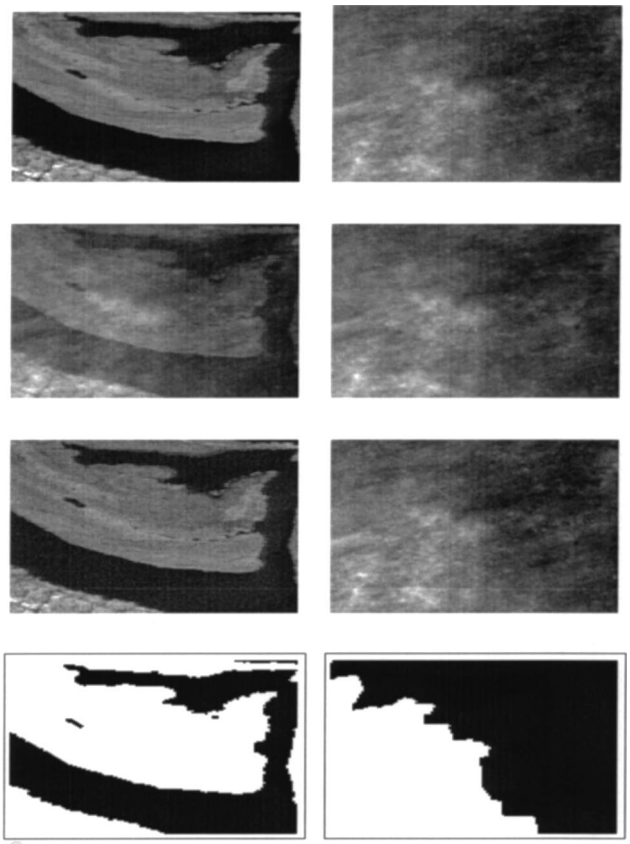

Fig. 7 From top to bottom: original sources, mixed sources, estimated sources, and segmented images.

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

$$
\begin{cases}
\mathcal{S}_k = \{ r \in \mathcal{S} | Z(r) = k \}, \\
N_k = |\mathcal{S}_k|.
\end{cases}
$$

8 Appendix B: Bartlett Decomposition

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

$$
W = B \frac{\Sigma_1^{\nu} v_k v_k^T}{\nu} B^T,
$$

where $\Sigma = BB^T$. This method involves νm simulations from univariate normal distribution leading to a high computational cost when ν increases.

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

Theorem Let W be $\mathcal{W}_{im}(\nu, \Sigma)$ and $\Sigma = BB^T$. Put $W = 1/\nu BVV^T B^T$, where V is a lower-triangular $m \times m$ matrix with positive diagonal elements. Then the elements v_{ij} $(1 \leqslant j \leqslant i \leqslant m)$ are independent, and each v_{ii}^2 is $\chi_{\nu - i + 1}^2$ $(i = 1, \ldots, m)$ 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*
- $W(m,m) = sqrt[2*gamrnd(tdf/2,1)]$;
- $B = [chol(\Sigma)]'$;
- $W = B*W$; and
- $W = W*W'/\nu$;

//---End of sampling------//

where *gamrnd* is a random generator from a gamma distribution, *randn* from a normal distribution and *chol* is the Cholesky factorization of a matrix.

References

1. A. Cichocki and R. Unbehaunen, ''Robust neural networks with online learning for blind identification and blind separation of sources,'' *IEEE Trans. Circuits Syst.* **43**(11), 894–906 (1996).
2. S. J. Roberts, ''Independent component analysis: Source assessment, and separation, a Bayesian approach,'' *IEE Proc. Vision Image Signal Process.* **145**(3), (1998).
3. T.-W. Lee, M. Lewicki, and T. Sejnowski, ''Unsupervised classification with non-Gaussian mixture models using ICA,'' *Advances in Neural Information Processing Systems*, pp. 508–512, MIT Press, Cambridge, MA (1999).
4. T.-W. Lee, M. Lewicki, and T. Sejnowski, ''Independent component analysis using an extended informax algorithm for mixed sub-Gaussian and super-Gaussian sources,'' *Neural Comput.* **11**, 409–433 (1999).
5. T.-W. Lee, M. Girolami, A. Bell, and T. Sejnowski, ''A unifying information-theoretic framework for independent component analysis,'' *Int. J. Comput. Math. Appl. Computation* **39**(11), 1–21 (2000).
6. I. Ziskind and M. Wax, ''Maximum likelihood localization of multiple sources by alternating projection,'' *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-36**(10), 1553–1560 (1988).
7. M. Wax, ''Detection and localization of multiple sources via the stochastic signals model,'' *IEEE Trans. Signal Process.* **39**(11), 2450–2456 (1991).
8. J.-F. Cardoso, ''Infomax and maximum likelihood for source separation,'' *IEEE Lett. Signal Process.* **4**(4), 112–114 (1997).
9. J.-L. Lacoume, ''A survey of source separation,'' in *Proc., First Int. Conf. on Independent Component Analysis and Blind Source Separation, ICA'99*, pp. 1–6, Aussois, France, January 11–15 (1999).
10. E. Oja, ''Nonlinear PCA criterion and maximum likelihood in independent component analysis,'' in *Proc., First Int. Conf. on Independent Component Analysis and Blind Source Separation, ICA'99*, pp. 143–149, Aussois, France, January 11–15 (1999).
11. R. B. MacLeod and D. W. Tufts, ''Fast maximum likelihood estimation for independent component analysis,'' in *Proc. First Int. Conf. on Independent Component Analysis and Blind Source Separation, ICA'99*, pp. 319–324, Aussois, France, January 11–15 (1999).
12. O. Bermond and J.-F. Cardoso, ''Approximate likelihood for noisy mixtures,'' in *Proc. First Int. Conf. on Independent Component Analysis and Blind Source Separation, ICA'99*, pp. 325–330, Aussois, France, January 11–15 (1999).
13. P. Comon, C. Jutten, and J. Herault, ''Blind separation of sources. 2. problems statement,'' *Signal Process.* **24**(1), 11–20 (1991).
14. C. Jutten and J. Herault, ''Blind separation of sources .1. an adaptive algorithm based on neuromimetic architecture,'' *Signal Process.* **24**(1), 1–10 (1991).
15. E. Moreau and B. Stoll, ''An iterative block procedure for the optimization of constrained contrast functions,'' in *Proc. First Int. Conf. on Independent Component Analysis and Blind Source Separation, ICA'99*, pp. 59–64, Aussois, France, January 11–15 (1999).
16. P. Comon and O. Grellier, ''Non-linear inversion of underdetermined mixtures,'' in *Proc. First Int. Conf. on Independent Component Analysis and Blind Source Separation, ICA'99*, pp. 461–465, Aussois, France, January 11–15 (1999).
17. J.-F. Cardoso and B. Laheld, ''Equivariant adaptive source separation,'' *IEEE Trans. Signal Process.* **44**(12), 3017–3030, (1996).
18. A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and Eric Moulines, ''A blind source separation technique based on second order statistics,'' *IEEE Trans. Signal Process.* **45**(2), 434–444 (1997).
19. S.-I. Amari and J.-F. Cardoso, ''Blind source separation—semiparametric statistical approach,'' *IEEE Trans. Signal Process.* **45**(11), 2692–2700 (1997).
20. J.-F. Cardoso, ''Blind signal separation: statistical principles,'' *Proc. IEEE. Special issue on blind identification and estimation* ■(10), 2009–2025 (1998).
21. J. J. Rajan and P. J. W. Rayner, ''Decomposition and the discrete karhunen-loeve transformation using a Bayesian approach,'' *IEE Proc. Vision Image Signal Process.* **144**(2), 116–123 (1997).
22. K. Knuth, ''Bayesian source separation and localization,'' in *SPIE'98 Proc.: Bayesian Inference for Inverse Problems*, A. Mohammad-Djafari, Ed., pp. 147–158 (July 1998).
23. K. Knuth and H. Vaughan, Jr., ''Convergent Bayesian formulation of blind source separation and an electromagnetic source estimation,'' in *MaxEnt 98 Proc.: Int. Workshop on Maximum Entropy and Bayesian Methods*, pp. 217–226, Garching, Germany (1998).
24. S. E. Lee and S. J. Press, ''Robustness of Bayesian factor analysis estimates,'' *Commun. Stat: Theory Meth.* **27**(8), (1998).
25. K. Knuth, ''A Bayesian approach to source separation,'' in *Proc. First Int. Workshop Independent Component Analysis and Signal Separation, ICA'99*, C. J. J.-F. Cardoso and P. Loubaton, Eds., Aussios, France (1999).
26. T.-W. Lee, M. Lewicki, M. Girolami, and T. Sejnowski, ''Blind source separation of more sources than mixtures using overcomplete representation,'' *IEEE Signal Process. Lett.* **6**(4), ■ (April 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., MaxEnt Workshops, Boise, Amer. Inst. Physics (July 1999).
28. P. Senecal and S. Amblard, ''MCMC methods for discrete source separation,'' in *Bayesian Inference and Maximum Entropy Methods*, pp. 350–360, Proc. of MaxEnt, Gif-sur-Yvette, France, Amer. Inst. Physics (July 2000).
29. H. Snoussi and A. Mohammad-Djafari, ''Bayesian unsupervised learning for source separation with mixture of Gaussians prior,'' *Int. J. VLSI Signal Processing Systems* (2002).
30. R. Choudrey and S. Roberts, ''Variational mixture of Bayesian independent component analyzers,'' *Neural Comput.* **15**(1), (2003).
31. R. Choudrey and S. Roberts, ''Variational Bayesian mixture of independent component analysers for finding self-similar areas in images,'' in *ICA*, Nara, Japan (April 2003).
32. R. Choudrey and S. Roberts, ''Bayesian ICA with hidden Markov model sources,'' in *ICA*, Nara, Japan (April 2003).
33. H. Attias, ''Blind separation of noisy mixture: An EM algorithm for independent factor analysis,'' *Neural Comput.* **11**, 803–851 (1999).
34. 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, pp. 293–302 (September 2001).
35. J. F. Cardoso, ''Informax and maximum likelihood for source separation,'' *IEEE Lett. Signal Process.* **4**, 112–114 (1997).
36. K. Knuth, ''A Bayesian approach to source separation,'' in *Proc., Independent Component Analysis Workshop*, pp. 283–288 (1999).
37. N. Peyrard, ''Convergence of MCEM and related algorithms for hidden Markov random field,'' Research Report 4146, INRIA (2001).
38. M. A. Tanner and W. H. Wong, ''The calculation of posterior distributions by data augmentation,'' *J. Am. Stat. Assoc.* **82**(398), 528–540 (1987).
39. H. Snoussi and A. Mohammad-Djafari, ''Bayesian separation of HMM sources,'' in *Bayesian Inference and Maximum Entropy Methods*, pp. 72–88, R. L. Fry, Ed., MaxEnt Workshops, Amer. Inst. Physics (August 2001).
40. H. Snoussi and A. Mohammad-Djafari, ''Information geometry and prior selection,'' in *Bayesian Inference and Maximum Entropy Methods*, C. Williams, Ed., pp. 307–327, MaxEnt Workshops, Amer. Inst. Physics (Aug. 2002).
41. S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs, Vol. 191, AMS, Oxford University Press, Oxford (2000).
42. J. W. Brewer, ''Kronecker products and matrix calculus in system theory,'' *IEEE Trans. Circuits Syst.* **CS-25**(9), 772–781 (1978).
43. C. Robert, *Méthodes de Monte-Carlo par chaines de Markov*, Economica, Paris, France (1996).
44. 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., July 2000, Gif-sur-Yvette, France, Proc. of MaxEnt, pp. 388–406, Amer. Inst. Physics.
45. O. Bermond, PhD thesis, École Nationale Supérieure des Télécommunications (2000).
46. B. D. Ripley, *Stochastic Simulation*, Wiley, New York (1987).

Hichem Snoussi received his diploma degree in electrical engineering from the Ecole Supérieure d'Electricité (Supélec), Gif-sur-Yvette, France, in 2000. He also received the his degree in signal processing from the Université de Paris-Sud, Orsay, France, in 2000. Since 2000, he has been working towards his PhD at the Laboratoire des Signaux et Systèmes, Centre National de la Recherche scientifique. His research interests include Bayesian technics for source separation, information geometry, and latent variable models.

Ali Mohammad-Djafari received the BS degree in electrical engineering from Polytechnique of Teheran, Iran, in 1975, the diploma degree (MSc) from Ecole Supérieure d'Electricité (Supélec), Gif-sur-Yvette, France, in 1977 and the ''Docteur-Ingénieur'' (PhD) degree and ''Doctorat d'Etat'' in Physics, from the Université Paris-Sud (UPS), Orsay, France, respectively, in 1981 and 1987. He was associate professor at UPS for two years (1981–1983). Since 1984, he has had a permanent position at the Centre National de la Recherche Scientifique (CNRS) and works at the Laboratoire des Signaux et Systèmes (L2S) at Supélec. From 1998 to 2002, he has been the head of the Signal and Image Processing division at this laboratory. Presently, he is Directeur de recherche and his main scientific interests are in developing new probabilistic methods based on Information Theory, Maximum Entropy and the Bayesian inference approaches for inverse problems in general, and more specifically signal and image reconstruction and restoration. The main application domain of his interests are Computed Tomography (X rays, PET, SPECT, MRI, microwave, ultrasound and eddy current imaging) either for medical imaging or for nondestructive testing (NDT) in industry.