A MEAN FIELD APPROXIMATION APPROACH TO BLIND SOURCE SEPARATION WITH L_p PRIORS

Mahieddine M. ICHIR and Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes, CNRS -Supélec - UPS
Plateau de Moulon, 91192, Gif-sur-Yvette, France
email: {ichir,djafari@lss.supelec.fr}, web: {mahieddine.ichir,djafari}.free.fr

ABSTRACT

In this paper we address the problem of Bayesian blind source separation with generalized p-Gaussian priors for the sources (also known as L_p priors). These kind of priors are useful when modeling sparse sources (spiky signals, wavelet coefficients ...) The corresponding posterior laws are non linear and either maximum a posteriori (MAP) or posterior mean estimates are computationally difficult to obtain especially for values of p approaching unity. In this work, we consider a mean field approximation approach to approximate the joint posterior distribution by a separable distribution on its parameters: unobservable sources, mixing matrix, noise covariance matrix and hyper-parameters (source scale parameters).

This approach requires, however, marginalisation of the log-likelihood with respect to these parameters. With appropriate prior assignments, this can be done explicitly for the mixing matrix, the noise covariance matrix and the scale parameters. For the sources, we consider a Kullback distance based approximation in order to obtain estimates of the first two moments of the sources. Simulation results are presented to support the proposed approach.

1. INTRODUCTION

Blind source separation (BSS) has emerged as an active area of research and finds application in various fields of engineering. It consists mainly in finding a set of unobservable sources from a set of their linear and instantaneous mixtures, formalized by:

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{x}_t$ is an m-column vector of the observed data at time t, $\boldsymbol{s}_t$ is an n-column vector of the unobserved sources at time t, $\boldsymbol{A}$ is the $m \times n$ mixing matrix and ε_t is the noise vector where it is assumed in the sequel that $\varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon)$[1].

The Bayesian solution to the BSS problem begins by writing the posterior joint distribution of the unknown parameters: the sources ($\boldsymbol{S} = \boldsymbol{s}_{1:T}$), the mixing matrix ($\boldsymbol{A}$) and the noise inverse covariance matrix (Σ_ε):

$$p(\boldsymbol{S}, \boldsymbol{A}, \Sigma_\varepsilon | \boldsymbol{X}) \propto p(\boldsymbol{X} | \boldsymbol{S}, \boldsymbol{A}, \Sigma_\varepsilon)\, \pi(\boldsymbol{S}, \boldsymbol{A}, \Sigma_\varepsilon) \tag{2}$$

where $p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{A}, \Sigma_\varepsilon)$ is the likelihood function and $\pi(\boldsymbol{S}, \boldsymbol{A}, \Sigma_\varepsilon)$ is the joint prior distribution of the parameters where we consider herein a separable prior on these parameters. An estimate is then defined, generally the maximum a posteriori or the posterior mean.

[1] for convenience, we work with *inverse* covariance matrices.

In our work, we are concerned with generalized p-Gaussian (gpG) priors for the sources of the form:

$$\pi(s_{i,t}) \propto \exp(-\lambda_i |s_{i,t}|^p), \quad 1 \leq p < 2 \tag{3}$$

for $i = 1, \ldots, n$, with $\pi(\boldsymbol{S}) = \prod_{i,t} \pi(s_{i,t})$, assuming spatial independence and time stationarity, and λ_i is a scale parameter. These priors have been used to model sparse signals, like the wavelet coefficients[6, 9, 3]. However, these kind of priors present some optimization difficulties, especially for values of p approaching unity.

A first approximation considered in this paper, relies on mean field approaches to BSS[8] in order to approximate the joint posterior distribution of the unknowns by a separable one. However it is not an ICA approach as in[8] in the sense that the approximating distribution of the sources in not separable, keeping thus the correlation feature of the latter. A second approximation is to approach this marginal distribution by a double exponential one based on the Kullback distance. The proposed approach is an alternative solution to the Monte Carlo Markov Chain solution considered in[3].

This paper is organized as follows: in section 2 we define the conjugate priors on the mixing matrix, the noise inverse covariance matrix and the sources scale parameters. In section 3 we briefly introduce the mean field approach and then give detailed expressions of the different approximating marginals of the parameters of interest in section 3.1. In 4 a simulation example is presented to support the proposed approach and we finally conclude in 5.

2. MIXING MATRIX, NOISE INVERSE COVARIANCE AND SCALE PARAMETER PRIORS

Without loss of generality, we consider the mixing matrix to be Gaussian:

$$\pi(\boldsymbol{A}|\mu_A, \Sigma_A) = \mathcal{N}(\mu_A, \Sigma_A) \tag{4}$$

The prior probability of the noise inverse covariance matrix Σ_ε is a Wishart distribution (a generalization of the χ^2 distribution for positive definite matrices):

$$\pi(\Sigma_\varepsilon | \nu, \Sigma) \propto |\Sigma_\varepsilon|^{\frac{\nu - m - 1}{2}} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(\Sigma_\varepsilon \Sigma^{-1}\right)\right) \tag{5}$$

where ν is the number of degrees of freedom of Σ_ε and Σ is a scale matrix. The expected value with respect to this prior is $\tilde{\Sigma}_\varepsilon = \nu\Sigma$. The Wishart prior is a conjugate prior that will allow us to account, in the BSS model, for a correlated Gaussian noise in the data.

The scale parameters associated to the sources λ_i are assigned Gamma priors of the form:

$$\pi(\lambda_i|\nu_0,\theta_0) \propto \lambda_i^{\nu_0-1}\exp(-\theta_0\lambda_i)\,\mathbb{I}_{\mathbb{R}_+} \tag{6}$$

for $i = 1,\ldots,n$.

3. THE MEAN FIELD APPROXIMATION TO SOURCE SEPARATION

The mean field approximation of a posterior distribution $p(\Theta|Data)$ of a set of parameters Θ begins by writing the Kullback distance between a *separable approximating distribution* $q(\Theta) = \prod_i q_i(\Theta_i)$ and that posterior:

$$\mathscr{D}(q|p) = \mathbb{E}_q\left[\log\frac{q}{p}\right] = \text{cte} - \mathscr{F}(q|\pi,L) \tag{7}$$

where[2]

$$\mathscr{F}(q|\pi,L) = -\sum_i \left\langle \log\frac{q_i}{\pi_i}\right\rangle_{q_i} + \langle \log L\rangle_q \tag{8}$$

where L is the likelihood function and π is a separable prior of the parameter set Θ (the sources, the mixing matrix and the noise inverse covariance matrix in our BSS problem). The objective, now, is to find, maximizing $\mathscr{F}(q|\pi,L)$, a set of separable approximating distributions $q_i(\Theta_i)$. This is done alternatively on each parameter conditionally on the others, where the solution to that variational problem is given by[7, 5]:

$$\hat{q}_i(\Theta_i) \;\propto\; \pi_i(\Theta_i)\exp(\psi(\Theta_i)) \tag{9}$$

where[3] $\psi(\Theta_i) = \langle \log L(\boldsymbol{X}|\Theta)\rangle_{q_{|i}}$ is a function of Θ_i obtained by marginalizing the likelihood function with respect to all the other parameters except Θ_i.

3.1 The expression of $\psi(\Theta_i)$ for BSS

Under the Gaussian noise assumption, the log-likelihood is given by:

$$\begin{aligned}\log L(\boldsymbol{x}_{1:T}|\Theta) =& \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\sum_t(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})^\dagger\Sigma_\varepsilon(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s}) \\ &+ \text{cte} \\ \equiv& \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\sum_t \boldsymbol{s}^\dagger\boldsymbol{A}^\dagger\Sigma_\varepsilon\boldsymbol{A}\boldsymbol{s} \\ &+ \sum_t \boldsymbol{x}^\dagger\Sigma_\varepsilon\boldsymbol{A}\boldsymbol{s} \qquad (10a) \\ \equiv& \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\sum_t \boldsymbol{A}_v^\dagger(\mathbb{I}\otimes\boldsymbol{s})\Sigma_\varepsilon(\mathbb{I}\otimes\boldsymbol{s}^\dagger)\boldsymbol{A}_v \\ &+ \sum_t \boldsymbol{x}^\dagger\Sigma_\varepsilon(\mathbb{I}\otimes\boldsymbol{s}^\dagger)\boldsymbol{A}_v \qquad (10b)\end{aligned}$$

where Θ stands for $\boldsymbol{S},\boldsymbol{A},\Sigma_\varepsilon$ and $\lambda_{1:n}$. (10a) and (10b) are two alternate expressions for the log-likelihood function where we dropped the time index t for convenience. $\boldsymbol{A}_v$ is the vector wise representation of a matrix defined by:

$$\boldsymbol{A}_v = [\boldsymbol{A}_{(1,.)},\ldots,\boldsymbol{A}_{(m,.)}]^\dagger \tag{11}$$

and $\otimes$ is the kronecker (tensor) product of matrices[1].

[2] we use the notation $\tilde{f} = \langle f(\theta)\rangle_q = \mathbb{E}_q[f(\theta)]$ to denote expectation.

[3] $\langle f(\theta_1,\ldots,\theta_n)\rangle_{q_{|i}} = \langle f(\theta_1,\ldots,\theta_n)\rangle_{q(\ldots,\theta_{i-1},\theta_{i+1},\ldots)} = \psi(\theta_i)$

3.1.1 *Approximate posterior of $\boldsymbol{A}_v$*

The marginal log-likelihood of the mixing matrix $\psi(\boldsymbol{A}_v)$ is given by:

$$\begin{aligned}\psi(\boldsymbol{A}_v) \;=\; & -\frac{1}{2}\sum_t \boldsymbol{A}_v^\dagger\Big(\tilde{\Sigma}_\varepsilon\otimes\tilde{\Sigma}_s^{-1} + \tilde{\Sigma}_\varepsilon\otimes\tilde{s}\tilde{b}\tilde{\boldsymbol{s}}^\dagger\Big)\boldsymbol{A}_v \\ & + \sum_t\big(\tilde{\Sigma}_\varepsilon\boldsymbol{x}\otimes\tilde{\boldsymbol{s}}\big)^\dagger\boldsymbol{A}_v \end{aligned} \tag{12}$$

where $\langle \boldsymbol{s}\boldsymbol{s}^\dagger\rangle_{q(s)} = \tilde{\Sigma}_s^{-1} + \tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^\dagger$ and $\tilde{\Sigma}_\varepsilon = \langle\Sigma_\varepsilon\rangle_{q(\Sigma_\varepsilon)}$. The *conjugate* prior of equation (4) enables us to write the approximate posterior as Gaussian:

$$q(\boldsymbol{A}) = \mathscr{N}(\tilde{\boldsymbol{A}}_v,\tilde{\Sigma}_A) \tag{13}$$

where $\langle \boldsymbol{A}_v\boldsymbol{A}_v^\dagger\rangle_{q(A)} = \tilde{\Sigma}_A^{-1} + \tilde{\boldsymbol{A}}_v\tilde{\boldsymbol{A}}_v^\dagger$, and:

$$\begin{aligned}\tilde{\Sigma}_A &= T\tilde{\Sigma}_\varepsilon\otimes\tilde{\Sigma}_s^{-1} + \sum_t\tilde{\Sigma}_\varepsilon\otimes\tilde{\boldsymbol{s}}_t\tilde{\boldsymbol{s}}_t^\dagger + \Sigma_A, \\ \tilde{\boldsymbol{A}}_v &= \tilde{\Sigma}_A^{-1}\Big[\sum_t\big(\tilde{\Sigma}_\varepsilon\boldsymbol{x}\otimes\tilde{\boldsymbol{s}}\big) + \Sigma_A\mu_A\Big]\end{aligned}$$

3.1.2 *Approximate posterior of Σ_ε*

The marginal log-likelihood of the noise inverse covariance matrix is given by:

$$\psi(\Sigma_\varepsilon) = \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{T}{2}\mathrm{Tr}\Big(\Sigma_\varepsilon\left\langle\boldsymbol{A}\tilde{\Sigma}_s^{-1}\boldsymbol{A}^\dagger\right\rangle_{q(A)}\Big) - \frac{1}{2}\mathrm{Tr}(\Sigma_\varepsilon\boldsymbol{Q}) \tag{14}$$

where

$$\begin{aligned}\boldsymbol{Q} = \sum_t\big(&\boldsymbol{x}\boldsymbol{x}^\dagger + \tilde{\boldsymbol{A}}\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^\dagger\tilde{\boldsymbol{A}}^\dagger \\ &+ (\mathbb{I}\otimes\tilde{\boldsymbol{s}}^\dagger)\tilde{\Sigma}_A^{-1}(\mathbb{I}\otimes\tilde{\boldsymbol{s}}) - \tilde{\boldsymbol{A}}\tilde{\boldsymbol{s}}\boldsymbol{x}^\dagger - \boldsymbol{x}\tilde{\boldsymbol{s}}^\dagger\tilde{\boldsymbol{A}}^\dagger\big)\end{aligned} \tag{15}$$

In a matrix form, $\boldsymbol{Q}$ can be equivalently written:

$$\boldsymbol{Q} = (\boldsymbol{X}-\boldsymbol{A}\boldsymbol{S})(\boldsymbol{X}-\boldsymbol{A}\boldsymbol{S})^\dagger + \boldsymbol{G}(\tilde{\Sigma}_A^{-1},\tilde{\boldsymbol{S}})$$

with

$$\boldsymbol{G}(\tilde{\Sigma}_A^{-1},\tilde{\boldsymbol{S}})[i,j] = \mathrm{Tr}\Big(\gamma_A^{i,j}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{S}}^\dagger\Big)$$

for $(i,j) = 1,\ldots,m$, where $\gamma_A^{i,j}$ is a n-square sub-matrix of $\tilde{\Sigma}_A^{-1}$:

$$\tilde{\Sigma}_A^{-1} = \begin{bmatrix}\gamma_A^{1,1} & \cdots & \gamma_A^{1,m} \\ \vdots & \ddots & \vdots \\ \gamma_A^{m,1} & \cdots & \gamma_A^{m,m}\end{bmatrix}$$

The log-likelihood (14) needs the evaluation of $\left\langle\boldsymbol{A}\tilde{\Sigma}_s^{-1}\boldsymbol{A}^\dagger\right\rangle_{q(A)}$. From the following lemma:

Lemma 1 *Let $\boldsymbol{Y}_v$ be the mn column vector representation of the $m\times n$ matrix $\boldsymbol{Y}$ defined by (11), and let the statistics of $\boldsymbol{Y}_v$ be given by $\langle\boldsymbol{Y}_v\boldsymbol{Y}_v^\dagger\rangle_{q(\boldsymbol{Y})} = \tilde{\Sigma}_y^{-1} + \tilde{\boldsymbol{Y}}_v\tilde{\boldsymbol{Y}}_v^\dagger$, then*

$$\begin{aligned}\left\langle\boldsymbol{Y}\boldsymbol{R}\boldsymbol{Y}^\dagger\right\rangle_{q(\boldsymbol{y})}[i,j] = \sum_{k=1}^n\sum_{l=1}^n \boldsymbol{R}[k,l]\tilde{\Sigma}_y^{-1}[n(i-1)+k,n(j-1)+l] \\ + \big(\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^\dagger\big)[i,j]\end{aligned}$$

for $(i,j)=1,\ldots,m$.

Then $\psi(\Sigma_\varepsilon)$ given in equation (14) rewrites:

$$\psi(\Sigma_\varepsilon)=\frac{T}{2}\log|\Sigma_\varepsilon|-\frac{T}{2}\mathrm{Tr}\Big(\Sigma_\varepsilon \boldsymbol{D}_{\boldsymbol{A}}(\tilde{\Sigma}_s^{-1},\tilde{\Sigma}_A^{-1},\tilde{\boldsymbol{A}})\Big)-\frac{1}{2}\mathrm{Tr}(\Sigma_\varepsilon \boldsymbol{Q}) \quad (16)$$

where $\boldsymbol{D}_{\boldsymbol{Y}}(\boldsymbol{R},\tilde{\Sigma}_{\boldsymbol{y}}^{-1},\tilde{\boldsymbol{Y}})=\langle \boldsymbol{Y}\boldsymbol{R}\boldsymbol{Y}^\dagger\rangle_{q(\boldsymbol{y})}$. With the prior defined in (5), the log posterior of Σ_ε is finally given by:

$$\log q(\Sigma_\varepsilon|\boldsymbol{X},\theta)=\frac{(T+\nu-m-1)}{2}\log|\Sigma_\varepsilon|-\frac{1}{2}\mathrm{Tr}(\Sigma_\varepsilon \boldsymbol{Q}_\varepsilon) \quad (17)$$

where $\boldsymbol{Q}_\varepsilon=\Sigma^{-1}+\boldsymbol{Q}+T\boldsymbol{D}_{\boldsymbol{A}}(\tilde{\Sigma}_s^{-1},\tilde{\Sigma}_A^{-1},\tilde{\boldsymbol{A}})$, defining a posteriori a Wishart distribution with a mean matrix $\tilde{\Sigma}_\varepsilon=(T+\nu)\boldsymbol{Q}_\varepsilon^{-1}$.

3.1.3 Approximate posterior of $\boldsymbol{S}$

In order to express the marginal log-likelihood of the sources, we first postulate the following:

Lemma 2 *Let $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be square symmetric matrices of dimensions m and mn respectively, and let η be a one dimensional n-column vector, then*

$$\frac{\partial\, Tr\big((\boldsymbol{Y}\otimes\eta\eta^\dagger)\boldsymbol{Z}\big)}{\partial\eta}=2\boldsymbol{F}(\boldsymbol{Y},\boldsymbol{Z})\eta$$

where

$$\boldsymbol{F}_{i,j}(\boldsymbol{Y},\boldsymbol{Z})=\sum_{k=1}^{m}\sum_{l=1}^{m}\boldsymbol{Y}[k,l]\boldsymbol{Z}[(k-1)n+i,(l-1)n+j]$$

for $(i,j)=1,\ldots,n$.

The marginal log-likelihood with respect to the sources is then given by:

$$\begin{aligned}\psi(\boldsymbol{s}) &= \boldsymbol{x}^\dagger\tilde{\Sigma}_\varepsilon\tilde{\boldsymbol{A}}\boldsymbol{s}-\frac{1}{2}\mathrm{Tr}\Big((\tilde{\Sigma}_\varepsilon\otimes\boldsymbol{s}\boldsymbol{s}^\dagger)\tilde{\Sigma}_A^{-1}\Big)-\frac{1}{2}\boldsymbol{s}^\dagger\tilde{\boldsymbol{A}}^\dagger\tilde{\Sigma}_\varepsilon\tilde{\boldsymbol{A}}\boldsymbol{s}\\ &= -\frac{1}{2}\boldsymbol{s}^\dagger\boldsymbol{Q}\boldsymbol{s}+\boldsymbol{v}^\dagger\boldsymbol{s} \quad (18)\end{aligned}$$

where $\boldsymbol{Q}=\tilde{\boldsymbol{A}}^\dagger\tilde{\Sigma}_\varepsilon\tilde{\boldsymbol{A}}+\boldsymbol{F}(\tilde{\Sigma}_\varepsilon,\tilde{\Sigma}_A^{-1})$ and $\boldsymbol{v}=\tilde{\boldsymbol{A}}^\dagger\tilde{\Sigma}_\varepsilon\boldsymbol{x}$. With the gpG prior of equation (3), the log-approximate distribution is given by:

$$\log q(\boldsymbol{s})=-\frac{1}{2}\boldsymbol{s}^\dagger\boldsymbol{Q}\boldsymbol{s}+\boldsymbol{v}^\dagger\boldsymbol{s}-\sum_{i=1}^{n}\lambda_i|s|_i^p,\quad 1\le p<2 \quad (19)$$

The two first moments of the sources with respect to this approximating distribution are not explicitly given, they are solutions to non linear multidimensional equations. We consider two approximations of these two quantities: 1) a first one based on a Gaussian approximation of the prior of $\boldsymbol{s}$; 2) a second one based on the minimization of the Kullback distance between the one dimensional Gaussian approximation and the double exponential distribution:

1. $\log\tilde{q}(\boldsymbol{s})=\log q(\boldsymbol{s}|p=2)=-\frac{1}{2}\boldsymbol{s}^\dagger\boldsymbol{Q}\boldsymbol{s}+\boldsymbol{v}^\dagger\boldsymbol{s}-\boldsymbol{s}^\dagger\boldsymbol{R}_\lambda\boldsymbol{s}$, which results in $\tilde{\Sigma}_{\tilde{q}}=\boldsymbol{Q}+2\boldsymbol{R}_\lambda$ and $\tilde{s}=\tilde{\Sigma}_{\tilde{q}}^{-1}\boldsymbol{v}$.
2. $\tilde{\Sigma}_s=\sqrt{2}\boldsymbol{U}$, where $\tilde{\Sigma}_{\tilde{q}}=\boldsymbol{U}^\dagger\boldsymbol{U}$.

3.1.4 Approximate posterior of the hyper-parameters

a) Scale parameter. In order to develop an expression for the marginal log-likelihood of the sources scale parameter, we first write the log-likelihood:

$$\log\pi(\boldsymbol{S}|\lambda_{1:n},p)=\sum_i\Big(pT\log\lambda_i-\lambda_i\sum_t|s_{i,t}|^p\Big) \quad (20)$$

Similar developments as those of equations (7) to (9) yield an expression for the log-approximate posterior:

$$\log q(\lambda_i)=(pT+\nu_0-1)\log\lambda_i-\lambda_i\Big(\theta_0+\sum_t\langle|s_{i,t}|^p\rangle_{q(s)}\Big) \quad (21)$$

defining thus a Gamma distribution where the expected value is given by:

$$\tilde{\lambda}_i=\langle\lambda_i\rangle_{q(\lambda_i)}=\frac{T/p+\nu_0}{\theta_0+\sum_t\langle|s_{i,t}|^p\rangle_{q(s)}} \quad (22)$$

where the quantity $\langle|s_{i,t}|^p\rangle_{q(s)}$ is approximated by means of monte carlo integration.

b) Power parameter. The power parameter p of the generalized p-Gaussian (gpG) prior distribution of equation (3) can be estimated by the method of moments[6]:

$$p=F^{-1}\left(\frac{\langle|s|\rangle_{\pi(s)}^2}{\langle s^2\rangle_{\pi(s)}}\right), \quad (23)$$

where $F(p)=\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}$ and $\Gamma(.)$ is the standard Gamma function. The absolute moment $\langle|s|\rangle_{\pi(s)}$ and the second moment $\langle s^2\rangle_{\pi(s)}$ can be evaluated by their empirical estimates or by Monte Carlo integration approximation.

4. SIMULATIONS

In order to support the proposed approach, two gpG signals have been generated with power parameters $p_1=1.2$ and $p_2=1.3$ resp. A square mixing matrix $\boldsymbol{A}=\big[[1,0.8];[0.8,1]\big]$ have been considered, and a spatially correlated noise with an inverse covariance matrix $\Sigma_\varepsilon=\big[[.13,-.67];[-.67,.13]\big]$ has been added to the observations. Figure (1) shows scatter plots of the observed mixtures as function of the original sources. As one can expect, the mean field approach depends on the initialization of its parameters, so all the parameters have been initialized to their mean values: $\boldsymbol{A}^0=\mu_A=\mathbb{I}_{m,n}$, $\Sigma_\varepsilon^0=\Sigma>>\mathbb{I}_m$ and the sources where initialized by[4] $\boldsymbol{S}^0=\boldsymbol{X}$. As a measure of performance and comparison, we have considered the performance index[4] given by:

$$PI(B=\hat{\boldsymbol{A}}^{-1}\boldsymbol{A})=\frac{1}{2}\left[\sum_i\Big(\sum_j\frac{|B_{ij}|^2}{\max_l|B_{il}|^2}-1\Big)+\sum_j\Big(\sum_i\frac{|B_{ij}|^2}{\max_l|B_{lj}|^2}-1\Big)\right] \quad (24)$$

As a convergence criteria, one would naturally evaluate the functional given in equation (8) since the objective is to

[4]this is equivalent by initializing the sources by $(\boldsymbol{A}^{0,\dagger}\boldsymbol{A}^0)^{-1}\boldsymbol{A}^{0,\dagger}\boldsymbol{X}$

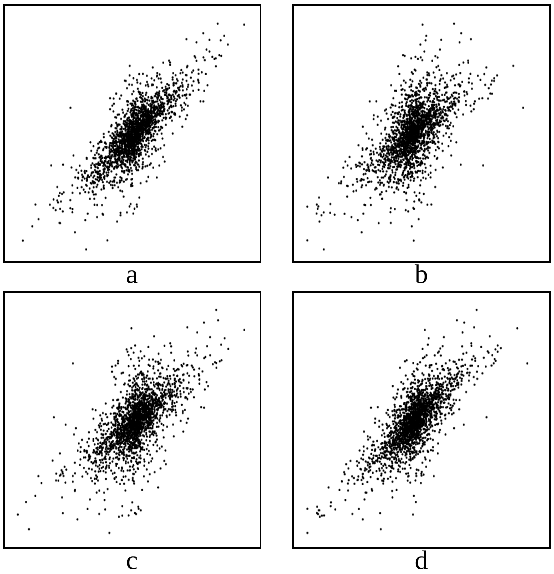

Figure 1: scatter plots of: (a). mixture X_1 vs. source S_1, (b). X_1 vs. S_2, (c). X_2 vs. S_1, (d). X_2 vs. S_2

look for the approximating distribution that maximizes this functional. However its evaluation is not so trivial as it needs explicit expressions of moments with respect to a Wishart prior. So we have choosen, for a stopping criteria, the stationnary points of successive differences of the mixing matrix norm. In figure (2) scatter plots of the estimated sources as function of the original ones are presented (a diagonal line is represented to visually evaluate the performances). Figure (3), represents the evolution of the performance index (PI) of equation (24) through the iterations.

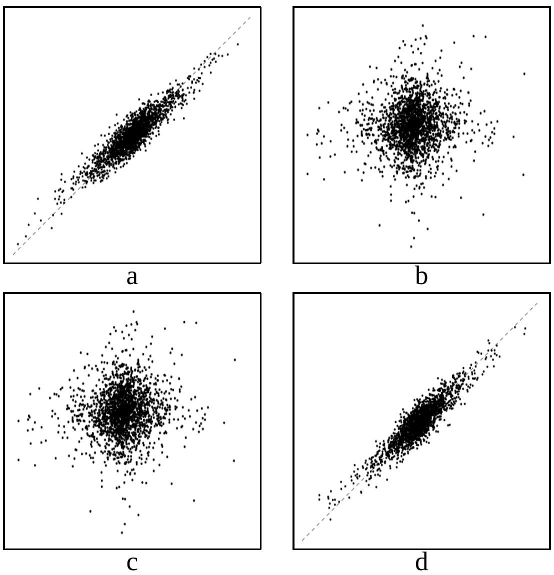

Figure 2: scatter plots of: (a). estimated source $\hat{S}_1$ vs. original source S_1, (b). $\hat{S}_1$ vs. S_2, (c). $\hat{S}_2$ vs. S_1, (d). $\hat{S}_2$ vs. S_2

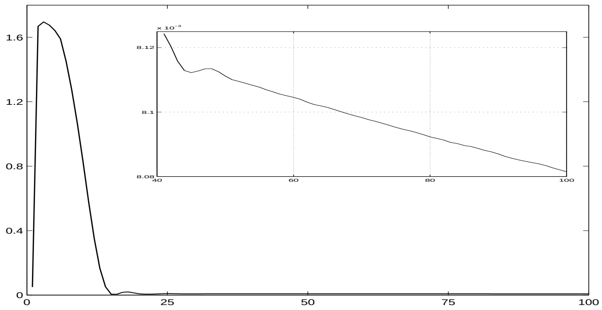

Figure 3: Evolution of the Performance Index (equation (24)) along the iterations (sub figure: zoom of the evolution of the PI at the last iterations).

5. CONCLUSION

In this paper, we have considered a mean field approximation to blind source separation under a Bayesian framework with L_p priors. These kind of priors are suited for modeling sparse signals such as the wavelet coefficients of piecewise regular signals. The mean field approach allowed us to establish a relatively simple but efficient algorithm. A matrix form of the noise covariance prior allowed us to account, in addition, for a spatially correlated Gaussian noise. We have shown, by a simulation example, that this is approach is quite promising. However, we think that some improvement can be made concerning the approximating distribution of the sources since the one presented is based on a unidimensional approximation. Even though the presented approach accounts for observations Gaussian noise (spatially correlated), we think that an additional denoising step should be processed on the estimated sources.

REFERENCES

[1] J. W. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Trans. on Circuits and Systems*, 9:772–781, 1978.

[2] Jean François Cardoso. Higher-order contrasts for independant component analysis. In *Neural Computation*, volume 11 of *MIT Letters*, pages 157–192. 1999.

[3] Mahieddine M. Ichir and Ali Mohammad-Djafari. Séparation de sources modélisées par des ondelettes. In *19eme Colloque sur le Traitement du Signal et des Images*, volume I, pages 185–188, ENST, Paris, Septembre 2003. GRETSI.

[4] Odile Macchi and Eric Moreau. Adaptive unsupervised separation of discrete sources. In *Signal Processing*, volume 73, pages 49–66, 1999.

[5] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, June 2002.

[6] Stephane G. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(7):674–693, Jully 1989.

[7] James W. Miskin. *Ensemble Learning for Independant Component Analysis*. PhD thesis, Selwyn College, Cambridge University, 2000.

[8] O. Winther P. A.d.F.R Højen-Sørensen and L. K. Hansen. Mean field implementation of bayesian ICA. 2001.

[9] Michael Zibulevsky and Barak A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionnary. *MIT Letters on Neural Computation*, 13:863–882, 2001.