

Separation of Non-negative Mixture of Non-negative Sources using a Bayesian Approach and MCMC Sampling

Saïd Moussaoui, David Brie, Ali Mohammad-Djafari, Cédric Carteret

Abstract—This paper considers the problem of blind source separation in the case where both the source signals and the mixing coefficients are non-negatives. The problem is referred to as non-negative source separation and the analysis is achieved in a Bayesian framework by taking the non-negativity of source signals and mixing coefficients as prior information. Since the main application concerns the analysis of spectral signals, to encode jointly non-negativity, sparsity and possible background in the sources, Gamma densities are used as priors. The source signals and the mixing coefficients are estimated by implementing a Monte Carlo Markov Chain (MCMC) for sampling their joint posterior density. Synthetic and experimental results motivate the problem of non-negative source separation and illustrate the effectiveness of the proposed method.

Index Terms—Bayesian estimation, Source separation, Non-negativity, Gamma distribution, Monte Carlo Markov Chains (MCMC), Spectroscopy.

I. INTRODUCTION

In analytical chemistry, it is often needed to process mixture data obtained by spectroscopy and/or chromatography analysis of multicomponent materials [1]–[3]. The processing aims at identifying the pure components of the materials and estimating the concentration of each component. These objectives are formalized as a source separation problem, where the linear instantaneous mixture model holds. In the case of optical spectroscopy, this model is validated according to Beer-Lambert law [4] (also termed as Beer’s law or Beer-Lambert-Bouguer law [5]). The pure component spectra are identified as the estimated sources and the concentrations as the mixing coefficients. The main constraint in this application is the non-negativity of both the source signals and the mixing coefficients. So, the problem is referred to as non-negative (positive) source separation. More generally, this non-negative source separation problem arises when one has to deal with non-negative mixtures of non-negative signals (spectra, images, hyperspectral data, etc.). In this paper, the case of spectral data is considered.

Manuscript submitted for publication in IEEE Transactions on Signal Processing, July 12, 2004. This work is supported by the “Region Lorraine” and the CNRS.

S. Moussaoui and D. Brie are with the CRAN CNRS UMR 7039, B.P. 239, University of Nancy 1, F-54506 Vandœuvre-lès-Nancy Cedex, France (e-mails: firstname.lastname@cran.uhp-nancy.fr)

A. Mohammad-Djafari is with the LSS CNRS-SUPÉLEC-UPS, University of Paris-Sud, 91192, Gif-sur-Yvette cedex, France (e-mail: djafari@lss.fr)

C. Carteret is with the LCPME CNRS UMR 7564, University of Nancy 1, 54600, Villers-lès-Nancy, France (e-mail: carteret@lcpce.cnrs-nancy.fr)

The linear instantaneous mixing model is expressed as

$$\mathbf{x}_t = \mathbf{A} \mathbf{s}_t + \mathbf{e}_t, \text{ for } t = 1, \dots, N, \quad (1)$$

where \mathbf{s}_t denotes the $n \times 1$ vector of the source signals $\{s_{(j,t)}\}_{j=1}^n$, \mathbf{x}_t the $m \times 1$ vector of the measured signals $\{x_{(i,t)}\}_{i=1}^m$, \mathbf{e}_t a $m \times 1$ vector of an additive noise (measurement errors and model uncertainties), \mathbf{A} is the $m \times n$ unknown mixing matrix containing the mixing coefficients $\{a_{(i,j)}\}_{i=1,j=1}^{m,n}$, and t is an observation variable. Having all the observations and using matrix notations, the mixing model is written as

$$\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E}, \quad (2)$$

where the matrices $\mathbf{X} \in \mathbb{R}^{m \times N}$, $\mathbf{S} \in \mathbb{R}_+^{n \times N}$ and $\mathbf{E} \in \mathbb{R}^{m \times N}$, contain respectively observations, source signals and noise sequences. The problem of blind source separation is then stated as follows: knowing the number of sources and having all the observations, estimate the sources signals and the mixing coefficients.

To achieve the separation, any prior knowledge and assumption about the mixing process and the source signals should be taken into account since this inverse problem is ill-posed, in the sense that the solution is not unique. Principal component analysis [6] (PCA), which is the most popular approach for the analysis of multivariate data, assumes that the signals to reconstruct are mutually uncorrelated, but this orthogonality constraint does not ensure neither the uniqueness nor the non-negativity of the solution. A more constraining assumption used for source separation is the mutual independence of the sources leading to the independent component analysis (ICA) concept [7], for which many algorithms has been developed (see the books [8]–[10] and the references therein). Assuming the mutual independence of the sources yields a solution which is unique (up to order and scale indeterminacies) but it does not ensure explicitly the non-negativity of both sources and mixing coefficients. Clearly, if the non-negative source signals are mutually statistically independent, they can be separated successfully by ICA methods and their non-negativity will be ensured implicitly (at least only few negative values appear in the estimates) as reported in [11], where a second order blind identification (SOBI) algorithm [12] was applied to the analysis of nuclear magnetic resonance (NMR) data. But when the source signals are not mutually independent or when their mutual independence is not observed due to the finite number of samples, the non-negativity information should be considered. It is possible to incorporate jointly non-negativity

and mutual independence of the sources, as proposed in [13], by using a nonlinear PCA method that optimizes a criterion that takes into account the source non-negativity, under the condition of *well grounded* sources. Unfortunately, since ICA methods produce an unmixing matrix, which is the (pseudo) inverse of the mixing matrix, the non-negativity of the mixing coefficients is not considered during the separation.

A second class of methods consists in minimizing a mean squares criterion under a non-negativity constraint, leading to algorithms differing on the way how this prior information is incorporated. In particular, the method presented in [14] performs an alternating least squares (ALS) estimation where the non-negativity is hardly imposed between successive iterations by setting to zero the negative values of source signals and mixing coefficients or by performing a non-negative least squares estimation [15]. The ALS method is the method which is widely used in the chemometrics community [16] where the problem is termed by multivariate curve resolution [17]. The non-negative matrix factorization (NMF) algorithm [18] is a second alternative of such approaches which achieves the decomposition by constructing a gradient descent algorithm over the objective function and updates iteratively sources and mixing coefficients by considering a particular multiplicative learning rule that ensures the estimates to be non-negatives. This method has also been applied to the case of noisy mixture signals as well as to the recovery of constituent spectra in chemical shift imaging [19]. Non-negative sparse coding (NNSC) method [20] also treats that type of problems by assuming non-negative sparse sources through the minimization of a penalized least square criterion while the non-negativity of the mixing coefficients is introduced in a similar way as in ALS method. Finally, positive matrix factorization (PMF) [21] is a more general method since it minimizes a compound regularized criterion that enforces positivity and sparseness of both source signals and mixing coefficients. However, its related optimization algorithm is numerically very expensive.

This paper addresses the problem of non-negative source separation in a Bayesian framework for an application to the analysis of mixtures in spectroscopy. The use of Bayesian theory to source separation is not new since it has been addressed in many papers [22]–[28] but, to our knowledge, its application to the separation of non-negative sources has only received few attention [29]–[31]. There are two main reasons that make Bayesian estimation approach very well suited to such an application. Firstly, Bayesian inference offers a very powerful theoretical framework to encode non-negativity information and, more generally, any additional prior knowledge on the mixing coefficients and the source signals. Secondly, due to the recent developments in Monte Carlo Markov chain (MCMC) methods [32]–[34] that enables to generate samples from the posterior density, various Bayesian estimators requiring integration or optimization can be used, even if the posterior law is not analytically tractable.

This paper is organized as follows: section II recalls the main idea of a Bayesian inference for source separation and presents the proposed probabilistic modeling for the analysis of spectral mixtures which consists in assigning Gamma density priors to both source signals and mixing

coefficient profiles. In section III, the Gibbs sampler [35] and a Metropolis-Hastings algorithm [36] are used to build a Markov chain which samples the joint posterior density from which source signals and mixing coefficients are estimated from the marginal distributions using the conditional mean estimator. All the Gibbs sampler steps and the required posterior conditional distributions are detailed. Finally, in section IV, some results, obtained with synthetic and experimental mixtures, states the problem of non-negative source separation and illustrate the usefulness of the proposed method.

II. BAYESIAN MODELING

The main idea of the Bayesian approach to source separation is to use not only the likelihood $p(\mathbf{X}|\mathbf{S}, \mathbf{A})$, but also any prior information on the source signals and the mixing process through the assignment of prior distributions, $p(\mathbf{S})$ and $p(\mathbf{A})$. According to the Bayes's theorem, the joint posterior density is expressed as

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{S}, \mathbf{A}) \cdot p(\mathbf{S}) \cdot p(\mathbf{A})}{p(\mathbf{X})}, \quad (3)$$

where the independence between \mathbf{A} and \mathbf{S} is assumed. Since $p(\mathbf{X})$ is a normalisation constant, one can write

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}) \cdot p(\mathbf{S}) \cdot p(\mathbf{A}). \quad (4)$$

From this posterior density, joint estimation of \mathbf{A} and \mathbf{S} can be achieved by using various Bayesian estimators [37]. However, the first task of the inference is to encode our knowledge on the noise sequences, source signals and mixing coefficients by appropriate probability distributions.

A. Noise Distribution and Likelihood

The noise sequences $\{e_{(i,t)}\}_{i=1,t=1}^{m,N}$, are assumed independent and identically distributed (i.i.d), independent of the source signals, stationary and Gaussian with zero mean and variances $\{\sigma_i^2\}_{i=1}^m$. Thus,

$$\mathbf{E} \sim \prod_{t=1}^N \prod_{i=1}^m \mathcal{N}(e_{(i,t)}; 0, \sigma_i^2), \quad (5)$$

and, noting $\boldsymbol{\theta}_1 = \{\sigma_i^2\}_{i=1}^m$, the likelihood is then expressed as

$$p(\mathbf{X}|\mathbf{A}, \mathbf{S}, \boldsymbol{\theta}_1) = \prod_{t=1}^N \prod_{i=1}^m \mathcal{N}\left(x_{(i,t)}; \sum_{k=1}^n a_{(i,k)} s_{(k,t)}, \sigma_i^2\right), \quad (6)$$

B. Priors on Source Signals and Mixing coefficients

The sources are assumed mutually statistically independent and each j -th source signal is supposed i.i.d and distributed as a Gamma distribution of parameters (α_j, β_j) . These parameters are considered constant for each source but may differ from one source to another. The Gamma density is used to take into account non-negativity and its parameters allow a better fit to the spectra distribution. To incorporate the mixing coefficient non-negativity, each column j of the mixing matrix is also assumed distributed as a Gamma distribution of parameters (γ_j, λ_j) . Each j -th column of the mixing matrix

corresponds to the evolution profile of the j -th source proportion in the mixture and its associated Gamma parameters are considered equal for each profile. The two-parameter Gamma density is expressed by

$$\mathcal{G}(z; a, b) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp\{-bz\} \mathbb{I}_{[0,+\infty]}(z). \quad (7)$$

where $\Gamma(a)$ is the Gamma function. The Gamma distribution is an exponential family distribution which is used for fitting non-negative data since $p(z < 0) = 0$. Recently, it has been applied for non-negative signal restoration (see for example [38]). The second advantage of the Gamma distribution is that its shape parameters allow to fit spectral data that may present some sparsity and possibly a background.

The prior densities on the source signals and the mixing matrix are then expressed by

$$p(\mathbf{S}|\boldsymbol{\theta}_2) = \prod_{t=1}^N \prod_{j=1}^n \mathcal{G}(s_{(j,t)}; \alpha_j, \beta_j), \quad (8)$$

$$p(\mathbf{A}|\boldsymbol{\theta}_3) = \prod_{i=1}^m \prod_{j=1}^n \mathcal{G}(a_{(i,j)}; \gamma_j, \lambda_j), \quad (9)$$

where the vectors $\boldsymbol{\theta}_2 = \{\alpha_i, \beta_j\}_{j=1}^n$ and $\boldsymbol{\theta}_3 = \{\gamma_j, \lambda_j\}_{j=1}^n$ contain the parameters of the Gamma distributions.

C. Posterior Density and Estimation Issues

Using Bayes's theorem and noting by $\boldsymbol{\theta}$ the vector containing the noise variances and the parameters of the Gamma densities, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3\}$, the posterior law is expressed as

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}) \propto \prod_{t=1}^N \prod_{i=1}^m \mathcal{N}\left(x_{(i,t)}; \sum_{k=1}^n a_{(i,k)} s_{(k,t)}, \sigma_i^2\right) \\ \times \prod_{t=1}^N \prod_{j=1}^n \mathcal{G}(s_{(j,t)}; \alpha_j, \beta_j) \times \prod_{i=1}^m \prod_{j=1}^n \mathcal{G}(a_{(i,j)}; \gamma_j, \lambda_j). \quad (10)$$

From this joint posterior density, various estimators can be used to estimate the sources and the mixing coefficients. The joint maximization of this posterior density with respect to \mathbf{S} and \mathbf{A} leads to the joint maximum *a posteriori* JMAP estimator. The estimation of the mixing matrix \mathbf{A} can be performed by marginalizing the posterior density with respect to \mathbf{S} , to get $p(\mathbf{A}|\mathbf{X})$ from which \mathbf{A} can be estimated. The optimization problems associated to these estimators can be achieved using either Gradient/Newton based algorithms providing that the posterior densities are analytically tractable or, if not, stochastic simulation tools. In this paper, this last solution is retained by sampling the posterior distribution using Gibbs algorithm and constructing the estimator from the samples of the Markov chain.

It is interesting to consider firstly the joint maximum *a posteriori* estimator. It corresponds to the joint maximization of the posterior density or equivalently to the minimization with respect to \mathbf{S} and \mathbf{A} of the objective function defined as

$$\Phi(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}) = -\log p(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}).$$

This objective function can be decomposed into three parts,

$$\Phi(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}) = \Phi_L(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}_1) + \Phi_{P_1}(\mathbf{S}|\boldsymbol{\theta}_2) + \Phi_{P_2}(\mathbf{A}|\boldsymbol{\theta}_3)$$

where the terms Φ_L , Φ_{P_1} , and Φ_{P_2} are given by

$$\Phi_L = \sum_{i=1}^m \frac{1}{2\sigma_i^2} \sum_{t=1}^N \left(x_{(i,t)} - \sum_{k=1}^n a_{(i,k)} s_{(k,t)}\right)^2, \quad (11)$$

$$\Phi_{P_1} = \sum_{j=1}^n \left((1 - \alpha_j) \sum_{t=1}^N \log s_{(j,t)} + \beta_j \sum_{t=1}^N s_{(j,t)} \right), \quad (12)$$

$$\Phi_{P_2} = \sum_{j=1}^n \left((1 - \gamma_j) \sum_{i=1}^m \log a_{(i,j)} + \lambda_j \sum_{i=1}^m a_{(i,j)} \right). \quad (13)$$

The first part Φ_L of the objective function is the mean squares criterion, while the last two parts are regularization terms that penalize the negative values of \mathbf{S} and \mathbf{A} respectively. This approach may be connected with previously proposed methods. Indeed, this criterion is an extension of the PMF method criterion, in which the Gamma parameters may differ for each source signal and mixing coefficient profile. The case $\{\alpha_j = 1\}_{j=1}^n$ corresponds to assigning an exponential distribution prior for source distribution and leads to a regularization criterion similar to that minimized in the NNSC method for sparse source estimation. As compared to these penalized least squares approaches, this Bayesian formulation has the advantage to give a well stated theoretical framework for estimating the hyperparameters.

In [31], the optimization of this criterion is performed using an alternating Gradient iterative descent procedure, updating, at each iteration, the source estimate using the latest estimate of the mixing coefficients, then the mixing matrix estimate is updated using the latest estimate of the source signals. The learning parameter of the Gradient algorithm is optimized at each iteration. The critical point with this optimization scheme comes from the initialization of the algorithm, since it is well known that the Gradient algorithm converges to the nearest stationary point of the criterion. Satisfactory results were obtained by initializing the source estimates from the observations or by the most mutually uncorrelated observations, but, to reduce the dependence with respect to the initial values, a stochastic optimization scheme is considered in this paper. In that respect, Gibbs sampler is used for sampling the posterior density and the estimation is achieved using the marginal posterior mean (MPM) estimator

$$(\hat{\mathbf{A}}, \hat{\mathbf{S}}) = \mathbb{E}_{p(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta})} \{\mathbf{S}, \mathbf{A}\}. \quad (14)$$

The choice of this estimator is motivated by its simpler implementation from the sampled posterior density.

As discussed previously, for an unsupervised learning, the hyperparameters of the prior distributions and the noise variances have also to be inferred. The joint posterior distribution including the hyperparameters is expressed as

$$p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{X}) \propto p(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}), \quad (15)$$

in which prior densities may be assigned to the hyperparameters $\{\sigma_i^2\}_{i=1}^m, \{\alpha_i, \beta_j\}_{j=1}^n$ and $\{\gamma_j, \lambda_j\}_{j=1}^n$.

III. MCMC SAMPLING AND ESTIMATION

A. Gibbs Sampler

The main objective of Gibbs sampling is to simulate a stationary ergodic Markov chain whose samples asymptotically follow the posterior density $p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X})$. Estimates of source signals and mixing coefficients are then calculated from the samples of this chain. The MCMC sampling procedure for source separation in the general case are firstly recalled and nextly the main steps of sampling for non-negative source separation are given.

To sample $p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X})$, at each new iteration r of the algorithm, the main steps consists in

1. Sampling the source signals $\mathbf{S}^{(r+1)}$ from

$$p(\mathbf{S} | \mathbf{X}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}) \propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}) p(\mathbf{S} | \boldsymbol{\theta}^{(r)}); \quad (16)$$

2. Sampling the mixing coefficients $\mathbf{A}^{(r+1)}$ from

$$p(\mathbf{A} | \mathbf{X}, \mathbf{S}^{(r+1)}, \boldsymbol{\theta}^{(r)}) \propto p(\mathbf{X} | \mathbf{S}^{(r+1)}, \mathbf{A}, \boldsymbol{\theta}^{(r)}) p(\mathbf{A} | \boldsymbol{\theta}^{(r)}); \quad (17)$$

3. Sampling the hyperparameters $\boldsymbol{\theta}^{(r+1)}$ from

$$p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}) \propto p(\mathbf{X} | \mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}, \boldsymbol{\theta}) \times p(\mathbf{S}^{(r+1)} | \boldsymbol{\theta}) p(\mathbf{A}^{(r+1)} | \boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (18)$$

There are three types of hyperparameters, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, supposed to be independent, so the third step of the sampler can be divided into three sub-steps

- 3.1 Sampling the noise variances $\boldsymbol{\theta}_1^{(r+1)}$ from

$$p(\boldsymbol{\theta}_1 | \mathbf{X}, \mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}) \propto p(\mathbf{X} | \mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}, \boldsymbol{\theta}_1) p(\boldsymbol{\theta}_1); \quad (19)$$

- 3.2 Sampling the source hyperparameters $\boldsymbol{\theta}_2^{(r+1)}$ from

$$p(\boldsymbol{\theta}_2 | \mathbf{S}^{(r+1)}) \propto p(\mathbf{S}^{(r+1)} | \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_2); \quad (20)$$

- 3.3 Sampling the mixing coefficient hyperparameters $\boldsymbol{\theta}_3^{(r+1)}$ from

$$p(\boldsymbol{\theta}_3 | \mathbf{A}^{(r+1)}) \propto p(\mathbf{A}^{(r+1)} | \boldsymbol{\theta}_3) p(\boldsymbol{\theta}_3). \quad (21)$$

B. MCMC Sampling of the Joint Posterior Density

We now describe the whole algorithm to implement the Gibbs sampler corresponding to the proposed inference for non-negative source separation. For the sake of simplicity the following notations $y_{(1:n)}, z_{(1:n,1:m)}$ are introduced to represent respectively $\{y_i\}_{i=1}^n$ and $\{z_{(i,j)}\}_{(i=1,j=1)}^{(n,m)}$.

After a random initialization of all the variables and at each iteration r of the algorithm

1. for $j = 1, \dots, n$ and $t = 1, \dots, N$, sample $s_{(j,t)}^{(r+1)}$ from

$$p\left(s_{(j,t)} | x_{(1:m,t)}, s_{(1:j-1,t)}^{(r+1)}, s_{(j+1:n,t)}^{(r)}, a_{(1:m,1:n)}^{(r)}\right);$$

2. for $i = 1, \dots, n$ and $j = 1, \dots, n$, sample $a_{(i,j)}^{(r+1)}$ from

$$p\left(a_{(i,j)} | x_{(1:m,t)}, a_{(i,1:j-1)}^{(r+1)}, a_{(i,j+1:n)}^{(r)}, s_{(1:n,1:T)}^{(r+1)}\right);$$

3. for $i = 1, \dots, m$, sample $(\sigma_i^2)^{(r+1)}$ from

$$p\left(\frac{1}{\sigma_i^2} | x_{(i,1:T)}, a_{(i,1:n)}^{(r+1)}, s_{(1:n,1:T)}^{(r+1)}\right);$$

4. for $j = 1, \dots, n$, sample $\alpha_j^{(r+1)}$ from

$$p\left(\alpha_j | s_{(j,1:N)}^{(r+1)}, \beta_j^{(r)}\right);$$

5. for $j = 1, \dots, n$, sample $\beta_j^{(r+1)}$ from

$$p\left(\beta_j | s_{(j,1:N)}^{(r+1)}, \alpha_j^{(r+1)}\right);$$

6. for $j = 1, \dots, n$, sample $\gamma_j^{(r+1)}$ from

$$p\left(\gamma_j | a_{(1:m,j)}^{(r+1)}, \lambda_j^{(r)}\right);$$

7. for $j = 1, \dots, n$, sample $\lambda_j^{(r+1)}$

$$p\left(\lambda_j | a_{(1:m,j)}^{(r+1)}, \gamma_j^{(r+1)}\right).$$

After r_{max} iterations, estimate source signals and mixing coefficients

$$\hat{s}_{(1:n,1:T)} = \frac{1}{r_{max} - r_{min}} \sum_{r=r_{min}+1}^{r_{max}} s_{(1:n,1:T)}^{(r)} \quad (22)$$

$$\hat{a}_{(1:m,1:n)} = \frac{1}{r_{max} - r_{min}} \sum_{r=r_{min}+1}^{r_{max}} a_{(1:m,1:n)}^{(r)} \quad (23)$$

The value r_{min} represents the number of iterations cor-

responding to the burn-in run of the Markov chain whose associated samples are discarded. Other posterior statistics such as variances, covariances may be computed from the retained samples of the Markov chain and their histograms can be represented.

C. Conditional Posterior Densities

All the required conditional posterior densities for MCMC sampling are detailed below. Firstly priors are assigned to source signals $s_{(1:n,1:N)}$, secondly to mixing coefficients $a_{(1:m,1:n)}$ and finally to noise variances $\sigma_{(1:m)}^2$ and Gamma density parameters $\alpha_{(1:n)}, \beta_{(1:n)}, \gamma_{(1:n)}, \lambda_{(1:n)}$.

1) *Source Signals*: At the r -th iteration of the Gibbs sampler, the conditional posterior density of each source signal $s_{(j,t)}$ is expressed as

$$\begin{aligned} & p\left(s_{(j,t)} \mid x_{(1:n,t)}, s_{(1:j-1,t)}^{(r+1)}, s_{(j+1:n,t)}^{(r)}, a_{(1:m,1:n)}^{(r)}\right) \\ & \propto p\left(x_{(1:n,t)} \mid s_{(j,t)}, s_{(1:j-1,t)}^{(r+1)}, s_{(j+1:n,t)}^{(r)}, a_{(1:m,1:n)}^{(r)}\right) \\ & \quad \times p\left(s_{(j,t)} \mid \alpha_j^{(r)}, \beta_j^{(r)}\right) \mathbb{I}_{[0,+\infty]}(s_{(j,t)}) \end{aligned} \quad (24)$$

which is proportional to

$$s_{(j,t)}^{\alpha_j^{(r)}-1} \exp\left\{-\frac{1}{2\left[\sigma_{s_j}^{likel}\right]^2}\left(s_{(j,t)} - \mu_{s_{(j,t)}}^{likel}\right)^2 - \beta_j^{(r)} s_{(j,t)}\right\} \times \mathbb{I}_{[0,+\infty]}(s_{(j,t)}) \quad (25)$$

where

$$\begin{cases} \left[\sigma_{s_j}^{likel}\right]^2 = \left[\sum_{i=1}^m \frac{\left[a_{(i,j)}^{(r)}\right]^2}{\left[\sigma_i^{(r)}\right]^2}\right]^{-1}, \\ \mu_{s_{(j,t)}}^{likel} = \frac{1}{\left[\sigma_{s_j}^{likel}\right]^2} \sum_{i=1}^m \frac{a_{(i,j)}^{(r)} \varepsilon_{(i,t)}^{-j}}{\left[\sigma_i^{(r)}\right]^2}, \\ \varepsilon_{(i,t)}^{-j} = x_{(i,t)} - \sum_{k=1}^{j-1} a_{(i,k)}^{(r)} s_{(k,t)}^{(r+1)} - \sum_{k=j+1}^n a_{(i,k)}^{(r)} s_{(k,t)}^{(r)}. \end{cases}$$

This conditional posterior density is not an usual pdf, therefore its sampling is achieved using the Metropolis-Hastings algorithm. An instrumental distribution is determined by rewriting the posterior law in the form

$$\begin{aligned} & p\left(s_{(j,t)} \mid x_{(1:n,t)}, s_{(1:j-1,t)}^{(r+1)}, s_{(j+1:n,t)}^{(r)}, a_{(1:m,1:n)}^{(r)}\right) \\ & \propto s_{(j,t)}^{\alpha_j^{(r)}-1} \exp\left\{-\frac{1}{2\left[\sigma_{s_j}^{post}\right]^2}\left(s_{(j,t)} - \mu_{s_{(j,t)}}^{post}\right)^2\right\} \\ & \quad \times \mathbb{I}_{[0,+\infty]}(s_{(j,t)}) \end{aligned} \quad (26)$$

with $\mu_{s_{(j,t)}}^{post} = \mu_{s_{(j,t)}}^{likel} - \beta_j^{(r)} \left[\sigma_{s_j}^{likel}\right]^2$ and $\sigma_{s_j}^{post} = \sigma_{s_j}^{likel}$. The mode of the posterior law is obtained by solving the following second order equation

$$s_{(j,t)}^2 - \mu_{s_j}^{post} s_{(j,t)} - \left[\sigma_{s_j}^{post}\right]^2 (\alpha_j^{(r)} - 1) = 0, \quad \text{with } s_{(j,t)} \geq 0, \quad (27)$$

whose resolution yields

$$\mu_{s_j}^{max} = \begin{cases} 0 & \text{if } \Delta < 0; \\ \max\left\{\frac{1}{2}\left(\mu_{s_j}^{post} + \sqrt{\Delta}\right), 0\right\} & \text{else,} \end{cases} \quad (28)$$

where $\Delta = \left(\mu_{s_j}^{post}\right)^2 + 4\left[\sigma_{s_j}^{post}\right]^2 (\alpha_j^{(r)} - 1)$. Note that the root $\left\{\frac{1}{2}\left(\mu_{s_j}^{post} - \sqrt{\Delta}\right)\right\}$, does not correspond to a maximum of the posterior law. Therefore, the instrumental density is taken as a truncated normal distribution of parameters $\mu_{s_j}^{inst} = \mu_{s_j}^{max}$ and $\sigma_{s_j}^{inst} = \sigma_{s_j}^{post}$

$$s_{(j,t)}^{(r+1)} \sim \mathcal{N}_+\left(s_{(j,t)}; \mu_{s_j}^{inst}, \left[\sigma_{s_j}^{inst}\right]^2\right). \quad (29)$$

The sampling from this distribution can be achieved by cumulative distribution function inversion technique [39] or by using an accept-reject method [40].

Note that constraining $\alpha_j = 1$ corresponds to taking an exponential prior for the j -th source distribution. The use of the Metropolis-Hastings algorithm is not necessary since the posterior density is a truncated normal of parameters equal to those of the proposed instrumental density.

2) *Mixing Coefficients*: The conditional posterior density of each mixing coefficient $a_{(i,j)}$ is expressed as

$$\begin{aligned} & p\left(a_{(i,j)} \mid x_{(i,1:N)}, a_{(i,1:j-1)}^{(r+1)}, a_{(i,j+1:n)}^{(r)}, s_{(1:n,1:T)}^{(r+1)}\right) \\ & \propto p\left(x_{(i,1:N)} \mid a_{(i,1:j-1)}^{(r+1)}, a_{(i,j+1:n)}^{(r)}, s_{(1:n,1:T)}^{(r+1)}\right) \\ & \quad \times p\left(a_{(i,j)} \mid \gamma_j^{(r)}, \lambda_j^{(r)}\right) \mathbb{I}_{[0,+\infty]}(a_{(i,j)}) \end{aligned} \quad (30)$$

which is proportional to

$$a_{(i,j)}^{\gamma_j^{(r)}-1} \exp\left\{-\frac{1}{2\left(\sigma_{a_{(i,j)}}^{likel}\right)^2}\left(a_{(i,j)} - \mu_{a_{(i,j)}}^{likel}\right)^2 - \lambda_j^{(r)} a_{(i,j)}\right\} \times \mathbb{I}_{[0,+\infty]}(a_{(i,j)}) \quad (31)$$

where

$$\begin{cases} \left[\sigma_{a_{(i,j)}}^{likel}\right]^2 = \frac{\left[\sigma_{a_{(i,j)}}^{(r)}\right]^2}{\sum_{t=1}^N s_{(j,t)}^{(r+1)}}; \\ \mu_{a_{(i,j)}}^{likel} = \frac{1}{\left[\sigma_{a_{(i,j)}}^{likel}\right]^2} \sum_{t=1}^N s_{(j,t)}^{(r+1)} \varepsilon_{(i,t)}^{-j}, \\ \varepsilon_{(i,t)}^{-j} = \left(x_{(i,t)} - \sum_{k=1}^{j-1} a_{(i,k)}^{(r+1)} s_{(k,t)}^{(r+1)} - \sum_{k=j+1}^n a_{(i,k)}^{(r)} s_{(k,t)}^{(r+1)}\right). \end{cases}$$

As for the source signals, this conditional posterior density is then rewritten in the form

$$\begin{aligned} & p\left(a_{(i,j)} \mid x_{(i,1:N)}, a_{(i,1:j-1)}^{(r+1)}, a_{(i,j+1:n)}^{(r)}, s_{(1:n,1:T)}^{(r+1)}\right) \\ & \propto a_{(i,j)}^{\gamma_j^{(r)}-1} \exp\left\{-\frac{1}{2\left[\sigma_{a_{(i,j)}}^{post}\right]^2}\left(a_{(i,j)} - \mu_{a_{(i,j)}}^{post}\right)^2\right\} \\ & \quad \times \mathbb{I}_{[0,+\infty]}(a_{(i,j)}) \end{aligned} \quad (32)$$

where $\mu_{a_{(i,j)}}^{post} = \mu_{a_{(i,j)}}^{likel} - \lambda_j^{(r)} \left[\sigma_{a_{(i,j)}}^{likel}\right]^2$ and $\sigma_{a_{(i,j)}}^{post} = \sigma_{a_{(i,j)}}^{likel}$. Its sampling is achieved using a Metropolis-Hastings algorithm. The instrumental density is calculated as in the case of the source signals.

3) *Noise Variances*: The posterior conditional density of each noise variance σ_i^2 is expressed as,

$$\begin{aligned} & p\left(\frac{1}{\sigma_i^2} \mid x_{(i,1:N)}, a_{(i,1:N)}^{(r+1)}, s_{(i,1:N)}^{(r+1)}\right) \\ & \propto \left(\frac{1}{\sigma_i^2}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_i^2} \sum_{t=1}^N \left(x_{(i,t)} - \sum_{k=1}^n a_{(i,k)}^{(r+1)} s_{(k,t)}^{(r+1)}\right)^2\right\} \\ & \quad \times p\left(\frac{1}{\sigma_i^2}\right). \end{aligned} \quad (33)$$

The prior for the noise variance σ_i^2 is an inverse Gamma, which corresponds to assigning a Gamma distribution for $\left(\frac{1}{\sigma_i^2}\right)$

$$\left(\frac{1}{\sigma_i^2}\right) \sim \mathcal{G}\left(\alpha_{\sigma_i^2}^{prior}, \beta_{\sigma_i^2}^{prior}\right), \quad (34)$$

leading to a posterior density given by

$$\left(\frac{1}{\sigma_i^2} \mid x_{(i,1:N)}, a_{(i,1:N)}^{(r+1)}, s_{(i,1:N)}^{(r+1)}\right) \sim \mathcal{G}\left(\frac{1}{\sigma_i^2}; \alpha_{\sigma_i^2}^{post}, \beta_{\sigma_i^2}^{post}\right),$$

with

$$\begin{cases} \alpha_{\sigma_i^2}^{post} = \frac{N}{2} + \alpha_{\sigma_i^2}^{prior}, \\ \beta_{\sigma_i^2}^{post} = \frac{1}{2} \sum_{t=1}^N \left(x_{(i,t)} - \sum_{k=1}^n a_{(i,k)}^{(r+1)} s_{(k,t)}^{(r+1)} \right)^2 + \beta_{\sigma_i^2}^{prior}. \end{cases}$$

The parameters $\alpha_{\sigma_i^2}^{prior}$ and $\beta_{\sigma_i^2}^{prior}$ are chosen according to an *a priori* noise level and variance. Note that this approach transforms the original problem of choosing σ_i^2 to that of choosing $\alpha_{\sigma_i^2}^{prior}$ and $\beta_{\sigma_i^2}^{prior}$. But the point is that this last choice is by no way as crucial as the choice of σ_i^2 is.

4) *Source Hyperparameters*: The sampled sources being given, their associated Gamma distribution parameters $\alpha_{(1:n)}$ and $\beta_{(1:n)}$ are sampled as follows.

The posterior density of each hyperparameter α_j is given as

$$\begin{aligned} p\left(\alpha_j \mid s_{(j,1:N)}^{(r+1)}, \beta_j^{(r)}\right) &\propto \prod_{t=1}^N \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} s_{(j,t)}^{\alpha_j-1} \cdot p(\alpha_j), \\ &\propto \frac{1}{\Gamma(\alpha_j)^N} \exp\left\{ \left(N \log \beta_j^{(r)} + \sum_{t=1}^N \log s_{(j,t)}^{(r+1)} \right) \alpha_j \right\} \\ &\quad \times p(\alpha_j). \end{aligned} \quad (35)$$

By assigning an exponential prior for α_j of parameters $\lambda_{\alpha_j}^{prior}$, this posterior density takes the form

$$p(\alpha_j \mid s_j, \mu_j) \propto \left(\frac{1}{\Gamma(\alpha_j)} \exp\left\{ \lambda_{\alpha_j}^{post} \alpha_j \right\} \right)^N, \quad (36)$$

where $\lambda_{\alpha_j}^{post} = \log \beta_j^{(r)} + \frac{1}{N} \sum_{t=1}^N \log s_{(j,t)}^{(r+1)} - \frac{1}{N} \lambda_{\alpha_j}^{prior}$. The sampling from this distribution is achieved using a Metropolis-Hastings algorithm. To obtain an instrumental density, a Gamma density

$$q(z) \propto z^{\alpha_z^q - 1} \exp\{-\beta_z^q z\}, \quad (37)$$

is firstly used to fit the term between brackets

$$g(z) = \frac{1}{\Gamma(z)} \exp\{\lambda_z z\}. \quad (38)$$

The parameters (α_z^q, β_z^q) of this function are determined in such a way that its mode and inflexion point are the same as that of the function $g(z)$. This Gamma density parameters are obtained as

$$\begin{cases} \alpha_z^q = 1 + \frac{z_{mode}^2}{(z_{mode} - z_{infl})^2}, \\ \beta_z^q = \frac{z_{mode}}{(z_{mode} - z_{infl})^2}, \end{cases} \quad (39)$$

where z_{mode} and z_{infl} are the mode and the superior inflexion point ($z_{infl} > z_{mode}$) of $g(z)$. The calculation of the first and second derivative of $g(z)$ yields these two non-linear equations

$$\begin{cases} \psi(z_{mode}) - \lambda_z = 0, \\ \psi^{(1)}(z_{infl}) - (\psi(z_{infl}) - \lambda_z)^2 = 0, \end{cases} \quad (40)$$

where ψ is the psi function, also called the digamma function and $\psi^{(1)}$ is its first derivative, also called trigamma function. These functions and their approximations are defined in [41, p.253] and the resolution of the two equations in (40) is performed using a numerical method for root finding [42, ch.9]. Finally, the posterior density $p(\alpha_j \mid s_j, \mu_j) = g(\alpha_j)^N$ is sampled using the Metropolis-Hastings algorithm with a Gamma instrumental density whose parameters are given by

$$\begin{cases} \alpha_{\alpha_j}^{inst} = N \left(\alpha_{\alpha_j}^q - 1 \right) + 1, \\ \beta_{\alpha_j}^{inst} = N \beta_{\alpha_j}^q. \end{cases} \quad (41)$$

Concerning the hyperparameter β_j , the posterior distribution $p\left(\beta_j \mid s_{(j,1:N)}^{(r+1)}\right)$ is expressed as

$$\begin{aligned} p\left(\beta_j \mid s_{(j,1:N)}^{(r+1)}, \alpha_j^{(r+1)}\right) &\propto \beta_j^{N \alpha_j^{(r+1)}} \\ &\quad \exp\left\{ -\beta_j \sum_{t=1}^N s_{(j,t)}^{(r+1)} \right\} \cdot p(\beta_j). \end{aligned} \quad (42)$$

Therefore, one can note that the conjugate prior for the parameter β_j is a Gamma density,

$$\beta_j \sim \mathcal{G}\left(\alpha_{\beta_j}^{prior}, \beta_{\beta_j}^{prior}\right), \quad (43)$$

leading to an *a posteriori* Gamma distribution

$$\left(\beta_j^{(r+1)} \mid s_{(j,1:N)}^{(r+1)}, \alpha_j^{(r+1)}\right) \sim \mathcal{G}\left(\alpha_{\beta_j}^{post}, \beta_{\beta_j}^{post}\right), \quad (44)$$

with parameters

$$\begin{cases} \alpha_{\beta_j}^{post} = 1 + N \alpha_j^{(r+1)} + \alpha_{\beta_j}^{prior}, \\ \beta_{\beta_j}^{post} = \sum_{t=1}^N s_{(j,t)} + \beta_{\beta_j}^{prior}. \end{cases} \quad (45)$$

To illustrate the proposed sampling algorithm for estimating the parameters of a Gamma density, an example is presented. A sequence of $N = 1000$ samples generated from a Gamma density of parameters $\alpha = 3$ and $\beta = 2$ is considered. Figure 1 shows one realization of the Markov chain and the evolution of the averaged acceptance rate of the Metropolis-Hastings algorithm. The good approximation of the conditional density of the parameters α results in a high acceptance rate of the Metropolis-Hastings algorithm and a fast convergence of the sampled parameters around the true value of the parameters ($\hat{\alpha} = 3.08 \pm 0.13, \hat{\beta} = 2.07 \pm 0.09$).

5) *Mixing Coefficient Hyperparameters*: Since the mixing coefficient profiles are also assigned by a gamma distribution prior, the parameters $\gamma_{(1:n)}$ and $\lambda_{(1:n)}$ are sampled in the same manner as for the source signal hyperparameters.

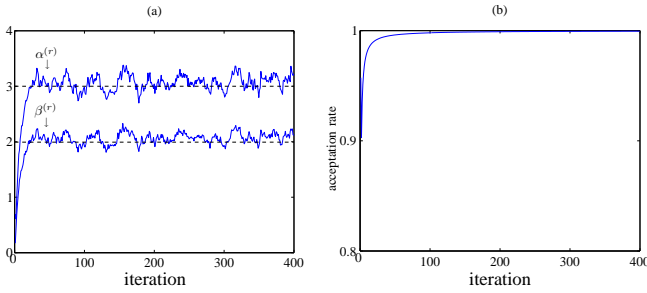


Fig. 1: (a) Generated Markov chains from the posterior density $p(\alpha, \beta)$ and (b) averaged acceptance rate of the Metropolis-Hastings algorithm

D. Comments on the MCMC Algorithm

As mentioned in subsection III-C, the posterior densities of the sources and the mixing coefficients are not usual, so their sampling is performed using a Metropolis-Hastings algorithm. In addition, the noise variances and the hyperparameters $\beta_{(1:n)}$ and $\lambda_{(1:n)}$ are sampled directly from their posterior conditional distributions while the hyperparameters $\alpha_{(1:n)}$ and $\gamma_{(1:n)}$ are sampled using a Metropolis-Hastings algorithm. So, the whole procedure results in an hybrid Gibbs-Metropolis-Hastings sampling. However, if an exponential prior is taken by constraining ($\alpha_{(1:n)} = 1$ and $\gamma_{(1:n)} = 1$), the sampling procedure does not require the use of the Metropolis-Hastings algorithm.

IV. EXPERIMENTS

To illustrate the problem of non-negative source separation and show the effectiveness of the proposed method, this section presents some numerical and experimental results. In a first time an experiment opens the discussion about the separation of non-negative sources using statistical independence and non-negativity assumptions. The first simulation is a situation where the independence assumption allows to achieve the separation and the second points out the need of taking into account the non-negativity. The next experiment concerns the analysis of a spectral mixture obtained by mixing experimentally three chemical species and measuring the resulting mixture data using a near infrared spectrometer.

A. Performance Measures

As a performance measure the performance index \mathcal{PI} defined by

$$\mathcal{PI} = \frac{1}{2} \sum_{i=1}^n \left\{ \left(\sum_{k=1}^n \frac{|g_{ik}|^2}{\max_{\ell} |g_{i\ell}|^2} - 1 \right) + \left(\sum_{k=1}^n \frac{|g_{ki}|^2}{\max_{\ell} |g_{\ell i}|^2} - 1 \right) \right\} \quad (46)$$

is used, where g_{ij} is the (i, j) th element of the global system matrix $\mathbf{G} = \mathbf{B}\mathbf{A}$, $\max_{\ell} |g_{i\ell}|$ stands for the maximum value among the elements in the i th row vector of \mathbf{G} and $\max_{\ell} |g_{\ell i}|$ represents the maximum value among the elements in the i th column vector of \mathbf{G} . It is zero for perfect signal separation. In practice, it takes small values when a good separation is achieved. This index assesses the overall separation performance and measures mainly the quality of the estimation of

the mixing matrix. However, it is very important to measure the accuracy of the reconstruction of each source signal. In that respect, one can use the residual cross-talk index defined as

$$\mathcal{CT}_{s_j} = \sum_{t=1}^N (s_{(j,t)} - \hat{s}_{(j,t)})^2, \quad (47)$$

where $\hat{s}_{(j,1:N)}$ is the estimate of the j -th source signal $s_{(j,1:N)}$ and the two signals have unit variance. In all the following results, the two performance criteria are expressed in dB.

B. Can ICA Methods Separate Non-negative Sources ?

In a first simulation, two mutually independent non-negative sequences (spectra of two speech signals) are mixed with non-negative mixing coefficients and an additive zero mean Gaussian noise is added in such a way to have a signal to noise ratio (SNR) of 20 dB. The mixing matrix is

$$\mathbf{A} = \begin{bmatrix} 0.60 & 0.40 \\ 0.40 & 0.60 \end{bmatrix}, \quad (48)$$

and the source signals with the resulting mixtures are shown in figure 2. The estimated empirical covariance matrix of the sources

$$\hat{\mathbf{R}}_s = \begin{bmatrix} 1.00 & 0.01 \\ 0.01 & 1.00 \end{bmatrix}, \quad (49)$$

shows that they are mutually uncorrelated.

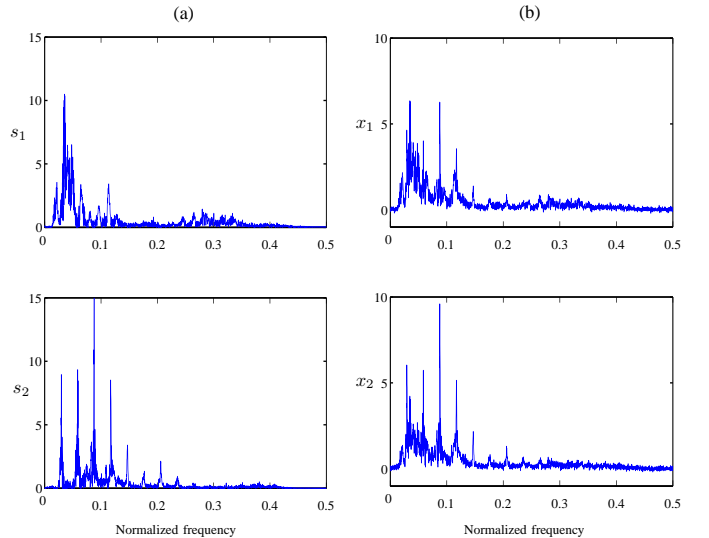


Fig. 2: (a) Source signals and (b) resulting mixtures

The analysis of this mixture using the independence assumption requires a first step of estimating the separating matrix from the centered mixture data using an ICA method. Among the available methods, SOBI [12], JADE [43] and FastICA [44] have been considered, however other ICA algorithms may be used as well. The estimation of the sources is then achieved by applying the separating matrix to the non-centered mixture data. The use of NNICA method does not require this first step. Table (I) summarizes the performances of the separation using different methods (where BPSS for *Bayesian Positive Source Separation* refers to the proposed

method). All the considered analysis methods succeeded to separate the two components with slightly different performance indexes, but we may note that even if the \mathcal{PI} is particularly similar in these methods, the \mathcal{CT} indexes are much better with BPSS as compared to the others. It turns out that the separation of non-negative independent sources can be achieved either using only the mutual independence or by incorporating the non-negativity of sources and mixing coefficients.

TABLE I: Speech spectra separation performances

	JADE	FastICA	NNICA	BPSS
$\mathcal{CT}_{\text{source1}}$	-13.04	-13.02	-12.81	-16.74
$\mathcal{CT}_{\text{source2}}$	-12.56	-12.59	-12.67	-16.56
\mathcal{PI}	-27.27	-27.64	-25.19	-26.93

C. Is there any Improvement Introduced by Considering Non-negativity ?

In a second simulation, the mixture data are obtained by constructing three synthetic spectra and simulating ten measures with mixing coefficients chosen in such a way to have an evolution profile similar to the component concentration behavior in chemical reactions. Figure (3) shows the source signals, their mixing coefficient profiles and the resulting mixture for a signal to noise ratio (SNR) equal to 20 dB.

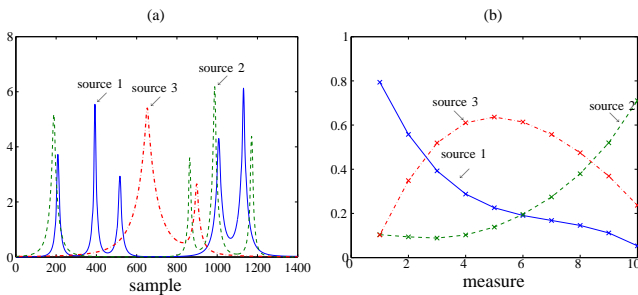


Fig. 3: (a) Original sources and (b) mixing coefficients

To assess the spatial correlation of the source signals, we calculated their empirical covariance matrix

$$\hat{\mathbf{R}}_s = \begin{bmatrix} 1.00 & 0.24 & -0.21 \\ 0.24 & 1.00 & -0.17 \\ -0.21 & -0.17 & 1.00 \end{bmatrix}. \quad (50)$$

The off-diagonal terms of this covariance matrix are non null, showing that the available samples of the source signals presents a significant spatial correlation, so the independence assumption required by usual ICA methods is not totally satisfied by these signals. Among the methods used, the best separation results, obtained by the FastICA method, are shown in figure (4). However, note the negative values of the estimated source signals which does not correspond to the very nature of the sources. The results obtained with SOBI algorithm are also given to illustrate the estimation of both negative sources and mixing coefficients. These results illustrate the need to take into account the non-negativity of both source signals and mixing coefficients.

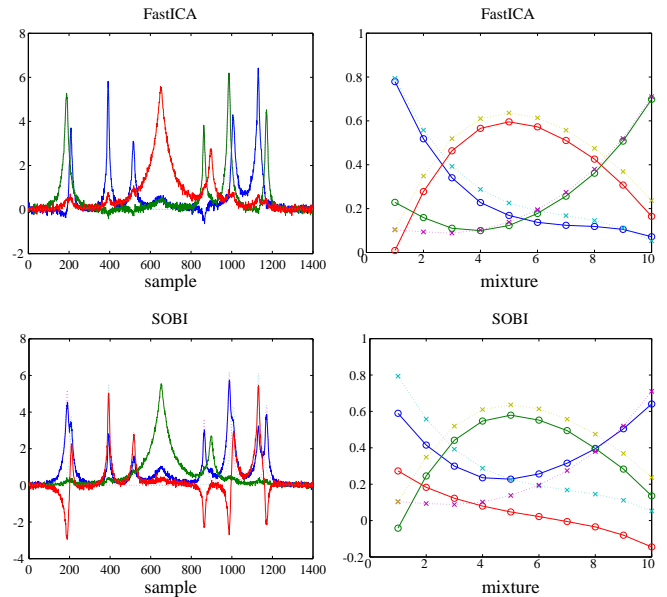


Fig. 4: Estimated sources (left) and mixing coefficients (right) using SOBI and FastICA methods (continuous line). True sources and mixing coefficients are shown in dotted lines

The proposed separation method is applied to the analysis of this mixture, yielding the results of figure (5). Both source signals and mixing coefficients are estimated successfully, without negative values.

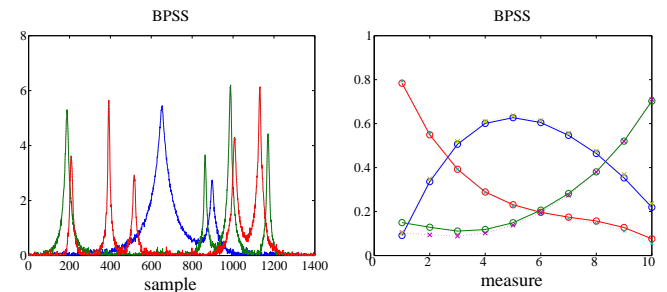


Fig. 5: Estimated sources (left) and mixing coefficients (right) using Gamma prior and MCMC sampling (continuous line). True sources and mixing coefficients are shown in dotted lines

Concerning the separation accuracy, table (II) summarizes the performance index reached by the different methods. Note particularly the superior performances of the proposed method. In this case, not only the \mathcal{PI} of the BPSS method is better but also all the \mathcal{CT} are much better by 10dB. This experiment allows to conclude that an improvement in the separation performances is significantly introduced by considering the non-negativity, which particularly illustrates the need of considering the non-negativity and motivates the usefulness of the Bayesian approach.

D. Near Infrared Data

To validate the proposed approach with real data, an experiment is performed in which the mixture data are obtained

TABLE II: Synthetic source separation performances

	JADE	FastICA	ICA-ALS	NMF	BPSS
$CT_{\text{source 1}}$	-13.99	-17.63	-17.22	-15.75	-22.46
$CT_{\text{source 2}}$	-18.56	-14.47	-14.82	-9.19	-24.62
$CT_{\text{source 3}}$	-16.05	-16.01	-16.53	-11.47	-27.29
PI	-15.10	-15.36	-15.85	-9.94	-27.23

from near infrared (NIR) spectroscopy measurements. Three known chemical species (cyclopentane, cyclohexane and n-pentane) are mixed experimentally with specified proportions. These species have been chosen for two main reasons. Firstly, their available spectra in the NIR frequency band are highly overlapping and as a consequence are spatially correlated. This precludes the use of standard ICA methods to achieve the separation. Secondly, these species do not interact when they are mixed, guaranteeing that no new component appears. Thus, the number of sources as well as their concentrations in the mixtures are known exactly. In addition, their individual spectra can be (and are) measured separately. Figure (6) shows the pure spectra of the chemical species and their concentration profiles.

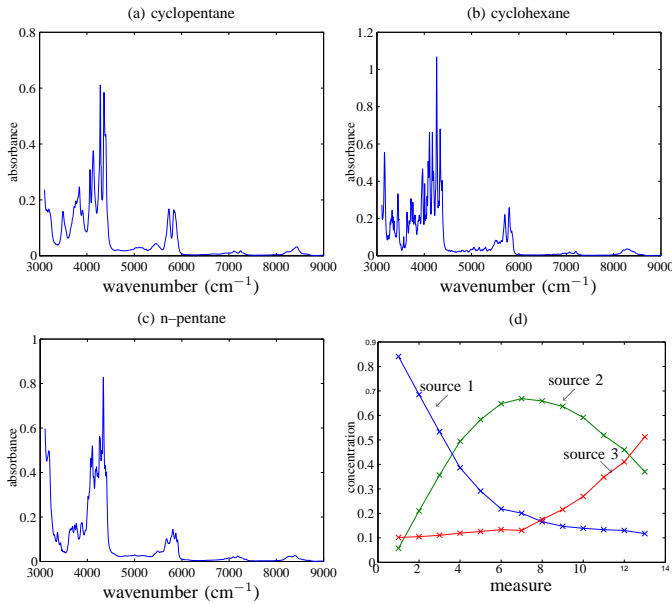


Fig. 6: (a,b,c) Constituent spectra and (d) concentration profiles

While processing the mixture data using different methods yields the performance index summarized in table III. These methods have been chosen because they give both non-negative sources and mixing coefficients. The ALS method is initialized using an ICA method. The more recent NMF method offers the advantage of being numerically faster. Comparing the results shows the effectiveness of the proposed inference.

TABLE III: NIR spectra separation accuracy

	ICA-ALS	NMF	BPSS
$CT_{\text{cyclopentane}}$	-14.20	-15.18	-33.23
$CT_{\text{cyclohexane}}$	-17.50	-23.43	-24.98
$CT_{\text{n-pentane}}$	-17.88	-14.01	-26.05
PI	-11.60	-8.10	-19.22

Figure 7 compares both the estimate and true sources in the more challenging bands (3500–4500 cm⁻¹ and 5000–6000 cm⁻¹), where the peaks of the different source spectra highly overlap, resulting in a significant cross-correlation. The source spectra are well reconstructed, which makes the identification of the components easier. Concerning the estimation of the

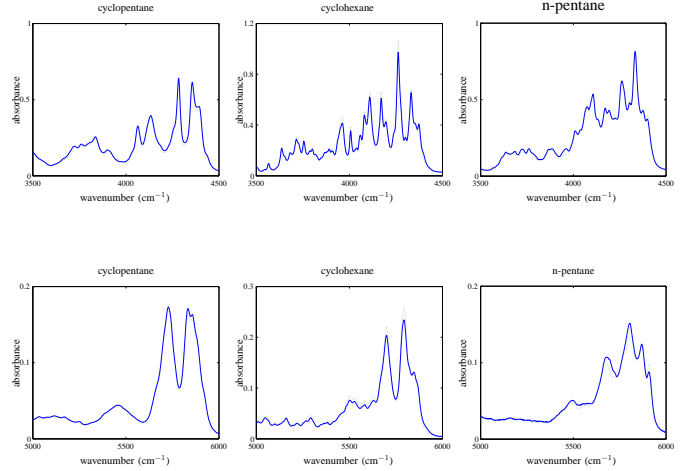


Fig. 7: Zoom of two subbands of the source spectra. The true spectra are shown in dotted lines.

concentrations, figure (8) shows the similarity of the estimated mixing coefficients with the true concentration profiles, but a small error still remains.

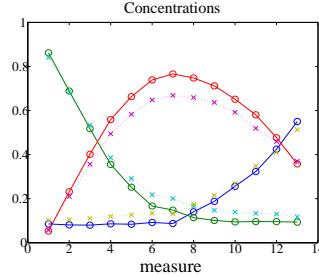


Fig. 8: Estimated (continuous curve) and true (dotted curve) concentration profiles

V. CONCLUSION

The problem of non-negative source separation has been addressed in this paper. The proposed Bayesian inference considers the non-negativity as prior information which is encoded through the assignment of Gamma distribution priors. The Gamma density is an exponential family distribution which is frequently used to represent non-negative data and its second advantage is that its shape allows to fit spectral signals that may present some sparsity and/or a possible background. The result that has been presented illustrate that such prior distribution is very suitable for the separation of spectral source signals. To achieve a better fit of the source signal distributions, the proposed approach can be straightforwardly extended to the more general model consisting in mixtures of Gamma or truncated normal distributions. A second result that has been discussed concerns the separation of non-negative

sources by using independent component analysis methods. It has been shown that the separation of positive sources by an ICA method is possible, but it is conditioned by the statistical independence of the non-negative sources. In that case, similar performances are obtained with the proposed approach. However, if the independence assumption is not totally satisfied by the sources, the non-negativity is an additional assumption that should be considered to improve the separation by an appropriate Bayesian analysis model.

REFERENCES

- [1] E. Malinowski, *Factor Analysis in Chemistry*. New York: John Wiley, 1991.
- [2] P. J. Gemperline and E. Cash, "Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with penalty function," *Analytical Chemistry*, vol. 75, pp. 4236–4243, 2003.
- [3] J. Shoonover, R. Max, and S. Zhang, "Multivariate curve resolution in the analysis of vibrational spectroscopy data files," *Applied Spectroscopy*, vol. 57, no. 5, pp. 154–170, 2003.
- [4] H. Pfeiffer and H. Liebhafsky, "The origins of Beer's law," *Journal of Chemical Education*, vol. 28, p. 123, 1951.
- [5] R. Ricci, M. Ditzler, and L. Nestor, "Discovering the Beer-Lambert law," *Journal of Chemical Education*, vol. 71, pp. 983–985, November 1994.
- [6] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933.
- [7] P. Comon, "Independent component analysis – a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [8] T.-W. Lee, *Independent Component Analysis – Theory and Applications*. Kluwer Academic, 1998.
- [9] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley, 2001.
- [10] S.-I. Amari and A. Cichocki, *Adaptive Blind Signal and Image Processing*. New York: John Wiley, 2002.
- [11] D. Nuzillard and J.-M. Nuzillard, "Application of blind source separation to 1-D and 2-D Nuclear Magnetic Resonance spectroscopy," *IEEE Signal Processing Letters*, vol. 5, no. 8, pp. 209–211, 1998.
- [12] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, pp. 434–444, 1997.
- [13] M. Plumbley, "Algorithms for non-negative independent component analysis," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 534–543, 2003.
- [14] R. Tauler and B. Kowalski, "Multivariate curve resolution applied to spectral data from multiple runs of an industrial process," *Analytical Chemistry*, vol. 65, pp. 2040–2047, 1993.
- [15] C. Lawson and R. Hanson, *Solving Least-Squares Problems*. Prentice-Hall, 1974.
- [16] R. Tauler, A. Izquierdo-Ridorsa, and E. Casassas, "Simultaneous analysis of several spectroscopic titrations with self-modelling curve resolution," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 3, pp. 293–300, 1993.
- [17] W. Lawton and E. Sylvestre, "Self-modeling curve resolution," *Technometrics*, vol. 13, pp. 617–633, 1971.
- [18] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [19] P. Sajda, S. Du, T. Brown, L. Parra, and R. Stoyanova, "Recovery of constituent spectra in 3D chemical shift imaging using non-negative matrix factorization," in *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2003)*, 2003, pp. 71–76.
- [20] P. Hoyer, "Non-negative sparse coding," in *Proceedings of IEEE Workshop on Neural Networks for Signal Processing (NNSP'2002)*, 2002, pp. 557–565.
- [21] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111–126, 1994.
- [22] S. Roberts, "Independent component analysis: Source assessment and separation, a Bayesian approach," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 145, no. 3, pp. 149–154, 1998.
- [23] K. Knuth, "A Bayesian approach for source separation," in *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation, (ICA'99)*, 1999, pp. 283–288.
- [24] A. Mohammad-Djafari, "A Bayesian approach to source separation," *American Institute of Physics (AIP) proceedings*, vol. 567, pp. 221–244, 1999.
- [25] S. Sénécal and P. Amblard, "Bayesian separation of discrete sources via Gibbs sampling," in *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2000)*, 2000, pp. 566–572.
- [26] D. Rowe, *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Boca Raton, Florida, USA: CRC Press, 2003.
- [27] A. Mohammad-Djafari and M. Ichir, "Wavelet domain blind image separation," in *SPIE, Mathematical Modeling, Wavelets X*, August 2003.
- [28] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images," *Journal of Electronic Imaging*, vol. 13, no. 2, 2004.
- [29] J. Miskin and D. MacKay, "Ensemble learning for blind source separation," in S. Roberts and R. Everson, editors, *Independent Component Analysis: Principles and Practice*. Cambridge University Press, 2001, pp. 209–233.
- [30] S. Roberts and R. Choudrey, "Data decomposition using independent component analysis with prior constraints," *Pattern Recognition*, vol. 36, pp. 1813–1825, 2003.
- [31] S. Moussaoui, A. Mohammad-Djafari, D. Brie, and O. Caspary, "A Bayesian method for positive source separation," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, 2004.
- [32] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1999.
- [33] C. Robert, *Monte Carlo Statistical Methods*. Berlin: Springer-Verlag, 1999.
- [34] W. Fitzgerald, "Markov chain Monte Carlo methods with applications to signal processing," *Signal Processing*, vol. 81, pp. 3–18, 2001.
- [35] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [36] W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [37] C. Robert, *The Bayesian Choice*, 2nd ed. Springer-Verlag, 2001.
- [38] I.-T. Hsiao, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction using mixture models as priors," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1466–1477, 2002.
- [39] A. Gelfand, A. Smith, and T. Lee, "Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling," *Journal of the American Statistical Association*, vol. 87, pp. 523–532, 1992.
- [40] C. Robert, "Simulation of a truncated normal variables," *Statistics and Computing*, vol. 5, pp. 121–125, 1995.
- [41] M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover Publications, 1972.
- [42] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge University Press, 1992.
- [43] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.
- [44] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.