



# Bayesian source separation of linear-quadratic and linear mixtures through a MCMC method

Leonardo Tomazeli Duarte, Christian Jutten, Saïd Moussaoui

## ► To cite this version:

Leonardo Tomazeli Duarte, Christian Jutten, Saïd Moussaoui. Bayesian source separation of linear-quadratic and linear mixtures through a MCMC method. MLSP 2009 - IEEE 19th International Workshop on Machine Learning for Signal Processing, Sep 2009, Grenoble, France. 6 p. hal-00424171

**HAL Id: hal-00424171**

**<https://hal.science/hal-00424171>**

Submitted on 14 Oct 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# BAYESIAN SOURCE SEPARATION OF LINEAR-QUADRATIC AND LINEAR MIXTURES THROUGH A MCMC METHOD

Leonardo T. Duarte<sup>§\*</sup>, Christian Jutten<sup>§†</sup> and Saïd Moussaoui<sup>‡</sup>

<sup>§</sup>GIPSA-lab, UMR CNRS 5216, Institut Polytechnique de Grenoble, France

<sup>‡</sup>IRCCyN, UMR CNRS 6597, Ecole Centrale Nantes, France

{leonardo.duarte,christian.jutten}@gipsa-lab.inpg.fr, said.moussaoui@irccyn.ec-nantes.fr

## ABSTRACT

In this work, we deal with source separation of linear - quadratic (LQ) and linear mixtures. By relying on a Bayesian approach, the developed method allows one to take into account prior informations such as the non-negativity and the temporal structure of the sources. Concerning the inference scheme, the implementation of a Gibbs' sampler equipped with latent variables simplifies the sampling steps. The obtained results confirm the effectiveness of the new proposal and indicate that it may be particularly useful in situations where classical ICA-based solutions fail to separate the sources.

## 1. INTRODUCTION

In Blind Source Separation (BSS), the goal is to retrieve a set of sources by using only mixed versions of these original signals. Usually, one assumes a linear mixing process and the separation is performed via Independent Component Analysis (ICA) [1]. However, despite the notorious results provided by this classical framework, its extension to the nonlinear case is desirable as it can broaden the range of BSS applications.

Unfortunately, BSS becomes more involved in its nonlinear extension [2]. For example, due to the degree of flexibility in a nonlinear model, the application of ICA does not guarantee source separation, that is, one may recover independent components that are still mixed versions of the sources. For such reasons, a more realistic approach is to consider restricted classes of nonlinear models for which source separation is still possible. A typical example in this context is the linear-quadratic (LQ) model [3, 4]. Besides the theoretical interest in the LQ model—it may pave the way for polynomial mixtures—this nonlinear model is useful in chemical sensing problems, such as in the design of gas electrode arrays [5].

Since the inversion of the LQ mixing model does not admit closed formulae in the general case, a major challenge in LQ-BSS concerns the definition of a suitable structure for the separating system. In [3, 4], this problem was dealt with by defining a recurrent separating system that was trained by

ICA-based cost functions. Nonetheless, despite its simplicity and its good performance, this approach can operate only when the sources and the mixing parameters lie within the stability region of the recurrent system. Even if the development of more elaborate recurrent networks [6] can extend the stability region, it seems that the resulting training algorithms may be quite complicate in these new situations.

Motivate by the above-mentioned problems, we propose in this work a Bayesian method for separating LQ mixtures. Indeed, as the Bayesian approach treats the BSS problem rather as a data representation problem, there is no need to define a separating system in this case. Furthermore, the Bayesian framework permits us to take into account prior information other than the statistical independence. For instance, we consider here two prior informations that are typical in chemical sensing applications, namely: 1) the bounds of the sources and of the mixing coefficient values are known in advance, and 2) the sources have a temporal structure. It is worth mentioning that, under minor changes, the developed method can also be applied to linear models. Finally, even for the nonlinear model treated here, a simple inference scheme can be set by using a MCMC method able to simulate the posterior distributions of the parameters. In that respect, the Gibbs sampler and some auxiliary variables are employed.

Concerning the organization of the paper, we start with the mixing model description. After that, in Section 3, we present the details of our proposal. Aiming to assess the gains brought by the proposed method, a set of simulations is conducted in Section 4. Finally, we present our conclusions in Section 5.

## 2. MIXING MODEL

Let  $x_{i,t}$  and  $s_{j,t}$  represent the  $i$ -th mixture and the  $j$ -th source at time  $t$ , respectively. The LQ mixing model is given by

$$x_{i,t} = \sum_{j=1}^{n_s} a_{i,j} s_{j,t} + \sum_{1 \leq j < k \leq n_s} b_{i,j,k} s_{j,t} s_{k,t} + n_{i,t} \quad (1)$$

$\forall i \in 1, \dots, n_c, \forall t \in 1, \dots, n_d,$

where  $a_{i,j}$  et  $b_{i,j,k}$  are the mixing parameters, and  $n_{i,t}$  corresponds to the noise term, which is assumed i.i.d and

\*L.T. Duarte is grateful to the CNPq (Brazil) for funding his PhD research.

†C. Jutten is with the Institut Universitaire de France.

Gaussian with unknown variance  $\sigma_i^2$ . The parameters  $n_s$ ,  $n_c$  and  $n_d$  correspond to the number of sources, sensors and available samples, respectively. Henceforth, all the unknown parameters will be represented by the vector<sup>1</sup>  $\theta = [s_{j,t}, a_{i,j}, b_{i,j,k}, \sigma_i^2, \mu_j, p_j]$ , and the following notation will be adopted:  $\theta_{-\theta_q}$  represents the vector containing all elements of  $\theta$  except  $\theta_q$ .

### 3. BAYESIAN SOURCE SEPARATION METHOD

In view of Eq. (1), the BSS problem treated in this work can be put as follow: given  $\mathbf{X}$  (matrix containing all  $x_{i,t}$ ), estimate the parameters of the vector  $\theta$ . In the Bayesian framework, this estimation task requires the expression of the posterior distribution  $p(\theta|\mathbf{X})$ , which, according to the Bayes' rule, can be written as  $p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)p(\theta)$ , where  $p(\mathbf{X}|\theta)$  is the likelihood function and  $p(\theta)$  denotes the prior distributions. Due to the assumption of white Gaussian noise in the observation model, the likelihood function is given by

$$p(\mathbf{X}|\theta) = \prod_{t=1}^{n_d} \prod_{i=1}^{n_c} \mathcal{N}_{x_{it}} \left( \sum_{j=1}^{n_s} a_{i,j} s_{j,t} + \sum_{1 \leq j < k \leq n_s} b_{i,j,k} s_{j,t} s_{k,t}; \sigma_i^2 \right), \quad (2)$$

where  $\mathcal{N}_{x_{ik}}(\mu; \sigma^2)$  corresponds to a Gaussian distribution of  $x_{ik}$  with mean  $\mu$  and variance  $\sigma^2$ . The expressions of prior distributions will be presented in the sequel.

#### 3.1. Definition of prior distributions

##### 3.1.1. Sources

If the limit values of the sources are known, one can incorporate this information by modeling them according to a truncated Gaussian distribution, i.e.

$$p(s_{j,t}|\mu_j, p_j, s_j^{\min}, s_j^{\max}) = \frac{\sqrt{\frac{p_j}{2\pi}} \exp\left(-\frac{p_j}{2} (s_{j,t} - \mu_j)^2\right) \mathbb{I}_{[s_j^{\min}, s_j^{\max}]}(s_{j,t})}{\Phi(\sqrt{p_j}(s_j^{\max} - \mu_j)) - \Phi(\sqrt{p_j}(s_j^{\min} - \mu_j))}, \quad (3)$$

where  $\mu_j, p_j$  are the unknown distribution parameters, and  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution.  $\mathbb{I}_{[s_j^{\min}, s_j^{\max}]}(s_{j,t})$  denotes the indicator function, which takes one in the interval  $[s_j^{\min}, s_j^{\max}]$  and zero otherwise.

The sources usually possess a time structure in real problems. Motivated by that, a second prior modeling<sup>2</sup> can be

<sup>1</sup>The parameters  $\mu_j$  and  $p_j$  correspond to the sources hyperparameters.

<sup>2</sup>Since the derivation of a Bayesian method is almost the same for the i.i.d. and the Markovian modeling, our calculations will be based in Eq. (3). Note however that, in the Markovian modeling, there is no need to estimate the term  $\mu_j$  that appears in this equation.

defined by substituting  $\mu_j = s_{j,t-1}$  in Eq. (3). The resulting prior is a first-order Markovian model quite similar to the classical AR(1) model driven by Gaussian noise, with the only difference that the recurrence is limited in the interval  $[s_j^{\min}, s_j^{\max}]$ . Both in the i.i.d. and in the Markovian modeling, we assume that the sources are mutually independent<sup>3</sup>, i.e.<sup>4</sup>  $p(\mathbf{S}) = \prod_{j=1}^{n_s} p(s_{j,:})$ .

##### 3.1.2. Sources hyperparameters

For the i.i.d. modeling, uniform priors are assigned for the sources hyperparameters, that is

$$p(\mu_j) \propto \mathbb{I}_{[\mu_j^{\min}, \mu_j^{\max}]}(\mu_j), \quad (4)$$

$$p(p_j) \propto \mathbb{I}_{[p_j^{\min}, p_j^{\max}]}(p_j), \quad (5)$$

where the parameters  $\mu_j^{\min}, \mu_j^{\max}, p_j^{\min}$  and  $p_j^{\max}$  should be assigned according to the available information. If, for example, the sources are expected to be concentrated near the minimum value, one can set  $\mu_j^{\min} < \mu_j^{\max} < s_j^{\min}$ . Conversely, if no additional information is available, one must increase the limits of both hyperparameters. Regarding the Markovian modeling, we have only one hyperparameter ( $p_j$ ) and the same prior of Eq. (5) is assigned for it.

##### 3.1.3. Mixing parameters

Before assigning priors for the mixing parameters, let us rewrite the LQ mixing model as

$$x_{i,t} = \sum_{m=1}^J c_{i,m} \bar{s}_{m,t} + n_{i,t}, \quad (6)$$

where  $J = n_s + \frac{n_s!}{2(n_s-2)!}$ . The vector  $\mathbf{c}_i = [c_{i,1}, \dots, c_{i,J}]$  stems from the concatenation of  $[a_{i,1}, \dots, a_{i,n_s}]$  and  $[b_{i,1,2}, \dots, b_{i,n_s-1,n_s}]$ ; and the vector  $\bar{\mathbf{s}}_t = [\bar{s}_{1,t}, \dots, \bar{s}_{J,t}]$  denotes the concatenation of the sources  $[s_{1,t}, \dots, s_{n_s,t}]$  and the linear quadratic terms  $[s_{1,t}s_{2,t}, s_{2,t}s_{3,t}, \dots, s_{n_s-1,t}s_{n_s,t}]$ . This new expression points out that the conditional distributions of  $a_{i,j}$  and  $b_{i,j,k}$  used in the Gibbs' sampler (see Section 3.3) assume similar expressions. Therefore, for sake of simplicity, both  $a_{i,j}$  and  $b_{i,j,k}$  will be represented by  $c_{i,m}$ .

In our method, the mixing coefficients  $c_{i,m}$  are modeled through uniform priors, i.e.

$$p(c_{i,m}) \propto \mathbb{I}_{[c_{i,m}^{\min}, c_{i,m}^{\max}]}(c_{i,m}). \quad (7)$$

The distribution bounds should be set based on the available information. An interesting aspect of this modeling is that it

<sup>3</sup>As discussed in [7], for instance, the independence assumption in the Bayesian approach is rather a working assumption, i.e., by assuming that, we are just omitting in our model a possible relation between the sources.

<sup>4</sup>In this work,  $c_{l,m:n}$  is an abbreviation for representing the elements  $c_{l,m}, \dots, c_{l,n}$ ; and  $c_{l,:}$  represents the elements  $c_{l,1}, \dots, c_{l,f}$  where  $f$  is the total number of elements.

renders possible to perform linear BSS using the same implementation of the LQ case. Indeed, if one sets  $c_{i,n_s+1:J}^{min} = c_{i,n_s+1:J}^{max} = 0$ , then the parameters that multiply the linear-quadratic terms become null, i.e. the resulting model becomes linear. Moreover, if  $s_j^{min} = 0$  and  $s_j^{max} \rightarrow \infty$ , our proposal becomes able to model non-negative prior as in non-negative matrix factorization (NMF).

### 3.1.4. Noise variances

A common approach [8] is to assign Gamma priors for the noise precisions  $r_i = 1/\sigma_i^2$ , that is

$$p(r_i) \propto r_i^{\alpha_{r_i}-1} \exp\left(\frac{-r_i}{\beta_{r_i}}\right) \mathbb{1}_{[0,+\infty[}(r_i). \quad (8)$$

This choice culminates in a conjugate pair, which eases the sampling step in the Gibbs sampler. Moreover, it is possible to set the hyperparameters  $\alpha_{r_i}$  and  $\beta_{r_i}$  to obtain a vague prior [8].

### 3.2. Bayesian inference and Gibbs sampler

Since we assume that all elements of  $p(\theta)$  are statistically independent (except the sources  $s_{j,:}$  and their hyperparameters  $\mu_j$  and  $p_j$ ), the posterior distribution  $p(\theta|\mathbf{X})$  can be rewritten as

$$p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta) \times \prod_{i=1}^{n_c} \prod_{m=1}^J p(c_{i,m}) \times \prod_{n=1}^{n_s} p(s_{j,:}|\mu_j, p_j) \times \prod_{n=1}^{n_s} p(\mu_j) \times \prod_{n=1}^{n_s} p(p_j) \times \prod_{i=1}^{n_c} p(\sigma_i^2). \quad (9)$$

The next step is to define an adequate algorithm for the estimation of  $\theta$  using this posterior distribution. A possible approach is based on calculating the Bayesian minimum mean square error (MMSE) estimator [9] which is defined as  $\theta_{MMSE} = \int \theta p(\theta|\mathbf{X}) d\theta$ .

Even though the integral present in the Bayesian MMSE estimator makes its exact calculation difficult, it is still possible to obtain good solutions through approximation methods. For example, if  $\theta^1, \theta^2, \dots, \theta^M$  represents a set of samples drawn from  $p(\theta|\mathbf{X})$ , then the Bayesian MMSE estimator can be approximated by  $\tilde{\theta}_{MMSE} = \frac{1}{M} \sum_{i=1}^M \theta^i$ . Therefore, in this methodology, which is called Monte Carlo integration, the implementation of the Bayesian MMSE estimator boils down to the task of finding an efficient way for sampling from the distribution  $p(\theta|\mathbf{X})$ .

In this work, the simulation of  $p(\theta|\mathbf{X})$  is conducted via the Gibbs' sampler [10], a Markov chain Monte Carlo (MCMC) method tailored to deal with high-dimensional distributions. One of the attractive features of the Gibbs' sampler is that it permits to simulate a joint distribution by sequentially sampling from the conditional distribution of each variable. This procedure can be summarized as follows:

1. Initialize all the parameters  $\theta_1^0, \theta_2^0, \dots, \theta_N^0$ ;

2. For  $p = 1$  to  $P$  do

$$\theta_1^p \sim p(\theta_1|\theta_2^{p-1}, \theta_3^{p-1}, \dots, \theta_N^{p-1}, \mathbf{X})$$

$$\theta_2^p \sim p(\theta_2|\theta_1^p, \theta_3^{p-1}, \dots, \theta_N^{p-1}, \mathbf{X})$$

$\vdots$

$$\theta_N^p \sim p(\theta_N|\theta_1^p, \theta_2^p, \dots, \theta_{N-1}^p, \mathbf{X})$$

end

The notation  $x \sim p(x)$  stands for the sampling operation, i.e.  $x$  is a sample obtained from the distribution  $p(x)$ . Therefore, the implementation of the Gibbs' sampler requires the conditional distribution of each unknown parameter of our problem. We shall obtain these expressions in the sequel.

### 3.3. Conditional distributions

After some manipulations, one can check that the conditional distribution of a given parameter  $\theta_q$  is given by

$$p(\theta_q|\theta_{-q}, \mathbf{X}) \propto p(\mathbf{X}|\theta)p(\theta_q). \quad (10)$$

Therefore, the calculation of the conditional distributions can be achieved by substituting the likelihood function and the prior distribution into this expression. In the sequel, this procedure will be done for each unknown parameter.

#### 3.3.1. Sources

It is not difficult to show that, by substituting expressions (2) and (3) into (10), one has

$$p(s_{j,t}|\theta_{-s_{j,t}}, \mathbf{X}) \propto \exp\left[-\sum_{i=1}^{n_c} \frac{1}{2\sigma_i^2} \left(\Psi_{i,j,t} s_{j,t} + \Omega_{ijt}\right)^2 - 0.5 p_j (s_{j,t} - \mu_j)^2\right] \mathbb{1}_{[s_j^{min}, s_j^{max}]}(s_{j,t}), \quad (11)$$

where

$$\Omega_{i,j,t} = x_{i,t} - \sum_{g=1, g \neq j}^{n_s} a_{i,g} s_{g,t} - \sum_{1 \leq g < k \leq n_s, g \neq j} b_{i,g,k} s_{g,t} s_{k,t}, \quad (12)$$

and

$$\Psi_{i,j,t} = a_{i,j} + \sum_{g=1, g \neq j}^{n_s} b_{i,j,g} s_{g,t}. \quad (13)$$

After expanding Eq. (11), one can obtain the expression of the conditional distribution of the sources, given by

$$p(s_{j,t}|\theta_{-s_{j,t}}, \mathbf{X}) \propto \exp\left(-\frac{(s_{j,t} - \mu_{Postj})^2}{2\sigma_{Postj}^2}\right) \mathbb{1}_{[s_j^{min}, s_j^{max}]}(s_{j,t}), \quad (14)$$

where  $\sigma_{Postj} = \sigma_{Lj}^2 \sigma_j^2 / (\sigma_{Lj}^2 + \sigma_j^2)$  and  $\mu_{Postj} = (\mu_{Lj} \sigma_j^2 + \mu_j \sigma_{Lj}^2) / (\sigma_{Lj}^2 + \sigma_j^2)$ , and

$$\sigma_{Lj}^2 = \left( \sum_{i=1}^{n_c} \frac{\Psi_{ijt}^2}{\sigma_i^2} \right)^{-1}, \quad (15)$$

$$\mu_{Lj} = \sigma_{Lj}^2 \sum_{i=1}^{n_c} \frac{\Omega_{ijt} \Psi_{ijt}}{\sigma_i^2}. \quad (16)$$

The expression in (11) corresponds to a truncated Gaussian distribution and its simulation can be easily conducted by the procedure proposed in [11].

### 3.3.2. Source hyperparameters

Let us start with the derivation of the conditional distribution of  $p_j$ . As the likelihood function (2) is not a function of  $p_j$ , it asserts that  $p(p_j | s_{j,:}, \mu_j) \propto p(s_{j,:} | p_j, \mu_j) p(p_j)$ , that is

$$p(p_j | s_{j,:}, \mu_j) \propto p_j^{\frac{n_d}{2}} \exp \left( -0.5 p_j \sum_{t=1}^{n_d} (s_{j,t} - \mu_j)^2 \right) \times \frac{\mathbb{I}_{[p_j^{min}, p_j^{max}]}(p_j)}{\Phi(\sqrt{p_j}(s_j^{max} - \mu_j)) - \Phi(\sqrt{p_j}(s_j^{min} - \mu_j))}. \quad (17)$$

This expression does not assume a standard form because of the nonlinearity in the denominator of the second term. This is also true for the distribution  $p(\mu_j | s_{j,:}, p_j)$  that appears in the i.i.d. modeling. These non-standard distributions could be simulated through the Metropolis-Hastings (MH) algorithm [10]. However, this sort of solution requires the definition of an instrumental function which is usually not a trivial task. Besides, the presence of MH would increase the complexity of our final solution.

In an alternative approach, the use of latent variables results in conditional distributions that assume standard forms. This idea, which was developed in [12], is based on the following transformation:

$$l_{j,t} = \mu_j + p_j^{-1/2} \times \Phi^{-1} \left( \frac{\Phi(\sqrt{p_j}(s_{j,t} - \mu_j)) - \Phi(\sqrt{p_j}(s_j^{min} - \mu_j))}{\Phi(\sqrt{p_j}(s_j^{max} - \mu_j)) - \Phi(\sqrt{p_j}(s_j^{min} - \mu_j))} \right). \quad (18)$$

It can be proved [12] that if  $s_{j,t}$  follows a truncated Gaussian with parameters  $\mu_j$  and  $p_j$ , then  $l_{j,t}$  is distributed according to a Gaussian distribution of mean  $\mu_j$  and precision  $p_j$ . Therefore, the definition of  $l_{j,t}$  turns the problem of estimating the parameters of a truncated Gaussian into the one of estimating the mean and precision of a Gaussian.

From the discussion of the last paragraph,  $l_{j,t}$  follows a Gaussian distribution, and therefore

$$p(l_{j,:} | \mu_j, p_j) = \prod_{t=1}^{n_d} \sqrt{\frac{p_j}{2\pi}} \exp \left( -\frac{p_j}{2} (l_{j,t} - \mu_j)^2 \right). \quad (19)$$

Using this equation and the prior distributions (4) and (5), one can show that the new conditional distributions of  $\mu_j$  and  $p_j$  are given by

$$p(\mu_j | p_j, l_{j,:}) \propto p(\mu_j) p(l_{j,:} | \mu_j, p_j) \propto \mathbb{I}_{[\mu_j^{min}, \mu_j^{max}]}(\mu_j) \times \exp \left( -\frac{p_j \cdot n_d}{2} \left( \mu_j - \frac{1}{n_s} \sum_{t=1}^{n_d} l_{j,t} \right)^2 \right) \quad (20)$$

$$p(p_j | \mu_j, l_{j,:}) \propto p(p_j) p(l_{j,:} | \mu_j, p_j) \propto \mathbb{I}_{[p_j^{min}, p_j^{max}]}(p_j) \times p_j^{\frac{n_d}{2}} \exp \left( -p_j \sum_{t=1}^{n_d} (l_{j,t} - \mu_j)^2 \right). \quad (21)$$

Now we have more tractable distribution, since  $p(\mu_j | p_j, l_{j,:})$  is a truncated Gaussian whereas  $p(p_j | \mu_j, l_{j,:})$  is a truncated Gamma. The simulation of these two distributions can be conducted through the method proposed in [11].

The original procedure of [12], described in the last paragraphs, can be readily extended for estimating  $p_j$  when the Markovian modeling is considered. Indeed, this can be done by observing that the innovation process  $s_{j,t} - s_{j,t-1}$  is distributed according to a truncated Gaussian whose limits depend on the time index. Therefore, the conditional distribution of  $p_j$  in this case is obtained by substituting  $\mu_j = s_{j,t-1}$  in Eq. (21). Also, the same substitution should be conducted in Eq. (18) for the calculation of the latent variables  $l_{j,t}$ .

### 3.3.3. Mixing parameters

The calculation of the conditional distributions can be done by substituting equations (7) and (2) into Eq. (10). Therefore, one obtains after some calculations

$$p(c_{i,m} | \theta_{-c_{i,m}}, \mathbf{X}) \propto \exp \left( -\frac{\rho_{i,m}^L}{2} (c_{i,m} - \nu_{i,m}^L)^2 \right) \mathbb{I}_{[c_{i,m}^{min}, c_{i,m}^{max}]}(c_{i,m}), \quad (22)$$

where

$$\rho_{i,m}^L = \sigma_i^2 \sum_{t=1}^{n_d} \bar{s}_{m,t}, \quad (23)$$

$$\nu_{i,m}^L = \frac{\sum_{t=1}^{n_d} \bar{s}_{m,t} \left( x_{i,t} - \sum_{g=1, g \neq m}^J c_{i,g} \bar{s}_{g,t} \right)}{\sum_{t=1}^{n_d} \bar{s}_{m,t}^2}. \quad (24)$$

Again, the resulting conditional distribution (Eq. (22)) is a truncated Gaussian distribution and can be simulated by the technique presented in [11].

### 3.3.4. Noise variances

The conditional distribution of the noise precision  $r_i = 1/\sigma_i^2$  is obtained by substituting (8) and (2) into Eq. (10), which



gives

$$p(r_i | \boldsymbol{\theta}_{-r_i}, \mathbf{X}) \propto r_i^{\frac{n_d}{2}} \exp(-0.5r_i\Theta_{i,t}) r_i^{\alpha_{r_i}-1} \exp\left(\frac{-r_i}{\beta_{r_i}}\right) \mathbb{1}_{[0,+\infty[}(r_i) \quad (25)$$

where  $\Theta_{i,t} = x_{i,t} - \sum_{j=1}^{n_s} a_{i,j}s_{j,t} - \sum_{i,j,k} b_{i,j,k}s_{j,t}s_{k,t}$ . This equation can be rewritten as

$$p(r_i | \boldsymbol{\theta}_{-r_i}, \mathbf{X}) \propto \exp\left(-r_i\left(0.5\Theta_{i,t} + \frac{1}{\beta_{r_i}}\right)\right) r_i^{\frac{n_d}{2} + \alpha_{r_i} - 1} \mathbb{1}_{[0,+\infty[}(r_i), \quad (26)$$

which is a Gamma distribution with parameters  $\alpha_i = \frac{n_d}{2} + \alpha_{r_i}$  and  $\beta_i^{-1} = 0.5\Theta_{i,t} + \beta_{r_i}^{-1}$ .

## 4. RESULTS

To access the performance of our proposal, we conduct a set of simulations with synthetic data. In a first scenario, we test our method in a linear source separation problem where the sources and the mixing coefficients are non-negative. Then, we address the case of linear-quadratic mixtures. In both situations, the following performance index is considered

$$SIR = \frac{1}{n_d \cdot n_s} \sum_{j=1}^{n_s} \sum_{t=1}^{n_d} 10 \log \left( \frac{E\{s_{j,t}^2\}}{E\{(s_{j,t} - \hat{s}_{j,t})^2\}} \right), \quad (27)$$

where  $\hat{s}_{j,t}$  is the estimation of the source  $j$  at time  $t$ . It is worth remembering that  $\hat{s}_{j,t}$  is given by  $\hat{s}_{j,t} = \frac{1}{P-B} \sum_{p=B}^P s_{j,t}^p$ , where  $s_{j,t}^p$  represents the  $p$ -th sample of  $s_{j,t}$  provided by the Gibbs' sampler;  $P$  and  $B$  denotes the number of total iterations and the length of the burn-in period<sup>5</sup>, respectively.

### 4.1. Separation of linear mixtures

To illustrate the performance of our proposal in a linear case, we tested it in situations where  $n_d = 300$ ,  $n_s = 3$ ,  $n_c = 3$ ; and the mixing matrix is given by  $\mathbf{A} = [1 \ 0.5 \ 0.5; 0.6 \ 1 \ 0.3; 0.8 \ 0.4 \ 1]$ . Three scenarios were considered: 1) the sources are realizations of truncated Gaussian distributions (matched case with our i.i.d. modeling); 2) the sources are realizations of truncated Gaussian Markovian process (matched case with our Markovian modeling); 3) the sources correspond to a sine wave, a ramp function and a sawtooth wave. In all these situations the signal-to-noise ratio (SNR) at each sensor was 20 dB. The total number of iterations of the Gibbs' samples was  $P = 10000$  with a burn-in period of  $B = 5000$ .

The results presented in Tab. 1 represent the mean SIR over 50 experiments. Despite a (not surprisingly) performance degradation when the Markovian prior is used for

**Table 1.** *SIR* (dB) for the separation of linear mixtures.

	Situation 1	Situation 2	Situation 3
Bayesian method			
i.i.d. modeling	17.50	17.69	15.55
Bayesian method			
Markovian modeling	12.60	17.07	18.37
FastICA	20.43	17.74	11.81

separating i.i.d. sources, our proposal was able to separate the sources. The FastICA algorithm [1] gave us better results in the first two situations. On the other hand, the application of this method on the third situation did not provide satisfactory results. This was due to the existence of two correlated sources in this scenario. It is worth remembering that, in contrast to the Bayesian approach, the FastICA searches for independent components and, therefore, it may fail when the sources are not independent.

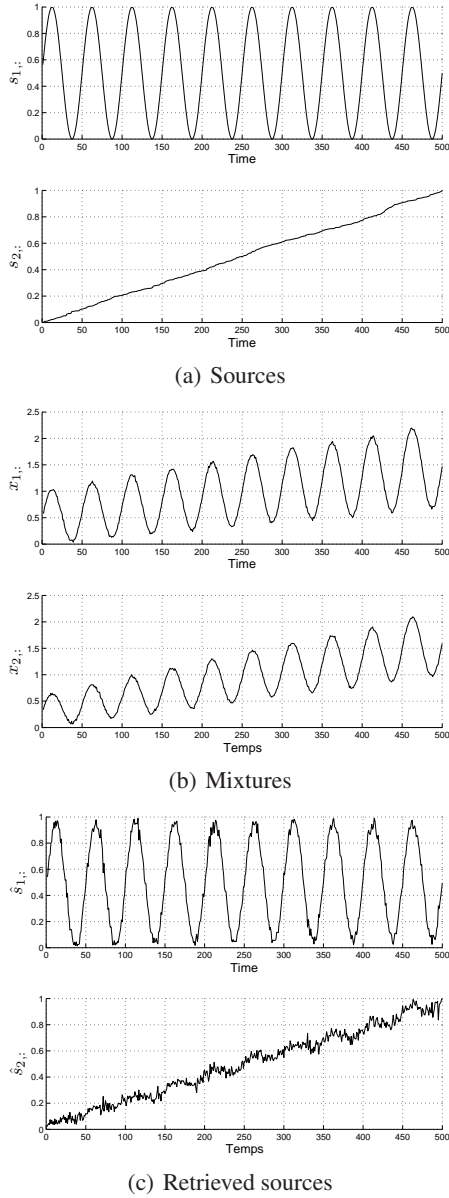
### 4.2. Separation of linear-quadratic mixtures

In a first moment, we considered a situation where  $n_d = 500$ ,  $n_s = 2$  and  $n_c = 2$ . The original sources and the mixtures are presented in Figs. 1(a) and 1(b), respectively. The mixing parameters were selected  $a_{1,1} = 1$ ,  $a_{1,2} = 0.5$ ,  $b_{1,1,2} = 0.2$ ,  $a_{2,1} = 0.5$ ,  $a_{2,2} = 1$ ,  $b_{2,1,2} = 0.2$ , and the SNR at each sensor was 30 dB. The hyperparameters related to the limit values of the prior distributions were<sup>6</sup>  $s_j^{min} = c_{i,m}^{min} = 0$  and  $s_j^{max} = c_{i,m}^{max} = 1$ . Concerning the Gibbs sampler parameters, the total number of iteration was  $P = 20000$  with a burn in period of  $B = 8000$ . In this situation, the obtained performed index were  $SIR = 26.63$  dB for the i.i.d. modeling and  $SIR = 27.47$  dB for the Markovian modeling. We also tested the ICA method proposed in [4] which was able to provide good approximations ( $SIR = 22.58$  dB). Despite the better performance, it worth mentioning that the gains brought by our method comes at the price of a greater computational effort.

We also tested our method in a second scenario similar to the first one with the only difference that the mixing parameters are now given by  $a_{1,1} = 1$ ,  $a_{1,2} = 0.7$ ,  $b_{1,1,2} = 0.6$ ,  $a_{2,1} = 0.6$ ,  $a_{2,2} = 1$ ,  $b_{2,1,2} = 0.6$ . Our method was able to retrieve the sources both for i.i.d. modeling ( $SIR = 22.17$  dB) and for the Markovian modeling ( $SIR = 23.71$  dB). To illustrate that, we shown in Fig. 1 the retrieved sources for the Markovian modeling. Note that, despite the noise amplification, which is typical in nonlinear systems, the retrieved signals are close to the sources. Conversely, in this new case, the method proposed in [4] failed to separate the sources since the mixing coefficients violate the stability condition of the recurrent separating system.

<sup>5</sup>Since the Markov chain associated with the Gibbs' sampler takes some iterations to reach the stationary distribution, the samples generate in an initial moment, the burn-in period, should be discarded.

<sup>6</sup>We set  $a_{1,1}^{min} = a_{2,2}^{min} = 1$  to avoid scaling ambiguities.



**Fig. 1.** Separation of LQ mixtures (Markovian modeling).

## 5. CONCLUSION

We proposed a Bayesian source separation method that can be used in linear-quadratic and linear mixing models. The application of Gibbs' sampler and the introduction of latent variables provided an algorithm that is easy to implement. Concerning the results, we observed that this proposal can be useful in some situations where ICA methods cannot be applied. One limitation of our solution concerns its computational complexity. Indeed, each iteration of the Gibbs' sampler performs  $n_s \times n_d$  simulations of univariate random variables. Therefore, the computational burden required by our proposal may become quite large in problems where the

number of sources and samples are considerable.

## 6. REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley & Sons, 2001.
- [2] C. Jutten and J. Karhunen, "Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures," *International Journal of Neural Systems*, vol. 14, pp. 267–292, 2004.
- [3] S. Hosseini and Y. Deville, "Blind separation of linear-quadratic mixtures of real sources using a recurrent structure," in *Proceedings of the 7th International Work-conference on Artificial And Natural Neural Networks, IWANN 2003*, 2003, pp. 289–296.
- [4] S. Hosseini and Y. Deville, "Blind maximum likelihood separation of a linear-quadratic mixture," in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2004*, 2004, pp. 694–701.
- [5] G. Bedoya, *Nonlinear blind signal separation for chemical solid-state sensor arrays*, Ph.D. thesis, Universitat Politècnica de Catalunya, 2006.
- [6] Y. Deville and S. Hosseini, "Recurrent networks for separating extractable-target nonlinear mixtures. part i: Non-blind configurations," *Signal Processing*, vol. 89, pp. 378–393, 2009.
- [7] C. Févotte and S. J. Godsill, "A bayesian approach for blind separation of sparse sources," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, pp. 2174–2188, 2006.
- [8] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4133–4145, 2006.
- [9] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, 1993.
- [10] C. P. Robert, *The Bayesian Choice*, Springer, 2007.
- [11] P. Damien and S. G. Walker, "Sampling truncated normal, beta, and gamma densities," *Journal of Computational and Graphical Statistics*, vol. 10, pp. 206–215, 2001.
- [12] W. Griffiths, "A Gibbs' sampler for a truncated multivariate normal distribution," Tech. Rep., University of Melbourne, 2002.