



HAL
open science

Bayesian Source Separation of Linear and Linear-quadratic Mixtures Using Truncated Priors

Leonardo Tomazeli Duarte, Christian Jutten, Saïd Moussaoui

► **To cite this version:**

Leonardo Tomazeli Duarte, Christian Jutten, Saïd Moussaoui. Bayesian Source Separation of Linear and Linear-quadratic Mixtures Using Truncated Priors. *Journal of Signal Processing Systems*, 2010, 65 (3), pp.311-323. 10.1007/s11265-010-0488-3 . hal-00484162

HAL Id: hal-00484162

<https://hal.science/hal-00484162>

Submitted on 19 May 2010

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 and Linear-quadratic Mixtures using Truncated Priors

Leonardo Tomazeli Duarte · Christian
Jutten · Saïd Moussaoui

Received: date / Accepted: date

Abstract In this work, we propose a Bayesian source separation method of linear-quadratic (LQ) and linear mixtures. Since our method relies on truncated prior distributions, it is particularly useful when the bounds of the sources and of the mixing coefficients are known in advance; this is the case, for instance, in non-negative matrix factorization. To implement our idea, we consider a Gibbs' sampler equipped with latent variables, which are set to simplify the sampling steps. Experiments with synthetic data point out that the new proposal performs well in situations where classical ICA-based solutions fail to separate the sources. Moreover, in order to illustrate the application of our method to actual data, we consider the problem of separating scanned images.

The work of L. T. Duarte is supported by Fapesp (Brazil). Most part of this paper was prepared while L. T. Duarte was a PhD student (founded by the CNPq-Brazil) at the Gipsa-lab (France). C. Jutten is also with the Institut Universitaire de France.

Leonardo T. Duarte
Laboratory of Signal Processing for Communications (DSPCom)
University of Campinas (UNICAMP)
C.P. 6101, CEP 13083-970, Campinas SP, Brazil
Tel.:+55(19)3521-3857
Fax.:+55(19)3521-3857
E-mail: leonardo.tomazeli.duarte@gmail.com

Christian Jutten
GIPSA-lab (UMR CNRS 5216)
Institut Polytechnique de Grenoble
BP 46, F-38402 Grenoble Cedex, France
Tel.:+33 (0)4 76 57 43 51
Fax.:+33 (0)4 76 57 47 90
E-mail: christian.jutten@gipsa-lab.grenoble-inp.fr

Saïd Moussaoui
IRCCyN (UMR CNRS 6597)
Ecole Centrale Nantes
BP 92101, F-44321 Nantes Cedex 03, France
Tel.:+33 (0)2 40 37 69 19
Fax.:+33 (0)2 40 37 69 30
E-mail: said.moussaoui@irccyn.ec-nantes.fr.

Keywords Source separation · Bayesian approach · Nonlinear mixtures · Truncated priors · Scanned images

1 Introduction

The problem of Blind Source Separation (BSS) concerns the recovery of a set of signals (sources) based on mixed versions of these original signals. The term “blind” is employed because separation is conducted by using a minimum amount of information about the sources and about the underlying mixing process. This problem has been attracting a great deal attention of the signal processing community, among others, for about two decades; one can find applications in a great diversity of domains, from astronomical imaging to biosignal processing [12,13].

Most of the existing BSS methods consider a linear mixing process and are based on the idea of Independent Component Analysis (ICA) [3,12]. In ICA, the sources are assumed to be mutually statistically independent. Given that such a property is lost after the mixing process, ICA tries to recover the sources by retrieving signals that are, again, independent.

Despite the impressive results provided by the conjunction ICA/linear models, the extension of this classical framework is quite desirable as it can broaden the range of BSS applications. On the other hand, the study of BSS in the case of, for instance, nonlinear mixing models becomes more involved [14,13]. For example, due to the high degree of flexibility present in a nonlinear model, the application of ICA in such a case does not guarantee source separation; that is, one may recover independent components that are still mixed versions of the sources. In view of such limitations, several papers [20,7,11] suggest that a better approach to nonlinear BSS is to consider constrained classes of nonlinear models, for which source separation is still possible.

In [10,11], Hosseini and Deville studied source separation in a particular constrained mixing model: the linear-quadratic (LQ) model. This kind of model can be seen as a second-order approximation of the mixing process, thus suggesting that LQ models could be a first step toward polynomial mixtures. In addition to such an interesting theoretical perspective, one can find applications to the LQ model in problems involving smart gas electrode arrays [2] and the separation of scanned images [16].

A challenging aspect in the development of BSS methods to LQ models is the definition of a suitable structure for the separating system. Indeed, differently from the linear case, the inversion of the LQ mixing model does not admit closed formulae in the general case. To overcome this problem, Hosseini and Deville [10,11] proposed a recurrent separating network that performs a sort of implicit inversion of the mixing system. Moreover, they proposed ICA-based methods to train the recurrent separating system.

Despite its simplicity and its good performance, the solution proposed in [10,11] presents a limitation. Actually, such an approach can operate only when the sources and the mixing parameters satisfy the stability conditions imposed by the recurrent separating system. In view of such a limitation, a recent work [5] showed that the development of more elaborate recurrent separating systems can extend this stability region. However, for such systems, the definition of training algorithms becomes quite complicate.

In this work, we propose an alternative method for separating LQ mixtures based on a Bayesian framework. Our main motivation is to define a method that does not suf-

fer from the above-mentioned risk of instability present in recurrent separating systems. 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. Another advantage brought by a Bayesian approach is the possibility of taking into account prior information other than the statistical independence. In this work, for instance, the following prior information can be taken into account: 1) the bounds of the sources and of the mixing coefficient values are known in advance, and 2) the sources have a temporal structure. These properties are incorporated via a modeling based on truncated priors. It is worth mentioning that, under minor changes, the developed method can also be applied to linear models and to non-negative matrix factorization.

Concerning the implementation of the proposed Bayesian method, an hierarchical model is considered. The inference scheme is based on Markov chain Monte Carlo (MCMC) methods and on the definition of a set of auxiliary variables for obtaining standard conditional distributions, which simplifies the posterior distribution sampling step.

The paper is organized as follows. In section 2, we introduce the mixing model considered in this work. Then, in Section 3, we introduce the proposed Bayesian method. Section 4 present a set of experiments to illustrate the gains brought by the proposed method. In a first moment, we consider simulations with synthetic data. Then, we illustrate the application of our method to the problem of separating scanned images. Finally, in Section 5, we present conclusions and perspectives for future works.

2 Problem description: the mixing model

Let $s_{j,t}$ denote the j -th source sample at time t . The LQ mixing model is thus given by

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

where $x_{i,t}$ corresponds to the i -th observed mixture at the instant t . The coefficient $a_{i,j}$ represents the linear mixing coefficients whereas $b_{i,j,k}$ denotes the quadratic terms. An independent and identically distributed (i.i.d) Gaussian noise sequence $n_{i,t}$, with zero mean and unknown variance σ_i^2 , is assumed at each sensor. The parameters n_s , n_c and n_d correspond to the number of sources, sensors and available samples, respectively.

The problem addressed in this work concerns thus the estimation of the sources $s_{j,t}$ from the observations $x_{i,t}$. Of course, since we are in a blind context, there are other unknown parameters in addition to $s_{j,t}$. In our notation, all these unknown parameters will be represented by the vector¹ $\boldsymbol{\theta} = [s_{j,t}, a_{i,j}, b_{i,j,k}, \sigma_i^2, \mu_j, p_j]$, and $\boldsymbol{\theta}_{-\theta_q}$ will denote the vector containing all elements of $\boldsymbol{\theta}$ except θ_q . Therefore, the estimation problem considered in this work can be stated as follows: given \mathbf{X} (data matrix containing all $x_{i,t}$), estimate the parameters of the vector $\boldsymbol{\theta}$. In the sequel, we discuss how to address this problem in a Bayesian context.

¹ The parameters μ_j and p_j correspond to the source parameters.

3 Bayesian source separation method

The central element in Bayesian estimation is the posterior distribution $p(\boldsymbol{\theta}|\mathbf{X})$, which, according to the Bayes' theorem, can be written as

$$p(\boldsymbol{\theta}|\mathbf{X}) \propto p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (2)$$

where $p(\mathbf{X}|\boldsymbol{\theta})$ denotes the likelihood function and $p(\boldsymbol{\theta})$ represents the prior distributions. The likelihood function provides a probabilistic model for the relation between the unknown parameters $\boldsymbol{\theta}$ and the observed data \mathbf{X} . Due to the assumption of white Gaussian noise in the observation model, the likelihood in our problem is given by

$$p(\mathbf{X}|\boldsymbol{\theta}) = \prod_{t=1}^{n_d} \prod_{i=1}^{n_c} \mathcal{N}_{x_{it}} \left(\sum_{j=1}^{n_s} a_{i,j} s_{j,t} + \sum_{1 \leq j < k \leq n_s} b_{i,j,k} s_{j,t} s_{k,t}; \sigma_i^2 \right), \quad (3)$$

where $\mathcal{N}_{x_{ik}}(\mu; \sigma^2)$ denotes a Gaussian distribution of x_{ik} with mean μ and variance σ^2 .

Concerning the prior distributions $p(\boldsymbol{\theta})$, they should be related to the available information at hand. In the sequel, we describe how these priors are defined in our method.

3.1 Definition of prior distributions

3.1.1 Sources

In this work, we consider a modeling based on truncated Gaussian distributions for the sources. This distribution is interesting when, for instance, the bounds of the sources are known. In a first moment, the following i.i.d. modeling is considered

$$p(s_{j,t}|\mu_j, p_j, s_j^{min}, s_j^{max}) = \frac{\sqrt{\frac{p_j}{2\pi}} \exp\left(-\frac{p_j}{2} (s_{j,t} - \mu_j)^2\right) \mathbb{1}_{[s_j^{min}, s_j^{max}]}(s_{j,t})}{\Phi\left(\sqrt{p_j} (s_j^{max} - \mu_j)\right) - \Phi\left(\sqrt{p_j} (s_j^{min} - \mu_j)\right)}, \quad (4)$$

where μ_j and p_j correspond to the distribution parameters, which, as we assume a hierarchical model, are assumed unknown, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The truncation points s_j^{min} and s_j^{max} are fixed and should be assigned by the user. In order to illustrate the distribution of Equation (4), we plot it in Figure 4 for different parameters and bound values. Note that the truncated Gaussian distribution provides a flexible prior as it can model a large range of distributions from almost uniform sources to signals that are concentrated near a certain value.

Very often in actual problems, the sources present a temporal structure. Motivated by that, a second prior modeling² can be defined by substituting $\mu_j = s_{j,t-1}$ in Equation (4). The resulting prior is a first-order Markovian model quite similar to the classical AR(1) model driven by Gaussian innovation sequence, 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 statistically independent, i.e. $p(\mathbf{S}) = \prod_{j=1}^{n_s} p(s_{j,:})$, where $s_{j,:}$ denotes the j -th source.

² 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 Equation (4). Note however that, in the Markovian modeling, there is no need to estimate the term μ_j that appears in this equation.

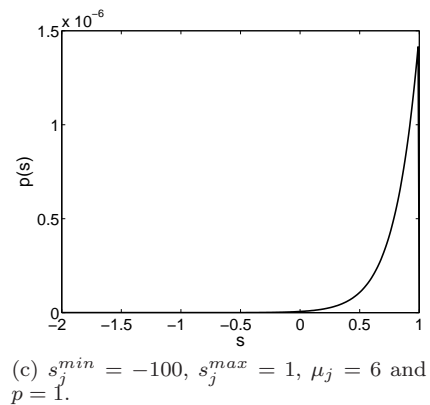
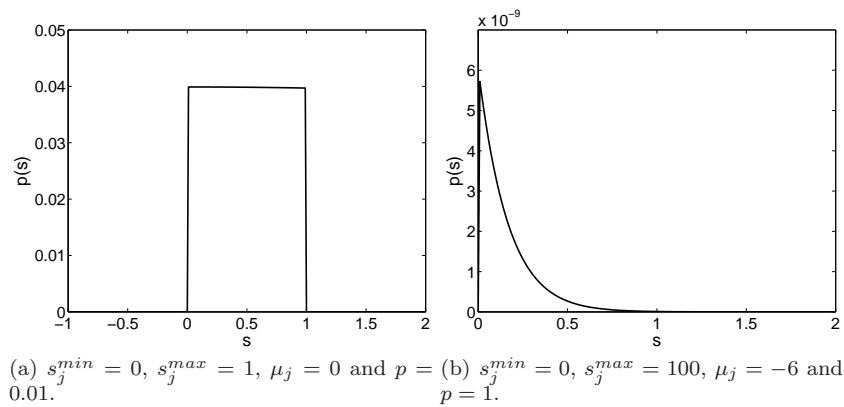


Fig. 1 Truncated Gaussian distribution: examples with different parameters.

3.1.2 Parameters related to the source's distributions

In our hierarchical models, the parameters associated with the prior distributions of the sources are also unknown and, thus, should be estimated. In this spirit, for the i.i.d. modeling expressed in Equation (4), the following uniform priors are considered

$$p(\mu_j) \propto \mathbb{1}_{[\mu_j^{min}, \mu_j^{max}]}(\mu_j), \quad (5)$$

$$p(p_j) \propto \mathbb{1}_{[p_j^{min}, p_j^{max}]}(p_j), \quad (6)$$

where the parameters $\mu_j^{min}, \mu_j^{max}, p_j^{min}$ and p_j^{max} should be defined based on available information. If, for example, the sources are expected to be concentrated around the inferior bound s_j^{min} , 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 to obtain an almost uniform distribution. Regarding the Markovian modeling, we only have one hyperparameter (p_j) with the same prior of Equation (6).

3.1.3 Mixing parameters

In an alternative notation, the LQ mixing model can be written as the following linear model

$$x_{i,t} = \sum_{m=1}^J c_{i,m} \bar{s}_{m,t} + n_{i,t}, \quad (7)$$

where $J = n_s + \frac{n_s!}{2(n_s-2)!}$. The vector $\mathbf{c}_i = [c_{i,1}, \dots, c_{i,J}]$ results from the concatenation of the linear terms $[a_{i,1}, \dots, a_{i,n_s}]$ and the quadratic terms $[b_{i,1,2}, \dots, b_{i,n_s-1,n_s}]$, and the vector $\bar{\mathbf{s}}_t = [\bar{s}_{1,t}, \dots, \bar{s}_{J,t}]$ is the concatenation of the sources $[s_{1,t}, \dots, s_{n_s,t}]$ and the quadratic terms $[s_{1,t}s_{2,t}, s_{2,t}s_{3,t}, \dots, s_{n_s-1,t}s_{n_s,t}]$. From this alternative notation, one can note that the conditional distributions of $a_{i,j}$ and $b_{i,j,k}$ used in the Gibbs' sampler (see Section 3.4) 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 the sequel.

Concerning the priors assigned for the mixing coefficients $c_{i,m}$, we consider the following uniform distribution

$$p(c_{i,m}) \propto \mathbb{1}_{[c_{i,m}^{min}, c_{i,m}^{max}]}(c_{i,m}). \quad (8)$$

Again, the distribution bounds $c_{i,m}^{min}$ and $c_{i,m}^{max}$ should be set according to any available information. It is worth mentioning here that one can perform linear BSS via the same implementation of the LQ case by simply setting $c_{i,n_s+1:J}^{min} = c_{i,n_s+1:J}^{max} = 0$, i.e. by canceling the quadratic terms. Besides, when $s_j^{min} = 0$ and $s_j^{max} \rightarrow \infty$, our proposal becomes able to perform non-negative matrix factorization (NMF), i.e. to provide a linear representation of the data in which both the mixing coefficients and the sources are non-negative.

3.1.4 Noise variances

Following the usual approach [17,6], Gamma priors were assigned for the noise precisions (inverse of variances) $r_i = 1/\sigma_i^2$, that is

$$p(r_i) \propto r_i^{\alpha r_i - 1} \exp\left(\frac{-r_i}{\beta r_i}\right) \mathbb{1}_{[0, +\infty[}(r_i). \quad (9)$$

This choice culminates in a conjugate pair, which eases the sampling step in the Gibbs' sampler, and accounts for positivity of the noise variance.

3.2 Bayesian inference

The next step of our development is to obtain the posterior distribution $p(\boldsymbol{\theta}|\mathbf{X})$. Since we assume that all elements of $\boldsymbol{\theta}$ are statistically independent (except the sources $s_{j,:}$ and their parameters μ_j and p_j), the distribution $p(\boldsymbol{\theta})$ factorizes and, as a consequence, Equation (2) can be written as

$$p(\boldsymbol{\theta}|\mathbf{X}) \propto p(\mathbf{X}|\boldsymbol{\theta}) \times \prod_{i=1}^{n_c} \prod_{m=1}^J p(c_{i,m}) \times \prod_{n=1}^{n_s} p(s_{j,:}|\mu_j, p_j) \times \prod_{n=1}^{n_s} p(\mu_j) \times \prod_{n=1}^{n_s} p(p_j) \times \prod_{i=1}^{n_c} p(\sigma_i^2). \quad (10)$$

There are several ways to define an estimator of $\boldsymbol{\theta}$ based on the posterior distribution $p(\boldsymbol{\theta}|\mathbf{X})$. In this work, this is done through the Bayesian minimum mean square error (MMSE) estimator [15] (also known as conditional mean estimator), which is given by

$$\boldsymbol{\theta}_{MMSE} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{X}) d\boldsymbol{\theta}. \quad (11)$$

The presence of an integral in the MMSE estimator makes its exact calculation difficult in our problem. Yet, as we shall see in the sequel, it is possible to obtain a good approximation through sampling methods.

Let us suppose that a set of samples drawn from $p(\boldsymbol{\theta}|\mathbf{X})$ is available; these samples will be represented by $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots, \boldsymbol{\theta}^M$. Then, the Bayesian MMSE estimator can be approximated by

$$\tilde{\boldsymbol{\theta}}_{MMSE} = \frac{1}{M} \sum_{i=1}^M \boldsymbol{\theta}^i. \quad (12)$$

Of course, in view of this approximation, 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 $p(\boldsymbol{\theta}|\mathbf{X})$. In this work, this is done via the Gibbs' sampler.

3.3 The Gibbs' sampler

The Gibbs' sampler is a Markov Chain Monte Carlo (MCMC) method specially tailored for simulating joint distributions [8]. The idea behind the Gibbs' sampler is to generate a set of samples by generating a Markov chain that admits the desired distribution ($p(\boldsymbol{\theta}|\mathbf{X})$ in our case) as stationary distribution. Considering that $x \sim p(x)$ stands for the sampling operation, i.e. x is a sample obtained from the distribution $p(x)$, then the generation of a Markov chain in the Gibbs' sampler can be summarized as follows [18]:

1. Initialize all the parameters $\theta_1^0, \theta_2^0, \dots, \theta_N^0$;
2. For $p = 1$ to P do

$$\begin{aligned} \theta_1^p &\sim p(\theta_1 | \theta_2^{p-1}, \theta_3^{p-1}, \dots, \theta_N^{p-1}, \mathbf{X}) \\ \theta_2^p &\sim p(\theta_2 | \theta_1^p, \theta_3^{p-1}, \dots, \theta_N^{p-1}, \mathbf{X}) \\ &\vdots \\ \theta_N^p &\sim p(\theta_N | \theta_1^p, \theta_2^p, \dots, \theta_{N-1}^p, \mathbf{X}) \end{aligned}$$

end

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. Therefore, the next step to implement a Gibbs' sampler is to obtain the conditional distribution of each unknown parameter of our problem. This will be done in the sequel.

3.4 Conditional distributions

For most of the unknown parameters in our problem, the expression of the conditional distribution can be obtained by observing that

$$p(\theta_q | \boldsymbol{\theta}_{-\theta_q}, \mathbf{X}) \propto p(\mathbf{X} | \boldsymbol{\theta}) p(\theta_q). \quad (13)$$

Thus, for a given parameter, 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.

3.4.1 Sources

The conditional distribution of the sources can be obtained by replacing (3) and (4) into (13):

$$p(s_{j,t} | \boldsymbol{\theta}_{-s_{j,t}}, \mathbf{X}) \propto \exp \left[- \sum_{i=1}^{n_c} \frac{1}{2\sigma_i^2} \left(\Psi_{i,j,t} s_{j,t} + \Omega_{i,j,t} \right)^2 - 0.5 p_j (s_{j,t} - \mu_j)^2 \right] \mathbb{1}_{[s_j^{\min}, s_j^{\max}]}(s_{j,t}), \quad (14)$$

where

$$\Omega_{i,j,t} = x_{i,t} - \sum_{g=1, g \neq j}^{n_s} a_{i,g} s_{g,t} - \sum_{1 \leq g < k \leq n_s, g \neq j} b_{i,g,k} s_{g,t} s_{k,t}, \quad (15)$$

and

$$\Psi_{i,j,t} = a_{i,j} + \sum_{g=1, g \neq j}^{n_s} b_{i,j,g} s_{g,t}. \quad (16)$$

After some manipulation, Equation (14) can be simplified as follows

$$p(s_{j,t} | \boldsymbol{\theta}_{-s_{j,t}}, \mathbf{X}) \propto \exp \left(- \frac{(s_{j,t} - \mu_{Post_j})^2}{2\sigma_{Post_j}^2} \right) \mathbb{1}_{[s_j^{\min}, s_j^{\max}]}(s_{j,t}), \quad (17)$$

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

$$\sigma_{L_j}^2 = \left(\sum_{i=1}^{n_c} \frac{\Psi_{ij,t}^2}{\sigma_i^2} \right)^{-1}, \quad (18)$$

$$\mu_{L_j} = \sigma_{L_j}^2 \sum_{i=1}^{n_c} \frac{\Omega_{ij,t} \Psi_{ij,t}}{\sigma_i^2}. \quad (19)$$

The conditional distribution (14) corresponds to a truncated Gaussian distribution and its simulation can be conducted by the procedure proposed in [4].

3.4.2 Parameters related to the priors attributed to the sources

The posterior distribution of the parameter p_j is given by

$$p(p_j|s_{j,:}, \mu_j) \propto p(s_{j,:}|p_j, \mu_j)p(p_j),$$

which, according to Equation (4), can be written as

$$p(p_j|s_{j,:}, \mu_j) \propto p_j^{\frac{n_d}{2}} \exp\left(-0.5p_j \sum_{t=1}^{n_d} (s_{j,t} - \mu_j)^2\right) \times \frac{\mathbb{1}_{[p_j^{min}, p_j^{max}]}(p_j)}{\Phi\left(\sqrt{p_j}(s_j^{max} - \mu_j)\right) - \Phi\left(\sqrt{p_j}(s_j^{min} - \mu_j)\right)}. \quad (20)$$

We point out here that such an expression does not assume a standard form because of the nonlinearity in the denominator of the second term. This is also the case for the distribution $p(\mu_j|s_{j,:}, p_j)$ that appears in the i.i.d. modeling. A first solution to simulate these non-standard distributions would be achieved via the Metropolis-Hastings (MH) algorithm [18]. However, the problem in such an approach is that it requires the definition of an adequate instrumental function which, very often, is not a trivial task.

In order to avoid using a MH step, we consider the alternative approach proposed in [9]. The idea here is to introduce the latent variables obtained by the following transformation:

$$l_{j,t} = \mu_j + p_j^{-1/2} \times \Phi^{-1}\left(\frac{\Phi\left(\sqrt{p_j}(s_{j,t} - \mu_j)\right) - \Phi\left(\sqrt{p_j}(s_j^{min} - \mu_j)\right)}{\Phi\left(\sqrt{p_j}(s_j^{max} - \mu_j)\right) - \Phi\left(\sqrt{p_j}(s_j^{min} - \mu_j)\right)}\right). \quad (21)$$

The interesting point here is that if $s_{j,t}$ follows a truncated Gaussian with parameters μ_j and variance $1/p_j$, then $l_{j,t}$ is distributed according to a Gaussian distribution of mean μ_j and variance $1/p_j$ (see [9] for details).

Since $l_{j,t}$ follows a Gaussian distribution, one can write

$$p(l_{j,:}|\mu_j, p_j) = \prod_{t=1}^{n_d} \sqrt{\frac{p_j}{2\pi}} \exp\left(-\frac{p_j}{2} (l_{j,t} - \mu_j)^2\right). \quad (22)$$

Using this equation and the prior distributions (5) and (6), one can show, after some simple calculations, that the new conditional distributions of μ_j and p_j are given by

$$p(\mu_j|p_j, l_{j,:}) \propto p(\mu_j)p(l_{j,:}|\mu_j, p_j) \propto \mathbb{1}_{[\mu_j^{min}, \mu_j^{max}]}(\mu_j) \times \exp\left(-\frac{p_j \cdot n_d}{2} \left(\mu_j - \frac{1}{n_s} \sum_{t=1}^{n_d} l_{j,t}\right)^2\right) \quad (23)$$

$$p(p_j|\mu_j, l_{j,:}) \propto p(p_j)p(l_{j,:}|\mu_j, p_j) \propto \mathbb{1}_{[p_j^{min}, p_j^{max}]}(p_j) \times p_j^{\frac{n_d}{2}} \exp\left(-p_j \sum_{t=1}^{n_d} (l_{j,t} - \mu_j)^2\right). \quad (24)$$

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 [4].

The original procedure of [9], 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 Equation (24). Also, the same substitution should be conducted in Equation (21) for the calculation of the latent variables $l_{j,t}$.

3.4.3 Mixing parameters

Analogously to the procedure conducted in Section (3.4.1), the conditional distributions of the mixing parameters can be obtained by substituting Equations (8) and (3) into Equation (13). It can be shown that

$$p(c_{i,m}|\boldsymbol{\theta}_{-c_{i,m}}, \mathbf{X}) \propto \exp\left(-\frac{\rho_{i,m}^L}{2}(c_{i,m} - \nu_{i,m}^L)^2\right) \mathbb{1}_{[c_{i,m}^{\min}, c_{i,m}^{\max}]}(c_{i,m}), \quad (25)$$

where

$$\rho_{i,m}^L = \sigma_i^2 \sum_{t=1}^{n_d} \bar{s}_{m,t}, \quad (26)$$

$$\nu_{i,m}^L = \frac{\sum_{t=1}^{n_d} \bar{s}_{m,t} \left(x_{i,t} - \sum_{g=1, g \neq m}^J c_{i,g} \bar{s}_{g,t}\right)}{\sum_{t=1}^{n_d} \bar{s}_{m,t}^2}. \quad (27)$$

Again, the resulting conditional distribution (Equation (25)) is a truncated Gaussian distribution and can be simulated by the technique presented in [4].

3.4.4 Noise variances

For the noise precisions $r_i = 1/\sigma_i^2$, the conditional distribution is obtained by substituting Equations (9) and (3) into Equation (13), which gives

$$p(r_i|\boldsymbol{\theta}_{-r_i}, \mathbf{X}) \propto r_i^{\frac{n_d}{2}} \exp(-0.5r_i\Theta_{i,t}) r_i^{\alpha_{r_i}-1} \exp\left(\frac{-r_i}{\beta_{r_i}}\right) \mathbb{1}_{[0,+\infty[}(r_i) \quad (28)$$

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

$$p(r_i|\boldsymbol{\theta}_{-r_i}, \mathbf{X}) \propto \exp\left(-r_i \left(0.5\Theta_{i,t} + \frac{1}{\beta_{r_i}}\right)\right) r_i^{\frac{n_d}{2} + \alpha_{r_i} - 1} \mathbb{1}_{[0,+\infty[}(r_i), \quad (29)$$

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

3.5 Algorithm description

The final algorithm is summarized in Table 1. In a first step, one must define the hyperparameters, which, as discussed before, should be done based on the available information. After that, it comes the running of the Gibbs' sampler, which is done by sampling the conditional distributions obtained in the last sections. Finally, the estimation of the unknown parameters is done by averaging the generated samples. However, as the Markov chain associated with the Gibbs' sampler takes some iterations to reach the stationary distribution, the samples generated in an initial moment, the burn-in period, must be discarded. The determination of the size of the burn-in period B is done through visual inspection.

Table 1 Bayesian source separation algorithm for LQ mixtures

1. Define hyperparameters $s_j^{min}, s_j^{max}, p_j^{min}, p_j^{max}, c_{im}^{min}, c_{im}^{max}$ and, for the i.i.d. case, μ_j^{min}, μ_j^{max} ;
2. Random initialization of the current samples θ^0 from their feasible domains;
3. Run Gibbs sampler
 - For $m = 1$ to M do
 - For $j = 1, \dots, n_s, t = 1, \dots, n_d$
 $s_{j,t}^m \sim p(s_{j,t} | \theta_{-s_{j,t}}, \mathbf{X})$ (Equation (17))
 - Define latent variables $l_{j,t}$ (Equation 21)
 - For $j = 1, \dots, n_s$
 $\mu_j^m \sim p(\mu_j | \theta_{-\mu_j}, \mathbf{L}_{j,:}, \mathbf{X})$ (Equation (23)) (only for the i.i.d. case)
 $p_j^m \sim p(p_j | \theta_{-p_j}, \mathbf{L}_{j,:}, \mathbf{X})$ (Equation (24))
 - For $i = 1, \dots, n_m, j = 1, \dots, J$
 $c_{i,j}^m \sim p(c_{i,j} | \theta_{-c_{i,j}}, \mathbf{X})$ (Equation (25))
 - For $i = 1, \dots, n_m$
 $r_i^m \sim p(r_i | \theta_{-r_i}, \mathbf{X})$ (Equation (29))
 - end
4. Infer the sources through the Bayesian MMSE estimation

$$\tilde{s}_{j,t} = \frac{1}{M-B} \sum_{m=B+1}^M s_{j,t}^m, \forall j, t, \quad (30)$$

where B denotes the number of iterations of the burn-in period.

4 Experimental Results

In order to test our method, we firstly conduct a set of simulations with synthetic data. In this test, we consider linear and LQ models. In both situations, the signal-to-interference ratio (SIR) is adopted as performance index

$$\text{SIR} = \frac{1}{n_d \cdot n_s} \sum_{j=1}^{n_s} \sum_{t=1}^{n_d} 10 \log \left(\frac{E\{s_{j,t}^2\}}{E\{(s_{j,t} - \hat{s}_{j,t})^2\}} \right), \quad (31)$$

where $\hat{s}_{j,t}$ is the estimation of the source j at time t (after scaling correction when needed). In a second moment, in order to test our method with actual data, we consider the problem of separating scanned images.

Table 2 SIR (dB) for the separation of linear mixtures.

	Situation 1	Situation 2	Situation 3
Bayesian method			
i.i.d. modeling	17.5	17.7	15.6
Bayesian method			
Markovian modeling	12.7	17.1	18.4
FastICA	20.4	17.8	11.8

4.1 Synthetic data

4.1.1 Separation of linear mixtures

We first test our method in a linear source separation problem where $n_d = 300$, $n_s = 3$, $n_c = 3$, and the mixing matrix is given by $\mathbf{A} = [1 \ 0.5 \ 0.5 ; 0.6 \ 1 \ 0.3 ; 0.8 \ 0.4 \ 1]$. The following 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 $M = 10000$ with a burn-in period of $B = 5000$.

The results in Table 2 represent the mean SIR over 50 experiments (the differences between the realization were the Markov chains initialization). Despite a expected performance degradation when the Markovian prior is used for separating i.i.d. sources, the proposed method performed satisfactorily in almost all simulations. Although the FastICA algorithm [12] gave us better results in the first two situations, its application to the third situation did not provide good 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.

4.1.2 Separation of linear-quadratic mixtures

First, we consider a situation where $n_d = 500$, $n_s = 2$ and $n_c = 2$. The original sources and the mixtures are presented in Figures. 2(a) and 2(b), respectively. The following mixing parameters were considered: $a_{1,1} = 1$, $a_{1,2} = 0.5$, $b_{1,1,2} = 0.2$, $a_{2,1} = 0.5$, $a_{2,2} = 1$, $b_{2,1,2} = 0.2$, and the SNR at each sensor was 30 dB. The hyperparameters related to the limit values of the prior distributions were³ $s_j^{min} = c_{i,m}^{min} = 0$ and $s_j^{max} = c_{i,m}^{max} = 1$. Concerning the Gibbs sampler parameters, the total number of iteration was $M = 20000$ with a burn in period of $B = 8000$. In this situation, the obtained performance indexes were SIR = 26 dB for the i.i.d. modeling and SIR = 27 dB for the Markovian modeling. We also tested the ICA method proposed in [11] which was able to provide good approximations (SIR = 22 dB). Despite the better performance, it is worth mentioning that the gains brought by our method comes at the price of a greater computational effort.

We then consider a scenario similar to the first one with the only difference that the mixing parameters are now given by $a_{1,1} = 1$, $a_{1,2} = 0.7$, $b_{1,1,2} = 0.6$, $a_{2,1} = 0.6$,

³ We set $a_{1,1}^{min} = a_{2,2}^{min} = 1$ to avoid scaling ambiguities.

$a_{2,2} = 1$, $b_{2,1,2} = 0.6$. Again, the proposed method was able to recover the sources both for i.i.d. modeling (SIR = 22 dB) and for the Markovian modeling (SIR = 23 dB). In Figure 2 the retrieved sources for the Markovian modeling are shown. Despite the noise amplification, which is expected in nonlinear systems, the estimated sources are close to the actual ones. On the other hand, in this second scenario, the ICA method proposed in [11] failed to separate the sources because the mixing coefficients violate the stability condition of the recurrent separating system.

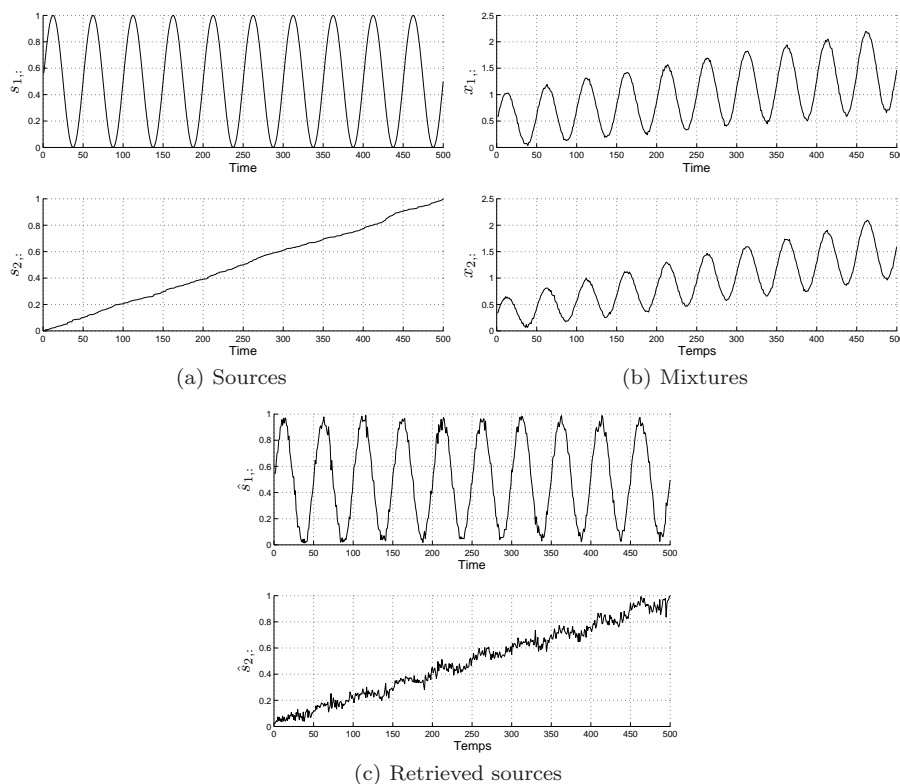
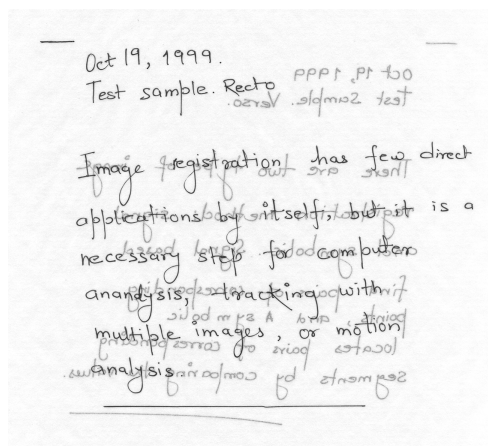


Fig. 2 Separation of LQ mixtures (Markovian modeling).

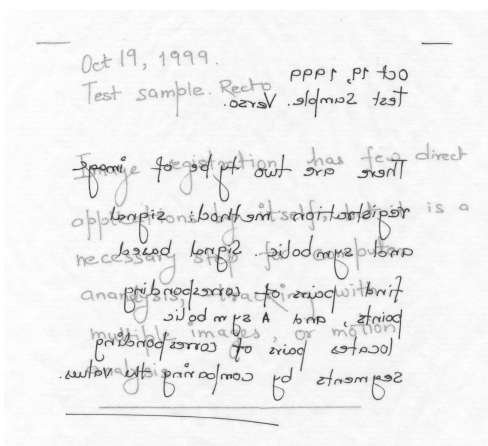
4.2 Separation of scanned images

Depending on the type of paper, image scanning of printed documents in a duplex mode (recto verso) may suffer from the so-called *show-through* effect [19]. Indeed, if the printed paper is not enough opaque, then the back-side of the image shows through in the front-side and vice-versa, causing a sort of interference in the scanned images.

This phenomenon is illustrated in Figure⁴ 3, which shows scanned versions of the original images (front-side and back-side).



(a) Front-side.



(b) Back-side (with vertical mirror inversion for pointing out its similarity with the front-side)

Fig. 3 Scanned images: the show-through effect.

Several papers [21, 16, 1] dealt with the show-through effect via the application of blind source separation methods. The problem formulation in this case is as follows: there are two sources and they correspond to the front-side and back-side images. The number of mixtures is also two and they correspond to the scanned images (recto and verso). Recently, Merrikh-Bayat et al. [16] suggested that the mixing process associated with image scanning can be satisfactorily approximated by the LQ model. Given that,

⁴ These images are available at <http://www.site.uottawa.ca/~edubois/documents/> and were considered in [21, 16].

Table 3 Estimated coefficients (scanned images)

	$a_{1,1}$	$a_{1,2}$	$a_{2,1}$	$a_{2,2}$	$b_{1,1,2}$	$b_{2,1,2}$
Bayesian method						
Linear model	0.76	0.26	0.28	0.73	–	–
Bayesian method						
LQ model	0.82	0.03	0.42	0.61	0.01	0.16

they applied the ICA-based source separation method proposed in [11] to retrieve the original images.

Aiming to illustrate the operation of our method in a scenario with actual data, we apply it to separate the scanned images of Figure 3. In particular, we focus on the sub-block of pixels (131×141) shown in the first line of Figure 4, which gives us $n_d = 18471$ samples of the mixed signals⁵. It is interesting to note here that some prior information is available in the considered situation. For example, because the scanned images are text documents, the sources are mostly white and, thus, are concentrated around⁶ one. This information can be taken into account in the definition of the hyperparameters. In our tests, for instance, we consider the i.i.d. modeling of Equation (4) with $p_j^{min} = 1$, $p_j^{max} = 2$, $\mu_j^{min} = 1$ and $\mu_j^{max} = 5$. Note that, since we are fixing the value of μ_j to the interval $[1, 5]$, the prior distributions of the sources tend to be concentrated around one.

In Figure 4, we plot the scanned images and the recovered images after applying the proposed Bayesian method for both linear and LQ cases. One can note that a slightly better performance is achieved when the LQ model is considered, although, for the back-side, there are some regions for which the linear model gives a better solution. In Table 3, we show the mixing coefficients estimated by the Bayesian methods. In all these cases, the total number of iterations was $M = 30000$ with a burn-in period of $B = 9000$. Concerning the execution time, the method took about 42 minutes⁷ in both situations (linear and LQ models).

For matter of comparison we also plot in Figure 4 the solution provided by the ICA-based method proposed in [11]. In this case, the recovered image of the front-side was close to the one provided by our Bayesian method (LQ model). Nonetheless, there is some background noise in the back-side image recovered by the ICA-based algorithm. Concerning the execution time, the LQ-ICA algorithm was much faster than our Bayesian method: it took about 4 minutes to converge.

5 Conclusion

In this work, we proposed a novel Bayesian source separation method that can be applied to linear-quadratic and linear mixing models. Our approach was based on the definition of truncated priors and, thus, is indicated for problems in which the bounds of the sources and of the mixing coefficients are known; a particular example in this context is non-negative factorization. Concerning the sources' prior modeling, we

⁵ The signals are obtained from column vectorization of the images.

⁶ The images are coded in the grayscale, where 0 and 1 represent, respectively, black and white.

⁷ The method was implemented in Matlab (Windows Vista) and the simulations were performed in a Intel Core 2 duo 2 GHz, 3048 GB RAM.

consider both an i.i.d. distribution and a Markovian one. Finally, the implementation of our method was done by setting a Gibbs' sampler and some auxiliary variables with the aim of simplifying the simulation of some conditional distributions.

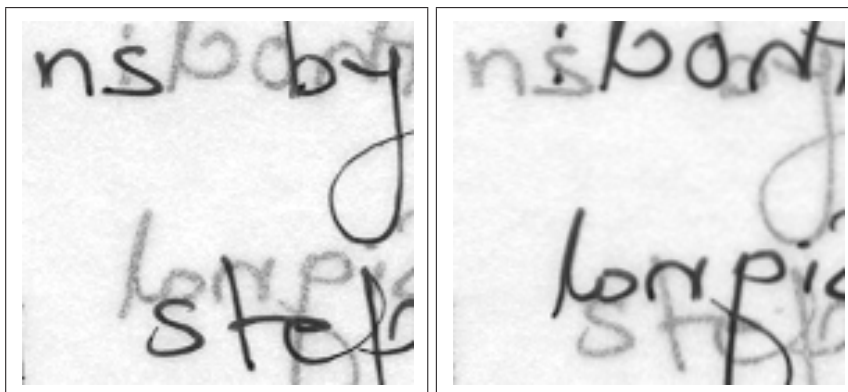
Simulations were conducted to check the viability of our proposal. In scenarios with artificial data, we observed that our method is particularly useful in some cases where existing methods fail. Furthermore, we tested our method in a real problem related to the separation of scanned images. The obtained results pointed out that our proposal is promising as it provided satisfactory solutions. Moreover, we believe that the remaining interference can be mitigated by the incorporation of a spatial filter into our Bayesian method, as done in [16]. This problem will be object of future works.

We could check that, despite the good results, the main limitation of our approach concerns its computational complexity. Indeed, as each iteration of the Gibbs' sampler performs $n_s \times n_d$ simulations of univariate random variables, the computational burden of our method may become heavy in medium and large-scale problems. A great number of works has been trying overcome such a limitation by considering methods such as the ones based on the variational approximation. This is another point that we will investigate in future works.

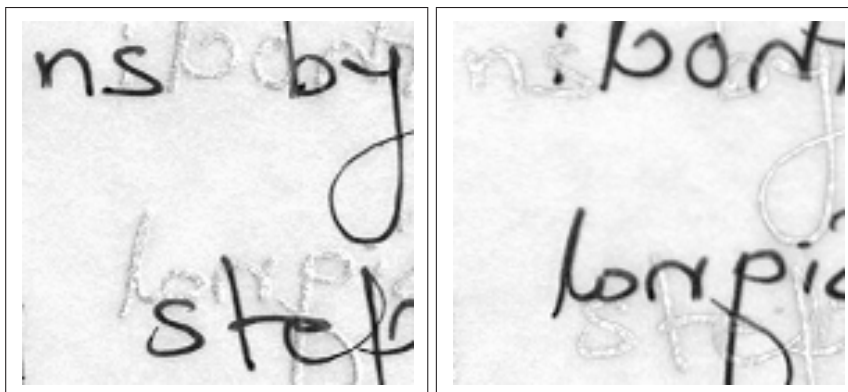
References

1. Almeida, L.B.: Separating a real-life nonlinear image mixture. *Journal of Machine Learning Research* **6**, 1199–1229 (2005)
2. Bedoya, G.: Nonlinear blind signal separation for chemical solid-state sensor arrays. Ph.D. thesis, Universitat Politècnica de Catalunya (2006)
3. Comon, P.: Independent component analysis, a new concept? *Signal Processing* **36**, 287–314 (1994)
4. Damien, P., Walker, S.G.: Sampling truncated normal, Beta, and Gamma densities. *Journal of Computational and Graphical Statistics* **10**, 206–215 (2001)
5. Deville, Y., Hosseini, S.: Recurrent networks for separating extractable-target nonlinear mixtures. part i: Non-blind configurations. *Signal Processing* **89**, 378–393 (2009)
6. Doucet, A., Wang, X.: Monte Carlo methods for signal processing. *IEEE Signal Processing Magazine* **22**, 152–170 (2005)
7. Eriksson, J., Koivunen, V.: Blind identifiability of class of nonlinear instantaneous ica models. In: *Proceedings of the XI European Signal Processing Conference (EUSIPCO)*, pp. 7–10 (2002)
8. Gilks, W., Richardson, S., Spiegelhalter, D. (eds.): *Markov chain Monte Carlo in practice*. Chapman & Hall (1995)
9. Griffiths, W.: A Gibbs' sampler for a truncated multivariate normal distribution. Tech. rep., University of Melbourne (2002)
10. Hosseini, S., Deville, Y.: 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*, pp. 289–296 (2003)
11. Hosseini, S., Deville, Y.: 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)*, pp. 694–701 (2004)
12. Hyvärinen, A., Karhunen, J., Oja, E.: *Independent component analysis*. John Wiley & Sons (2001)
13. Jutten, C., Comon, P. (eds.): *Séparation de sources 2: au-delà de l'aveugle et applications*. Hermès Science Publications (2007). In French
14. Jutten, C., Karhunen, J.: Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. *International Journal of Neural Systems* **14**, 267–292 (2004)
15. Kay, S.M.: *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall (1993)

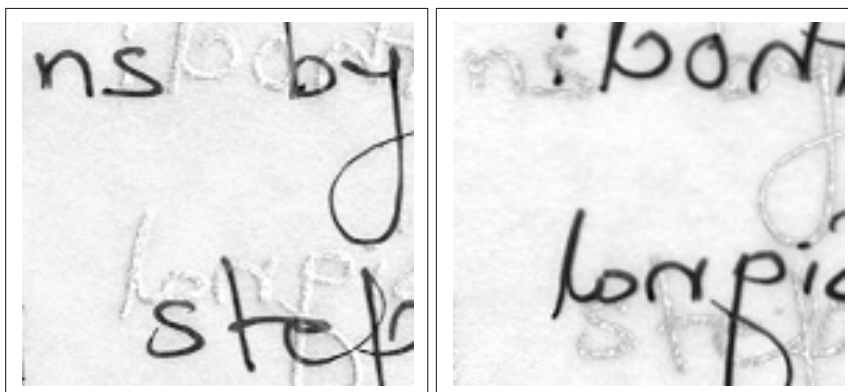
-
16. Merrikh-Bayat, F., Babaie-Zadeh, M., Jutten, C.: A nonlinear blind source separation solution for removing the show-through effect in the scanned documents. In: Proceedings of the EUSIPCO 2008 (2008)
 17. Moussaoui, S., Brie, D., Mohammad-Djafari, A., Carteret, C.: Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. *IEEE Transactions on Signal Processing* **54**, 4133–4145 (2006)
 18. Robert, C.P.: *The Bayesian Choice*. Springer (2007)
 19. Sharma, G.: Show-through cancellation in scans of duplex printed documents. *IEEE Transactions on Image Processing* **10**(5), 736–754 (2001). DOI 10.1109/83.918567
 20. Taleb, A., Jutten, C.: Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing* **47**(10), 2807–2820 (1999). DOI 10.1109/78.790661
 21. Tonazzini, A., Salerno, E., Bedini, L.: Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. *International Journal of Document Analysis* **10**, 17–25 (2007)



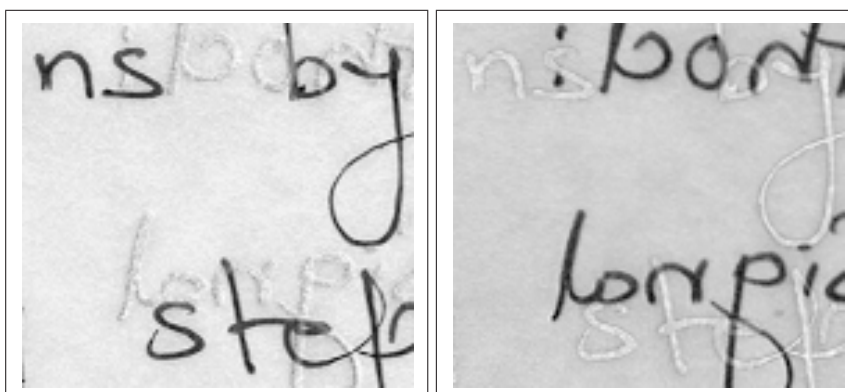
Scanned images (mixtures).



Images recovered by the proposed method (linear model).



Images recovered by the proposed method (LQ model).



Images recovered by the LQ-ICA method of [11].

Fig. 4 Results: separation of scanned images.