Bayesian separation of spectral sources under non-negativity and full additivity constraints

Nicolas Dobigeon, Saïd Moussaoui, Jean-Yves Tourneret, Cedric Carteret

▶ **To cite this version:**

Nicolas Dobigeon, Saïd Moussaoui, Jean-Yves Tourneret, Cedric Carteret. Bayesian separation of spectral sources under non-negativity and full additivity constraints. Signal Processing, 2009, Special Section: Visual Information Analysis for Security, 8 (12), pp.2657-2669. 10.1016/j.sigpro.2009.05.005 . hal-00455580

HAL Id: hal-00455580
https://hal.science/hal-00455580

Submitted on 14 Feb 2022

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.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author's version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 3104

To link to this article
URL : http://dx.doi.org/10.1016/j.sigpro.2009.05.005

To cite this version :
Dobigeon, Nicolas and Moussaoui, Said and Tourneret, Jean-Yves and Carteret, Cédric (2009) *Bayesian separation of spectral sources under non-negativity and full additivity constraints.* Signal Processing, vol. 89 (n° 12). pp. 2657-2669. ISSN 0165-1684

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr.

Bayesian separation of spectral sources under non-negativity and full additivity constraints

Nicolas Dobigeon [a,b,*], Saïd Moussaoui [c], Jean-Yves Tourneret [a], Cédric Carteret [d]

[a] University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France
[b] University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA
[c] IRCCyN/ECN - CNRS UMR 6597, 1 rue de la Noë, BP 92101, 44321 Nantes cedex 3, France
[d] University of Nancy, LCPME - CNRS UMR 7564, 405 rue de Vandoeuvre, 54600 Villers-lès-Nancy, France

Signal Processing 89 (2009) 2657–2669
doi:10.1016/j.sigpro.2009.05.005

A B S T R A C T

This paper addresses the problem of separating spectral sources which are linearly mixed with unknown proportions. The main difficulty of the problem is to ensure the full additivity (sum-to-one) of the mixing coefficients and non-negativity of sources and mixing coefficients. A Bayesian estimation approach based on Gamma priors was recently proposed to handle the non-negativity constraints in a linear mixture model. However, incorporating the full additivity constraint requires further developments. This paper studies a new hierarchical Bayesian model appropriate to the non-negativity and sum-to-one constraints associated to the sources and the mixing coefficients of linear mixtures. The estimation of the unknown parameters of this model is performed using samples obtained with an appropriate Gibbs algorithm. The performance of the proposed algorithm is evaluated through simulation results conducted on synthetic mixture data. The proposed approach is also applied to the processing of multi-component chemical mixtures resulting from Raman spectroscopy.

Keywords:
Spectral source separation
Non-negativity constraint
Full additivity constraint
Bayesian inference
Markov chain Monte Carlo methods

1. Introduction

Blind source separation (BSS) is a signal processing problem arising in many applications where one is interested by extracting signals that are observed as mixtures [1]. Pioneering works dealing with this problem have focused on the mutual statistical independence of the sources, which led to the well known independent component analysis (ICA) [2–5]. However, when the sources and the mixing coefficients have to satisfy specific constraints the resulting constrained source separation problem becomes more complicated. Therefore appropriate separation algorithms have to be developed to handle these constraints. When the sources are actually independent, ICA provides estimates of the sources and mixing coefficients which implicitly satisfy these constraints. However, these algorithms, that try to maximize the independence between the estimated sources, have not been designed for correlated sources.

Non-negativity is a physical constraint which has retained a growing attention during the last decade. For instance, Plumbley and his co-authors have addressed the case of non-negative independent sources and proposed the non-negative independent component analysis algorithm [6]. The case of both non-negative sources and non-negative mixing coefficients has been handled by using non-negative matrix factorization (NMF) algorithms [7] and a Bayesian positive source separation (BPSS) algorithm [8]. By adding a source sparsity constraint, a method ensuring the sparseness of the sources (referred to as non-negative sparse coding) has been presented in [9]. A Bayesian approach allowing one to perform the separation

* Corresponding author at: University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France.
Tel.: +335 61 58 84 82.
E-mail address: nicolas.dobigeon@enseeiht.fr (N. Dobigeon).
URL: http://www.enseeiht.fr/~dobigeon (N. Dobigeon).

of sparse sources has also been proposed in [10] using a T-student distribution. Cauchy hyperbolic priors have been introduced in [11] without considering the non-negativity constraint.

This paper addresses a source separation problem in the case of linear instantaneous mixtures where the source signals are non-negative and the mixing coefficients satisfy non-negativity and full additivity constraints. These constraints have been observed in many applications. These applications include analytical chemistry for the analysis of kinetic reactions monitored by spectroscopy [12] or image processing for the analysis of hyperspectral images [13]. A Bayesian framework appropriate to constrained source separation problem is first proposed. Prior distributions encoding non-negativity and full additivity constraints are assigned to the source signals and mixing coefficients. However, the standard Bayesian estimators resulting from these priors have no simple closed form expression. As a consequence, Markov chain Monte Carlo (MCMC) methods are proposed to generate samples according to the full posterior distribution of the unknown parameters. Estimators of the mixing coefficients and the source signals are then constructed from these generated samples. The paper is organized as follows. Section 3 defines a hierarchical Bayesian model (HBM) for the addressed constrained source separation problem. In particular, prior distributions are introduced such that they are concentrated on a simplex and they satisfy the positivity and full additivity constraints. Section 4 describes a Gibbs sampling strategy that allows one to overcome the computational complexity inherent to this HBM. Simulations conducted on synthetic mixture data are presented in Section 5. As a consequence, the performance of the proposed Bayesian estimation algorithm can be appreciated for constrained source separation problems. The interest of the proposed Bayesian approach is also illustrated by the analysis of real experimental data reported in Section 6. Conclusions and perspectives are reported in Section 7.

2. Problem statement

The linear mixing model studied in this paper assumes that the observed signal is a weighted sum of M unknown sources. In the case of spectral mixture data this model can be expressed by

$$y_{i,j} = \sum_{m=1}^{M} c_{i,m} s_{m,j} + e_{i,j}, \tag{1}$$

where $y_{i,j}$ is the observed spectrum at time/spatial index i ($i = 1,\ldots,N$) in the jth spectral band ($j = 1,\ldots,L$), N is the number of observed spectra, M is the number of mixture components and L is the number of spectral bands. The coefficient $c_{i,m}$ is the contribution of the mth component in the ith mixture and $e_{i,j}$ is an additive noise modeling measurement errors and model uncertainties. The linear mixing model can be represented by the following matrix formulation:

$$\mathbf{Y} = \mathbf{CS} + \mathbf{E}, \tag{2}$$

where the matrices $\mathbf{Y} = [y_{i,j}]_{i,j} \in \mathbb{R}^{N\times L}$, $\mathbf{C} = [c_{i,m}]_{i,m} \in \mathbb{R}^{N\times M}$, $\mathbf{S} = [s_{m,j}]_{m,j} \in \mathbb{R}^{M\times L}$ and $\mathbf{E} = [e_{i,j}]_{i,j} \in \mathbb{R}^{N\times L}$ contain, respectively, the observed spectra, the mixing coefficients, the spectral sources and the additive noise components. The noise sequences $\mathbf{e}_i = [e_{i,1},\ldots,e_{i,L}]^{\mathrm{T}}$ ($i = 1,\ldots,N$) are assumed to be independent and identically distributed (i.i.d.) according to zero-mean Gaussian distributions with covariance matrices $\sigma^2_{\mathrm{e},i}\mathbf{I}_L$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix. Note that this last assumption implies that the noise variances are the same in all the spectral bands. This reasonable assumption has been considered in many recent works including [8,11]. It could be relaxed at the price of increasing the computational complexity of the proposed algorithm [14].

In the framework of spectral data analysis, it is obvious from physical considerations that both the mixing coefficients and the source signals satisfy the following non-negativity constraints:

$$s_{m,j} \geq 0 \quad \text{and} \quad c_{i,m} \geq 0, \quad \forall (i,m,j). \tag{3}$$

Moreover, in many applications, the mixing coefficients have also to satisfy the full additivity constraint[1]:

$$\sum_{m=1}^{M} c_{i,m} = 1, \quad \forall i. \tag{4}$$

These applications include spectroscopy for the analysis of kinetic reactions [15] and hyperspectral imagery where the mixing coefficients correspond to abundance fractions [16].

The separation problem addressed in this paper consists of jointly estimating the abundances and the spectral sources under the non-negativity and the full additivity constraints. There are several methods allowing one to address the estimation problem under non-negativity constraint. These methods include NMF methods [17] and its variants [1]. From a Bayesian point of view an original model was proposed in [8] where Gamma priors are used to encode the positivity of both the sources and the mixing coefficients. This paper goes a step further by including the additivity of the mixing coefficients in the Bayesian model. Note that this constraint allows one to resolve the scale indeterminacy inherent to the linear mixing model even if non-negativity constraint is imposed. Indeed, this full additivity constraint enforces the ℓ_1 norm of each concentration vector $\mathbf{c}_i$ to be equal to $\|\mathbf{c}_i\|_1 = \sum_{m=1}^{M} |c_{i,m}| = 1$.

3. Hierarchical Bayesian model

The unknown parameter vector for the source separation problem described previously is $\boldsymbol{\Theta} = (\mathbf{S}, \mathbf{C}, \boldsymbol{\sigma}^2_{\mathrm{e}})$ where $\mathbf{S}$ and $\mathbf{C}$ are the source and concentration matrices and $\boldsymbol{\sigma}^2_{\mathrm{e}} = (\sigma_{e,1},\ldots,\sigma_{e,N})^{\mathrm{T}}$ contains the noise variances. Following the Bayesian estimation theory, the inference of the unknown parameters from the available data $\mathbf{Y}$ is based on the posterior distribution $f(\boldsymbol{\Theta}|\mathbf{Y})$, which is related to the observation likelihood $f(\mathbf{Y}|\boldsymbol{\Theta})$ and the parameter priors

[1] This condition is also referred to as *sum-to-one* constraint in the literature.

$f(\boldsymbol{\Theta})$ via the Bayes' theorem

$$f(\boldsymbol{\Theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\Theta})f(\boldsymbol{\Theta}),$$

where $\propto$ means "proportional to". The observation likelihood and the parameter priors are detailed in the sequel.

3.1. Observation likelihood

The statistical assumptions on the noise vector $\mathbf{e}_i$ and the linear mixing model described in (1) allow one to write:

$$\mathbf{y}_i|\mathbf{S}, \mathbf{c}_i, \sigma^2_{\mathrm{e},i} \sim \mathcal{N}(\mathbf{S}^{\mathsf{T}}\mathbf{c}_i, \sigma^2_{\mathrm{e},i}\mathbf{I}_L), \tag{5}$$

where $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,L}]^{\mathsf{T}}$, $\mathbf{c}_i = [c_{i,1}, \ldots, c_{i,M}]^{\mathsf{T}}$ and $\mathcal{N}(\cdot,\cdot)$ denotes the Gaussian distribution. By assuming the mutual independence between the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_N$, the likelihood of $\mathbf{Y}$ is

$$f(\mathbf{Y}|\mathbf{C}, \mathbf{S}, \boldsymbol{\sigma}^2_{\mathrm{e}}) \propto \frac{1}{\prod_{i=1}^N \sigma^L_{\mathrm{e},i}} \exp\left(-\sum_{i=1}^N \frac{\|\mathbf{y}_i - \mathbf{S}^{\mathsf{T}}\mathbf{c}_i\|_2^2}{2\sigma^2_{\mathrm{e},i}}\right), \tag{6}$$

where $\|\mathbf{x}\|_2 = (\mathbf{x}^{\mathsf{T}}\mathbf{x})^{1/2}$ stands for the standard ℓ_2 norm.

3.2. Parameter priors

3.2.1. Concentrations

In order to ensure the non-negativity and additivity constraints, the concentrations are assigned a Dirichlet prior distribution. This distribution is frequently used in statistical inference for positive variables summing to one. The Dirichlet probability density function (pdf) is defined by

$$\mathcal{D}(\mathbf{c}_i|\delta_1, \ldots, \delta_M) = \frac{\Gamma\left(\sum_{m=1}^M \delta_m\right)}{\prod_{m=1}^M \Gamma(\delta_m)} \left(\prod_{m=1}^M c_{i,m}^{\delta_m - 1}\right) \times \mathbf{1}_{\{c_{i,m}\geq 0; \sum_{m=1}^M c_{i,m}=1\}}(\mathbf{c}_i), \tag{7}$$

where $\delta_1, \ldots, \delta_M$ are the Dirichlet distribution parameters, $\Gamma(\cdot)$ is the Gamma function and $\mathbf{1}_A(.)$ denotes the indicator function defined on the set A:

$$\begin{cases} \mathbf{1}_A(\mathbf{x}) = \mathbf{1} & \text{if } \mathbf{x} \in A, \\ \mathbf{1}_A(\mathbf{x}) = \mathbf{0} & \text{otherwise.} \end{cases} \tag{8}$$

According to this prior, the expected value of the mth spectral source abundance is $\mathbb{E}[c_{i,m}] = \delta_m / \sum_{m=1}^M \delta_m$. We assume here that the abundances are *a priori* equiprobable (reflecting the absence of knowledge regarding these parameters) which corresponds to identical parameters $\{\delta_m = 1, \forall m = 1, \ldots, M\}$. An interesting reparametrization can be introduced here to handle the full additivity constraint. This reparametrization consists of splitting the concentration vectors into two parts[2]:

$$\mathbf{c}_i = [\mathbf{a}_i^{\mathsf{T}}, c_{i,M}]^{\mathsf{T}}, \tag{9}$$

where $\mathbf{a}_i^{\mathsf{T}} = [c_{i,1}, \ldots, c_{i,M-1}]$ and $c_{i,M} = 1 - \sum_{m=1}^{M-1} c_{i,m}$. It induces a new unknown parameter vector $\boldsymbol{\Theta} = \{\mathbf{A}, \mathbf{S}, \boldsymbol{\sigma}^2_{\mathrm{e}}\}$ (the same notation is used for this new parameter vector to avoid defining new variables). The proposed prior for $\mathbf{a}_i$, $i = 1, \ldots, N$ is a uniform distribution on the following simplex:

$$\mathbb{S} = \left\{\mathbf{a}_i; a_{i,m} \geq 0, \forall m = 1, \ldots, M-1, \sum_{m=1}^{M-1} a_{i,m} \leq 1\right\}. \tag{10}$$

By assuming *a priori* mutual independence between the vectors $\mathbf{a}_i$, the prior distribution for the matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]^{\mathsf{T}}$ reduces to

$$f(\mathbf{A}) \propto \prod_{i=1}^N \mathbf{1}_{\mathbb{S}}(\mathbf{a}_i). \tag{11}$$

3.2.2. Source signals

To take into account the non-negativity constraint, the two parameter Gamma distribution seems to be a good candidate thanks to its flexibility, i.e. the pdf has many different shapes depending on the values of its parameters (see [8] for motivations). This distribution encodes positivity and covers a wide range of distribution shapes.[3] The assumption of independent source samples leads to a prior distribution for each spectral source expressed as

$$f(\mathbf{s}_m|\alpha_m, \beta_m) = \left[\frac{\beta_m^{\alpha_m}}{\Gamma(\alpha_m)}\right]^L \prod_{j=1}^L [s_{m,j}^{\alpha_m - 1} \exp(-\beta_m s_{m,j}) \mathbf{1}_{\mathbb{R}^+}(s_{m,j})]. \tag{12}$$

Note that this distribution generalizes the exponential prior presented in [19,20] and has the advantage of providing a wider variety of distributions (see also Section 5.5 for additional details regarding the exponential prior). Finally, by assuming the mutual independence between the spectral sources, we obtain the following prior distribution for $\mathbf{S}$:

$$f(\mathbf{S}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{m=1}^M f(\mathbf{s}_m|\alpha_m, \beta_m), \tag{13}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_M]^{\mathsf{T}}$ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_M]^{\mathsf{T}}$ are the source hyperparameter vectors.

3.2.3. Noise variances

Conjugate priors which are here inverse Gamma (IG) distributions are chosen for each noise variance $\sigma^2_{\mathrm{e},i}$ [21, Appendix A]:

$$\sigma^2_{\mathrm{e},i}|\rho_{\mathrm{e}}, \psi_{\mathrm{e}} \sim \mathcal{IG}\left(\frac{\rho_{\mathrm{e}}}{2}, \frac{\psi_{\mathrm{e}}}{2}\right), \tag{14}$$

where $\mathcal{IG}(a, b)$ denotes the IG distribution with parameters a and b. Note that choosing conjugate distributions as priors makes the Bayesian analysis easier [22, Chapter 2]. By assuming the independence between the noise variances $\sigma^2_{\mathrm{e},i}$, $i = 1, \ldots, N$, the prior distribution of $\boldsymbol{\sigma}^2_{\mathrm{e}}$ is

$$f(\boldsymbol{\sigma}^2_{\mathrm{e}}|\rho_{\mathrm{e}}, \psi_{\mathrm{e}}) = \prod_{i=1}^N f(\sigma^2_{\mathrm{e},i}|\rho_{\mathrm{e}}, \psi_{\mathrm{e}}). \tag{15}$$

The hyperparameter ρ_{e} will be fixed to $\rho_{\mathrm{e}} = 2$ whereas ψ_{e} is an adjustable hyperparameter as in [23].

[2] From a practical point of view, it is interesting to note that the component of $\mathbf{a}_i$ to be discarded will be randomly chosen at each iteration of the algorithm introduced in Section 4.

[3] A more general model would consist of using a mixture of Gamma distributions as in [18]. However, the Gamma distribution which leads to a simple Bayesian model has been preferred here for simplicity.

3.3. Hyperparameter priors

The hyperparameter vector associated with the prior distributions previously introduced is $\boldsymbol{\Phi} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \psi_e\}$. Obviously, the BSS performances depend on the values of these hyperparameters. In this paper, we propose to estimate them within a fully Bayesian framework by assigning them non-informative prior distributions. This naturally introduces a second level of hierarchy within the Bayes' paradigm, resulting in a so-called hierarchical Bayesian model [24, p. 299].

3.3.1. Source hyperparameters

Conjugate exponential densities with parameters λ_{α_m} have been chosen as prior distributions for the hyperparameters α_m [21, Appendix A]:

$$\alpha_m|\lambda_{\alpha_m} \sim \mathcal{E}(\lambda_{\alpha_m}). \tag{16}$$

Conjugate Gamma distributions with parameters $(\alpha_{\beta_m}, \beta_{\beta_m})$ have been elected as prior distributions for the hyperparameters β_m [21, Appendix A]:

$$\beta_m|\alpha_{\beta_m}, \beta_{\beta_m} \sim \mathcal{G}(\alpha_{\beta_m}, \beta_{\beta_m}). \tag{17}$$

The fixed hyperparameters $\{\alpha_{\beta_m}, \beta_{\beta_m}, \lambda_{\alpha_m}\}_m$ have been chosen to obtain flat priors, i.e. with large variances: $\alpha_{\beta_m} = 2, \beta_{\beta_m} = 10^{-2}$ and $\lambda_{\alpha_m} = 10^{-2}$.

3.3.2. Noise variance hyperparameters

The prior for ψ_e is a non-informative Jeffreys' prior which reflects the lack of knowledge regarding this hyperparameter

$$f(\psi_e) \propto \frac{1}{\psi_e} \mathbf{1}_{\mathbb{R}^+}(\psi_e). \tag{18}$$

Assuming the independence between the hyperparameters, the prior distribution of the hyperparameter vector $\boldsymbol{\Phi} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \psi_e\}$ can be written as

$$\begin{aligned} f(\boldsymbol{\Phi}) \propto & \prod_{m=1}^{M} [\lambda_{\alpha_m} \exp(-\lambda_{\alpha_m}\alpha_m)\mathbf{1}_{\mathbb{R}^+}(\alpha_m)] \\ & \times \prod_{m=1}^{M} [\beta_m^{\alpha_{\beta_m}-1} \exp(-\beta_{\beta_m}\beta_m)\mathbf{1}_{\mathbb{R}^+}(\beta_m)] \frac{1}{\psi_e} \mathbf{1}_{\mathbb{R}^+}(\psi_e). \end{aligned} \tag{19}$$

3.4. Posterior distribution of $\boldsymbol{\Theta}$

The posterior distribution of the unknown parameter vector $\boldsymbol{\Theta} = \{\mathbf{A}, \mathbf{S}, \boldsymbol{\sigma}_e^2\}$ can be computed from the following hierarchical structure:

$$f(\boldsymbol{\Theta}|\mathbf{Y}) \propto \int f(\mathbf{Y}|\boldsymbol{\Theta}) f(\boldsymbol{\Theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi})\, d\boldsymbol{\Phi}, \tag{20}$$

where $f(\mathbf{Y}|\boldsymbol{\Theta})$ and $f(\boldsymbol{\Phi})$ have been defined in (6) and (19). Moreover, by assuming the independence between $\mathbf{A}$, $\mathbf{S}$ and $\boldsymbol{\sigma}_e^2$, the following result can be obtained:

$$f(\boldsymbol{\Theta}|\boldsymbol{\Phi}) = f(\mathbf{A}) f(\mathbf{S}|\boldsymbol{\sigma}_s^2) f(\boldsymbol{\sigma}_e^2|\rho_e, \psi_e), \tag{21}$$

where $f(\mathbf{A})$, $f(\mathbf{S}|\boldsymbol{\sigma}_s^2)$ and $f(\boldsymbol{\sigma}_e^2|\rho_e, \psi_e)$ have been defined previously. This hierarchical structure, depicted in the directed acyclic graph (DAG) of Fig. 1, allows one to integrate out the hyperparameters ψ_e and $\boldsymbol{\beta}$ from the joint distribution $f(\boldsymbol{\Theta}, \boldsymbol{\Phi}|\mathbf{Y})$, yielding

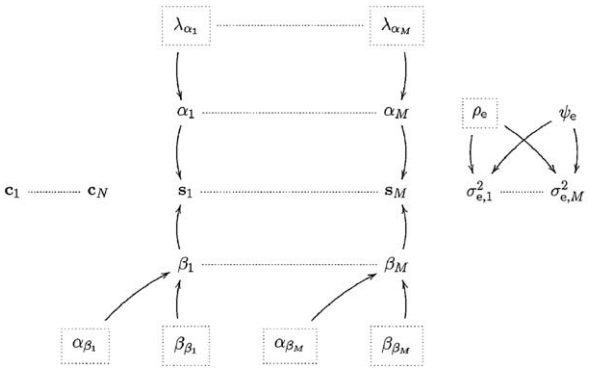

Fig. 1. DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes).

$$\begin{aligned} f(\mathbf{A}, \mathbf{S}, \boldsymbol{\sigma}_e^2, \boldsymbol{\alpha}|\mathbf{Y}) \propto & \prod_{i=1}^{N} \left[\frac{\mathbf{1}_{\mathbb{S}}(\mathbf{a}_i)}{\sigma_{e,i}^{L+2}} \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{S}^T\mathbf{c}_i\|^2}{2\sigma_{e,i}^2} \right) \right] \\ & \times \prod_{m=1}^{M} \left[\frac{\Gamma(L\alpha_m + \alpha_{\beta_m} + 1)}{(\sum_{j=1}^{L} s_{m,j} + \beta_{\beta_m})^{L\alpha_j + \alpha_{\beta_m} + 1}} \right] \\ & \times \prod_{m=1}^{M} \left[\left(\prod_{j=1}^{L} \frac{s_{m,j}}{\Gamma(\alpha_m)} \right)^{\alpha_m - 1} \mathbf{1}_{\mathbb{R}_+^L}(\mathbf{s}_m) \right]. \end{aligned} \tag{22}$$

The posterior distribution in (22) is clearly too complex to derive the classical Bayesian estimators of the unknown parameters, such as the minimum mean square error (MMSE) estimator or the maximum *a posteriori* (MAP) estimator. To overcome the difficulty, it is quite common to make use of MCMC methods to generate samples asymptotically distributed according to the exact posterior of interest [24]. The simulated samples are then used to approximate integrals by empirical averages for the MMSE estimator and to estimate the maximum of the posterior distribution for the MAP estimator. The next section proposes a Gibbs sampling strategy for the BSS of the spectral mixtures under the positivity and full additivity constraints.

4. Gibbs sampler

The Gibbs sampler is an iterative sampling strategy that consists of generating samples (denoted $\widetilde{\cdot}^{(t)}$) distributed according to the conditional distribution of each parameter. This section describes a Gibbs sampling strategy generating samples $(\widetilde{\mathbf{A}}^{(t)}, \widetilde{\mathbf{S}}^{(t)}, \{\widetilde{\boldsymbol{\sigma}}^2\}^{(t)}, \widetilde{\boldsymbol{\alpha}}^{(t)})$ asymptotically distributed according to (22). The main steps of the algorithm (denoted as Algorithm 1) are detailed from Sections 4.1–4.3.

Algorithm 1. Gibbs sampling algorithm for blind spectral source separation

- *Initialization*:
  (1) sample the hyperparameter $\widetilde{\psi}_e^{(0)}$ from the pdf in (18),
  (2) for $i = 1, \ldots, N$, sample the noise variance $\{\widetilde{\sigma}_{e,i}^2\}^{(0)}$ from the pdf in (14),

(3) for $m = 1, \ldots, M$, sample the hyperparameter $\widetilde{\alpha}_m^{(0)}$ from the pdf in (16),
(4) for $m = 1, \ldots, M$, sample the hyperparameter $\widetilde{\beta}_m^{(0)}$ from the pdf in (17),
(5) for $m = 1, \ldots, M$, sample the source spectrum $\widetilde{\mathbf{s}}_m^{(t)}$ from the pdf in (12).
(6) Set $t \leftarrow 1$,

- *Iterations*: for $t = 1, 2, \ldots,$ do
  1. for $i = 1, \ldots, N$, sample the concentration vector $\widetilde{\mathbf{a}}_i^{(t)}$ from the pdf in (25),
  2. sample the hyperparameter $\widetilde{\psi}_{\mathrm{e}}^{(t)}$ from the pdf in (26),
  3. for $i = 1, \ldots, N$, sample the noise variance $\{\widetilde{\sigma}_{\mathrm{e},i}^2\}^{(t)}$ from the pdf in (27),
  4. for $m = 1, \ldots, M$, sample the hyperparameter $\widetilde{\alpha}_m^{(t)}$ from the pdf in (28),
  5. for $m = 1, \ldots, M$, sample the hyperparameter $\widetilde{\beta}_m^{(t)}$ from the pdf in (29),
  6. for $m = 1, \ldots, M$, sample the source spectrum $\widetilde{\mathbf{s}}_m^{(t)}$ from the pdf in (30).
  7. Set $t \quad t+1$.

4.1. Generation according to $f(\mathbf{A}|\mathbf{S}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{Y})$

Straightforward computations yield for each observation

$$f(\mathbf{a}_i|\mathbf{S}, \sigma_{\mathrm{e},i}^2, \mathbf{y}_i) \propto \exp\left[-\frac{(\mathbf{a}_i - \boldsymbol{\mu}_i)^{\mathsf{T}} \boldsymbol{\Lambda}_i^{-1} (\mathbf{a}_i - \boldsymbol{\mu}_i)}{2}\right] \mathbf{1}_{\mathbb{T}}(\mathbf{a}_i), \quad (23)$$

where

$$\begin{cases} \boldsymbol{\Lambda}_i = \left[\dfrac{1}{\sigma_{\mathrm{e},i}^2} (\mathbf{S}_{-M,\cdot}^{\mathsf{T}} - \mathbf{s}_M \mathbf{u}^{\mathsf{T}})^{\mathsf{T}} (\mathbf{S}_{-M,\cdot}^{\mathsf{T}} - \mathbf{s}_M \mathbf{u}^{\mathsf{T}})\right]^{-1}, \\ \boldsymbol{\mu}_i = \boldsymbol{\Lambda}_i \left[\dfrac{1}{\sigma_{\mathrm{e},i}^2} (\mathbf{S}_{-M,\cdot}^{\mathsf{T}} - \mathbf{s}_M \mathbf{u}^{\mathsf{T}})^{\mathsf{T}} (\mathbf{y}_i - \mathbf{s}_M)\right], \end{cases} \quad (24)$$

with $\mathbf{u} = [1, \ldots, 1]^{\mathsf{T}} \in \mathbb{R}^{M-1}$ and where $\mathbf{S}_{-M,\cdot}$ denotes the matrix $\mathbf{S}$ from which the Mth row has been removed. As a consequence, $\mathbf{a}_i|\mathbf{S}, \sigma_{\mathrm{e},i}^2, \mathbf{y}_i$ is distributed according to a truncated Gaussian distribution on the simplex $\mathbb{S}$:

$$\mathbf{a}_i|\mathbf{S}, \sigma_{\mathrm{e},i}^2, \mathbf{y}_i \sim \mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i). \quad (25)$$

When the number M of spectral sources is relatively small, the generation of $\mathbf{a}_i|\mathbf{S}, \sigma_{\mathrm{e},i}^2, \mathbf{y}_i$ can be achieved using a standard Metropolis Hastings (MH) step. By choosing the Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)$ as proposal distribution for this MH step, the acceptance ratio of the MH algorithm reduces to 1 if the candidate is inside the simplex $\mathbb{S}$ and 0 otherwise. For higher dimension problems, the acceptance ratio of the MH algorithm can be small, leading to poor mixing properties. In such cases, an alternative strategy based on a Gibbs sampler can be used (see [25,26]).

4.2. Generation according to $f(\boldsymbol{\sigma}_{\mathrm{e}}^2|\mathbf{A}, \mathbf{S}, \mathbf{Y})$

To sample according to $f(\boldsymbol{\sigma}_{\mathrm{e}}^2|\mathbf{A}, \mathbf{S}, \mathbf{Y})$, it is very convenient to generate samples from $f(\boldsymbol{\sigma}_{\mathrm{e}}^2, \psi_{\mathrm{e}}|\mathbf{A}, \mathbf{S}, \mathbf{Y})$ by using the two following steps:

4.2.1. Generation according to $f(\psi_{\mathrm{e}}|\boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{A}, \mathbf{S}, \mathbf{Y})$

The conditional distribution is expressed as the following IG distribution:

$$\psi_{\mathrm{e}}|\boldsymbol{\sigma}_{\mathrm{e}}^2, \rho_{\mathrm{e}} \sim \mathcal{IG}\left(\frac{N\rho_{\mathrm{e}}}{2}, \frac{1}{2}\sum_{i=1}^{N} \frac{1}{\sigma_{\mathrm{e},i}^2}\right). \quad (26)$$

4.2.2. Generation according to $f(\boldsymbol{\sigma}_{\mathrm{e}}^2|\psi_{\mathrm{e}}, \mathbf{A}, \mathbf{S}, \mathbf{Y})$

After a careful examination of $f(\boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{A}, \psi_{\mathrm{e}}|\mathbf{S}, \mathbf{Y})$, it can be deduced that the conditional distribution of the noise variance in each observation spectrum is the following IG distribution:

$$\sigma_{\mathrm{e},i}^2|\psi_{\mathrm{e}}, \mathbf{a}_i, \mathbf{S}, \mathbf{y}_i \sim \mathcal{IG}\left(\frac{\rho_{\mathrm{e}} + L}{2}, \frac{\psi_{\mathrm{e}} + \|\mathbf{y}_i - \mathbf{S}\mathbf{c}_i^{\mathsf{T}}\|^2}{2}\right). \quad (27)$$

4.3. Generation according to $f(\mathbf{S}|\mathbf{A}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{Y})$

This generation can be achieved thanks to the three following steps, as in [8].

4.3.1. Generation according to $f(\boldsymbol{\alpha}|\boldsymbol{\beta}, \mathbf{S}, \mathbf{A}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{Y})$

From the joint distribution $f(\mathbf{A}, \mathbf{S}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{Y})$, we can express the posterior distribution of α_m $(m = 1, \ldots, M)$ as

$$f(\alpha_m|\mathbf{s}_m, \beta_m) \propto \prod_{j=1}^{L} \left[\frac{\beta_m^{\alpha_m}}{\Gamma(\alpha_m)} s_{m,j}^{\alpha_m}\right] e^{-\lambda_{\alpha_m} \alpha_m} \mathbf{1}_{\mathbb{R}^+}(\alpha_m). \quad (28)$$

This posterior is not easy to simulate as it does not belong to a known distribution family. Therefore, an MH step is required to generate samples $\widetilde{\alpha}_m^{(t)}$ distributed according to (28). The reader is invited to consult [8] for more details regarding the choice of the instrumental distribution in order to obtain a high acceptance rate for the MH algorithm.

4.3.2. Generation according to $f(\boldsymbol{\beta}|\boldsymbol{\alpha}, \mathbf{S}, \mathbf{A}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{Y})$

Similarly, the posterior distribution of the hyperparameter vector $\boldsymbol{\beta}$ can be determined by looking at the joint distribution $f(\mathbf{A}, \mathbf{S}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{Y})$. In this case, the posterior distribution of the individual hyperparameter β_m $(m = 1, \ldots, M)$ is the following Gamma distribution:

$$\beta_m|\alpha_m, \mathbf{s}_m \sim \mathcal{G}\left(1 + L\alpha_m + \alpha_{\alpha_m}, \sum_{j=1}^{L} s_{m,j} + \beta_{\alpha_m}\right). \quad (29)$$

4.3.3. Generation according to $f(\mathbf{S}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{Y})$

Finally, the posterior distribution of the source observed in the jth spectral band is

$$f(s_{m,j}|\alpha_m, \beta_m, \mathbf{A}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \mathbf{Y}) \propto s_{m,j}^{\alpha_m - 1} \mathbf{1}_{\mathbb{R}^+}(s_{m,j}) \times \exp\left[-\frac{(s_{m,j} - \mu_{m,j})^2}{2\delta_m^2} - \beta_m s_{m,j}\right], \quad (30)$$

with

$$\begin{cases} \delta_m^2 = \left[\sum\limits_{i=1}^{N} \dfrac{c_{i,m}^2}{\sigma_{\mathrm{e},i}^2}\right]^{-1}, \\ \mu_{m,j} = \dfrac{1}{\delta_m^2} \sum\limits_{i=1}^{N} \dfrac{c_{i,m} \varepsilon_{i,j}^{(-m)}}{\sigma_{\mathrm{e},i}^2}, \end{cases} \quad (31)$$

where $\varepsilon_{i,j}^{(-m)} = y_{i,j} - \sum_{k\neq m} c_{i,k}s_{k,j}$. The generation of samples distributed according to (30) is achieved by using an MH algorithm whose proposal is a positive truncated normal distribution [8]. The generation according to the positive truncated Gaussian distribution can be achieved thanks to an accept–reject scheme with multiple proposal distributions (see [25,27,28] for details).

5. Experimental results with synthetic data

This section presents some experiments performed on synthetic data to illustrate the performance of the proposed Bayesian spectral unmixing algorithm.

5.1. Mixture synthesis

The spectral sources have been simulated to get signals similar to absorption spectroscopy data. Each spectrum is obtained as a superposition of Gaussian and Lorentzian functionals with randomly chosen parameters (location, amplitude and width) [8]. Fig. 2(left) shows an example of $M = 3$ source signals of $L = 1000$ spectral bands. For this application, a "spectral" band corresponds to a given value of the wavelength λ (expressed in nanometers). The mixing coefficients have been chosen to obtain evolution profiles similar to component abundance variation in a kinetic reaction, as depicted in Fig. 2 (top, right). The abundance fraction profiles have been simulated for $N = 10$ observation times, which provides $N = 10$ observation spectra. An i.i.d. Gaussian sequence has been added to each observation with appropriate standard deviation to have a signal to noise ratio (SNR) equal to 20 dB. One typical realization of the observed spectra is shown in Fig. 2 (bottom, right).

5.2. Separation with non-negativity and full additivity constraints

Fig. 3 summarizes the result of a Monte Carlo simulation with 100 runs where the mixing matrix has been kept unchanged, while new sources and noise sequences have been generated at each run. Fig. 3a shows a comparison between the true concentrations (cross) and their MMSE estimates (circles) obtained for a Markov chain of $N_{\mathrm{MC}} = 1000$ iterations including $N_{\mathrm{b-i}} = 200$ burn-in iterations. These estimates have been computed according to the MMSE principle ($i = 1, \ldots, M$):

$$\hat{\mathbf{a}}_i = \frac{1}{N_r}\sum_{t=1}^{N_r} \widetilde{\mathbf{a}}_i^{(N_{\mathrm{b-i}})} + t, \tag{32}$$

where $N_r = N_{\mathrm{MC}} - N_{\mathrm{b-i}}$ is the number of iterations used for the estimation. The estimated abundances are clearly in good agreement with the actual abundances and the estimates satisfy the positivity and full additivity constraints. By comparing Figs. 2 (left) and 3 (top), it can be observed that the source signals have also been correctly estimated.

It is interesting to note that the proposed algorithm generates samples distributed according to the posterior distribution of the unknown parameters $f(\mathbf{A}, \mathbf{S}, \boldsymbol{\sigma}_{\mathrm{e}}^2, \boldsymbol{\alpha}|\mathbf{Y})$. These samples can be used to obtain the posterior distributions of the concentrations or the source spectra.

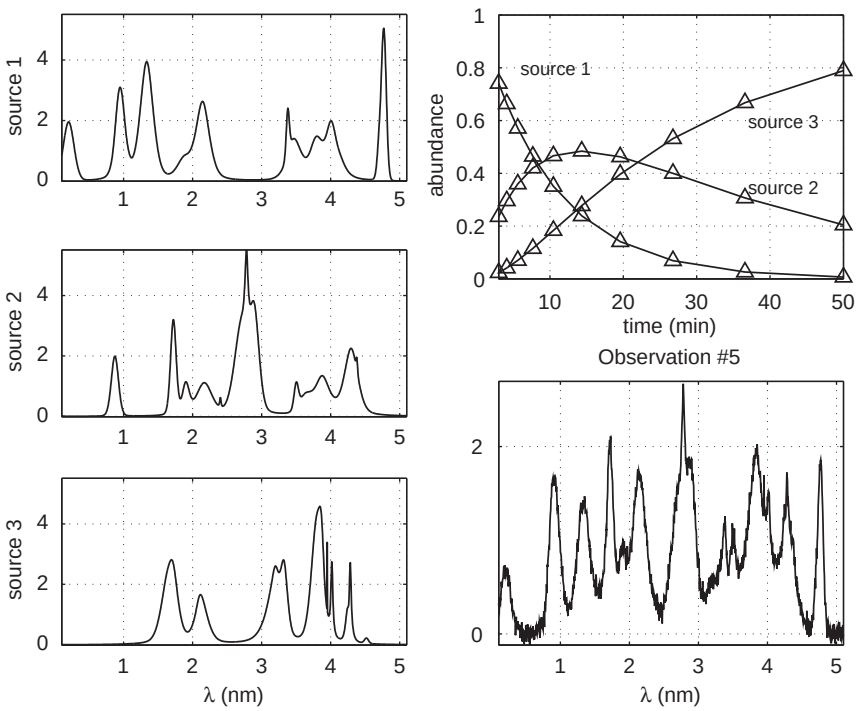

Fig. 2. Top: example of $M = 3$ simulated spectral sources where the x-axis corresponds to the wavelength expressed in nm and the y-axis corresponds to the absorbance of the spectra. Bottom, left: abundance evolution profiles. Bottom, right: one typical realization of the observed spectra.

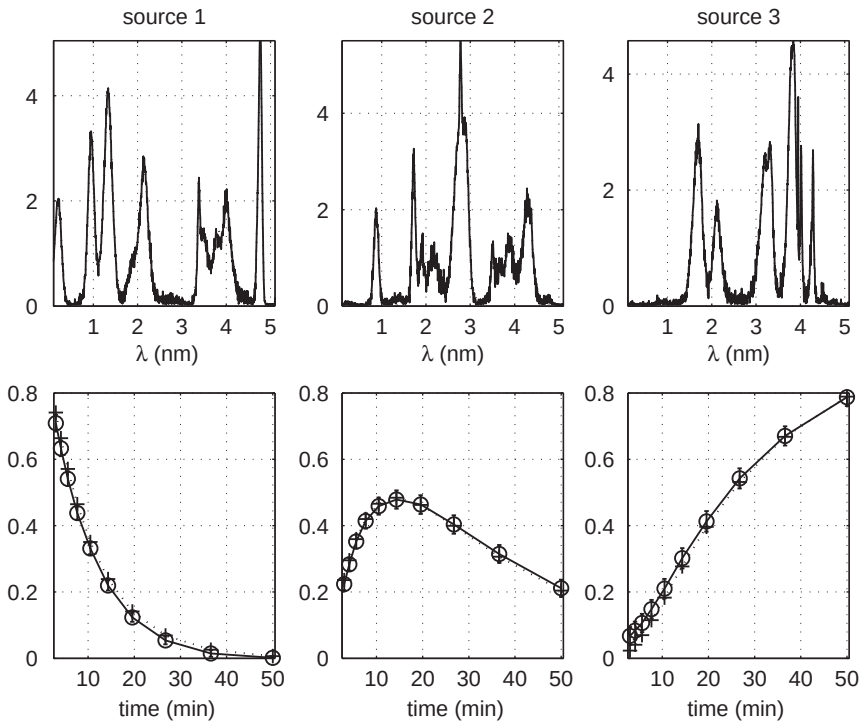

Fig. 3. Top: simulated (dotted) and estimated (continuous line) source spectra. Bottom: simulated values (cross) and MMSE estimates (circles) of the abundances. Error bars indicate the estimated 95% confidence intervals from the simulated Markov chain.

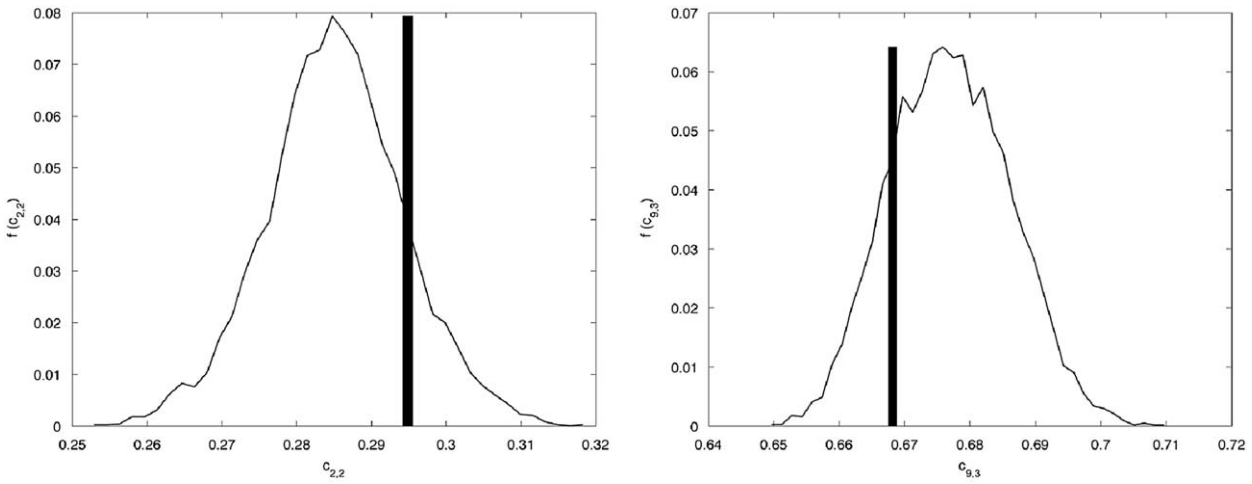

Fig. 4. Left (resp. right): posterior distribution of the concentration of the second (resp. third) spectral component in the mixture observed at index time $i = 2$ (resp. $i = 9$). The actual values appear as black bars.

As an example, typical posterior distributions for two mixing coefficients are depicted in Fig. 4. These posteriors are in good agreement with the theoretical posterior distributions in (25), i.e. truncated Gaussian distributions.

5.3. Monitoring sampler convergence

An important issue when using MCMC methods is convergence monitoring. The Gibbs sampler detailed in Section 4 generates random samples $(\widetilde{\mathbf{A}}^{(t)}, \widetilde{\mathbf{S}}^{(t)}, \widetilde{\boldsymbol{\sigma}}^{2(t)}, \widetilde{\boldsymbol{\alpha}}^{(t)})$ asymptotically distributed according to the posterior distribution in (22). The quantities of interest, i.e. the concentration coefficients and the source spectra, are then approximated by empirical averages according to (32). However, two essential parameters have to be tuned: the length N_{MC} of the constructed Markov chain and the length $N_{\mathrm{b-i}}$ of the burn-in period, i.e. the number of simulated samples to be discarded before computing the averages. This section reports some works conducted to

ensure the convergence of the proposed algorithm and the accuracy of the estimation for the unknown parameters.

First, the burn-in period $N_{\mathrm{b-i}} = 200$ has been determined thanks to the popular potential scale reduction factor (PSRF). The PSRF was introduced by Gelman and Rubin [29] and has been widely used in the signal processing literature (see for instance [30–32]). It consists of running several parallel Markov chains and computing the following criterion:

$$\hat{\rho} = \left(1 - \frac{1}{N_r}\right)\left[1 + \frac{1}{(N_r - 1)}\frac{B(\kappa)}{W(\kappa)}\right], \tag{33}$$

where W and B are the within and between-sequence variances of the parameter κ, respectively. Different choices for κ can be used for our source separation problem. Here, we consider the parameters $\sigma^2_{\mathrm{e},i}$ $(i = 1, \ldots, N)$ as recommended in [33]. Table 1 shows the PSRF obtained for the $N = 10$ observation times computed from $M = 10$ Markov chains. All these values of $\sqrt{[]}\hat{\rho}$ confirm the good convergence of the sampler since a recommendation for convergence assessment is $\sqrt{[]}\hat{\rho} < 1.2$ [34, p. 332].

The Markov chain convergence can also be monitored by a graphical supervision of the generated samples of the noise variances. As an illustration, the outputs of 10 Markov chains for one of the parameter $\sigma^2_{\mathrm{e},i}$ are depicted in Fig. 5. All the generated samples converge to a similar value after a short burn-in period (200 iterations, in this example).

Once the number of burn-in iterations has been fixed, the number of iterations necessary to obtain accurate estimates of the unknown parameters via (32) has to be adjusted. This paper proposes to evaluate N_r with appropriate graphical evaluations (see [35, p. 28] for motivations). Fig. 6 shows the reconstruction error associated to the different spectra defined as

$$e_r^2(p) = \frac{1}{NL}\sum_{i=1}^{N} \|\mathbf{y}_i - (\hat{\mathbf{c}}_i^{(p)}\hat{\mathbf{S}}^{(p)})^{\mathsf{T}}\|^2, \tag{34}$$

where $\hat{\mathbf{c}}_i^{(p)}$ and $\hat{\mathbf{S}}^{(p)}$ are the MMSE estimates of the abundance vector $\mathbf{c}_i$ and the source matrix $\mathbf{S}$ computed after $N_{\mathrm{b-i}} = 200$ burn-in iterations and $N_r = p$ iterations. The number of iterations N_r required to compute the empirical averages following the MMSE estimator (32) can be fixed to ensure the reconstruction error is below a predefined threshold. Fig. 6 shows that a number of iterations $N_r = 500$ is sufficient to ensure a good estimation of the quantities of interest $\mathbf{C}$ and $\mathbf{S}$.

Table 1
Potential scale reduction factors of $\sigma^2_{\mathrm{e},i}$ computed from $M = 10$ Markov chains.

Obs. index	$\sqrt{[]}\hat{\rho}$
1	1.0048
2	1.0013
3	1.0027
4	0.9995
5	1.0097
6	1.0078
7	1.0001
8	0.9994
9	1.0080
10	1.0288

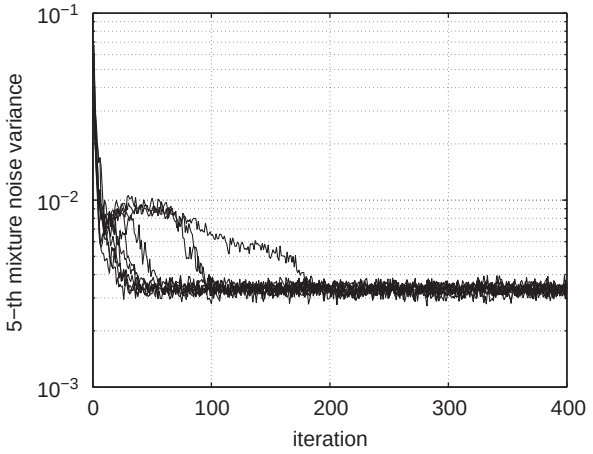

Fig. 5. Outputs of $M = 10$ Markov chains for the parameter $\sigma^2_{\mathrm{e},5}$.

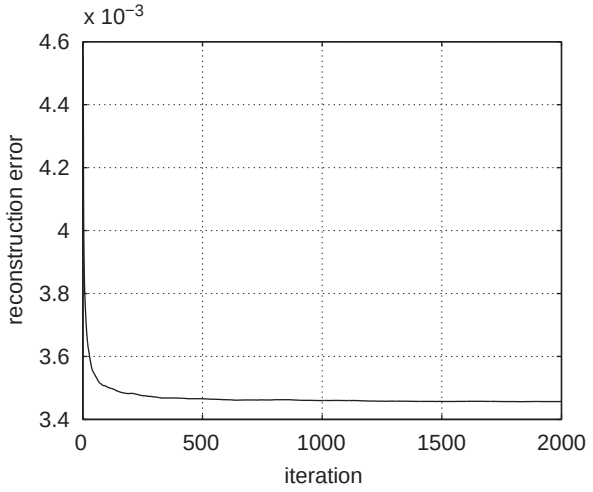

Fig. 6. Evolution of the reconstruction error with respect to the iteration number (with a burn-in of $N_{\mathrm{b-i}} = 200$ iterations).

5.4. Comparison with other BSS algorithms

The proposed Bayesian approach has been compared with other standard BSS methods. Synthetic mixtures have been processed by the non-negative ICA (NN-ICA) algorithm proposed by Plumbley and Oja [36], the iterative NMF method described in [7] and the Bayesian positive source separation algorithm introduced in [8].

All these methods do not include the full additivity constraint. To evaluate the relevance of this additional constraint, *ad hoc* re-scaled versions of these methods have also been considered. Simulations have been conducted by applying the 4 algorithms using 100 Monte Carlo runs, each run being associated to a randomly generated source. Table 2 shows the normalized mean square errors (NMSEs) for the estimated sources and

Table 2
Estimation performance for different BSS algorithms (100 Monte Carlo runs).

	NMSE (**S**)	NMSE (**C**)	Av. diss (**S**) (%)	Time (min)
Proposed approach	0.0071	0.0024	13.9	44
BPSS	0.0121	0.0025	13.4	45
Re-scaled BPSS	0.0126	0.0023	13.4	45
NN-ICA	0.0613	0.0345	20.0	3
Re-scaled NN-ICA	0.0602	0.0384	19.4	3
NMF	0.2109	1.9149	22.6	1
Re-scaled NMF	0.0575	0.0496	24.5	1

abundance matrices as defined in [37]

$$\text{NMSE}\ (\mathbf{S}) = \sum_{m=1}^{M} \frac{\|\mathbf{s}_m - \hat{\mathbf{s}}_m\|^2}{\|\mathbf{s}_m\|^2},$$
$$\text{NMSE}\ (\mathbf{C}) = \sum_{i=1}^{N} \frac{\|\mathbf{c}_i - \hat{\mathbf{c}}_i\|^2}{\|\mathbf{c}_i\|^2}. \tag{35}$$

In addition, the estimation performances have been compared in terms of dissimilarity. Denoted $\text{diss}(\cdot,\cdot)$, it measures how the estimated source spectrum differs from the reference one [38] and is defined by

$$\text{diss}(\mathbf{s}_m, \hat{\mathbf{s}}_m) = \sqrt{[]}1 - \text{corr}(\mathbf{s}_m, \hat{\mathbf{s}}_m)^2, \tag{36}$$

where $\text{corr}(\mathbf{s}_m, \hat{\mathbf{s}}_m)$ is the correlation coefficient between $\mathbf{s}_m$ and its estimate $\hat{\mathbf{s}}_m$. Consequently the average dissimilarity over the M sources is reported in Table 2.

These results demonstrate that an *ad hoc* re-scaling of the results obtained by NMF techniques is not always an efficient means to improve the estimation performance. Indeed, the *ad hoc* re-scaled version of NMF provides lower MSEs than the corresponding standard algorithms. On the other hand, this constraint does not significantly improve the NN-ICA or the BPSS algorithms. As far as the Bayesian algorithms are concerned, they clearly provide better estimation performance than the non-Bayesian approaches. However, the proposed fully constrained algorithm clearly outperforms the two BPSS algorithms, especially regarding the source estimation.

The computation times required by each of these algorithms are also reported in Table 2 for a MATLAB implementation on a 2.2 GHz Intel Core 2. This shows that the complexities of the proposed method and BPSS algorithms are quite similar and higher than the complexities of the NN-ICA and MNF algorithms. This seems to be the price to pay to obtain significantly better estimation performances.

5.5. Modified Bayesian models with other source priors

As it has been mentioned previously, several distributions can be chosen as priors for the source spectra, provided these distributions have positive supports. The previous HBM studied in Section 3.2.2 is based on Gamma distributions as source priors. However, simpler models can be obtained for instance by choosing exponential priors with different scale parameters $\sigma^2_{\text{s},m}$:

$$f(\mathbf{s}_m|\sigma^2_{\text{s},m}) \propto \prod_{j=1}^{L} \frac{1}{\sigma^2_{\text{s},m}} \exp\left[-\frac{s_{m,j}}{2\sigma^2_{\text{s},m}}\right] \mathbf{1}_{\mathbb{R}^+}(s_{m,j}), \tag{37}$$

or positive truncated Gaussian distribution with different hidden variances $\sigma^2_{\text{s},m}$:

$$f(\mathbf{s}_m|\sigma^2_{\text{s},m}) \propto \prod_{j=1}^{L} \frac{1}{\sigma^2_{\text{s},m}} \exp\left[-\frac{s^2_{m,j}}{2\sigma^2_{\text{s},m}}\right] \mathbf{1}_{\mathbb{R}^+}(s_{m,j}). \tag{38}$$

The resulting Bayesian algorithms are simpler since only one hyperparameter $\sigma^2_{\text{s},m}$ has to be adjusted for each source.

For both choices, conjugate IG distributions $\mathcal{IG}(\rho_{\text{s}}/2, \psi_{\text{s}}/2)$ are chosen as prior distributions for the hyperparameters $\sigma^2_{\text{s},m}$, $m = 1,\ldots,M$. After integrating out the hyperparameter vector $\boldsymbol{\Phi} = \{\psi_{\text{e}}, \boldsymbol{\sigma}^2_{\text{s}}\}$, the posterior distribution in (22) can be expressed as

$$f(\mathbf{A},\mathbf{S},\boldsymbol{\sigma}^2_{\text{e}}|\mathbf{Y}) \propto \prod_{m=1}^{M} T(\mathbf{s}_m, \rho_{\text{s}}, \psi_{\text{s}}) \mathbf{1}_{\mathbb{R}^L_+}(\mathbf{s}_m) \times \prod_{i=1}^{N} \left[\frac{1}{\sigma^{L+2}_{e,i}} \exp\left[-\frac{\|\mathbf{y}_i - \mathbf{S}^{\text{T}}\mathbf{c}_i\|^2}{2\sigma^2_{\text{e},i}}\right] \mathbf{1}_{\mathbb{S}}(\mathbf{a}_i)\right]. \tag{39}$$

The scalar $T(\mathbf{s}_m, \rho_{\text{s}}, \psi_{\text{s}})$ depends on the prior distribution used for the source spectra

$$T(\mathbf{s}_m, \rho_{\text{s}}, \psi_{\text{s}}) = \begin{cases} [\psi_{\text{s}} + \|\mathbf{s}_m\|_1]^{-(L+\rho_{\text{s}})/2} & \text{for exponential priors,} \\ [\psi_{\text{s}} + \|\mathbf{s}_m\|_2^2]^{-(L+\rho_{\text{s}})/2} & \text{for truncated Gaussian priors.} \end{cases} \tag{40}$$

In the Gibbs sampling strategy presented in Section 4, the generation according to $f(\mathbf{S}|\mathbf{A},\boldsymbol{\sigma}^2_{\text{e}},\mathbf{Y})$ in Section 4.3 is finally achieved using the following two steps:

- generation according to $f(\boldsymbol{\sigma}^2_{\text{s}}|\mathbf{S},\mathbf{A},\boldsymbol{\sigma}^2_{\text{e}},\mathbf{Y})$:

  $$\sigma^2_{\text{s},m}|\mathbf{s}_m \sim \mathcal{IG}(L + \rho_{\text{s}}, \psi_{\text{s}} + \|\mathbf{s}_m\|^b_{\ell_b}), \tag{41}$$

  where $b = 1$ for the exponential prior and $b = 2$ otherwise,
- generation according to $f(\mathbf{S}|\boldsymbol{\sigma}^2_{\text{s}},\mathbf{A},\boldsymbol{\sigma}^2_{\text{e}},\mathbf{Y})$:

  $$\mathbf{s}_m|\boldsymbol{\sigma}^2_{\text{s}},\mathbf{A},\boldsymbol{\sigma}^2_{\text{e}},\mathbf{Y} \sim \mathcal{N}^+(\boldsymbol{\lambda}_m, \delta^2_m \mathbf{I}_L), \tag{42}$$

  where $\boldsymbol{\lambda}_m$ and δ^2_m, similar to (31), are derived following the model in [8].

Table 3 reports the NMSEs (computed from 100 Monte Carlo runs following (35)) for the sources and concentration matrices estimated by the different Bayesian algorithms. The results are significantly better when employing the Gamma distribution, which clearly indicates that the Gamma prior seems to be the best choice to model the distribution of the sources when analyzing spectroscopy data.

6. Separation of chemical mixtures monitored by Raman spectroscopy (RS)

Calcium carbonate is a chemical material used commercially for a large variety of applications such as filler for plastics or paper. Depending on operating conditions,

Table 3
NMSE for different source priors (100 Monte Carlo runs).

	Gamma	Truncated Gaussian	Exponential
NMSE (**S**)	0.0071	0.0269	0.0110
NMSE (**C**)	0.0024	0.0089	0.0029

calcium carbonate crystallizes as calcite, aragonite or vaterite. Calcite is the most thermodynamically stable of the three, followed by aragonite or vaterite. Globally, the formation of calcium carbonate by mixing two solutions containing, respectively, calcium and carbonate ions takes place in two well distinguished steps. The first step is the precipitation one. This step is very fast and provides a mixture of calcium carbonate polymorphs.[4] The second step (a slow process) represents the phase transformation from the unstable polymorphs to the stable one (calcite). The physical properties of the crystallized product depend largely on the polymorphic composition, so it is necessary to quantify these polymorphs when they are mixed. Several techniques based on infrared spectroscopy (IR), X-ray diffraction (XRD) or Raman spectroscopy can be used to determine the composition of $CaCO_3$ polymorph mixtures. However, contrary to XRD and IR, RS is a faster method since it does not require a sample preparation and is a promising tool for an online polymorphic composition monitoring. In our case, the crystallization process of calcium carbonate is carried out in 5 mol/L NaCl solutions, which correspond to a real industrial situation. Under the industrial conditions, the calcite is the desired product.

The main purpose of this experiment is to show how the proposed constrained BSS method can be used for processing Raman spectroscopy data to study the relation between polymorphs and temperature and to explore favorable conditions for calcite formation in saline solutions.

6.1. Mixture preparation and data acquisition

Calcium chloride and sodium carbonate separately dissolved in sodium chloride solutions of the same concentration (5 mol/L) were rapidly mixed to precipitate calcium carbonate. A 100 mL solution containing 0.625 M of Na_2CO_3 and 5 M of NaCl was added to a 2.5 L solution containing 0.025 M of $CaCl_2$ and 5 M of NaCl (the precipitation is carried out under stoichiometric conditions). A preliminary investigation detailed in [39] suggested that the temperature and the aging time are the most important factors that can affect the polymorphic composition. Therefore the experiments were operated in a temperature range between 20 and 70 °C and retaining several aging times of the precipitated mixture. A sample was collected 2 min after the beginning of the experiment to determine the polymorphic composition at the end of the precipitation step. Then, samples were collected at regular time intervals to follow the polymorph transformation.

Raman Spectra were collected on a Jobin–Yvon T64000 spectrometer equipped with an optical microscope, a threefold monochromator, and a nitrogen-cooled CCD camera. The excitation was induced by a laser beam of argon spectra physic laser stability 2017 at a wavelength of 514.5 nm. The beam was focused using a long-frontal $\times 50$ objective (numerical aperture 0.5) on an area of about $3\,\mu m^2$. The laser power on the sample was approximately 20 mW and the acquisition time was 1 min. The spectral resolution was $3\,cm^{-1}$, with a wavenumber precision better than $1\,cm^{-1}$. The Raman spectra were collected at five points, which were randomly distributed throughout the mixture. The average of all spectra was considered as the Raman spectrum of the corresponding mixture for the considered temperature value and aging time. Raman spectra were collected 2 min after the beginning of the experiment for various temperatures ranging between 20 and 70 °C in order to determine the influence of temperature on the polymorph precipitation. Moreover for each temperature, Raman spectra were collected at regular time intervals for monitoring phase transformation. Finally, a total of $N = 37$ Raman spectra of $L = 477$ wavelengths have been obtained.

6.2. Data preprocessing

The Raman spectra of the polymorph mixture are firstly processed using a background removal approach proposed in [40]. In this method, the baseline is represented by a polynomial whose parameters are estimated by minimizing a truncated quadratic cost function. This method requires the specification of the polynomial order and the threshold of the quadratic cost function truncation. This method was applied for each spectrum separately with a fifth order polynomial and a threshold chosen by trial and error. Fig. 7 shows the Raman spectra at the beginning of the phase transformation step, after background removal.

6.3. Polymorph mixture separation under non-negativity and full additivity constraints

The number of sources to be recovered is fixed to $M = 3$ according to the prior knowledge on the mixture composition. The iteration number is fixed to 1000 iterations where the first 200 samples are discarded since they correspond to the burn-in period of the Gibbs sampler. Fig. 8 illustrates the estimated spectra using the proposed approach incorporating the non-negativity and the full additivity constraints.

From a spectroscopic point of view and according to the positions of the vibrational peaks, the identification of the three components is very easy: the first source corresponds to calcite, the second spectrum to aragonite and the third one to vaterite. A measure of the dissimilarity between the estimated spectra and the measured pure spectra of the three components gives 4.56% for calcite, 0.65% for aragonite and 4.76% for vaterite. These

[4] The ability of a chemical substance to crystallize with several types of structures, depending on a physical parameter, such as temperature, is known as polymorphism. Each particular form is said a polymorph.

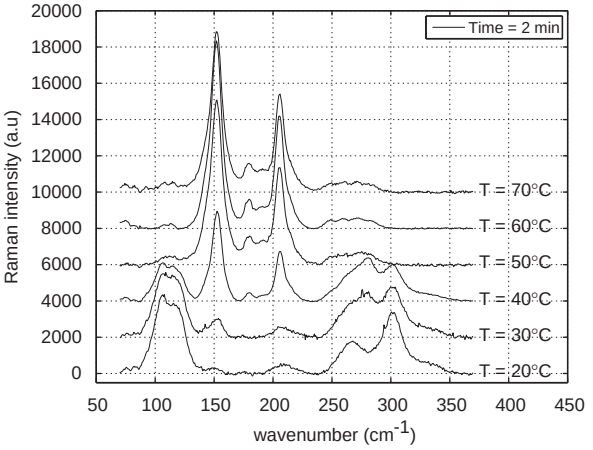

Fig. 7. Mixture spectra at the beginning of the phase transformation.

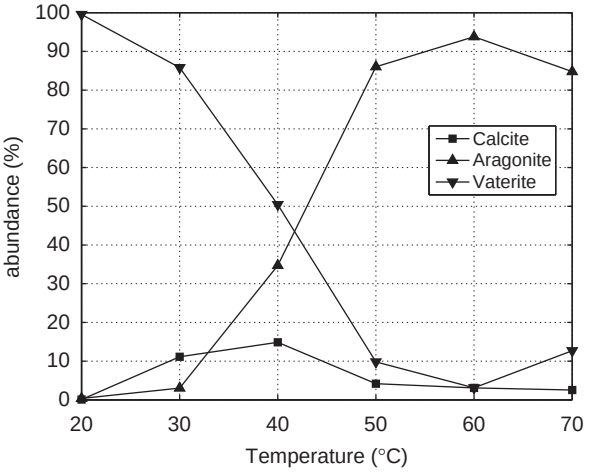

Fig. 9. Three component abundances at the beginning of the phase transformation for different temperature values.

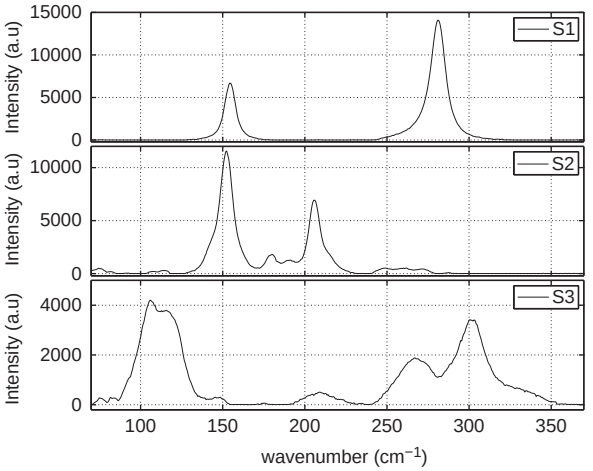

Fig. 8. Estimated sources.

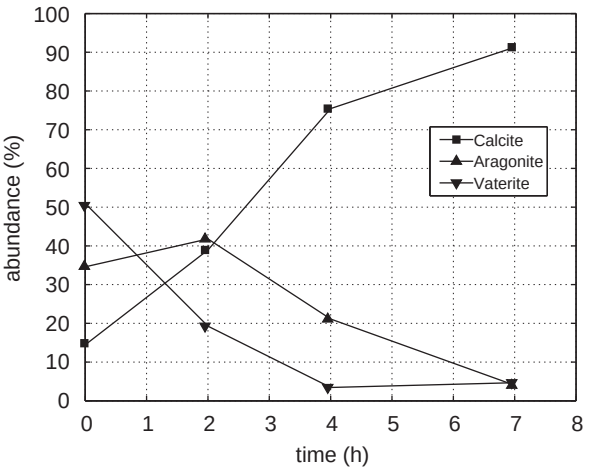

Fig. 10. Evolution of the three component abundances for $T = 40\,^{\circ}\mathrm{C}$.

results show that the proposed method can be applied successfully without imposing any prior information on the shape of the pure spectra.

The evolution of the polymorph proportions versus temperature is shown in Fig. 9. Pure vaterite is observed at 20 °C and a quite pure aragonite is obtained at 60 °C. However, between 20 and 60 °C ternary mixtures are observed. The content of calcite is maximal at 40 °C. Let us now consider the phase transformation evolution at this temperature value. The concentration profile versus precipitation time at 40 °C is reported in Fig. 10. At the beginning of the phase transformation (2 min), the ternary mixture is composed of 50% vaterite, 35% aragonite and 15% calcite. After 2 h, the vaterite is transformed to aragonite and calcite. After 7 h, vaterite and aragonite are almost totally transformed to calcite. So, aging time promotes the formation of calcite which is in agreement with some results reported in the literature [41,42].

6.4. Polymorph mixture analysis using other BSS algorithms

The dataset resulting from this experiment is also used to compare the performances of standard BSS methods taking into account the non-negativity constraint and their re-scaled versions ensuring the full additivity constraint. Table 4 summarizes the performances of the considered separation algorithms in terms of normalized mean square errors, dissimilarity measures and computation times. It can be noticed that the proposed approach provides source estimates with better accuracy than the other methods. In addition to the good estimation quality, the second advantage of the proposed method is its ability to scale the sources during the estimation algorithm. Thus it does not require any post-processing of the estimation results. However, as previously highlighted, the price to pay for having such results is the computational times required by the proposed MCMC-based estimation method.

Table 4
Estimation performance for different BSS algorithms on real spectroscopic data.

	NMSE ($\mathbf{S}$)	Diss ($\mathbf{S}$)	Time (s)
Proposed approach	0.0072	3.34	146
BPSS	0.0118	4.87	205
Re-scaled BPSS	0.0124	4.87	205
NNICA	0.1007	11.82	29
Re-scaled NNICA	0.3996	11.82	29
NMF	0.0093	4.25	26
Re-scaled NMF	0.0109	4.25	26

7. Conclusion

This paper studied Bayesian algorithms for separating linear mixtures of spectral sources under non-negativity and full additivity constraints. These two constraints are required in some applications such as hyperspectral imaging and spectroscopy to get meaningful solutions. A hierarchical Bayesian model was defined based on priors ensuring the fulfillment of the constraints. Estimation of the sources as well as the mixing coefficients was then performed by using samples distributed according to the joint posterior distribution of the unknown model parameters. A Gibbs sampler strategy was proposed to generate samples distributed according to the posterior of interest. The generated samples were then used to estimate the unknown parameters. The performance of the algorithm was first illustrated by means of simulations conducted on synthetic signals. The application to the separation of chemical mixtures resulting from Raman spectroscopy was finally investigated. The proposed Bayesian algorithm provided very promising results for this application. Particularly, when the computational times is not a study constraint, the proposed method clearly outperforms other standard NMF techniques, which can give approximative solutions faster. Perspectives include the development of a similar methodology for unmixing hyperspectral images. Some results were already obtained for the unmixing of known sources. However, the joint estimation of the mixing coefficients (abundances) and the sources (endmembers) is a still an open and challenging problem.

Acknowledgments

This work was supported by the French CNRS/GDR-ISIS. The authors would like to thank the anonymous referees and the associate editor for their constructive comments improving the quality of the paper. They also would like to thank Prof. Petar M. Djurić from Stony Brook University for helping them to fix the English grammar.

References

[1] A. Cichocki, S.-I. Amari, Adaptive Blind Signal and Image Processing—Learning Algorithms and Applications, Wiley, Chichester, England, 2002.
[2] P. Comon, C. Jutten, J. Hérault, Blind separation of sources, part II: problems statement, Signal Processing 24 (1) (1991) 11–20.
[3] P. Comon, Independent component analysis—a new concept?, Signal Processing 36 (3) (1994) 287–314.
[4] J.-F. Cardoso, Blind signal separation: statistical principles, Proceedings of the IEEE 9 (10) (1998) 2009–2025.
[5] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[6] M.D. Plumbley, Algorithms for non-negative independent component analysis, IEEE Transactions on Neural Networks 14 (3) (2003) 534–543.
[7] P. Sajda, S. Du, T.R. Brown, R. Stoyanova, D.C. Shungu, X. Mao, L.C. Parra, Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain, IEEE Transactions on Medical Imaging 23 (12) (2004) 1453–1465.
[8] S. Moussaoui, D. Brie, A. Mohammad-Djafari, C. Carteret, Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling, IEEE Transactions on Signal Processing 54 (11) (2006) 4133–4145.
[9] P.O. Hoyer, Non-negative matrix factorization with sparseness constraints, Journal of Machine Learning Research 5 (2004) 1457–1469.
[10] C. Févotte, S.J. Godsill, A Bayesian approach for blind separation of sparse sources, IEEE Transactions on Audio, Speech, Language Processing 14 (6) (2006) 2174–2188.
[11] H. Snoussi, J. Idier, Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures, IEEE Transactions on Signal Processing 54 (9) (2006) 3257–3269.
[12] E.R. Malinowski, Factor Analysis in Chemistry, third ed., Wiley, New York, 2002.
[13] C.-I. Chang, Hyperspectral Data Exploitation: Theory and Applications, Wiley, Hoboken, NJ, 2007.
[14] N. Dobigeon, J.-Y. Tourneret, A.O. Hero, Bayesian linear unmixing of hyperspectral images corrupted by colored Gaussian noise with unknown covariance matrix, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal (ICASSP), Las Vegas, Nevada, USA, 2008.
[15] A. de Juan, E. Casassas, R. Tauler, Soft-modelling of analytical data, in: Wiley encyclopedia of analytical chemistry: instrumentation and applications, of Soft-modelling of analytical data, vol. 11, Wiley Interscience, New York, 2000, pp. 9800–9837.
[16] N. Dobigeon, J.-Y. Tourneret, C.-I. Chang, Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery, IEEE Transactions on Signal Processing 56 (7) (2008) 2684–2695.
[17] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.
[18] I.-T. Hsiao, A. Rangarajan, G. Gindi, Bayesian image reconstruction for transmission tomography using mixture model priors and deterministic annealing algorithms, Journal of Electronic Imaging 12 (1) (2003) 7–16.
[19] J. Miskin, D. MacKay, Ensemble learning for blind source separation, in: S. Roberts, R. Everson (Eds.), Independent Component Analysis: Principles and Practice, Cambridge University Press, Cambridge, 2001, pp. 209–233.
[20] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, Blind unmixing of linear mixtures using a hierarchical Bayesian model. Application to spectroscopic signal analysis, in: Proceedings of the IEEE/SP Workshop Statistical Signal Processing, Madison, USA, 2007, pp. 79–83.
[21] C.P. Robert, The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation, Springer Texts in Statistics, second ed., Springer, New York, USA, 2007.
[22] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.
[23] E. Punskaya, C. Andrieu, A. Doucet, W. Fitzgerald, Bayesian curve fitting using MCMC with applications to signal segmentation, IEEE Transactions on Signal Processing 50 (3) (2002) 747–758.
[24] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 1999.
[25] C.P. Robert, Simulation of truncated normal variables, Statistics and Computing 5 (2) (1995) 121–125.
[26] N. Dobigeon, J.-Y. Tourneret, Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex, Technical Report, IRIT/ENSEEIHT/TéSA, March 2007, URL ⟨http://www.enseeiht.fr/~dobigeon/papers/Dobigeon_TechReport_2007b.pdf⟩.
[27] J. Geweke, Efficient simulation from the multivariate normal and Student-T distributions subject to linear constraints, in: E.M. Keramidas (Ed.), Computing Science and Statistics, Proceedings of the 23th Symposium on the Interface, Interface Foundation of North America, Inc., Fairfax, VA, 1991, pp. 571–578.

[28] V. Mazet, D. Brie, J. Idier, Simulation of positive normal variables using several proposal distributions, in: Proc. IEEE Workshop on Statistical Signal Processing (SSP), Bordeaux, France, 2005, pp. 37–42.
[29] A. Gelman, D.B. Rubin, Inference from iterative simulation using multiple sequences, Statistical Science 7 (4) (1992) 457–511.
[30] P.M. Djurić, J.-H. Chun, An MCMC sampling approach to estimation of nonstationary hidden Markov models, IEEE Transactions on Signal Processing 50 (5) (2002) 1113–1123.
[31] N. Dobigeon, J.-Y. Tourneret, J.D. Scargle, Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model, IEEE Transactions on Signal Processing 55 (2) (2007) 414–423.
[32] N. Dobigeon, J.-Y. Tourneret, M. Davy, Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach, IEEE Transactions on Signal Processing 55 (4) (2007) 1251–1263.
[33] S. Godsill, P. Rayner, Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler, IEEE Transactions on Speech and Audio Processing 6 (4) (1998) 352–372.
[34] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman & Hall, London, 1995.
[35] C.P. Robert, S. Richardson, Markov chain Monte Carlo methods, in: C.P. Robert (Ed.), Discretization and MCMC Convergence Assessment, Springer, New York, 1998, pp. 1–25.
[36] M.D. Plumbley, E. Oja, A nonnegative PCA algorithm for independent component analysis, IEEE Transactions on Neural Networks 15 (1) (2004) 66–76.
[37] J.K. Tugnait, Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria, IEEE Transactions on Signal Processing 45 (3) (1997) 658–672.
[38] S. Moussaoui, C. Carteret, D. Brie, A. Mohammad-Djafari, Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods, Chemometrics and Intelligent Laboratory Systems 81 (2) (2006) 137–148.
[39] C. Carteret, A. Dandeu, S. Moussaoui, H. Muhr, B. Humbert, E. Plasari, Polymorphism studied by lattice phonon raman spectroscopy and statistical mixture analysis method. Application to calcium carbonate polymorphs during batch crystallization, Crystal Growth & Design 9 (2) (2009) 807–812.
[40] V. Mazet, C. Carteret, D. Brie, J. Idier, B. Humbert, Background removal from spectra by designing and minimising a non-quadratic cost function, Chemometrics and Intelligent Laboratory Systems 76 (2) (2005) 121–133.
[41] M.J. Kitamura, Controlling factor of polymorphism in crystallization process, Journal of Crystal Growth 237 (2002) 2205–2214.
[42] A. Dandeu, B. Humbert, C. Carteret, H. Muhr, E. Plasari, J.-M. Bossoutrot, Raman spectroscopy—a powerful tool for the quantitative determination of the composition of polymorph mixtures: application to $CaCO_3$ polymorph mixtures, Chemical Engineering & Technology 29 (2) (2006) 221–225.