Application of Bayesian non-negative source separation to mixture analysis in spectroscopy

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

Application of Bayesian Non-negative Source Separation to Mixture Analysis in Spectroscopy

Saïd Moussaoui*, David Brie*, Cédric Carteret† and
Ali Mohammad-Djafari**

*Centre de Recherche en Automatique de Nancy (CRAN), CNRS UMR 7039,
B.P. 239, Université Henri Poincaré, Nancy 1, 54506, Vandœuvre-lès-Nancy, France
email: \{firstname.lastname\}@cran.uhp-nancy.fr
†Laboratoire de Chimie Physique et Microbiologie pour l'Environnement (LCPME),
CNRS UMR 7564, Université Henri Poincaré, Nancy 1, 54600, Villers–lès–Nancy, France
e-mail: carteret@lcpe-cnrs.fr
**Laboratoire de Signaux et Systèmes (LSS), CNRS-SUPÉLEC-UPS,
Université Paris-Sud, 91192, Gif-sur-Yvette cedex, France
e-mail: djafari@lss.supelec.fr

Abstract. In this paper we present an application of Bayesian non-negative source separation to the analysis of spectral mixtures obtained from the analysis of multicomponent substances. The processing aims are formalized as a non-negative source separation problem. The proposed Bayesian inference for the analysis is introduced and the main steps of the estimation algorithm are outlined. Finally, some results obtained with simulated and experimental data are presented.

PROBLEM STATEMENT

The analysis of multicomponent chemical substances using spectroscopic techniques yields data which are mixtures of the pure component spectra. The processing aims at identifying the unknown pure components and determining their concentrations [1]. According to Beer-Lambert-Bouguer law [2], the mixing model is linear instantaneous

$$X_i(\nu_k) = \sum_{j=1}^{n} A_{(i,j)} S_j(\nu_k) + E_i(\nu_k), \quad \text{for } k = 1,\ldots,N, \tag{1}$$

where $X_i(\nu_k)$, $S_j(\nu_k)$ represent respectively the i-th observation and the j-th pure component absorption at wavelength ν_k (this measurement variable can also correspond to a wavenumber, chemical shift, etc.) and $\left\{A_{(i,j)}\right\}_{j=1}^{n}$ represent the mixing coefficients which are proportional to the concentration of the n pure components in the i-th mixture. The additif noise term $E_i(\nu_k)$ represents measurement errors and model uncertainties. By varying a physical parameter such as temperature and pressure, the amount of each pure component in the substance changes due to chemical reaction or molecular interactions. For m different values of the physical parameter, the observation spectra are expressed using matrix notations as

$$\underline{X} = \underline{A}\,\underline{S} + \underline{E}, \tag{2}$$

where $\underline{X}$ is the $m \times N$ data matrix with i-th observation spectra of N wavelengths in each row and $\underline{A}$ is the $m \times n$ mixing matrix whose columns are proportional to the concentration profile of the n components. $\underline{S}$ is a $n \times N$ matrix of the n spectra of the n pure components, in its rows, and $\underline{E}$ is a $m \times N$ matrix of the additive noise sequences. The problem of mixture analysis in spectroscopy is then stated as follows: knowing the number of components and having all the observations, estimate the pure component spectra and their concentrations. These objectives are formalized as a particular source separation problem in which the sources are identified as the pure component spectra and the concentration profiles are deduced from the mixing coefficients.

Two main constraints are associated to this problem: all the source signals are non-negatives

$$S_j(\nu_k) \geq 0, \forall\ j,k, \tag{3}$$

and all the mixing coefficients are non-negatives

$$A_{(i,j)} \geq 0, \forall\ i,j. \tag{4}$$

So, mixture analysis in spectroscopy corresponds to a non-negative source separation problem. In chemometrics the problem is termed by self-modeling curve resolution [3] and the mostly used methods consist in minimizing the mean squares error criterion under the non-negativity constraint, leading to algorithms differing on the manner how the non-negativity is introduced. In particular, alternating least squares (ALS) method [4] performs an estimation where the non-negativity is hardly imposed between successive iterations by setting to zero the negative estimates or by performing a non-negative least squares estimation [5]. The second method named non-negative matrix factorization (NMF), which has been presented recently [6], achieves the decomposition by constructing a gradient descent algorithm over the objective function and updates iteratively sources and mixing coefficients by considering a particular multiplicative learning rule that ensures the estimates to be non-negatives.

In this paper we address the problem of non-negative source separation in a Bayesian framework. we firstly present an approach that we proposed in [7, 8] and finally, we discuss some results obtained when applying these methods to the separation of a simulated non-negative mixture and to the analysis of spectral data obtained from an infrared (IR) spectroscopy experiment.

BAYESIAN NON-NEGATIVE SOURCE SEPARATION

The main idea of a Bayesian approach for source separation is to formalize any available knowledge on the source signals and the mixing coefficients through the assignment of prior distributions $p(\underline{S})$ and $p(\underline{A})$. According to Bayes's theorem and considering the likelihood $p(\underline{X}|\underline{S},\underline{A})$ and these prior distribution, we obtain the posterior density expressed as

$$p(\underline{S},\underline{A}|\underline{X}) \propto p(\underline{X}|\underline{S},\underline{A}) \cdot p(\underline{S}) \cdot p(\underline{A}). \tag{5}$$

From this posterior density, joint estimation of $\underline{S}$ and $\underline{A}$ can be achieved by using various Bayesian estimators. However, the main task of the inference is to encode the available knowledge by appropriate probability distribution functions.

Bayesian Separation Model

The noise sequences are assumed independent and identically distributed (i.i.d), independent of the source signals, stationary and Gaussian with zero mean and variances $\left\{\sigma_i^2\right\}_{i=1}^m$. Therefore, the likelihood is given as

$$p\left(\underline{X}|\underline{A},\underline{S},\underline{\theta}_1\right)=\prod_{k=1}^{N}\prod_{i=1}^{m}\mathcal{N}\left(X_i(\nu_k);\sum_{\ell=1}^{n}A_{(i,\ell)}S_\ell(\nu_k),\sigma_i^2\right), \tag{6}$$

where $\underline{\theta}_1=\left\{\sigma_i^2\right\}_{i=1}^m$ and $\mathcal{N}\left(z;\mu,\sigma^2\right)$ refers to a normal distribution of the variable z with mean μ and variance σ^2. The sources are assumed mutually statistically independent and each j-th source signal is supposed i.i.d and distributed as a Gamma distribution of parameters $\left(\alpha_j,\beta_j\right)$. The Gamma density is used to take into account the non-negativity and its parameters allow to fit the spectra distribution that may present some sparsity and possibly a background. To incorporate the mixing coefficient non-negativity, each column j of the mixing matrix is also assumed distributed as a Gamma distribution of parameters $\left(\gamma_j,\lambda_j\right)$. The two-parameter Gamma density is expressed by

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

where $\Gamma(a)$ is the Gamma function. The prior densities of the source signals and the mixing matrix are then given by

$$p(\underline{S}|\underline{\theta}_2)=\prod_{k=1}^{N}\prod_{j=1}^{n}\mathcal{G}(S_j(\nu_k);\alpha_j,\beta_j), \tag{8}$$

$$p(\underline{A}|\underline{\theta}_3)=\prod_{i=1}^{m}\prod_{j=1}^{n}\mathcal{G}(A_{(i,j)};\gamma_j,\lambda_j), \tag{9}$$

where $\underline{\theta}_2=\left\{\alpha_i,\beta_j\right\}_{j=1}^n$ and $\underline{\theta}_3=\left\{\gamma_j,\lambda_j\right\}_{j=1}^n$. Using Bayes's theorem and noting by $\underline{\theta}$ the vector containing the hyperparameters $\underline{\theta}=\{\underline{\theta}_1,\underline{\theta}_2,\underline{\theta}_3\}$, the posterior law is given as

$$\begin{aligned}p\left(\underline{S},\underline{A}|\underline{X},\underline{\theta}\right)\propto&\prod_{k=1}^{N}\prod_{i=1}^{m}\mathcal{N}\left(X_i(\nu_k);\sum_{j=1}^{n}A_{(i,j)}S_j(\nu_k),\sigma_i^2\right)\\&\times\prod_{k=1}^{N}\prod_{j=1}^{n}\mathcal{G}(S_j(\nu_k);\alpha_j,\beta_j)\times\prod_{i=1}^{m}\prod_{j=1}^{n}\mathcal{G}(A_{(i,j)};\gamma_j,\lambda_j).\end{aligned} \tag{10}$$

For an unsupervised learning, the hyperparameters $\underline{\theta}$ have also to be inferred. The joint posterior distribution including the hyperparameters is expressed as

$$p\left(\underline{S},\underline{A},\underline{\theta}|\underline{X}\right)\propto p\left(\underline{S},\underline{A}|\underline{X},\underline{\theta}\right)\cdot p\left(\underline{\theta}\right), \tag{11}$$

in which prior densities are assigned to the hyperparameters $\underline{\theta}$.

MCMC Sampling and Estimation

The estimation of the source signals and the mixing coefficients is performed by sampling the joint posterior distribution and constructing the estimator from the samples of the Markov chain. The estimation is achieved using the marginal posterior mean (MPM) estimator

$$\left(\hat{\underline{A}}, \hat{\underline{S}}\right) = \mathbb{E}_{p(\underline{S},\underline{A}|\underline{X},\underline{\theta})}\left\{\underline{S},\underline{A}\right\}, \tag{12}$$

and the simulation of the posterior density $p(\underline{S},\underline{A},\underline{\theta}|\underline{X})$ is performed using an hybrid Metropolis-Hastings-Gibbs sampling algorithm. The main steps of the sampling scheme are firstly outlined and then the conditional posterior densities are given.

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

1. source signals $\underline{S}^{(r+1)}$ from $p\left(\underline{S}|\underline{X},\underline{A}^{(r)},\underline{\theta}^{(r)}\right)$,
2. mixing coefficients $\underline{A}^{(r+1)}$ from $p\left(\underline{A}\middle|\underline{X},\underline{S}^{(r+1)},\underline{\theta}^{(r)}\right)$,
3. noise variances $\underline{\theta}_1^{(r+1)}$ from $p\left(\underline{\theta}_1\middle|\underline{X},\underline{S}^{(r+1)},\underline{A}^{(r+1)}\right)$,
4. source hyperparameters $\underline{\theta}_2^{(r+1)}$ from $p\left(\underline{\theta}_2\middle|\underline{S}^{(r+1)}\right)$,
5. mixing coefficient hyperparameters $\underline{\theta}_3^{(r+1)}$ from $p\left(\underline{\theta}_3\middle|\underline{A}^{(r+1)}\right)$.

All the variable are randomly initialized at the first iteration of the sampler and the MPM estimator is implemented by averaging the retained samples of the Markov chain (the first samples corresponding to the burn-in run are discarded).

Conditional Posterior Densities

The scaler version of the sampling scheme is implemented and each component of $\underline{S}$, $\underline{A}$ and $\underline{\theta}$ is sampled conditionally to the most recent other components. All the required conditional posterior densities for MCMC sampling are detailed below. Firstly priors are assigned to source signals $S_j(v_k)$, secondly to mixing coefficients $A_{(i,j)}$ and finally to the hyperparameters.

Source Signals. At the r-th iteration of the sampler, the conditional posterior density of each source signal is given as

$$\begin{aligned} &p\left(S_j(v_k)\middle|X_{(1:n)}(v_k), S^{(r+1)}_{(1:j-1)}(v_k), S^{(r)}_{(j+1:n)}(v_k), A^{(r)}_{(1:m,1:n)}, \underline{\theta}_1^{(r)}, \alpha_j^{(r)}, \beta_j^{(r)}\right) \\ &\qquad \propto S_j(v_k)^{\alpha_j^{(r)}-1} \exp\left[-\frac{1}{2\left[\sigma_{S_j}^{post}\right]^2}\left(S_j(v_k)-\mu_{S_j}^{post}(v_k)\right)^2\right] \mathbb{I}_{[0,+\infty]}\left(S_j(v_k)\right) \end{aligned} \tag{13}$$

where $\mu_{S_j}^{post}(v_k) = \mu_{S_j}^{likel}(v_k) - \beta_j^{(r)} \left[\sigma_{S_j}^{post}\right]^2$, and

$$\left\{\begin{array}{rcl} \left[\sigma_{S_j}^{post}\right]^2 & = & \left[\sum_{i=1}^{m} \frac{\left[A_{(i,j)}^{(r)}\right]^2}{\left[\sigma_i^{(r)}\right]^2}\right]^{-1}, \\ \mu_{S_{(j,t)}}^{likel} & = & \frac{1}{\left[\sigma_{S_j}^{likel}\right]^2} \sum_{i=1}^{m} \frac{A_{(i,j)}^{(r)} \varepsilon_i^{-j}(v_k)}{\left[\sigma_i^{(r)}\right]^2}, \\ \varepsilon_i^{-j}(v_k) & = & X_i(v_k) - \sum_{\ell=1}^{j-1} A_{(i,\ell)}^{(r)} S_\ell^{(r+1)}(v_k) - \sum_{\ell=j+1}^{n} A_{(i,\ell)}^{(r)} S_\ell^{(r)}(v_k). \end{array}\right.$$

This distribution is not usual, so its sampling is achieved using a Metropolis-Hastings algorithm. An instrumental density is derived from this posterior law as a truncated normal distribution of variance set to $\sigma_{S_j}^{post}$ and mean equal to the mode of this posterior law. The sampling from the truncated normal distribution can be achieved by cumulative distribution function inversion technique or by using an accept-reject method [9]. An interesting point with this instrumental distribution is that constraining $\alpha_j = 1$ corresponds to taking an exponential prior for the j-th source distribution. The use of the Metropolis-Hastings algorithm is avoided since the conditional posterior density is a truncated normal of parameters equal to those of the proposed instrumental density.

Mixing Coefficients. The conditional posterior density of each mixing coefficient is

$$p\left(A_{(i,j)} \middle| X_i(v_{1:N}), A_{(i,1:j-1)}^{(r+1)}, A_{(i,j+1:n)}^{(r)}, \underline{S}^{(r+1)}, \underline{\theta}_1^{(r)}, \gamma_j^{(r)}, \lambda_j^{(r)}\right)$$
$$\propto A_{(i,j)}^{\gamma_j^{(r)}-1} \exp\left\{-\frac{1}{2\left[\sigma_{A_{(i,j)}}^{post}\right]^2}\left(A_{(i,j)} - \mu_{A_{(i,j)}}^{post}\right)^2\right\} \mathbb{I}_{[0,+\infty]}\left(A_{(i,j)}\right) \quad (14)$$

where $\mu_{A_{(i,j)}}^{post} = \mu_{A_{(i,j)}}^{likel} - \lambda_j^{(r)} \left[\sigma_{A_{(i,j)}}^{likel}\right]^2$, and

$$\left\{\begin{array}{rcl} \left[\sigma_{A_{(i,j)}}^{post}\right]^2 & = & \frac{\left[\sigma_{A_{(i,j)}}^{(r)}\right]^2}{\sum_{k=1}^{N} S_j^{(r+1)}(v_k)}, \\ \mu_{A_{(i,j)}}^{likel} & = & \frac{1}{\left[\sigma_{A_{(i,j)}}^{likel}\right]^2} \sum_{k=1}^{N} S_j^{(r+1)}(v_k) \varepsilon_i^{-j}(v_k), \\ \varepsilon_i^{-j}(v_k) & = & \left(X_i(v_k) - \sum_{\ell=1}^{j-1} A_{(i,\ell)}^{(r+1)} S_\ell^{(r+1)}(v_k) - \sum_{\ell=j+1}^{n} A_{(i,\ell)}^{(r)} S_\ell^{(r+1)}(v_k)\right), \end{array}\right.$$

The sampling of this distribution is achieved using a Metropolis-Hastings algorithm and the same derivation of the instrumental density as for the source signals.

Noise Variances. The conditional posterior conditional density of each noise variance σ_i^2 is expressed by

$$p\left(\frac{1}{\sigma_i^2}\middle|X_i\left(\nu_{(1:N)}\right),A_{(i,1:N)}^{(r+1)},\underline{S}^{(r+1)}\right)$$
$$\propto\left(\frac{1}{\sigma_i^2}\right)^{\frac{N}{2}}\exp\left[-\frac{1}{2\sigma_i^2}\sum_{t=1}^{N}\left(X_i(\nu_k)-\sum_{\ell=1}^{n}A_{(i,\ell)}^{(r+1)}S_\ell^{(r+1)}(\nu_k)\right)^2\right]\cdot p\left(\frac{1}{\sigma_i^2}\right). \quad (15)$$

The conjugate prior of σ_i^{-2} is a Gamma distribution of parameters $\left(\alpha_{\sigma_i^2}^{prior},\beta_{\sigma_i^2}^{prior}\right)$. Therefore, its conditional posterior density is also of Gamma distribution of parameters $\alpha_{\sigma_i^2}^{post}=\frac{N}{2}+\alpha_{\sigma_i^2}^{prior}$ and $\beta_{\sigma_i^2}^{post}=\frac{1}{2}\sum_{k=1}^{N}\left(X_i(\nu_k)-\sum_{\ell=1}^{n}A_{(i,\ell)}^{(r+1)}S_\ell^{(r+1)}(\nu_k)\right)^2+\beta_{\sigma_i^2}^{prior}$.

Source signal hyperparameters. The posterior density of each hyperparameter α_j is given as

$$p\left(\alpha_j|S_j^{(r+1)}\left(\nu_{(1:N)}\right),\beta_j^{(r)}\right)$$
$$\propto\frac{1}{\Gamma(\alpha_j)^N}\exp\left[\left(N\log\beta_j^{(r)}+\sum_{t=1}^{N}\log S_j^{(r+1)}(\nu_k)\right)\alpha_j\right]\cdot p(\alpha_j), \quad (16)$$

which by assigning an exponential prior to α_j with parameter $\lambda_{\alpha_j}^{prior}$, takes the form

$$p\left(\alpha_j|S_j^{(r+1)}\left(\nu_{(1:N)}\right),\beta_j^{(r)}\right)\propto\left(\frac{1}{\Gamma(\alpha_j)}\exp\left[\lambda_{\alpha_j}^{post}\alpha_j\right]\right)^N\mathbb{I}_{[0,+\infty]}(\alpha_j), \quad (17)$$

where $\lambda_{\alpha_j}^{post}=\log\beta_j^{(r)}-\frac{1}{N}\lambda_{\alpha_j}^{prior}+\frac{1}{N}\sum_{k=1}^{N}\log S_j^{(r+1)}(\nu_k)$. The sampling from this distribution is achieved using a Metropolis-Hastings algorithm where an instrumental distribution is derived as a Gamma distribution with parameters calculated from the mode and the superior inflexion point of this distribution [8]. Concerning the hyperparameter β_j, its conditional posterior distribution is given as

$$p\left(\beta_j|S_j^{(r+1)}\left(\nu_{(1:N)}\right),\alpha_j^{(r+1)}\right)\propto\beta_j^{N\alpha_j^{(r+1)}}\exp\left[-\beta_j\sum_{k=1}^{N}S_j^{(r+1)}(\nu_k)\right]\cdot p(\beta_j). \quad (18)$$

The conjugate prior assigned to β_j is a Gamma density of parameters $\left(\alpha_{\beta_j}^{prior},\beta_{\beta_j}^{prior}\right)$. Therefore, its conditional posterior density is also a Gamma distribution with parameters $\alpha_{\beta_j}^{post}=\left(N\alpha_j^{(r+1)}+\alpha_{\beta_j}^{prior}+1\right)$ and $\beta_{\beta_j}^{post}=\left(\sum_{k=1}^{N}S_j^{(r+1)}\left(\nu_{(1:N)}\right)+\beta_{\beta_j}^{prior}\right)$.

Mixing coefficient hyperparameters. The mixing coefficient hyperparameters are sampled using the same manner as the hyperparameters of the source signals.

EXPERIMENTS

To illustrate the usefulness of the proposed method, we present two results obtained with numerical and experimental mixtures. To measure the estimation performances we use the performance index, noted PI, and the cross-talk, noted CT, which are defined by

$$\mathrm{PI} = \frac{1}{2}\sum_{i=1}^{n}\left\{\left(\sum_{k=1}^{n}\frac{|G_{(i,k)}|^2}{\max\limits_{\ell}|G_{(i,\ell)}|^2}-1\right)+\left(\sum_{k=1}^{n}\frac{|G_{(k,i)}|^2}{\max\limits_{\ell}|G_{(\ell,i)}|^2}-1\right)\right\} \tag{19}$$

$$\mathrm{CT}_{S_j} = \sum_{k=1}^{N}\left(S_j(\nu_k)-\hat{S}_j(\nu_k)\right)^2, \tag{20}$$

where $G_{(i,j)}$ are elements of the matrix $\underline{G} = \underline{\hat{A}}^{\#}\underline{A}$. The PI measures the overall separation performances and indicates mainly the mixing matrix estimation quality, while the cross-talk assesses the accuracy of the source signal reconstruction. In the following, the two indexes are expressed in dB.

The first data set is obtained by mixing three simulated non-negative signals that are similar to real spectra. The mixing coefficients are also chosen in such a way to get an evolution similar to what we get in chemical reactions. Figure 1 shows the source signals, the mixing coefficients and the resulting mixtures for an SNR of 20 dB.

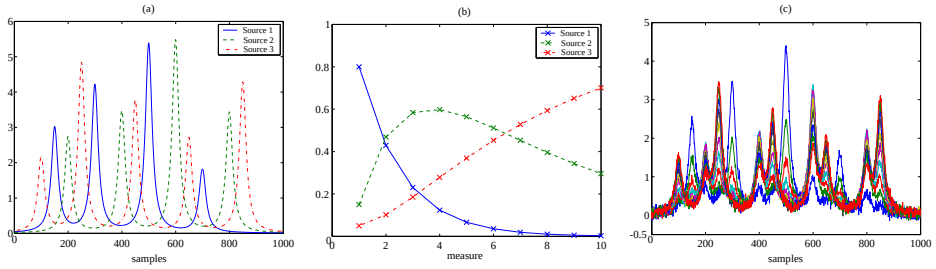

FIGURE 1. (a) Source signals, (b) mixing coefficients and (c) resulting mixtures for SNR = 20 dB

The second experiment consists in mixing three known chemical species (cyclopentane, cyclohexane and n–pentane) and the mixture data are obtained by near infrared (NIR) spectroscopy measurements. These species have been chosen because their spectra in the NIR frequency band are highly overlapping which makes the separation difficult and they do not interact when they are mixed, guaranteing that no new component appears. The pure spectra and concentration are shown in figure (2).

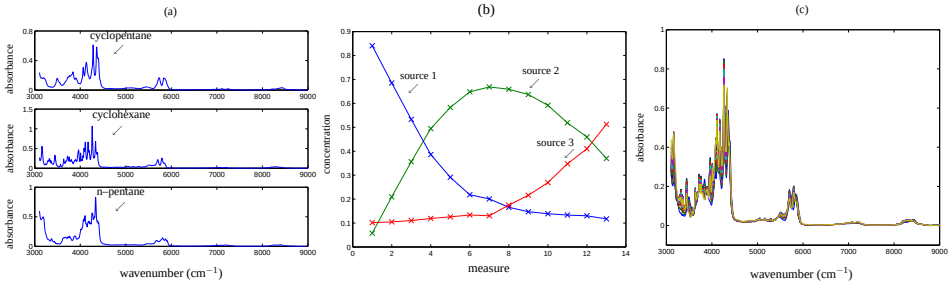

FIGURE 2. (a) Constituent spectra, (b) concentration profiles and measured mixture data

TABLE 1. Comparison of the separation performances using different methods

	Simulated data				Experimental data			
	NNICA	**NMF**	**ALS**	**BPSS**	**NNICA**	**NMF**	**ALS**	**BPSS**
$CT_{Source\ 1}$	-12.95	-13.64	-17.65	-20.99	-4.82	-14.20	-15.18	-33.23
$CT_{Source\ 2}$	-10.70	-11.70	-11.11	-19.94	-5.64	-17.50	-23.43	-24.98
$CT_{Source\ 3}$	-19.93	-19.83	-21.76	-19.11	-4.77	-17.88	-14.01	-26.05
PI	-10.03	-9.60	-12.01	-18.41	-1.02	-11.60	-8.10	-19.22

Table 1 summarizes the performances of the analysis of these mixtures using different methods. NNICA is the non-negative independent component analysis method [10] and BPSS (for Bayesian positive source separation) refers to the proposed approach. These results show the superior performances of the Bayesian separation approach.

CONCLUSION

The problem of non-negative source separation has been addressed in this paper. The Bayesian inference allows to consider the non-negativity as prior information which is encoded through the assignment of Gamma distribution priors. The result that has been presented illustrate that such prior distribution is very suitable for the separation of spectral source signals. To achieve a better fit of the source signal distributions, the proposed approach can be straightforwardly extended to a more general model consisting in mixtures of Gamma or truncated normal distributions.

ACKNOWLEDGMENTS

This work is supported by the "Region Lorraine" and the CNRS.

REFERENCES

1. Malinowski, E., *Factor Analysis in Chemistry*, John Willey, 1991.
2. Ricci, R., Ditzler, M., and Nestor, L., *Journal of Chemical Education*, **71**, 983–985 (1994).
3. Lawton, W., and Sylvestre, E., *Technometrics*, **13**, 617–633 (1971).
4. Tauler, R., Kowalski, B., and Fleming, S., *Analytical Chemistry*, **65**, 2040–2047 (1993).
5. Lawson, C., and Hanson, R., *Solving Least-Squares Problems*, Prentice-Hall, 1974.
6. Lee, D., and Seung, H., *Nature*, **401**, 788–791 (1999).
7. Moussaoui, S., Mohammad-Djafari, A., Brie, D., and Caspary, O., "A Bayesian Method for Positive Source Separation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, Montreal, Canada, 2004.
8. Moussaoui, S., Brie, D., Mohammad-Djafari, A., and Carteret, C., Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling (June 2004), CRAN, Internal Report.
9. Robert, C., *Statistics and Computing*, **5**, 121–125 (1995).
10. Plumbley, M., *IEEE Transactions on Neural Networks*, **14**, 534–543 (2003).