Bayesian approach for blind separation of underdetermined mixtures of sparse sources

Cédric Févotte, Simon J. Godsill, and Patrick J. Wolfe

Cambridge University Engineering Dept., Cambridge, CB2 1PZ, UK,
{cf269,sjg,pjw47}@eng.cam.ac.uk,
WWW home page: http://www-sigproc.eng.cam.ac.uk/~cf269/

Abstract. We address in this paper the problem of blind separation of underdetermined mixtures of sparse sources. The sources are given a Student t distribution, in a transformed domain, and we propose a bayesian approach using Gibbs sampling. Results are given on synthetic and audio signals.

1 Introduction

Blind Source Separation (BSS) consists in estimating n signals (the sources) from the sole observation of m mixtures of them (the observations). In this paper we consider linear instantaneous mixtures of time series: at each time index, the observations are a linear combination of the sources at the same time index. Moreover, we are interested in the underdetermined case ($m < n$). This case is very difficult to handle because contrary to (over)determined mixtures ($m \geq n$), estimating the mixing system (a single matrix in the linear instantaneous case) is not sufficient for reconstructing the sources, since for $m < n$ the mixing matrix is not invertible. Then, it appears that separation of underdetermined mixtures requires important prior information on the sources to allow their reconstruction.

In this paper we address the case of sparse sources, meaning that only a few samples are significantly non-zero. The use of sparsity to handle source separation problems has arisen in several papers, see for instance [1, 2]. In these papers, source time series are assumed to have a sparse representation on a given or learnt dictionary, possibly overcomplete. The aim of methods then becomes the estimation of the coefficients of the sources on the dictionary and not the time series in themselves. The time series are then reconstructed from the estimated coefficients.

More specifically, in [3, 4] the coefficients of the representations of the sources in the dictionary are given a discrete mixture a Gaussian distributions with 2 or 3 states (one Gaussian with very small variance, the other(s) with big variance) and a probabilistic framework is presented for the estimation of the mixing matrix and the sources. In particular, in [4], the authors use EM optimisation and present results with speech signals decomposed on a MDCT orthogonal basis [5]. The use of an orthogonal basis provides equivalence between representations in the time domain and transformed domain, and separation can be simply performed in the transformed domain instead of the time domain. The use of an overcomplete dictionary is very appealing because it allows sparser representations but leads to much more tricky calculations [1].

Motivated by the successful results of Student t modeling for audio restoration in [6], we address in this paper separation of Student t distributed sources, which leads to sparse modeling when the degrees of freedom is low. We emphasize that we will work in the transformed domain: the observations and sources we consider have arisen from the decomposition of some corresponding time series on a dictionary, which is restricted at this point to be an *orthogonal basis* (to satisfy equivalence between time and transformed domains). The method we present is a bayesian approach: a Gibbs sampler is derived to sample from the posterior conditional distribution of the parameters (which include the sources and the mixing matrix).

The paper is organised as follows: section 2 introduces notations and assumptions, in section 3 we derive the posterior distributions of the parameters to be estimated and section 4 presents results on synthetic and audio signals. Conclusions and perspectives are given in section 5.

2 Model and assumptions

2.1 Model

We consider the following standard linear model, $\forall t = 0, \ldots, N-1$:

$$\mathbf{x}_t = \mathbf{A}\,\mathbf{s}_t + \mathbf{n}_t \tag{1}$$

where $\mathbf{x}_t = [x_{1,t}, \ldots, x_{m,t}]^T$ is vector of size m containing the observations, $\mathbf{s}_t = [s_{1,t}, \ldots, s_{n,t}]^T$ is a vector of size n containing the sources, $\mathbf{n}_t = [n_{1,t}, \ldots, n_{m,t}]^T$ is a vector of size m containing noise. Variables without time index t denote whole sequences of samples, *e.g*, $\mathbf{x} = [\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}]$ and $x_1 = [x_{1,0}, \ldots, x_{1,N-1}]$.

2.2 Assumptions

1) We assume that each source sequence s_i is independently and identically distributed (i.i.d), with Student t distribution $t(\alpha_i, \lambda_i)$:

$$p(s_{i,t}) = K\left(1 + \frac{1}{\alpha_i}\left(\frac{s_{i,t}}{\lambda_i}\right)^2\right)^{-\frac{\alpha_i+1}{2}} \tag{2}$$

α_i is the "degrees of freedom", λ_i is a scale parameter and K is a normalizing constant. With $\lambda = 1$ and $\alpha = 1$, the Student t distribution is equal to the standard Cauchy distribution, and it tends to the standard Gaussian distribution when $\alpha \to +\infty$. Fig. 1 plots Student t densities for several values of α. For small α, the Student t has "fatter tails" than the normal distribution. A nice property of the Student t distribution is the fact that it can be expressed as a Scaled Mixture of Gaussians [7], such that

$$p(s_{i,t}) = \int_0^{+\infty} \mathcal{N}(s_{i,t}|0, v_{i,t})\,\mathcal{IG}\left(v_{i,t}|\frac{\alpha_i}{2}, \frac{2}{\alpha_i\,\lambda_i^2}\right) dv_{i,t} \tag{3}$$

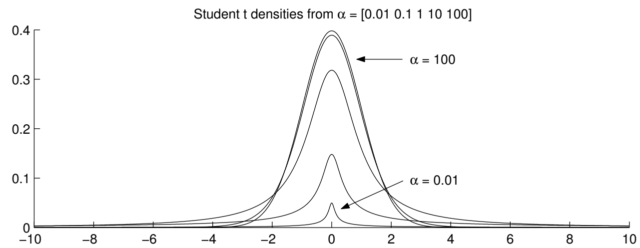

Fig. 1. Student t densities for $\lambda = 1$ and $\alpha = [0.01, 0.1, 1, 10, 100]$ - From $\alpha = 100$ $t(\alpha,1)$ is very close to $\mathcal{N}(0,1)$.

where $\mathcal{N}(x|0,v)$ denotes the normal distribution with mean 0 and variance v and $\mathcal{IG}(x|\gamma,\beta)$ denotes Inverted Gamma distribution, defined by $\mathcal{IG}(x|\gamma,\beta) = (x^{-(\gamma+1)})/(\Gamma(\gamma)\,\beta^{\gamma})\,\exp(-1/(\beta\,x))$, for $x \geq 0$.

$p(s_{i,t})$ can thus be interpreted as a marginal of the joint distribution $p(s_{i,t}, v_{i,t})$, defined by:

$$p(s_{i,t}, v_{i,t}) = p(s_{i,t}|v_{i,t})\,p(v_{i,t}|\alpha_i, \lambda_i) \tag{4}$$

with:

$$p(s_{i,t}|v_{i,t}) = \mathcal{N}(s_{i,t}|0, v_{i,t}) \quad \text{and} \quad p(v_{i,t}|\alpha_i, \lambda_i) = \mathcal{IG}\left(v_{i,t}|\frac{\alpha_i}{2}, \frac{2}{\alpha_i\,\lambda_i^2}\right) \tag{5}$$

The fact that $s_{i,t}$ is Gaussian conditionally upon the variance $v_{i,t}$ of the Gaussian distribution in Eq. (3) is a very convenient property which will help us deriving posterior distributions of the parameters in the implementation of the Gibbs sampler. The Student t can be interpreted as an infinite sum of Gaussians, which contrasts with the finite sums of Gaussians used in [3, 4].

In the following, we note $\mathbf{v}_t = [v_{1,t}, \ldots, v_{n,t}]^T$, $\mathbf{v} = [\mathbf{v}_0, \ldots, \mathbf{v}_{N-1}]$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]$.

2) We assume that the source sequences are mutually independent, such that $p(\mathbf{s}) = \prod_{i=1}^{n} p(s_i)$.

3) We assume that $\mathbf{n}$ is a i.i.d Gaussian noise with covariance $\sigma^2\,\mathbf{I}_m$, and σ unknown.

We now present a Markov chain Monte Carlo approach to estimate the set of parameter of interest $\{\mathbf{A}, \mathbf{s}, \sigma\}$ together with the set $\{\mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}\}$.

3 Derivations for the Gibbs sampler

We propose to generate samples from the posterior distribution $p(\mathbf{A}, \mathbf{s}, \sigma, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}|\mathbf{x})$ of whole set of parameters. We use a Gibbs sampler which only requires to derive the expression of the posterior distribution of each parameter conditionally upon the data $\mathbf{x}$ and the other parameters, see details for instance in [8].

3.1 Likelihood

With the gaussian noise assumption, the likelihood of one sample of the observations is written:

$$p(\mathbf{x}_t|\mathbf{A}, \mathbf{s}_t, \sigma) = \mathcal{N}(\mathbf{x}_t|\mathbf{A}\,\mathbf{s}_t, \sigma^2\,\mathbf{I}_m) \tag{6}$$

where $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. With the i.i.d source assumption, the likelihood of the observations is written:

$$p(\mathbf{x}|\mathbf{A}, \mathbf{s}, \sigma) = \prod_{t=0}^{N-1} \mathcal{N}(\mathbf{x}_t|\mathbf{A}\,\mathbf{s}_t, \sigma^2\,\mathbf{I}_m) \tag{7}$$

$$= \frac{1}{(2\,\pi\,\sigma^2)^{\frac{N\,m}{2}}} \exp(-\frac{1}{2\,\sigma^2} \sum_{t=0}^{N-1} \|\mathbf{x}_t - \mathbf{A}\,\mathbf{s}_t\|_F^2) \tag{8}$$

3.2 Expression of $p(\mathbf{s}|\mathbf{A}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$

We have:

$$p(\mathbf{s}|\mathbf{A}, \sigma, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto p(\mathbf{x}|\mathbf{A}, \mathbf{s}, \sigma)\, p(\mathbf{s}|\mathbf{v}) \tag{9}$$

With $p(\mathbf{s}|\mathbf{v}) = \prod_{t=0}^{N-1} p(\mathbf{s}_t|\mathbf{v}_t) = \prod_{t=0}^{N-1} \mathcal{N}(\mathbf{s}_t|0, \mathrm{diag}\,(\mathbf{v}_t))$ [1] and with Eq. (7), we have

$$p(\mathbf{s}|\mathbf{A}, \sigma, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{t=0}^{N-1} \mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{\mathbf{s}_t}, \boldsymbol{\Sigma}_{\mathbf{s}_t}) \tag{10}$$

where $\boldsymbol{\Sigma}_{\mathbf{s}_t} = \left(\frac{1}{\sigma^2}\,\mathbf{A}^T\,\mathbf{A} + \mathrm{diag}\,(\mathbf{v}_t)^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_{\mathbf{s}_t} = \frac{1}{\sigma^2}\,\boldsymbol{\Sigma}_{\mathbf{s}_t}\,\mathbf{A}^T\,\mathbf{x}_t$.

3.3 Expression of $p(\mathbf{A}|\mathbf{s}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$

Let $\mathbf{r}_1, \ldots, \mathbf{r}_m$ denote the transposed rows of $\mathbf{A}$, such that $\mathbf{A}^T = [\mathbf{r}_1\ \ldots\ \mathbf{r}_m]$. Let $\mathbf{S}_t$ and $\mathbf{a}$ denote the $m \times n\,m$ matrix and the $n\,m \times 1$ vector defined by

$$\mathbf{S}_t = \begin{bmatrix} \mathbf{s}_t^T & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{s}_t^T \end{bmatrix} \quad \text{and} \quad \mathbf{a} = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{bmatrix} \tag{11}$$

By construction, we have:

$$\mathbf{A}\,\mathbf{s}_t = \mathbf{S}_t\,\mathbf{a} \tag{12}$$

Of course, the estimation of $\mathbf{a}$ is equivalent to the estimation of $\mathbf{A}$, and we have:

$$p(\mathbf{a}|\mathbf{s}, \sigma, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto p(\mathbf{x}|\mathbf{a}, \mathbf{s}, \sigma)\, p(\mathbf{a}) \tag{13}$$

[1] $\mathrm{diag}\,(\mathbf{u})$ is the diagonal matrix whose main diagonal is given by $\mathbf{u}$

Without further information on the mixing matrix, we assume uniform prior and set $p(\mathbf{a}) \propto 1$. With Eq. (7) and (12), we have then $p(\mathbf{a}|\mathbf{s}, \sigma, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto \prod_{t=0}^{N-1} \mathcal{N}(\mathbf{x}_t|\mathbf{S}_t\,\mathbf{a}, \sigma^2\,\mathbf{I}_m)$ and it follows that:

$$p(\mathbf{a}|\mathbf{s}, \sigma, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \mathcal{N}(\mathbf{a}|\boldsymbol{\mu}_{\mathbf{a}}, \boldsymbol{\Sigma}_{\mathbf{a}}) \tag{14}$$

with $\boldsymbol{\Sigma}_{\mathbf{a}} = \sigma^2 \left(\sum_{t=0}^{N-1} \mathbf{S}_t^T \mathbf{S}_t\right)^{-1}$ and $\boldsymbol{\mu}_{\mathbf{a}} = \frac{1}{\sigma^2}\,\boldsymbol{\Sigma}_{\mathbf{a}} \sum_{t=0}^{N-1} \mathbf{S}_t^T\,\mathbf{x}_t$. To fix the well known BSS indeterminacies on gain and permutations, we set in practice the first row of $\mathbf{A}$ to ones and only estimate the other rows.

3.4 Expression of $p(\sigma|\mathbf{A}, \mathbf{s}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$

As before we have:

$$p(\sigma|\mathbf{A}, \mathbf{s}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto p(\mathbf{x}|\mathbf{A}, \mathbf{s}, \sigma)\, p(\sigma) \tag{15}$$

Using Jeffreys prior $p(\sigma) = 1/\sigma$ and expression (8) of the likelihood, we have:

$$p(\sigma|\mathbf{A}, \mathbf{s}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto \sigma^{-(2\,\gamma_\sigma+1)}\, \exp\left(-\frac{1}{\beta_\sigma\,\sigma^2}\right) \tag{16}$$

with $\gamma_\sigma = \frac{m\,N}{2}$ and $\beta_\sigma = 2/\sum_{t=0}^{N-1} \|\mathbf{x}_t - \mathbf{A}\,\mathbf{s}_t\|_F^2$. It appears that $\sigma|\mathbf{A}, \mathbf{s}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}$ can be drawn from $\sqrt{\mathcal{IG}(\gamma_\sigma, \beta_\sigma)}$.

3.5 Expression of $p(\mathbf{v}|\mathbf{A}, \mathbf{s}, \sigma, \boldsymbol{\alpha}, \boldsymbol{\lambda})$

Since the data $\mathbf{x}$ does not depend on the parameters $\{\mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\lambda}\}$, their posterior distributions only depend on the prior distributions. The posterior distribution of $\mathbf{v}$ is then:

$$p(\mathbf{v}|\mathbf{A}, \mathbf{s}, \sigma, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto p(\mathbf{s}|\mathbf{v})\, p(\mathbf{v}|\boldsymbol{\alpha}, \boldsymbol{\lambda}) \tag{17}$$

With $p(\mathbf{s}|\mathbf{v})\, p(\mathbf{v}|\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{t=0}^{N-1} \prod_{i=1}^{n} p(s_{i,t}|v_{i,t})\, p(v_{i,t}|\alpha_i, \lambda_i)$ one can show that:

$$p(s_{i,t}|v_{i,t})\, p(v_{i,t}|\alpha_i, \lambda_i) \propto \mathcal{IG}\left(v_{i,t}|\gamma_{v_i}, \beta_{v_{i,t}}\right) \tag{18}$$

with $\gamma_{v_i} = (\alpha_i + 1)/2$ and $\beta_{v_{i,t}} = 2/(s_{i,t}^2 + \alpha_i\,\lambda_i^2)$. Thus:

$$p(\mathbf{v}|\mathbf{A}, \mathbf{s}, \sigma, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{t=0}^{N-1} \prod_{i=1}^{n} \mathcal{IG}\left(v_{i,t}|\gamma_{v_i}, \beta_{v_{i,t}}\right) \tag{19}$$

3.6 Expression of $p(\boldsymbol{\alpha}|\mathbf{A}, \mathbf{s}, \sigma, \mathbf{v}, \boldsymbol{\lambda})$

We have:

$$p(\boldsymbol{\alpha}|\mathbf{A}, \mathbf{s}, \sigma, \mathbf{v}, \boldsymbol{\lambda}) \propto p(\mathbf{v}|\boldsymbol{\alpha}, \boldsymbol{\lambda})\, p(\boldsymbol{\alpha}) \tag{20}$$

With $p(\mathbf{v}|\boldsymbol{\alpha},\boldsymbol{\lambda})\,p(\boldsymbol{\alpha}) = \prod_{i=1}^{n}\prod_{t=0}^{N-1} p(v_{i,t}|\alpha_i,\lambda_i)p(\alpha_i)$, one can show that

$$p(\boldsymbol{\alpha}|\mathbf{A},\mathbf{s},\sigma,\mathbf{v},\boldsymbol{\lambda}) \propto \prod_{i=1}^{n} \frac{P_i^{-(\frac{\alpha_i}{2}+1)}}{\Gamma(\frac{\alpha_i}{2})^N} \left(\frac{\alpha_i\,\lambda_i^2}{2}\right)^{\frac{\alpha_i\,N}{2}} \exp\left(-\frac{\alpha_i\,\lambda_i^2}{2}\,S_i\right) p(\alpha_i) \quad (21)$$

with $S_i = \sum_{t=0}^{N-1}\frac{1}{v_{i,t}}$ and $P_i = \prod_{t=0}^{N-1} v_{i,t}$. In practice we choose a uniform prior on α_i and set $p(\boldsymbol{\alpha}) \propto 1$. As the distribution of $\boldsymbol{\alpha}|\mathbf{A},\mathbf{s},\sigma,\mathbf{v},\boldsymbol{\lambda}$ is not straightforward to sample from and since the precise value α_i for each source is unlikely to be important, we sample $\boldsymbol{\alpha}$ from a uniform grid of discrete values with probability mass given by Eq. (21).

3.7 Expression of $p(\boldsymbol{\lambda}|\mathbf{A},\mathbf{s},\boldsymbol{\sigma},\mathbf{v},\boldsymbol{\alpha})$

Finally, the posterior distribution of the scale parameters is given by:

$$p(\boldsymbol{\lambda}|\mathbf{A},\mathbf{s},\sigma,\mathbf{v},\boldsymbol{\alpha}) \propto p(\mathbf{v}|\boldsymbol{\alpha},\boldsymbol{\lambda})\,p(\boldsymbol{\lambda}) \quad (22)$$

With $p(\mathbf{v}|\boldsymbol{\alpha},\boldsymbol{\lambda})\,p(\boldsymbol{\lambda}) = \prod_{i=1}^{n}\left(\prod_{t=0}^{N-1} p(v_{i,t}|\alpha_i,\lambda_i)\right)p(\lambda_i)$, one can show that:

$$p(\boldsymbol{\lambda}|\mathbf{A},\mathbf{s},\sigma,\mathbf{v},\boldsymbol{\lambda}) \propto \prod_{i=1}^{n} \lambda_i^{\alpha_i\,N}\,\exp\left(-\frac{\alpha_i\,S_i}{2}\,\lambda_i^2\right)p(\lambda_i) \quad (23)$$

With Jeffreys prior $p(\lambda_i) = 1/\lambda_i$, it appears that $\lambda_i|\mathbf{A},\mathbf{s},\sigma,\mathbf{v},\boldsymbol{\lambda}$ can be drawn from $\sqrt{\mathcal{G}(\gamma_{\lambda_i},\beta_{\lambda_i})}$, with $\gamma_{\lambda_i} = (\alpha_i\,N)/2$ and $\beta_{\lambda_i} = 2/(\alpha_i\,S_i)$, and where $\mathcal{G}(\gamma,\beta)$ is the Gamma distribution, whose density is written $\mathcal{G}(x|\gamma,\beta) = x^{\gamma-1}/(\Gamma(\gamma)\,\beta^{\gamma})\,\exp(-x/\beta)$, for $x \geq 0$.

4 Results

Synthetic signals We present results of the method over a mixture of $n = 2$ Student t sources of length $N = 1000$ with $m = 3$ observations. The mixing matrix is arbitrarly chosen as $\mathbf{A} = [1\ 1\ 1; 1\ -0.5\ 0.2]$. The sources are simulated with $\boldsymbol{\alpha} = [0.9\ 0.7\ 0.8]$ and $\boldsymbol{\lambda} = [0.03\ 0.003\ 0.002]$. The values of $\boldsymbol{\alpha}$ are chosen according to a range of values that seem to fit reasonably well MDCT coefficients of several types of audio signals. The values of the scale parameters $\boldsymbol{\lambda}$ are of little importance. Noise was added on the observations with variance $\sigma = 0.1$, which leads to 35dB and 30dB SNR on each observation. We ran 10000 iterations of the Gibbs sampler. The convergence of $\mathbf{r}_2$ [2] (initialized to zeros), σ (initialised with random value between 0 and 1), $\boldsymbol{\alpha}$ (initialised to ones) and $\boldsymbol{\lambda}$ (initialised to $[0.01\ \ 0.01\ \ 0.01]$) is shown on Fig. 2.

Estimated sources were computed as mean estimates of the 2000 last sampled values of $\mathbf{s}$ (that is after convergence of all the values of $\mathbf{r}_2$ is obtained).

[2] We recall that $\mathbf{r}_1$ is set to ones.

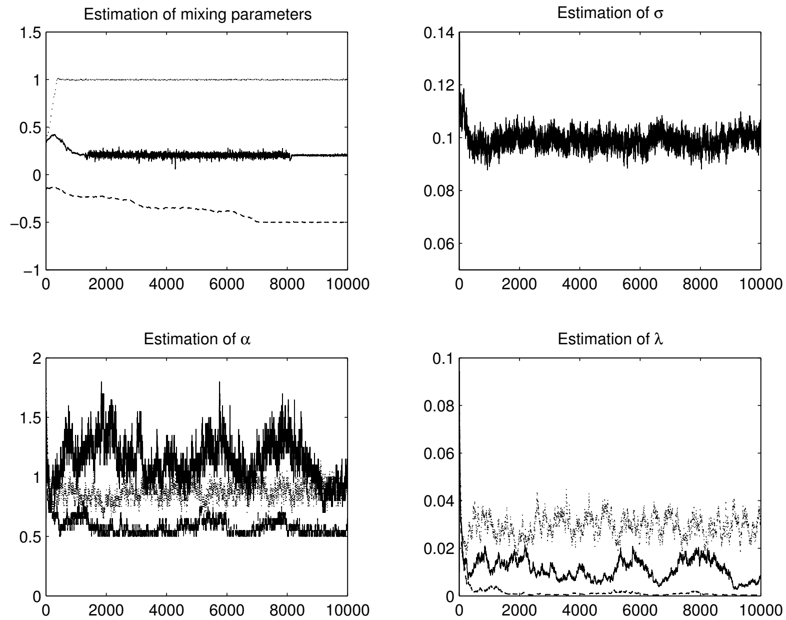

Fig. 2. Estimation of $\mathbf{A}$, σ, $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$ with Gibbs sampler

Sources estimates are compared to the original ones by computing the evaluation criteria described in [9]: Source to Distortions Ratio (global criterion), Source to Interference Ratio, Source to Noise Ratio, Source to Artifact Ratio. We obtain (values in dB):

	SDR	SIR	SNR	SAR
$\hat{s}_1$	30.6	40.2	33.7	34.6
$\hat{s}_2$	40.36	47.6	41.6	52.1
$\hat{s}_3$	26.57	44.8	27.0	37.7

With 30 dB corresponding to hearing threshold, the estimates are very good. Furthermore, one can see from Fig. 2 that mixing parameters converge to the exact values of the mixing matrix. The noise variance σ^2 converge to its true value within only a few samples. Besides, the sampled values of $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ show high variance, but considering the quality of the sources estimates, their precise values are of little importance.

Audio signals We have applied our model and method a mixture of three musical signals (s_1 = cello, s_2 = percussions, s_3 = piano) with two observations and $\approx$ 20dB SNR on each observation. Separation was performed on MDCT coefficients of the original signals ($\approx 3s$ sampled at 8000Hz) with window length equal to 128 samples (16ms). 5000 iterations of the sampler were run, convergence was obtained after $\approx$ 2000 iterations. Mixing matrix was chosen as in previous section. Audio samples can be listened to at `http://www-sigproc.eng.cam.ac.uk/~cf269/ica04/sound_files.html`. The obtained performance criteria are:

	SDR	SIR	SNR	SAR
$\hat{s}_1$	11.6	16.6	29.2	13.6
$\hat{s}_2$	1.3	10.8	27.2	2.1
$\hat{s}_3$	4.1	8.7	28.7	6.5

5 Conclusion

Good results of section 4 show the relevance of the bayesian approach to handle separation of underdetermined mixtures of sparse sources. The quality of the audio estimates is average due to a high amount of artifacts, but interference rejection is good. However there is room for improvement. Indeed, the method can be extended to overcomplete dictionaries, and other prior distributions of the coefficients of the decomposition of the sources can be used. For example generalised Gaussian distributions family can be used easily as they can be expressed as scaled mixtures of Gaussians too.

Motivated by these promising results, the next step is to study what kind of prior and dictionary can be used with a particular type of signal.

Acknowledgements C. Févotte and S. J. Godsill acknowledge the partial support of EU RTN MOUMIR (HP-99-108). P. J. Wolfe and S. J. Godsill acknowledge partial support from EPSRC ROPA Project 67958 "High Level Modelling and Inference for audio signals using Bayesian atomic decompositions". Many thanks to Laurent Daudet for providing us with MDCT code.

References

1. Zibulevsky, M., Pearlmutter, B.A., Bofill, P., Kisilev, P.: Blind source separation by sparse decomposition. In Roberts, S.J., Everson, R.M., eds.: Independent Component Analysis: Principles and Practice. Cambridge University Press (2001)
2. Lewicki, M.S., Sejnowski, T.J.: Learning overcomplete representations. Neural Computations **12** (2000) 337–365
3. Olshausen, B.A., Millman, K.J.: Learning sparse codes with a mixture-of-Gaussians prior. In S. A. Solla, T.K.L., ed.: Advances in Neural Information Processing Systems. MIT press (2000) 841–847
4. Davies, M., Mitianoudis, N.: A simple mixture model for sparse overcomplete ICA. IEE Proceedings on Vision, Image and Signal Processing (2004)
5. Mallat, S.: A wavelet tour of signal processing. Academic Press (1998)
6. Wolfe, P.J., Godsill, S.J., Ng, W.J.: Bayesian variable selection and regularisation for time-frequency surface estimation. J. R. Statist. Soc. **B** (2004)
7. Andrews, D.F., Mallows, C.L.: Scale mixtures of normal distributions. J. R. Statist. Soc. **B** (1974) 99–102
8. Gilks, W.R., Richardson, S., Spiegelhalter, D.J.: Markov Chain Monte Carlo in Practice. Chapman & Hall (1996)
9. Gribonval, R., Benaroya, L., Vincent, E., Févotte, C.: Proposals for performance measurement in source separation. In: Proc. 4^{th} Symposium on Independent Component Analysis and Blind Source Separation (ICA'03), Nara, Japan (2003)