Blind Separation of Sparse Sources Using Jeffrey's Inverse Prior and the EM Algorithm

Cédric Févotte and Simon J. Godsill

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

Abstract. In this paper we study the properties of the Jeffrey's inverse prior for blind separation of sparse sources. This very sparse prior was previously used for Wavelet-based image denoising. In this paper we consider separation of 3×3 and 2×3 noisy mixtures of audio signals, decomposed on a MDCT basis. The hierarchical formulation of the inverse prior allows for EM-based computation of MAP estimates. This procedure happens to be fast when compared to a standard more complex Markov chain Monte Carlo method using the flexible Student t prior, with competitive results obtained.

Blind Source Separation (BSS) consists of estimating n signals (the sources) from the sole observation of m mixtures of them (the observations). If many efficient approaches exist for (over)determined ($m \geq n$) non-noisy linear instantaneous, in particular within the field of Independent Component Analysis, the general linear instantaneous case, with mixtures possibly noisy and/or underdetermined ($m < n$) is still a very challenging problem.

A now common approach for BSS, in particular for underdetermined mixtures, consists of exploiting source sparsity assumptions. Sparsity means that only "few" expansion coefficients of the sources on a given basis are significantly different from zero and its use to handle source separation problem (possibly underdetermined) was introduced in the seminal papers [1,2].

In [3,4] we modeled the expansion coefficients of the sources by identically and independently distributed (i.i.d) Student t processes and a Gibbs sampler (a standard MCMC simulation method) was proposed to sample from the posterior distribution of the mixing matrix, the input noise variance, the source coefficients and hyperparameters of the Student t distributions. The method was successfully applied to determined and underdetermined noisy audio mixtures, decomposed on a MDCT basis (a local cosine basis). In this paper, we give the source coefficients the Jeffrey's inverse prior $p(x) \propto 1/|x|$. This prior was used for image denoising and sparse regression in [5,6]. It provides very sparse signal estimates and, as shown in [5] in the context of denoising, is good compromise between soft and hard thresholding. Though Jeffrey's prior corresponds to an improper probability density function, it admits a hierarchical formulation which leads to proper posterior densities, and allows for efficient EM-based computa-

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 593–600, 2006.

tion of Maximum A Posteriori (MAP) estimates of the source coefficients, the mixing matrix and the noise variance.

The paper is organized as follows: Section 1 introduces notations and assumptions. Section 2 presents the different priors used for the source coefficients, the mixing matrix, and the input noise variance. Section 3 gives the EM updates of each of the latter parameters. Section 4 provides separation results for determined and underdetermined mixtures of audio sources. The Jeffrey's prior is shown to have good denoising properties, and the proposed method happens to be fast when compared to the more complex MCMC approach using the flexible Student t prior proposed in [4], with competitive results obtained. Section 5 draws some conclusions.

1 Notations

1.1 Mixture and Aim

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

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ is a 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 and $\mathbf{n}(t) = [n_1(t), \ldots, n_m(t)]^T$ is a vector of size m containing additive noise. Variables without time index t denote whole sequences of samples, *e.g*, $\mathbf{x} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$ and $x_1 = [x_1(1), \ldots, x_1(N)]$.

The aim of the following work is to estimate the sources $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ up to the standard BSS indeterminacies on gain and order, that is, compute $\hat{\mathbf{s}}$ and $\hat{\mathbf{A}}$ such that ideally $\hat{\mathbf{A}} = \mathbf{A}\,\mathbf{D}\,\mathbf{P}$ and $\hat{\mathbf{s}} = \mathbf{P}^T\,\mathbf{D}^{-1}\,\mathbf{s}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix.

1.2 Time Domain / Transform Domain

Let $x \in \mathbb{R}^{1\times N} \rightarrow \tilde{x} \in \mathbb{R}^{1\times N}$ denote a bijective linear transform, preferably orthonormal. Denoting for $k = 1, \ldots, N$, $\tilde{\mathbf{x}}_k = [\tilde{x}_{1,k}, \ldots, \tilde{x}_{1,m}]^T$ and $\tilde{\mathbf{n}}_k$, $\tilde{\mathbf{s}}_k$ similarly, by linearity of the t-f transform we have

$$\tilde{\mathbf{x}}_k = \mathbf{A}\,\tilde{\mathbf{s}}_k + \tilde{\mathbf{n}}_k \tag{2}$$

Furthermore, the t-f transform being bijective, solving the problem defined by Eq. (1) in the time domain is equivalent to solving Eq. (2) in the transform domain.

1.3 Some Assumptions

We make the following assumptions:

- $\mathbf{A}$ is full-column rank,
- The source coefficients $\{\tilde{s}_{i,k}\}$ are assumed to be *sparse* and mutually independent,

– The noise $\mathbf{n}_t$ is assumed to be i.i.d Gaussian with covariance $\sigma^2\,\mathbf{I}_n$. When the transform $x \rightarrow \tilde{x}$ is orthonormal, $\tilde{\mathbf{n}}_k$ is equivalently i.i.d Gaussian with covariance $\sigma^2\,\mathbf{I}_n$.

2 Priors

2.1 Source Coefficients

The source coefficients are given Jeffrey's inverse prior:

$$p(\tilde{s}_{i,k}) \propto \frac{1}{|\tilde{s}_{i,k}|} \tag{3}$$

This is a very heavy-tailed prior, symmetrical and centered at 0, and thus a relevant model for sparsity. This prior is *scale-invariant*: if $p(x) \propto 1/x$ and if $y = a\,x$ where a is a constant, then, by applying the rule for the change of variable in a pdf, $p(y) \propto 1/y$. Thus, oppositely to Student t or Laplace priors, the inverse prior does not require to update any scale parameter for the sources. As noted in [5], this prior is so heavy-tailed that it is actually *improper* (its integral is not finite). Improper priors are common in Bayesian inference, and can be used as long as they lead to proper posterior distributions. In fact the prior $p(\tilde{s}_{i,k}) \propto 1/|\tilde{s}_{i,k}|$ used with a Gaussian likelihood leads to an improper posterior distribution for the sources. Thus, as in [5,6], we rather use the following hierarchical formulation of the inverse prior:

$$p(\tilde{s}_{i,k}|v_{i,k}) = \mathcal{N}(s_{i,k}|0, v_{i,k}) \quad \text{with} \quad p(v_{i,k}) \propto 1/v_{i,k} \tag{4}$$

where $\mathcal{N}(x|\mu, \sigma^2)$ denotes the density of the normal distribution. With these assumptions, elementary integration yields

$$\int_0^\infty p(\tilde{s}_{i,k}|v_{i,k})\, p(v_{i,k})\; dv_{i,k} \propto 1/|\tilde{s}_{i,k}| \tag{5}$$

The inverse prior $p(v_{i,k}) \propto 1/v_{i,k}$ being itself a limiting case of the inverted-Gamma distribution, the prior $p(\tilde{s}_{i,k}) \propto 1/|\tilde{s}_{i,k}|$ may be regarded as a special case of the Student t prior which was used for source separation purposes in [4]. In the following we will denote $\mathbf{v}_k = [v_{1,k}, \ldots, v_{n,k}]^T$ and $\mathbf{v} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$.

2.2 Noise Variance

σ^2 is given a inverted-Gamma conjugate prior:

$$p(\sigma^2|\alpha_\sigma, \beta_\sigma) = \mathcal{IG}(\sigma^2|\alpha_\sigma, \beta_\sigma) \tag{6}$$

where $\mathcal{IG}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{-(\alpha+1)}\, \exp(-\frac{\beta}{x}),\ x \in [0, +\infty)$. The inverted-Gamma distribution has a unique mode, which is found at $x = \beta/(\alpha + 1)$.

2.3 Mixing Matrix

Let $\mathbf{r}_1, \ldots, \mathbf{r}_m$ be the $n \times 1$ vectors denoting the transposed rows of $\mathbf{A}$, such that $\mathbf{A}^T = [\mathbf{r}_1 \ \ldots \ \mathbf{r}_m]$. We give each row $\mathbf{r}_i$ a Gaussian conjugate prior with zero mean:

$$p(\mathbf{r}_j|\sigma^2) = \mathcal{N}(\mathbf{r}_j|0, \sigma_r^2 \, \mathbf{I}_n) \tag{7}$$

3 EM Updates

We now describe how to find a MAP estimates of the parameters $\boldsymbol{\theta} = \{\tilde{\mathbf{s}}, \mathbf{A}, \sigma^2\}$, using the Expectation Maximization algorithm (EM) [7]. With the variances $\mathbf{v}$ treated as missing data, the EM algorithm is based on the alternate evaluation and maximization of the following function:

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}') &= \int_{\mathbf{v}} \log p(\boldsymbol{\theta}|\mathbf{v}, \tilde{\mathbf{x}}) \; p(\mathbf{v}|\boldsymbol{\theta}', \tilde{\mathbf{x}}) \; d\mathbf{v} \\ &= \int_{\mathbf{v}} \log p(\tilde{\mathbf{s}}, \mathbf{A}, \sigma|\mathbf{v}, \tilde{\mathbf{x}}) \; p(\mathbf{v}|\tilde{\mathbf{s}}') \; d\mathbf{v} \\ &= \int_{\mathbf{v}} \log p(\tilde{\mathbf{s}}|\mathbf{A}, \sigma, \mathbf{v}, \tilde{\mathbf{x}}) \; p(\mathbf{v}|\tilde{\mathbf{s}}') \; d\mathbf{v} + \log p(\mathbf{A}|\tilde{\mathbf{s}}, \sigma^2, \tilde{\mathbf{x}}) + \log p(\sigma^2|\tilde{\mathbf{s}}, \tilde{\mathbf{x}}) \end{aligned} \tag{8}$$

One iteration of the EM algorithm is as follows:

$$\begin{aligned} \textbf{E-step} \quad & \text{Evaluate } Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) \\ \textbf{M-step} \quad & \boldsymbol{\theta}^{(l+1)} = \text{Argmax } Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) \end{aligned}$$

The derivation of the update steps for every parameter is mainly a matter of finding the modes of the posterior distributions involved in Eq. (8). In the following we skip derivations, details of the calculations of the posterior distributions can be found in [4].

3.1 Missing Data Posterior Distribution

The E-step requires integration over the posterior distribution of the missing data $\mathbf{v}$:

$$p(\mathbf{v}|\tilde{\mathbf{s}}) = \prod_{i,k} p(v_{i,k}|\tilde{s}_{i,k}) \quad \text{with} \quad p(v_{i,k}|\tilde{s}_{i,k}) = \mathcal{IG}\left(v_{i,k}|\frac{1}{2}, \frac{\tilde{s}_{i,k}^2}{2}\right) \tag{9}$$

3.2 Update of $\tilde{\mathbf{s}}$

The posterior distribution of $\tilde{\mathbf{s}}$ is $p(\tilde{\mathbf{s}}|\mathbf{A}, \sigma^2, \mathbf{v}, \tilde{\mathbf{x}}) = \prod_k p(\tilde{\mathbf{s}}_k|\mathbf{A}, \sigma^2, \mathbf{v}_k, \tilde{\mathbf{x}}_k)$ with

$$p(\tilde{\mathbf{s}}_k|\mathbf{A}, \sigma^2, \mathbf{v}_k, \tilde{\mathbf{x}}_k) = \mathcal{N}(\tilde{\mathbf{s}}_k|\boldsymbol{\mu}_{\tilde{\mathbf{s}}_k}, \boldsymbol{\Sigma}_{\tilde{\mathbf{s}}_k}) \tag{10}$$

where $\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}_k} = \left(\frac{1}{\sigma^2}\,\mathbf{A}^T\,\mathbf{A} + \mathrm{diag}\,(\mathbf{v}_k)^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_{\tilde{\mathbf{s}}_k} = \frac{1}{\sigma^2}\,\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}_k}\,\mathbf{A}^T\,\tilde{\mathbf{x}}_k$. The value of $\tilde{\mathbf{s}}_k$ is simply updated to the mode $\boldsymbol{\mu}_{\tilde{\mathbf{s}}_k}$ of the posterior distribution, with $\mathrm{diag}\,(\mathbf{v}_k)^{-1}$ being integrated over the missing data posterior distribution:

$$\tilde{\mathbf{s}}_k = \left(\mathbf{A}^T\,\mathbf{A} + \sigma^2\,\mathrm{diag}\left(\mathbf{c}_k^{'}\right)\right)^{-1}\,\mathbf{A}^T\,\tilde{\mathbf{x}}_k \tag{11}$$

where

$$\mathbf{c}_k = [c_{1,k}, \ldots, c_{n,k}]^T \quad \text{and} \quad c_{i,k} = \mathrm{E}\left\{\frac{1}{v_{i,k}}|\tilde{s}_{i,k}\right\} = \frac{1}{\tilde{s}_{i,k}^2} \tag{12}$$

3.3 Update of A

The rows of $\mathbf{A}$ are a posteriori mutually independent with

$$p(\mathbf{r}_j|\tilde{\mathbf{s}}, \sigma^2, \tilde{\mathbf{x}}) = \mathcal{N}(\mathbf{r}_j|\boldsymbol{\mu}_{\mathbf{r}_j}, \boldsymbol{\Sigma}_{\mathbf{r}}) \tag{13}$$

where $\boldsymbol{\Sigma}_{\mathbf{r}} = \left(\frac{1}{\sigma^2}\sum_k \tilde{\mathbf{s}}_k\,\tilde{\mathbf{s}}_k^T + \frac{1}{\sigma_r^2}\,\mathbf{I}_n\right)^{-1}$ and $\boldsymbol{\mu}_{\mathbf{r}_j} = \frac{1}{\sigma^2}\,\boldsymbol{\Sigma}_{\mathbf{r}}\,\sum_k \tilde{x}_{j,k}\,\tilde{\mathbf{s}}_k$. Row j is updated to the mode $\boldsymbol{\mu}_{\mathbf{r}_j}$ of the posterior distribution: [1]

$$\mathbf{r}_j = \left(\sum_k \tilde{\mathbf{s}}_k\,\tilde{\mathbf{s}}_k^T + \frac{\sigma^2}{\sigma_r^2}\,\mathbf{I}_n\right)^{-1}\,\sum_k \tilde{x}_{j,k}\,\tilde{\mathbf{s}}_k \tag{14}$$

3.4 Update of σ^2

The posterior distribution $p(\sigma^2|\tilde{\mathbf{s}}, \tilde{\mathbf{x}})$ is written

$$p(\sigma^2|\tilde{\mathbf{s}}, \tilde{\mathbf{x}}) = \mathcal{IG}(\sigma^2|\alpha, \beta) \tag{15}$$

with $\alpha = \frac{(N-n)m}{2} + \alpha_\sigma$ and

$$\beta = \sum_{j=1}^{m}\left(\left(\sum_k \tilde{x}_{j,k}^2\right) - \left(\sum_k \tilde{x}_{j,k}\tilde{\mathbf{s}}_k^T\right)\left(\sum_k \tilde{\mathbf{s}}_k\tilde{\mathbf{s}}_k^T\right)^{-1}\left(\sum_k \tilde{x}_{j,k}\tilde{\mathbf{s}}_k\right)\right) + \beta_\sigma \tag{16}$$

σ^2 is updated to the mode of the distribution:

$$\sigma^2 = \frac{\beta}{\alpha + 1} \tag{17}$$

4 Results

We present results of separation of 3×3 and 2×3 mixtures of audio sources. Results are discussed in Section 5.

[1] In practice the columns of $\mathbf{A}$ are normalized to 1 after each iteration to solve the BSS indeterminacy on gain.

4.1 Determined Mixture

We first study a mixture of $n = 3$ audio sources (speech, piano, guitar) with $m = 3$ observations. The mixing matrix is given in Table 1. We set $\sigma = 0.1$, which corresponds to approximatively 9.5dB noise on each observation. The signals are sampled at 8kHz with length $N = 65356$ ($\approx 8s$). We used a MDCT basis to decompose the observations, using a sine bell and 50% overlap, yielding a time resolution of 64ms (half the window length). The MDCT is a local cosine transform known to provide sparse representations of audio signals [8]. Time resolutions of 32ms and 128ms were also tried but led to slightly poorer results. The proposed method was compared with the method in [4], which uses a Student t prior on the sources and computes MMSE estimate of the parameters using a Gibbs sampler. We also show the results provided by the standard ICA algorithm JADE [9] applied to $\tilde{\mathbf{x}}$, which estimates a separating matrix via joint-diagonalization of a set of cumulant matrices, and apply the obtained matrix to the data, without denoising of the sources estimates. In our EM algorithm $\mathbf{A}$ is initialized with the JADE estimate and σ^2 is initialized to 1. We used noninformative priors $\sigma_r = +\infty$ and $\alpha_\sigma = \beta_\sigma = 0$ for $\mathbf{A}$ and σ^2.

Table 1 shows the mixing matrices estimated by the two methods. Table 2 provides separation quality criteria for the sources estimates. The criteria used are described in [10]. Basically, the SDR (Source to Distortion Ratio) provides an overall separation performance criterion, the SIR (Source to Interferences Ratio) measures the level of interferences from the other sources in each source estimate, SNR (Source to Noise Ratio) measures the error due to the additive noise on the sensors and the SAR (Source to Artifacts Ratio) measures the level of artifacts in the source estimates. The higher are the ratios, the better is quality of estimation. We point out that the performance criteria are invariant to a change of basis, so that figures can be computed either on the time sequences ($\hat{\mathbf{s}}$ compared to $\mathbf{s}$) or the MDCT coefficients ($\hat{\tilde{\mathbf{s}}}$ compared to $\tilde{\mathbf{s}}$). The estimated sources can be listened to at `http://www-sigproc.eng.cam.ac.uk/~cf269/ica06/sound_files.html`, which is perhaps the best way to assess the audio quality of the results.

Table 1. Estimates of $\mathbf{A}$ for the determined mixture

Original matrix	Jeffrey + EM
$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0.8 & 1.3 & -0.9 \\ 1.2 & -0.7 & 1.1 \end{bmatrix}$	$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 \\ 0.8090 & 1.3097 & -0.8922 \\ 1.1921 & -0.7341 & 1.0593 \end{bmatrix}$
t + MCMC	JADE
$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 \\ 0.7914 & 1.3063 & -0.9004 \\ \pm(0.0049) & (\pm 0.0049) & (\pm 0.0047) \\ 1.1922 & -0.6980 & 1.1079 \\ \pm(0.0063) & \pm(0.0050) & \pm(0.0045) \end{bmatrix}$	$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 \\ 0.8403 & 1.3430 & -0.9085 \\ 1.1543 & -0.9408 & 0.8823 \end{bmatrix}$

Table 2. Performance criteria for the determined mixture

	$\hat{s}_1$				$\hat{s}_2$				$\hat{s}_3$			
	SDR	SIR	SAR	SNR	SDR	SIR	SAR	SNR	SDR	SIR	SAR	SNR
Jeffrey + EM	10.6	34.5	10.6	29.2	14.0	36.3	14.0	32.1	11.7	43.4	11.8	29.9
t + MCMC	11.8	40.3	12.5	20.4	15.1	47.0	15.9	22.9	13.1	43.4	13.9	20.7
JADE	5.4	32.7	-	5.4	7.5	17.1	-	8.1	5.9	14.9	-	6.6

4.2 Undetermined Mixture

We now consider the more difficult case consisting of discarding one observation of the previous mixture, thus yielding an underdetermined problem. Results are found in Tables 3 and 4. In the EM algorithm the mixing matrix was initialized with the result of the simple clustering method described in [1] [2] which yielded $\mathbf{A}_{init} = [1\ 1\ 1; 0.6348\ 1.4405\ -0.9216]$.

Table 3. Estimates of $\mathbf{A}$ for the underdetermined mixture

Jeffrey + EM	t + MCMC
$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 \\ 0.7083 & 1.5630 & -0.9109 \end{bmatrix}$	$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 1 \\ 0.7715 & 1.3072 & -0.9089 \\ (\pm 0.0059) & (\pm 0.0060) & (\pm 0.0052) \end{bmatrix}$

Table 4. Performance criteria for the underdetermined mixture

	$\hat{s}_1$				$\hat{s}_2$				$\hat{s}_3$			
	SDR	SIR	SAR	SNR	SDR	SIR	SAR	SNR	SDR	SIR	SAR	SNR
Jeffrey + EM	1.0	19.5	1.1	26.6	5.9	16.2	6.4	29.8	9.6	21.0	10.0	29.9
t + MCMC	0.7	14.1	1.4	14.6	6.4	23.3	6.7	19.8	11.4	28.5	13.9	15.3

5 Conclusions

Not surprisingly the performances of both method are better in the determined case than in the underdetermined one, notably in terms of SIRs. This is because when $\mathbf{A}$ is square, $\tilde{\mathbf{s}}_k$ given by Eq. (11) is simply the application of the weighted pseudo-inverse of $\mathbf{A}$ to $\tilde{\mathbf{x}}_k$. Tables 2 and 4 also show that Jeffrey's inverse prior leads to higher SNRs than the Student t + MCMC method. This is because the inverse prior leads to much sparser representations and actually sets many coefficients to zero (see [5]). This was confirmed by computing sparsity indexes on the obtained source coefficients estimates. Though the SARs obtained with both methods are quite similar in each case, when listening to the source estimates

[2] This method simply consists in projecting the observations on the sphere and run a K-Means algorithm to identify the slope of the clusters generated by the mixing matrix columns.

those obtained with the Jeffrey's + EM approach tend to suffer from stronger musical noise. We believe this is because the residual error originating from the additive noise on the observations masks part of the artifacts in the Student t + MCMC estimates.

One major advantage of the Jeffrey's + EM approach is computational cost: it takes 9 min to achieve convergence (50 iterations) in the underdetermined case, when it takes 6h to run the 2000 iterations necessary to obtain convergence of the Student t + MCMC method (and the MMSE estimates were computed from the next 500 samples). But on the other hand, as in many cases, the EM approach happened to be very sensitive to initialization and could lead to local maxima, while the MCMC approach scans all the posterior distribution of the parameters and thus provides reliable estimates, independently of the initializations.

Acknowledgement. This work was supported by the European Commission funded Research Training Network HASSIP (HPRN-CT-2002-00285).

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. Jourjine, A., Rickard, S., Yilmaz, O.: Blind separation of disjoint orthogonal signals: Demixing n sources from 2 mixtures. In: Proc. ICASSP. Volume 5., Istanbul, Turkey (2000) 2985–2988
3. Févotte, C., Godsill, S.J., Wolfe, P.J.: Bayesian approach for blind separation of underdetermined mixtures of sparse sources. In: Proc. 5th International Conference on Independent Component Analysis and Blind Source Separation (ICA 2004), Granada, Spain (2004) 398–405
4. Févotte, C., Godsill, S.J.: A Bayesian approach for blind separation of sparse sources. IEEE Trans. Speech and Audio Processing (In press) Preprint available at `http://www-sigproc.eng.cam.ac.uk/~cf269/`.
5. Figueiredo, M.A.T., Nowak, R.D.: Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior. IEEE Trans. Image Processing **10** (2001) 1322–1331
6. Figueiredo, M.A.T.: Adaptive sparseness for supervised learning. IEEE Trans. Pattern Analysis and Machine Intelligence **25** (2003) 1150–1159
7. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, Series B **39** (1977) 1–38
8. Mallat, S.: A wavelet tour of signal processing. Academic Press (1998)
9. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non Gaussian signals. IEE Proceedings-F **140** (1993) 362–370
10. Gribonval, R., Benaroya, L., Vincent, E., Févotte, C.: Proposals for performance measurement in source separation. In: Proc. 4$^{\text{th}}$ Symposium on Independent Component Analysis and Blind Source Separation (ICA'03), Nara, Japan (2003)