UNDERDETERMINED BLIND SEPARATION OF SPARSE SOURCES WITH INSTANTANEOUS AND CONVOLUTIVE MIXTURES

David Luengo, Ignacio Santamaría, Luis Vielva, and Carlos Pantaleón
Departamento de Ingeniería de Comunicaciones (DICOM)
Universidad de Cantabria, ETSIIT, Santander, Spain
Phone: +34 942 201392 Ext. 26
Fax: +45 942 201488
E-mail: david@gtas.dicom.unican.es
Web: http://gtas.dicom.unican.es

Abstract. We consider the underdetermined blind source separation problem with linear instantaneous and convolutive mixtures when the input signals are sparse, or have been rendered sparse. In the underdetermined case the problem requires solving three subproblems: detecting the number of sources, estimating the mixing matrix, and finding an adequate inversion strategy to obtain the sources. This paper solves the first two problems. We assume that the number of sources is unknown, and estimate it by means of an information theoretic criterion (MDL). Then the mixing matrix is expressed in spheric coordinates and we estimate sequentially the angles and amplitudes of each column, and their order. The performance of the method is illustrated through simulations.

INTRODUCTION

The blind source separation (BSS) problem deals with the extraction of the complete set of input signals (whose number is possibly unknown), using only a set of output signals (observations) generated as an unknown mixture of the inputs, and some statistical knowledge of the sources (e.g. their independency is typically assumed). The solution of the problem involves solving three subproblems: detecting the number of sources, estimating the mixing matrix, and finding an adequate inversion strategy to obtain the sources. To a great extent this solution depends on the type of problem considered.

The BSS problem can be classified according to the way in which the observations are obtained from the sources. When the mixing process is nonlinear the separation is very complicated or even impossible, so most authors consider a *linear mixture*. Linear mixtures can be further classified as *instantaneous* (memoryless system) or *convolutive* (system with memory).

0-7803-8178-5/03/$17.00 © 2003 IEEE

Another possible classification in the linear case is according to the size of the mixing matrix. If there are more equations than unknowns, the system is *overdetermined* and a solution that minimizes the L_2 norm can be found. If the number of unknowns equals the number of equations, the system is *determined* and has an exact solution. Finally, when we have more unknowns than equations, the system is said to be *underdetermined*, and the solution has to be found using additional a priori information.

We consider the underdetermined blind source separation problem with linear instantaneous and convolutive mixtures when the input signals are sparse, or have been rendered sparse. This paper deals with the first two subproblems: detecting the number of sources, and estimating the mixing matrix. The classical method to solve the first subproblem, proposed by Wax and Kailath [10], is based on the minimization of some information theoretic criterion, such as Akaike's Information Criteria (AIC) or Schwartz and Rissanen's Minimum Description Length (MDL). However, this method requires an overdetermined system, so it cannot be used directly. In this paper we exploit the sparsity of the sources to propose a method based on constructing an autocorrelation matrix from the angles of the observations to which the MDL may be applied.

The next step is to estimate the mixing matrix. Many algorithms have been proposed to solve this problem for the determined linear instantaneous mixture using statistical principles (see [2] for a review) or geometrical concepts (e.g. see [7]). However, several articles [4, 1] have shown recently that it is possible to solve the problem using simple algorithms when the sources are sparse. For many interesting signals, which are not sparse, such as speech and music, [1] shows that it is possible to apply some transformation to render them sparse. The linear convolutive mixture is harder to separate, and has received less attention. This problem is usually approached in the frequency domain (e.g. see [3]), since there it is reduced to solving a set of independent instantaneous problems.

In this article we present a unified approach in the time domain to identify the mixing matrix for linear instantaneous and convolutive mixtures. The algorithm, an extension of the one presented in [6], is based on expressing the mixing matrix in spheric coordinates and estimating sequentially the angle and norm of each column, and their correct order, exploiting the sparsity of the sources. In the next section the mathematical models for the mixture and the sources are presented. In the third section the detection of the number of sources is considered and simulations are presented to show its detection probability. In the fourth section the algorithm for the estimation of the mixing matrix is described and an example is also presented. Finally, the last section shows the conclusions.

PROBLEM STATEMENT

Mathematical Model

We consider first the *linear instantaneous mixture*. Assuming that we have m observations and q sources ($q > m$) the problem may be formulated as

$$\mathbf{Y} = \mathbf{HS} + \mathbf{W} = \mathbf{X} + \mathbf{W}; \tag{1}$$

where $\mathbf{Y} = [\mathbf{y}(0),\ \ldots,\ \mathbf{y}(N-1)]$ is an $m \times N$ matrix, formed by stacking N successive observations $\mathbf{y}(n) = [y_1(n),\ \ldots,\ y_m(n)]^T$; $\mathbf{H} = [\mathbf{h}(0),\ \ldots,\ \mathbf{h}(q-1)]$ is the $m \times q$ mixing matrix, with $\mathbf{h}(k) = [h_1(k),\ \ldots,\ h_m(k)]^T$; $\mathbf{S} = [\mathbf{s}(0),\ \ldots, \mathbf{s}(N-1)]$ is a $q \times N$ matrix with the input signals, where $\mathbf{s}(n) = [s_0(n),\ \ldots,\ s_{q-1}(n)]^T$; $\mathbf{W} = [\mathbf{w}(0),\ \ldots,\ \mathbf{w}(N-1)]$ is an $m \times N$ additive white Gaussian noise (AWGN) matrix, with $\mathbf{w}(n) = [w_1(n),\ \ldots,\ w_m(n)]^T$, and where $\mathbf{w}(n) \sim \mathcal{N}(0, \sigma_w \mathbf{I})$; and $\mathbf{X} = \mathbf{HS} = [\mathbf{x}(0),\ \ldots,\ \mathbf{x}(N-1)]$ is the $m \times N$ matrix of observations in the absence of noise, with $\mathbf{x}(n) = [x_1(n),\ \ldots,\ x_m(n)]^T$.

The *convolutive mixture* can be formulated also as (1), but replacing each element of $\mathbf{H}$ and $\mathbf{s}(n)$ by a vector. In this case the mixing matrix can be expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{11}^T & \cdots & \mathbf{h}_{1q}^T \\ \vdots & & \vdots \\ \mathbf{h}_{m1}^T & \cdots & \mathbf{h}_{mq}^T \end{bmatrix};$$

where $\mathbf{h}_{ij}^T = [h_{ij}(0),\ \ldots,\ h_{ij}(L_j-1)]$ is the channel's impulse response from the j-th source to the i-th observation. Since the channels from a given source to every observation are not necessarily of the same length, we take $L_j = \max_i L_{ij}$, and fill with zeros the shorter channels. Hence, now the dimension of the mixing matrix is $m \times D$, where $D = \sum_{i=1}^{q} L_i$, and we can have an underdetermined system even if there are less sources than observations. In the instantaneous case $L_i = 1$ for all i, so $D = q$. The source vector at the n-th instant is now given by $\mathbf{s}(n) = [\mathbf{s}_0(n)^T,\ \ldots,\ \mathbf{s}_{q-1}(n)^T]^T$, with $\mathbf{s}_k(n) = [s_k(n),\ \ldots,\ s_k(n-L_k+1)]^T$, and $\mathbf{S}$ is a $D \times N$ matrix.

Model of the Sources

In order to be able to quantify the sparsity of the sources, we consider the following probabilistic model for their pdfs [4]:

$$f_{S_i}(s_i) = p_i \delta(s_i) + (1-p_i) p_{S_i}(s_i), \tag{2}$$

where p_i is the sparsity factor for the i-th source, and $p_{S_i}(s_i)$ is the pdf of the source when it is active. We further assume that the sources are independent and Gaussian, with zero mean and equal variance σ_s^2, and a common sparsity factor p. Hence, (2) becomes

$$f_S(s) = p\delta(s) + \frac{1-p}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{s^2}{2\sigma_s^2}\right). \tag{3}$$

From (1), it is clear that the observation vector at a given moment in the instantaneous case can be expressed as a linear combination of the columns of the mixing matrix, with an additive noise:

$$\mathbf{y}(n) = \sum_{k=0}^{q-1} \mathbf{h}(k) s_k(n) + \mathbf{w}(n). \tag{4}$$

Hence, if at a given instant only the i-th source is active, then (in the absence of noise) the observation vector is collinear with $\mathbf{h}(i)$, i.e. (4) becomes $\mathbf{y}(n) = \mathbf{h}(i)s_i(n)$. In the presence of noise, the directions of the observation vector are simply spread around the true direction given by the angle of $\mathbf{h}(i)$. A similar expression can be obtained for the convolutive mixture when we have only one nonzero element out of the D samples of $\mathbf{s}(n)$.

Recently it has been shown that it is possible to solve the problem using simple algorithms when the sources are sparse [4, 1].Although the performance of these algorithms decreases remarkably when the sources are not sparse, in [1] it has been shown that under the adequate transformation many interesting signals can be rendered sparse. The general idea of these techniques is to exploit the clustering of the observations around the corresponding directions of the columns of the mixing matrix $\mathbf{H}$. In figure 1 we show an example for $m = 2$, $n = 5$, $N = 10000$, $\mathbf{H}$ given by (12), $p = 0.75$, and SNR=25 dB.

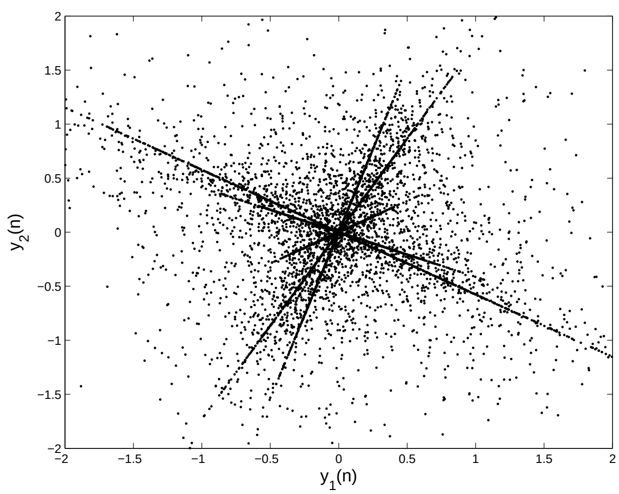

Figure 1: Scatter plot of the observations for $m = 2$, $n = 5$, $N = 10000$, $\mathbf{H}$ given by (12), $p = 0.75$, and SNR=25 dB.

Parameterization of the Mixing Matrix

To implement the algorithm we need to parameterize adequately the mixing matrix. We consider each column of $\mathbf{H}$ as a vector in an m-dimensional space and express it in spheric coordinates as in [8]. In this article we show only the case $m = 2$, since it allows a proper visualization of the algorithm and the results obtained can be easily extended to an arbitrary number of observations. In this case the mixing matrix becomes

$$\mathbf{H} = \begin{bmatrix} \|\mathbf{h}(0)\|\cos\theta_0 & \cdots & \|\mathbf{h}(D-1)\|\cos\theta_{D-1} \\ \|\mathbf{h}(0)\|\sin\theta_0 & \cdots & \|\mathbf{h}(D-1)\|\sin\theta_{D-1} \end{bmatrix}, \tag{5}$$

where D is the number of columns of $\mathbf{H}$, $\theta_k = \arctan(h_2(k)/h_1(k))$ is the angle of each column, and $\|\mathbf{h}(k)\| = \mathbf{h}(k)^T\mathbf{h}(k)$ their L_2 norm ($k = 0, \ldots, D-1$).

From (5) we notice immediately the four steps required to identify $\mathbf{H}$: estimating its dimension, identifying the angle of each column of the mixing matrix, estimating their norms, and ordering the columns of the resulting matrix. In the instantaneous case only the first two steps are necessary, as any scale factor and permutation in the columns of $\mathbf{H}$ is admissible. This is not true for the convolutive case, where the four steps are required. In the next two sections we describe the different stages of the algorithm.

DETECTION OF THE NUMBER OF SOURCES

Overdetermined Case

The classical way of detecting the number of narrowband sources embedded in a set of observations contaminated by noise is using an information theoretic criteria such as the AIC or MDL [10]. Both of them select the number of signals which minimizes a cost function composed of the log-likelihood function and an additional term which penalizes the complexity of the model. We use the MDL, since it provides a consistent estimator, whereas the AIC tends to overestimate the number of signals [10]. The cost function for the MDL is given by

$$\mathrm{MDL} = -\ln p(\mathbf{y}|\hat{\boldsymbol{\phi}}) + \frac{n_p}{2}\ln N, \tag{6}$$

where $\hat{\boldsymbol{\phi}}$ is a vector with the maximum likelihood (ML) estimate of the parameters of the model, and n_p is the total number of parameters for each model order.

Since we are considering the Gaussian linear case, the covariance matrix of the observations is $\mathbf{R}_y = \mathbf{R}_x + \sigma_w^2\mathbf{I} = \mathbf{H}\mathbf{R}_s\mathbf{H}^T + \sigma_w^2\mathbf{I}$, where $\mathbf{R}_s$ is the covariance matrix of the sources, and we consider a real mixing matrix. Thus, assuming $\mathbf{H}$ is of full column rank, and in the overdetermined case ($m > D$), theoretically the $m - D$ smallest eigenvalues are equal to σ_w^2. Hence, determining the number of sources is equivalent to estimating the multiplicity of the smallest eigenvalue, since in practice all the eigenvalues are different. The approach taken in [10] considers the family of covariance matrices $\mathbf{R}_y^{(k)} = \mathbf{R}_x^{(k)} + \sigma^2\mathbf{I}$, where $\mathbf{R}_x^{(k)}$ is a semidefinite positive matrix of rank k, σ^2 an unknown scalar representing the noise variance, and the order of the model

k ranging from 1 to $m-1$. The MDL criterion in (6) then becomes [10]:

$$\mathrm{MDL}(k) = -\ln\left(\frac{\prod_{i=k+1}^{m} l_i^{1/(m-k)}}{\frac{1}{m-k}\sum_{i=k+1}^{m} l_i}\right)^{(m-k)N} + \frac{k(2m-k)}{2}\ln N, \qquad (7)$$

where $l_1 > l_2 \cdots > l_m$ are the eigenvalues of the sample covariance matrix, defined by

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}(n)\mathbf{y}(n)^T. \qquad (8)$$

And the model order is selected as the one which minimizes (7).

Underdetermined Case

In the underdetermined case we cannot use directly (7). However, we may build a vector with the angle of each observation $\boldsymbol{\theta} = [\theta(0),\ \ldots,\ \theta(N-1)]^T$, where $\theta(n) = \arctan(y_2(n)/y_1(n))$. From this vector we construct an estimation of the pdf of the angles (histogram), which resembles the power spectral density (psd) of D tones in AWGN. Noting the close relationship between a pdf and a psd we can obtain an autocorrelation function whose Fourier transform is a scaled version of the desired pdf [5]. This procedure was already used in [9] to estimate the mixing matrix in the frequency domain using ESPRIT. The autocorrelation function from the angle vector is [9]

$$R_\theta[k] = \frac{1}{2\pi N}\sum_{n=0}^{N-1} \exp(j\tilde{\theta}[n]k) \qquad k = 0,\ \ldots,\ M; \qquad (9)$$

where $\tilde{\theta}[n] = 2\theta[n]$ is a transformation required to expand $\theta[n]$ to the range $[-\pi, \pi]$. Then we build the correlation matrix for the model of order k as

$$\mathbf{C}_\theta^{(k)} = \begin{bmatrix} R_\theta[0] & R_\theta^*[1] & \cdots & R_\theta^*[k-1] \\ R_\theta[1] & R_\theta[0] & \cdots & R_\theta^*[k-2] \\ \vdots & \vdots & \ddots & \vdots \\ R_\theta[k-1] & R_\theta[k-2] & \cdots & R_\theta[0] \end{bmatrix}. \qquad (10)$$

where * denotes the complex conjugate. Equation (10) is the equivalent of (8) for our model. Using (10) we may minimize (7) to estimate the size of $\mathbf{H}$.

Simulations

An example of the behaviour of the method is presented in figure 2. The probability of detection for an instantaneous mixture is evaluated using the same mixing matrix $\mathbf{H}$ as in [6]:

$$\mathbf{H} = \begin{bmatrix} \cos(\pi/4) & 0.3\cos(-7\pi/12) & 0.7\cos(2\pi/9) \\ \sin(\pi/4) & 0.3\sin(-7\pi/12) & 0.7\sin(2\pi/9) \end{bmatrix} \qquad (11)$$

For the convolutive mixture we consider two sources with subchannels of length 3 and 2 respectively. The amplitudes of each column of the mixing matrix $\mathbf{H}$ are 0.7, 0.4, 1.0, 0.2 and 0.6. And their angles are $\pi/3$, $7\pi/8$, $-\pi/6$, $\pi/6$ and $-3\pi/5$. The resulting mixing matrix is

$$\mathbf{H} = \begin{bmatrix} 0.3500 & -0.3696 & 0.8660 & 0.1732 & -0.1854 \\ 0.6062 & 0.1531 & -0.5000 & 0.1000 & -0.5706 \end{bmatrix}. \qquad (12)$$

In figure 2 we can see that detecting the correct number of sources with probability almost one is harder for the convolutive mixture. However, for sparsity factors above 0.75 and an SNR over 25 the performance of the method is satisfactory for both mixtures.

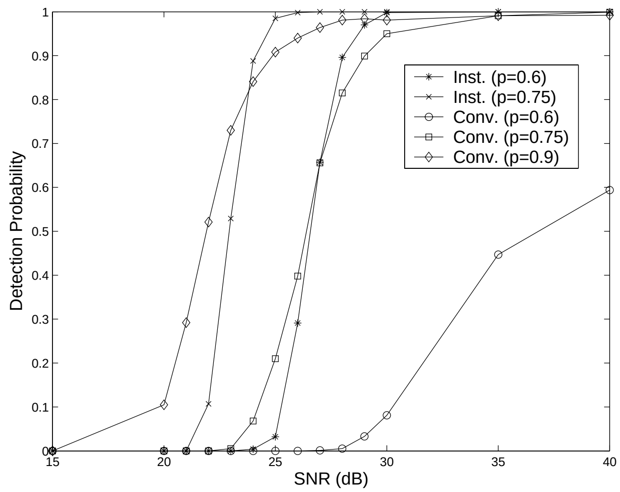

Figure 2: Probability of detection as a function of the SNR for an instantaneous mixture given by (11), and a convolutive mixture given by (12). Sparsity factor $p = 0.75$, and number of points $N = 10000$.

ESTIMATION OF THE MIXING MATRIX

PDF of the Observations

The algorithm for estimating $\mathbf{H}$ exploits the fact that there is typically only one active source. In this case, and for $m = 2$, the pdf of the observations is a bivariate Gaussian given by

$$f_{\mathbf{Y}}(\mathbf{y}(n)) = \frac{1}{2\pi|\mathbf{R}_y|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}(n)^T\mathbf{R}_y^{-1}\mathbf{y}(n)\right), \qquad (13)$$

where $\mathbf{R}_y = \sigma_s^2\mathbf{h}(k)\mathbf{h}(k)^T + \sigma_w^2\mathbf{I}$ is the correlation matrix of the observations when the k-th element of $\mathbf{s}(n)$ is the only active one. If we have N_k samples for which only the k-th source is active, the global pdf is their product. Hence, the log-likelihood function in terms of the angle and norm is

$$
\begin{aligned}
\ln f_{\mathbf{Y}}(\mathbf{y}) \quad = \quad & -\frac{N_k}{2}\ln(\|\mathbf{h}(k)\|^2\sigma_s^2+\sigma_w^2)+\frac{\sigma_s^2\|\mathbf{h}(k)\|^2}{2\sigma_w^2(\|\mathbf{h}(k)\|^2\sigma_s^2+\sigma_w^2)} \\
& \sum_{n=1}^{N_k}(y_1(n)\cos\theta_k+y_2(n)\sin\theta_k)^2, \qquad (14)
\end{aligned}
$$

where the constant terms that do not depend on the angle or norm have been omitted. Assuming the noise variance σ_w^2 has been estimated previously in the absence of signal, and setting σ_s^2 arbitrarily to 1, we can obtain the ML estimates of the angle and norm of each column from (14).

Identification of the Angles

The first step of the algorithm is estimating the D peaks corresponding to the order of the model. First, we construct the psd $S_\theta(w)$ as the Fourier transform of (9) as in [9]. But instead of using ESPRIT to estimate the peaks, which has a high computational cost, we use a simpler *grid-search algorithm* adapted from [6]. Using this algorithm we also obtain a set of samples collinear with each of the estimated angles (the set of samples whose angle is within $\Delta\theta$ radians of the estimated angle), which is used for the estimation of the amplitudes and to obtain the correct column order.

Estimation of the Amplitudes

So far we have identified the mixing matrix up to a scale and permutation indeterminacy. In the case of the instantaneous linear mixture this is enough. However, for convolutive mixtures we need to estimate the relative amplitudes of the columns and their order (at least of the columns corresponding to each subchannel). Taking the derivative of (14) and equating it to zero, we obtain the ML estimator for the norms

$$
\|\hat{\mathbf{h}}(k)\| = \sqrt{\frac{(\mathbf{y}_1\cos\theta_k+\mathbf{y}_2\sin\theta_k)^T(\mathbf{y}_1\cos\theta_k+\mathbf{y}_2\sin\theta_k)-N_k\sigma_w^2}{N_k\sigma_s^2}},
$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are respectively the sets of first and second components of the observation vectors collinear with the k-th column of $\mathbf{H}$.

Column Ordering

The permutation indeterminacy can be eliminated exploiting the temporal correlation between consecutive input vectors. The ordering method is based on the observation that, in the absence of noise, a nonzero sample of the i-th source surrounded by $L_i - 1$ zeros is consecutively collinear with the L_i columns of the mixing matrix corresponding to its impulse response. Of course, it is also required that the rest of the sources be inactive during $2L_i - 1$ samples. Hence, we can estimate the column ordering considering L_i consecutive output samples for each source which are collinear with some column of $\mathbf{H}$, and setting the most likely column ordering as the one which appears most often. Under moderate noise conditions the samples still belong to the collinear set of each angle, and the algorithm remains valid.

Simulations

In this section we show an example of the performance of the algorithm for the instantaneous mixture with $\mathbf{H}$ given by (11), and the convolutive mixture with $\mathbf{H}$ given by (12). The performance is measured by the normalized MSE, i.e. the L_2 norm of the error divided by the sum of the L_2 norm of each column of the mixing matrix. The MSE for each SNR and register's length is obtained averaging 2500 simulations. The results are shown in figure 3. The MSE decreases as the SNR increases, up to a saturation value obtained for SNRs over 30 dB, and whose value depends on the number of samples. The results for the instantaneous mixture (dashed line) are 2–5 dB better than for the convolutive one (continuous line). Note that due to its difficulty the underdetermined case has not been widely studied, so there is no standard algorithm against which we can compare our results.

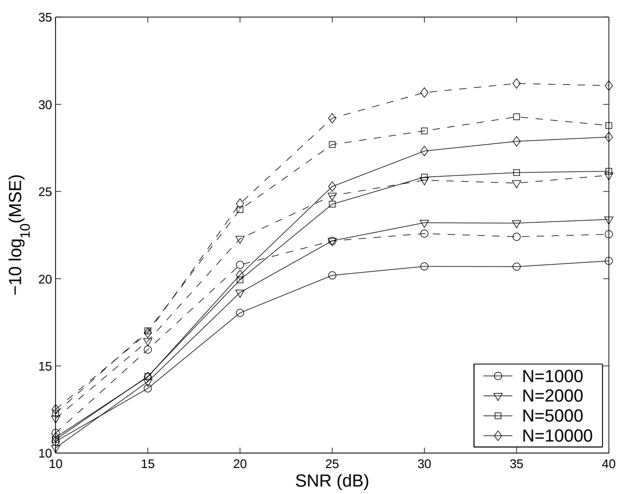

Figure 3: Normalized MSE (dB) as a function of the SNR for different values of N, $p = 0.8$ and $\mathbf{H}$ given by (11) (dashed line), and (12) (continuous line).

CONCLUSIONS

We have considered the estimation of the mixing matrix for the underdetermined noisy BSS problem with linear instantaneous and convolutive mixtures when the sources are sparse. First, we estimate the size of the mixing matrix using an information theoretic criterion (MDL). Then, expressing this matrix in spheric coordinates and exploiting the sparsity of the sources, we estimate sequentially the angles of each column, their amplitudes, and their order. Future lines of work include studying the performance of the algorithm for a greater number of observations and with real data, and considering a similar method in the frequency domain.

ACKNOWLEDGEMENTS

This work has been partially financed by the Spanish MCYT (Ministerio de Ciencia y Tecnología) under grant TIC2001-0751-C04-03.

REFERENCES

[1] P. Bofill and M. Zibulevsky, "Underdetermined Blind Source Separation Using Sparse Representations," **Signal Processing**, vol. 81, no. 11, pp. 2353–2362, Nov. 2001.

[2] J. F. Cardoso, "Blind Signal Separation: Statistical Principles," **Proceedings of the IEEE**, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[3] A. Dapena and L. Castedo, "Blind Source Separation in the Frequency Domain: A Novel Solution to the Amplitude and the Permutation Indeterminacies," in **Proc. IWANN**, 2001, vol. II, pp. 603–610.

[4] D. Erdogmus, L. Vielva and J. C. Príncipe, "Nonparametric Estimation and Tracking of the Mixing Matrix for Underdetermined Blind Source Separation," in **Proc. ICA**, 2001.

[5] S. M. Kay, "Model-Based Probability Density Function Estimation," **IEEE Signal Processing Letters**, vol. 5, no. 12, pp. 318–320, Dec. 1998.

[6] D. Luengo, I. Santamaría, J. I. nez, L. Vielva and C. Pantaleón, "A Fast Blind SIMO Channel Identification Algorithm for Sparse Sources," **IEEE Signal Processing Letters**, vol. 10, no. 5, pp. 148–151, May 2003.

[7] A. Mansour, N. Ohnishi and C. G. Puntonet, "Blind Multiuser Separation of Instantaneous Mixture Algorithm Based on Geometrical Concepts," **Signal Processing**, vol. 82, pp. 1155–1175, 2002.

[8] L. Vielva, Y. Pereiro, D. Erdogmus and J. C. Príncipe, "Inversion Techniques for Underdetermined BSS in an Arbitrary Number of Dimensions," in **Proc. ICA**, 2003.

[9] L. Vielva, I. Santamaría, C. Pantaleón, J. I. nez, D. Erdogmus and J. C. Príncipe, "Estimation of the Mixing Matrix for Underdetermined Blind Source Separation Using Spectral Estimation Techniques," in **Proc. XI European Signal Processing Conf. (EUSIPCO)**, 2002, pp. 557–560.

[10] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," **IEEE Trans. on Acoustics, Speech, and Signal Processing**, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.