A general solution to blind inverse problems for sparse input signals ☆

David Luengo[a],*, Ignacio Santamaría[b], Luis Vielva[b]

[a]*Dpto. de Teoría de la Señal y Comunicaciones (TSC), Universidad Carlos III de Madrid, Leganés (Madrid) 28911, Spain*
[b]*Dpto. de Ingeniería de Comunicaciones (DICOM), Universidad de Cantabria, Santander (Cantabria) 39005, Spain*

Abstract

In this paper, we present a computationally efficient algorithm which provides a general solution to blind inverse problems for sparse input signals. The method takes advantage of the clustering typical of sparse input signals to identify the channel matrix, solving four problems sequentially: detecting the number of input signals (i.e. clusters), estimating the directions of the clusters, estimating their amplitudes, and ordering them. Once the channel matrix is known, the pseudoinverse can be used as the canonical solution to obtain the input signals. When the input signals are not sparse enough, the algorithm can be applied after a linear transformation of the signals into a domain where they show a good degree of sparsity. The performance of the algorithm for the different types of problems considered is evaluated using Monte Carlo simulations.

---

*Corresponding author.

E-mail addresses: luengod@ieee.org (D. Luengo), nacho@gtas.dicom.unican.es (I. Santamaría), luis@gtas.dicom.unican.es (L. Vielva).

1. Introduction

Blind deconvolution (BDE), blind equalization (BEQ), and blind source separation (BSS) are three closely related problems where the ultimate goal is to estimate the input signals using only the noisy output signals and some statistical assumptions about the inputs, but without explicit knowledge of the channel. In BDE, the aim is to obtain the input signal of an unknown linear time-invariant (LTI) system when the noisy output signals are available [16]. In BEQ, the input signals are drawn from a known finite alphabet, and the objective is to obtain the sequence of input symbols that minimizes the probability of error [8]. Finally, in BSS the input signals are usually considered to be independent, and the goal is to recover a replica of them (possibly subject to a global scale and rotation factor) imposing the restriction of maximum independence of the reconstructed signals [7].

In any of these applications, prior to the estimation of the input signals, the system's transfer function must be identified, either explicitly or implicitly. This problem is known as *blind channel identification* (BCI) [8]. We consider systems which can have multiple inputs and outputs, with linear relations between the input and the output signals (i.e. a *linear mixture*), and finite impulse response (FIR) subchannels between all inputs and outputs. Hence, the solution of the BCI problem in general amounts to estimating a matrix: the *channel's* or *mixing matrix*. The method presented in this paper solves the BCI problem, and then uses Moore–Penrose's *pseudoinverse* [10] as the canonical solution to invert the channel's matrix and obtain the input signals.

The algorithm presented in the sequel can be applied to SIMO and MIMO systems, as well as SISO and MISO systems with oversampling. When the system has memory the output is often named a *convolutive mixture*, whereas for memoryless systems it is usually called an *instantaneous mixture*. Depending on the number of inputs and outputs, we can distinguish three cases: overdetermined (more outputs than inputs), determined (the same number), and underdetermined (less outputs than inputs). Our method deals in a unified way with SIMO and MIMO systems, instantaneous and convolutive mixtures, and the overdetermined, determined and underdetermined cases.

In order to do so we impose a condition on the input signals: *sparsity*. In the overdetermined and determined cases the requirement of sparsity is not essential to be able to identify the mixture. For example, in BSS it is well-known that in the determined case the input signals can be separated (up to a permutation and a global scale indeterminacy) as long as at most one of them is Gaussian and the mixing matrix is nonsingular [6]. In the underdetermined case, sparsity is necessary to obtain good estimates of the input signals, even if the mixing matrix is known [5]. However, many interesting signals satisfy this requisite (e.g. some biomedical signals, or signals from seismic deconvolution and nondestructive evaluation), and for many other ones which are not sparse enough (e.g. audio, speech or images) linear transformations such as the Fourier transform or expansions using an overcomplete basis can be used to increase their sparsity [5,31]. In this paper we assume that the input signals already satisfy the requirement of sparsity. The main idea of the

algorithm is to exploit the clustering of the output signals, which occurs typically when the input signals are sparse, to solve any blind signal processing problem sequentially in five stages:

(1) Detecting the number of input signals.
(2) Identifying the directions of the cluster related to each input signal.
(3) Estimating the norm of the basis vector associated to each cluster.
(4) Sorting appropriately the clusters.
(5) Inverting the channel matrix to obtain the input signals.

The first step can be considered a "preprocessing" stage, necessary to estimate the dimension of the problem. Steps (2)–(4) solve the BCI problem, providing the channel matrix required to estimate the input signals. Finally, the last step inverts the mixture, achieving the desired identification of the input signals. Note that for certain problems one or more steps may not be required. For example, in some applications the number of input signals may be known. Moreover, for systems without memory any permutation and global scale factor in the input signals is usually acceptable [6], so steps (3) and (4) can be omitted.

The paper is organized as follows. In Section 2 the mathematical model is presented, including a parameterization of the mixing matrix which allows the partition of the BCI problem into three sequential subproblems. Next, in Section 3, the probabilistic model for the sources and the output signals is introduced. In Section 4 the five stages of the algorithm are shown, and its performance is evaluated. Then, Section 5 presents a brief discussion of potential applications, and finally the conclusions are shown in Section 6.

2. Mathematical model of the mixture

2.1. Linear mixture

We consider a system with *q sources* and *m observations* or *measurements*. The observations are obtained from the sources as the output of a linear system plus additive white Gaussian noise (AWGN). Hence, we have a system of *m* linear equations (*output signals*) with *l* unknowns (*input signals*). The number of input signals ($l \geqslant q$) depends on the type of problem: $l = q$ for a memoryless system, and $l > q$ for a system with memory. In any case, $m > 1$ and $l > 1$, and, regardless of the type of problem studied, we can always construct a MIMO system. Hence, the information available for each sample can be expressed as

$$\vec{y}(n) = \vec{H}\vec{s}(n) + \vec{w}(n). \tag{1}$$

Assuming a data set composed of *N* samples, $\{\vec{y}(n)\}_{n=0}^{N-1}$, all the information at our disposal can be grouped together in a single equation as

$$\vec{Y} = \vec{H}\vec{S} + \vec{W} = \vec{X} + \vec{W}, \tag{2}$$

where $\vec{Y} = [\vec{y}(0), \dots, \vec{y}(N-1)]$ is the $m \times N$ *output matrix*, constructed stacking N consecutive output vectors, $\vec{y}(n) = [y_1(n), \dots, y_m(n)]^{\mathrm{T}}$; $\vec{H}$ is the $m \times l$ *mixing matrix*, which provides the channel's transfer function, and has a structure that depends on the type of problem considered; $\vec{S} = [\vec{s}(0), \dots, \vec{s}(N-1)]^{\mathrm{T}}$ is the $l \times N$ *input matrix*, which contains the input signals, and which has also a problem-dependent structure; $\vec{W} = [\vec{w}(0), \dots, \vec{w}(N-1)]$ is the $m \times N$ *AWGN matrix*, with $\vec{w}(n) = [w_1(n), \dots, w_m(n)]^{\mathrm{T}}$, and where $w_i(n) \sim \mathcal{N}(0, \sigma_{w_i}^2)$, meaning that each component is Gaussian with zero mean and variance $\sigma_{w_i}^2$; and, finally, $\vec{X} = \vec{H}\vec{S} = [\vec{x}(0), \dots, \vec{x}(N-1)]$ is the $m \times N$ output matrix in the absence of noise, with $\vec{x}(n) = [x_1(n), \dots, x_m(n)]^{\mathrm{T}}$.

2.2. Parameterization of the mixing matrix

In the previous subsection we have shown the mathematical model for a linear mixture. The mixing matrix, $\vec{H}$, and the input vector, $\vec{s}(n)$, have a structure which is problem dependent. However, regardless of the application and the structure of $\vec{H}$ and $\vec{s}$, we can consider a columnwise representation of the mixing matrix as

$$\vec{H} = [\vec{h}_1, \dots, \vec{h}_l], \tag{3}$$

where $\vec{h}_i$ denotes the ith column of $\vec{H}$, and $h_i(k)$ its kth element. Similarly, the ith element of the input vector for a given sample will be denoted as $s_i(n)$, regardless of the memory of the problem and the number of sources.

It is well-known that the output vector at the nth sample can be expressed as a linear combination of the columns of $\vec{H}$ [5,9,10]:

$$\vec{y}(n) = \sum_{i=1}^{l} s_i(n)\vec{h}_i + \vec{w}(n). \tag{4}$$

Hence the columns of the mixing matrix, $\vec{h}_i$, can be seen as *basis vectors* in an m-dimensional space, and $s_i(n)$ as the portion of each basis vector contained in a given output vector. Thus, identifying $\vec{H}$ is equivalent to estimating the optimum set of basis vectors.

Instead of tackling this problem directly (i.e. estimating each element of $\vec{h}_i$) we are going to solve the equivalent problem of estimating the direction and magnitude of each basis vector, which amounts to solving a clustering problem. Although clustering techniques are not new in BSS problems (e.g. see [29] for an algorithm which uses a clustering technique, the E-M algorithm and ICA to solve a biomedical problem), usually the methods proposed do not exploit explicitly the sparsity inherent in many applications.

In this paper we present the case $m = 2$, and indicate how to extend the algorithm to the case $m > 2$. In order to do so, we express each basis vector in polar coordinates as

$$\vec{h}_i = r_i[\cos(\theta_i)\sin(\theta_i)]^{\mathrm{T}}, \tag{5}$$

where r_i is the magnitude of the ith basis vector, given by

$$r_i = \sqrt{h_i(1)^2 + h_i(2)^2}, \tag{6}$$

and θ_i is the angle:

$$\theta_i = \arctan \frac{h_i(2)}{h_i(1)}. \tag{7}$$

This parameterization allows us to solve the BCI problem in four sequential stages. First of all, the number of basis vectors has to be estimated, i.e. we have to establish the dimension of the problem (number of clusters). Then, we have to estimate the direction of each basis vector (i.e. the orientation of the clusters). If we are considering an instantaneous mixture the other two steps are not required, since any permutation and global scale factor in the basis vectors is generally admissible. When we have a convolutional mixture, two additional stages must be performed: estimating the magnitude of each basis vector, and ordering the vectors to avoid permutations in the columns of $\vec{H}$.

3. Statistical model of the input and output signals

3.1. Model of the sources

The algorithms proposed to solve each stage of the BCI problem make use of an important feature of many input signals: their sparsity. A source is said to be sparse if it is inactive at least 50% of the time (although typical inactivity periods in many applications can range from 75% to 95% of the time). We are going to characterize the sources statistically using their probability density function (PDF). Although some authors consider a Laplacian PDF to model sparse input signals [5,31], we are going to consider the model for the PDF used in [9,26], which allows a greater flexibility in the selection of different PDFs depending on the type of problem. According to it, the PDF for each individual input signal is

$$p_{S_i}(s_i) = p_i \delta(s_i) + (1 - p_i) f_{S_i}(s_i), \tag{8}$$

where p_i is the sparsity factor for the ith input signal, which indicates the probability of the source being inactive, $f_{S_i}(s_i)$ is the PDF of the ith source when it is active, and $i = 1, \ldots, l$. When the PDF of each source is Gaussian, (8) becomes the well-known Bernouilli–Gaussian (BG) model, widely used in nondestructive evaluation or seismic deconvolution [16]:

$$p_{S_i}(s_i) = p_i \delta(s_i) + \frac{1 - p_i}{\sqrt{2\pi\sigma_{s_i}^2}} \exp\left(-\frac{s_i^2}{2\sigma_{s_i}^2}\right). \tag{9}$$

Although the BG model is the one used throughout the article, (8) allows the use of any PDF of interest for the sources, such as Laplacian or uniform PDFs. Note that, since we are going to consider that the l input signals are independent, the PDF of the input vector is the product of (8) for all the input signals.

Now, we notice that, when the sparsity factor is high, there are many samples for which only one input signal is different from zero (i.e. active). Hence, if the kth

source is the only active one, the output signal can be written as

$$\vec{y}(n) = s_k(n)\vec{h}_k + \vec{w}(n). \tag{10}$$

Thus, in the absence of noise the output vector is aligned with the kth column of $\vec{H}$ (i.e. the direction of $\vec{y}(n)$ is given by the kth basis vector, $\vec{h}_k$). When the output signals are corrupted by noise, the direction of $\vec{y}(n)$ will be spread around the true direction given by $\vec{h}_k$. In moderate/high signal to noise ratio (SNR) situations, this results in a clustering of the output vectors around the basis vectors, which can be exploited to identify them [5,9]. Fig. 1 shows a typical scatter plot of the components of the output vector, which displays the clustering characteristic of sparse input signals. The data, generated synthetically using the BG model and a mixing matrix given by

$$\vec{H} = \begin{bmatrix} 0.3500 & -0.3696 & 0.8600 & 0.1732 & -0.1854 \\ 0.6062 & 0.1531 & -0.5000 & 0.1000 & -0.5706 \end{bmatrix} \tag{11}$$

closely resemble the time series typical of applications such as seismic deconvolution (see Fig. 5 for a time-domain representation using a different $\vec{H}$).

3.2. Model of the output signals

In order to develop the different stages of the algorithm we require a statistical model of the output signals. Since the algorithm is based on the clustering of the
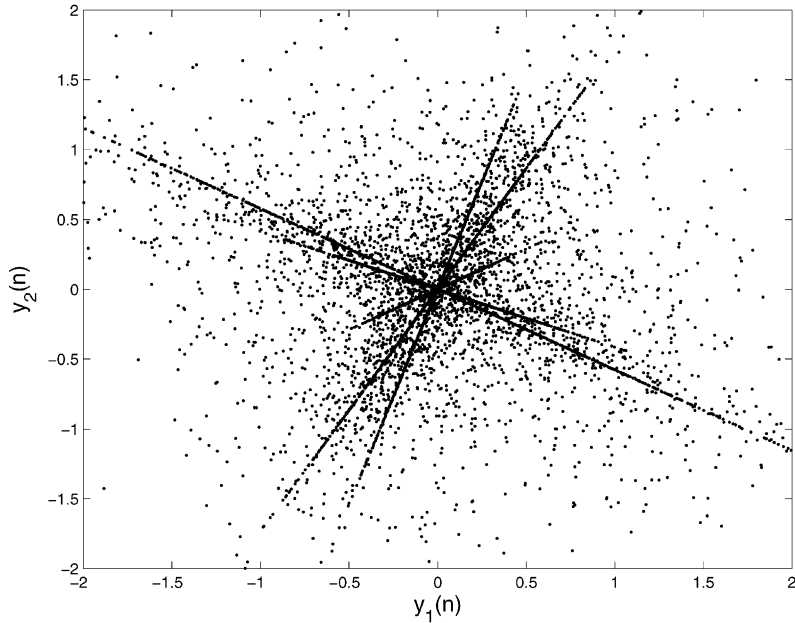

Fig. 1. Scatter plot of the output signals mixed with (11) using the BG model for the sources with $\sigma_s^2 = 1$, $p = 0.75$, $N = 10000$, and SNR = 30 dB.

outputs around the basis vectors when there is only one nonzero input, we just need a model for this case. Considering an equal variance for all the sources, σ_s^2, and for all the samples of the noise vector, σ_w^2, the PDF of $\vec{y}(n)$ can be easily obtained for $m = 2$. In the absence of noise, and assuming that only the kth input signal is active, the PDF of each of the components of the output vector is simply a scaled version of the PDF of s_k. Since both outputs follow a deterministic relation, the PDF of the output vector is

$$
\begin{aligned}
f_{\vec{X}}(\vec{x}(n)) &= \frac{1}{|h_k(1)|} f_{S_k}\left(\frac{x_1(n)}{h_k(1)}\right) \delta\left(x_2(n) - \frac{h_k(2)}{h_k(1)} x_1(n)\right) \\
&= \frac{1}{|h_k(2)|} f_{S_k}\left(\frac{x_2(n)}{h_k(2)}\right) \delta\left(x_1(n) - \frac{h_k(1)}{h_k(2)} x_2(n)\right).
\end{aligned} \tag{12}
$$

Since the noise is white and independent of the sources, the PDF of the noisy output vector is simply (12) convolved with the PDF of the noise. Considering AWGN and the BG model, we obtain a zero-mean bivariate Gaussian PDF for the output vector characterized by an autocorrelation matrix [15]

$$
\vec{R}_y = \sigma_s^2 \vec{h}_k \vec{h}_k^{\mathrm{T}} + \sigma_w^2 \vec{I}. \tag{13}
$$

If we have N_k samples for which this happens (i.e. for which $\vec{x}(n)$ is aligned with $\vec{h}_k$), the global PDF is their product. Hence, the log-likelihood function in terms of the magnitude and angle of the kth column is [15]

$$
\begin{aligned}
\ln f_{\vec{Y}}(\vec{y}) = &-\frac{N_k}{2} \ln(r_k^2 \sigma_s^2 + \sigma_w^2) \\
&+ \frac{\sigma_s^2 r_k^2}{2\sigma_w^2(r_k^2 \sigma_s^2 + \sigma_w^2)} \sum_{n_k} (y_1(n_k)\cos\theta_k + y_2(n_k)\sin\theta_k)^2,
\end{aligned} \tag{14}
$$

where the constant terms that do not depend on the angle or magnitude of the kth basis vector have been omitted, and $n_k = \{n : \arctan(x_2(n)/x_1(n)) = \theta_k\}$.

4. Description of the algorithm

In this section we describe in detail the five stages of the algorithm: detecting the number of input signals, estimating the direction of each basis vector of the mixing matrix, estimating their magnitudes, ordering them, and inverting the mixture to obtain the input signals. As discussed previously, in the case of a memoryless system the third and fourth steps are not required, since a global scale factor and a permutation in the input signals are admissible. If the system has memory the five steps are essential.

4.1. Detection of the number of input signals

The standard way of detecting the number of narrowband input signals embedded in a set of observations contaminated by noise is using information theoretic criteria

such as Akaike's information criterion (AIC) or Schwartz and Rissanen's minimum description length (MDL) principle [30]. Both of them select the number of signals which minimizes a cost function composed of the log-likelihood function plus an additional term which penalizes the complexity of the model. However, the approach presented in [30], based on the eigenvalues of the sample covariance matrix, requires more outputs than inputs, and consequently cannot be applied directly to the underdetermined case. Several modifications and improvements of this method have been presented, and some other algorithms are also available, but all of them require $l < m$.

Nevertheless, in [15] the algorithm of [30] has been extended to the case $m = 2$. Noting the similarities between a power spectral density (PSD) and a PDF [12], an autocorrelation matrix can be constructed from the PDF of the angle, and used as the sample covariance matrix for the algorithm presented in [30]. The steps required to detect the number of sources are the following:

(1) Obtain an $N \times 1$ vector of angles from the output signals:

$$\widetilde{\theta}(n) = \arctan \frac{y_2(n)}{y_1(n)}, \tag{15}$$

where $-\pi < \widetilde{\theta}(n) \leqslant \pi$, and $n = 0, \ldots, N - 1$.

(2) Noting the similarity between a PDF and a PSD, we may obtain an "autocorrelation function" (ACF) for the angles as the inverse Fourier transform (IFT) of the estimated PDF of θ [12]. Using a train of impulses at the angles of the output signals as the estimated PDF,

$$p_\Theta(\theta) = \frac{1}{N} \sum_{n=0}^{N-1} \delta(\theta - \widetilde{\theta}(n)) \tag{16}$$

and taking its IFT, the ACF of the angles becomes

$$\hat{R}_\Theta[k] = \frac{1}{2\pi N} \sum_{n=0}^{N-1} \exp(jk\widetilde{\theta}(n)), \tag{17}$$

i.e. samples of the characteristic function for $k = 0, \ldots, N - 1$.

(3) Construct the global autocorrelation matrix (ACM) using (17), so that its (i,j)th element is given by $\vec{R}_\Theta = \hat{R}_\Theta[i - j] = \hat{R}_\Theta^*[j - i]$.

(4) Now, for increasing model orders ($i = 1, \ldots, M$), apply an information theoretic criterion (ITC) using the first i columns and rows of the ACM:

$$\text{ITC}(i) = -\ln f_{\vec{Y}}(\vec{y}|\hat{\phi}^{(i)}) + C(N)v(i), \tag{18}$$

where the first term is the log-likelihood function conditioned by the parameter set of the ith hypothesis, $\hat{\phi}^{(i)}$, and the second term penalizes the complexity of the model. It is composed of $C(N)$, which is a function that takes into account the size of the data set, and $v(i)$, which is the number of free parameters of the ith hypothesis. The two most commonly used ITCs are the AIC [2], for which $C(N) = 1$, and the MDL [19,20], for which $C(N) = \frac{1}{2}\ln N$. In this case, the
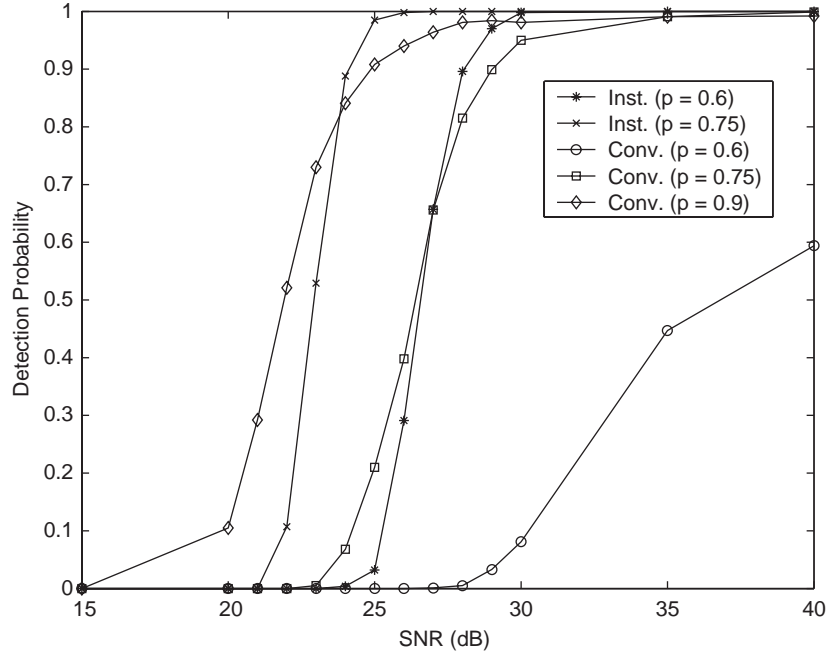
Fig. 2. Probability of detection for an instantaneous mixture given by (20), and a convolutive mixture given by (11), with $p = 0.75$, and $N = 10000$.

number of free parameters for both of them is $v(i) = i(2M - i)$ [30]. Since the AIC has been shown to yield estimators which are not consistent [30], we use the MDL, which can be expressed as a function of the eigenvalues of the ACM [30]:

$$\text{MDL}(i) = -\ln\left(\frac{\prod_{j=i+1}^{M}\lambda_j^{1/(M-i)}}{(1/(M-i))\sum_{j=i+1}^{M}\lambda_j}\right)^{(M-i)N} + \frac{i(2M-i)}{2}\ln N, \qquad (19)$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ are the eigenvalues of the ACM.

(5) The number of input signals is selected as the model order which minimizes (19).

The probability of detection achieved is shown in Fig. 2 for the BG model and two mixing matrices. The performance for an instantaneous mixture is tested using the following 2×3 mixing matrix [14,15]:

$$\vec{H} = \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & 0.3\cos\left(\frac{-7\pi}{12}\right) & 0.7\cos\left(\frac{2\pi}{9}\right) \\ \sin\left(\frac{\pi}{4}\right) & 0.3\sin\left(\frac{-7\pi}{12}\right) & 0.7\sin\left(\frac{2\pi}{9}\right) \end{bmatrix}$$

$$= \begin{bmatrix} 0.707 & -0.077 & 0.536 \\ 0.707 & -0.289 & 0.450 \end{bmatrix}. \qquad (20)$$

The performance for a convolutive mixture is tested using the 2×5 mixing matrix given by (11).

Although this method provides good results for moderate/high SNRs and $m = 2$, it cannot be directly extended to the case $m > 2$. In these cases, a simple approach based on setting a threshold in the multidimensional PDF of the angles can be considered. This approach shows a satisfactory performance for moderate/high SNRs, but requires the setting of a subjective threshold. A better alternative, if a clustering method such as the competitive one presented in [13] is used for the next step, is to start with a high number of basis vectors and consider some merging strategy (e.g. two vectors merge when they differ in less than a given angle). The final number of basis vectors equals the number of estimated input signals. This method presents the advantage of providing a joint solution to the first two problems: detecting the number of signals and estimating the directions of the basis vectors. However, the issue of convergence of any clustering algorithm should be carefully considered to ensure consistent solutions.

4.2. Estimation of the direction of the basis vectors

There are several ways to estimate the directions of the basis vectors (i.e. the columns of the mixing matrix), but they are all based on the alignment between the output vectors and the basis vectors when only one input signal is different from zero. In [5] a potential function based clustering approach is used. In [9] an approach based on Parzen windowing is shown to provide very good results. The competitive clustering approach presented in [13] can also be used. However, in this paper we consider two alternatives: the estimation from the PSD considered in the previous section, and an histogram-based estimator.

In the previous section we noted the close relation between a PDF and a PSD, and constructed an ACF (17). Taking the Fourier transform of (17) we obtain a PSD function, and can apply any of the rich variety of spectral estimation techniques available [22]. This approach has already been considered in [28], where the ESPRIT method was used to estimate the peaks corresponding to each basis vector. However, although this method provides very good results, it also requires a high computational cost. Thus, as a cost-efficient alternative, we are going to use a histogram-based estimator. This approach was considered in [9] and discarded because of its poor results. Nevertheless, it can be greatly improved if we consider the ML estimator of the angles inside each bin, instead of the center of the selected bin. The method proceeds as follows:

(1) Construct a histogram of angles in the range $[0, \pi]$ from the set of angles estimated previously for each output signal, $\tilde{\theta}(n)$. An example of a typical estimated PDF for the mixture given by (11) is shown in Fig. 3.
(2) Select the m highest peaks of the histogram, establishing a strategy to avoid the detection of false peaks due to noise.
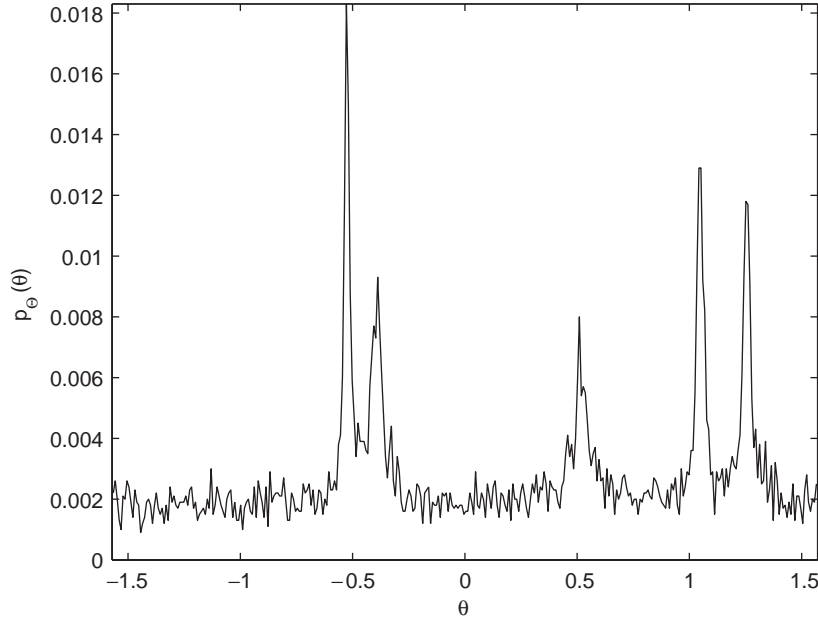
Fig. 3. Example of the estimated PDF (histogram) of the angle for a convolutive mixture given by (11) with $p = 0.9$, and SNR $= 30$ dB.

(3) Apply the ML estimator for the angles inside each of the selected bins. If we consider a BG model, it can be easily seen from (14) that the ML estimator for the angle of the kth basis vector is [15]

$$\hat{\theta}_k = \frac{1}{2} \arctan\left(\frac{2\vec{y}_1^{\mathrm{T}}\vec{y}_2}{\vec{y}_1^{\mathrm{T}}\vec{y}_1 - \vec{y}_2^{\mathrm{T}}\vec{y}_2}\right), \tag{21}$$

where $\vec{y}_1$ and $\vec{y}_2$ are the vectors with the first and second components of the N_k output signals whose angle falls inside the kth selected bin.

Note that, when the exact PDF of the input signals is unknown or the ML estimator cannot be obtained, we can estimate the angles simply averaging the estimates which fall inside each selected bin, making this stage PDF-independent. This approach provides good results for $m = 2$, but presents an increasing difficulty and computational cost as m increases (searches in $(m-1)$-dimensional histograms are required). The same happens for the potential function and Parzen windowing approaches, and the approach based on the PSD (although an estimator for $m = 3$ has been proposed in [27]). The only viable alternative for high m seems to be a clustering approach such as the one presented in [13]. However, the adequate initialization of the basis vectors for this algorithm is a delicate task and remains an open problem.

4.3. Estimation of the amplitude of the basis vectors

So far we have identified the mixing matrix up to a scale and a permutation indeterminacy. In the case of an instantaneous mixture the BCI problem is solved, and the only remaining step is inverting the mixture to obtain the input signals. However, for convolutive mixtures we need to estimate the relative amplitudes of the columns and their order. From the previous section we have a set of samples approximately aligned with each of the l basis vectors. Hence, we can easily apply the ML estimator of their magnitudes for each bin, which is readily obtained from (14) [15]:

$$r_k = \sqrt{\frac{(\vec{y}_1 \cos \theta_k + \vec{y}_2 \sin \theta_k)^{\mathrm{T}}(\vec{y}_1 \cos \theta_k + \vec{y}_2 \sin \theta_k) - N_k \sigma_w^2}{N_k \sigma_s^2}}, \tag{22}$$

where $\vec{y}_1$ and $\vec{y}_2$ are the vectors obtained in the previous stage. This approach can be easily extended for $m > 2$. Its main restriction is that it is dependent on the PDF of the input signals, which may not be precisely known for some applications. In those cases, when the noise is zero mean and independent of the input signals, we note that

$$E\{\vec{y}(n_k)^{\mathrm{T}}\vec{y}(n_k)\} = \sigma_s^2 r_k^2 + \sigma_w^2, \tag{23}$$

where $E\{\cdot\}$ denotes the mathematical expectation, taken over the set of outputs aligned with the kth basis vector. Hence, in these cases the sample mean can be used to estimate the magnitude of each column of $\vec{H}$:

$$r_k = \sqrt{\frac{\sum_{n_k} \vec{y}(n_k)^{\mathrm{T}}\vec{y}(n_k) - N_k \sigma_w^2}{N_k \sigma_s^2}}. \tag{24}$$

4.4. Ordering the basis vectors

The permutation indeterminacy can be removed by exploiting the temporal correlation between consecutive input vectors. The ordering method is based on the fact that, in the absence of noise, a nonzero sample of the ith source ($1 \leqslant i \leqslant q$) surrounded by $l_i - 1$ zeros is sequentially aligned with the l_i consecutive columns of the mixing matrix related to its impulse response. Obviously, the other sources must also be inactive during those samples. Hence, we can estimate the order of the basis vectors considering the set of output samples which are sequentially aligned with l_i different basis vectors, and setting the most likely column order as the one which appears most often. In a certain sense this is the ML estimator of the column order, since we are estimating the most likely order of the basis vectors based on the empirical PDF of their order, and works very well under moderate/high SNR conditions.

At this point the BCI problem has been solved, both for the instantaneous and the convolutive mixtures. As an example of the performance of the whole BCI algorithm, Fig. 4 shows the MSE obtained for the outputs for the instantaneous mixture using (20) and the convolutive mixture using (11). The results for the
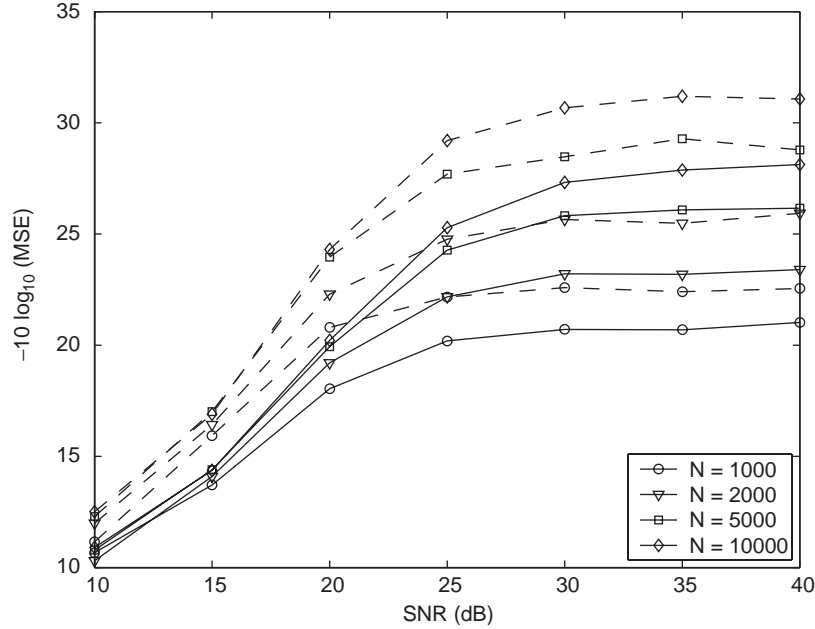
Fig. 4. Normalized MSE (dB) as a function of the SNR for $p = 0.8$ and $\vec{H}$ given by (20) for the instantaneous mixture (dashed line) and (11) for the convolutive mixture (continuous line).

convolutive mixture are 2–5 dB worse than for the instantaneous mixture due to the increased number of sources, and the additional variance introduced by the magnitude estimation step.

4.5. Inverting the mixture

In the determined case the input signals are completely characterized by the mixing matrix. In the overdetermined case, the pseudoinverse provides the solution with minimun L_2 norm of the error, and hence it is commonly used. In the underdetermined case, the pseudoinverse is the solution with minimum L_2 norm [10], and thus can be considered the canonical inversion strategy. However, it has been shown in [26] that much better inversion strategies can be developed. For example, in [26] a Bayesian inversion strategy, which has a high computational cost, has been developed, altogether with several heuristic criteria. In this paper we use one such simple heuristic criterion for the inversion: the output signals which are aligned with some basis vector are inverted using only the corresponding column of $\vec{H}$, whereas the rest of the outputs are inverted using the pseudoinverse. An example of the inversion of the mixture is shown in Fig. 5 for the instantaneous case, where the close resemblance of both signals, in spite of a scale factor and the appearance of noise peaks, can be appreciated.
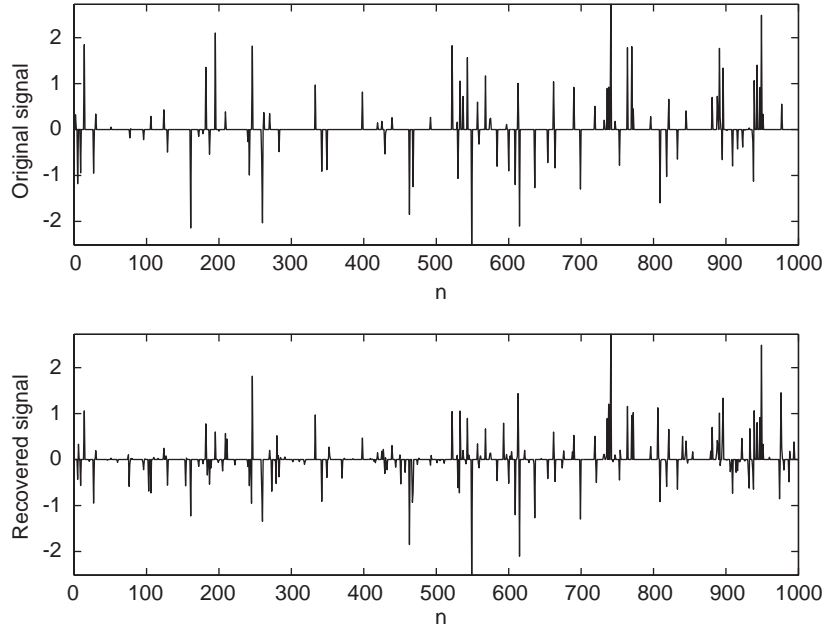
Fig. 5. Example of an original input signal and the recovered signal for an instantaneous mixture given by (20) with $p = 0.9$ and SNR $= 30$ dB.

5. Applications

In Section 2.1 it was pointed out that the structure of $\vec{H}$ and $\vec{s}(n)$ depended on the problem at hand. In this section we describe briefly the different possible structures and some associated applications.

5.1. SISO systems with oversampling

In this case there is a single source, $s(n)$, and channel, $h(n)$. Multiple output signals are obtained using an oversampling ratio m. Now the elements of the input vector are $s_i(n) = s(n - i)$, and the elements of the channel's matrix are $h_{ij} = h(i - 1 + (j - 1)m)$ for $i,j = 1, \ldots, N$.

This approach is very common in BEQ, where a SIMO system can be constructed from a SISO problem by oversampling [8]. Using this technique Tong et al. proposed an algorithm based on subspace methods for identifying $\vec{H}$ using only second-order statistics of the output signals [24,25]. Since then, there have been several extensions of this idea (see for example [1,17,23]). These methods provide very good results, but have a high computational cost. The algorithm described in this article could be applied to obtain a low cost solution of the problem. The main challenge in this case is finding a domain where the input signals are sparse enough.

5.2. SIMO systems with convolutive mixtures

In this case, the ith element of the input vector at time n is again $s_i(n) = s(n - i)$, and the elements of the mixing matrix are $h_{ij} = h_i(n - j - 1)$, where $h_i(\cdot)$ denotes the subchannel from the input to the ith observation.

This situation is typical in many BDE problems, such as seismic deconvolution [16] and nondestructive evaluation [21], where the output of the system in response to an input signal is measured using several sensors placed at different locations. The standard approach to this problem is to perform maximum likelihood (ML) deconvolution [16], but due to its high computational cost simpler methods may be preferred in some cases. Besides, in these applications the input signals are often sparse enough to apply the techniques of the article in the time domain, without transforming them into any other domain.

5.3. MIMO systems with instantaneous mixtures

The simplest MIMO systems are those in which we have an instantaneous mixture. In this case, the nth sample of the observation vector is simply a weighted linear combination of the input signals. Thus, each element of the mixing matrix, h_{ij}, represents the contribution of the jth source to the ith observation.

This is the most widely used model in blind source separation (BSS), because its simplicity allows the obtention of good solutions under certain assumptions. The determined case has been widely studied, and excellent solutions are available based on statistical principles, independent component analysis, and information theoretic criteria (see for example [4,6,11]). The underdetermined case is more challenging and has received little attention until recently, when several methods have been proposed [5,9,31]. The algorithm presented in this paper follows a similar approach, and can be considered an extension of the methods presented separately for BSS and BCI in [9,14,15].

5.4. MIMO systems with convolutive mixtures

The last case is a combination of the two previous problems: a system with $q > 1$ sources, and memory. This problem appears typically in BSS with convolutive mixtures [7], and in BEQ of MIMO communication systems [8,18], and is usually solved in the frequency domain [3]. The algorithm presented in this article is able to solve the MIMO problem with convolutive mixtures in the same way as the other three problems, in the time domain, as long as the input signals are sparse enough.

6. Conclusions

In this paper we have presented a computationally efficient algorithm for solving inverse problems when the input signals are sparse, which can be applied to blind deconvolution, blind equalization, and blind source separation. The proposed

method takes advantage of the sparsity of the input signals and a parameterization of the columns of the mixing matrix (basis vectors) in polar coordinates to solve the problem in five sequential stages: detecting the number of input signals, estimating the directions of the basis vectors, estimating their amplitudes, ordering the basis vectors, and inverting the mixture. Explicit formulas have been provided for the BG model and $m = 2$, and considerations for the extension to different PDFs and $m > 2$ have been done. Future research lines include the extension of the method to an arbitrary number of output signals, to different PDFs of the sources or even sources with unknown PDFs, and for nonlinear and post-nonlinear mixtures, possibly using spectral clustering techniques.

References

[1] K. Abed-Meraim, E. Moulines, P. Loubaton, Prediction error methods for second-order blind identification, IEEE Trans. Signal Process. 45 (3) (1997) 694–705.

[2] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control AC-19 (1974) 716–723.

[3] S. Araki, R. Mukai, S. Makino, T. Nishikawa, H. Saruwatari, The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech, IEEE Trans. Speech Audio Process. 11 (2) (2003) 109–116.

[4] A.J. Bell, T.J. Sejnowski, An information–maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.

[5] P. Bofill, M. Zibulevsky, Underdetermined blind source separation using sparse representations, Signal Process. 81 (11) (2001) 2353–2362.

[6] J.F. Cardoso, Blind signal separation: statistical principles, Proc. IEEE 86 (10) (1998) 2009–2025.

[7] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing, Wiley, New York, 2002.

[8] Z. Ding, Y.G. Li, Blind Equalization and Identification, Marcel Dekker, New York, 2001.

[9] D. Erdogmus, L. Vielva, J.C. Príncipe, Nonparametric estimation and tracking of the mixing matrix for underdetermined blind source separation, in: Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA), San Diego, CA, USA, 2001, pp. 189–193.

[10] G.H. Golub, C.F. Van Loan, Matrix Computations, 3rd ed., The John Hopkins University Press, Baltimore, MD, USA, 1996.

[11] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (4–5) (2000) 411–430.

[12] S.M. Kay, Model-based probability density function estimation, IEEE Signal Process. Lett. 5 (12) (1998) 318–320.

[13] D. Luengo, C. Pantaleón, I. Santamaría, L. Vielva, J. Ibáñez, Multiple composite hypothesis testing: a competitive approach, J. VLSI Signal Process. 37 (2004) 319–331.

[14] D. Luengo, I. Santamaría, J. Ibáñez, L. Vielva, C. Pantaleón, A fast blind SIMO channel identification algorithm for sparse sources, IEEE Signal Process. Lett. 10 (5) (2003) 148–151.

[15] D. Luengo, I. Santamaría, L. Vielva, C. Pantaleón, Underdetermined blind separation of sparse sources with instantaneous and convolutive mixtures, in: Proceedings of the XIII IEEE International Workshop on Neural Networks for Signal Processing (NNSP), Toulouse, France, 2003, pp. 279–288.

[16] J.M. Mendel, Maximum Likelihood Deconvolution, Springer, Berlin, 1990.

[17] E. Moulines, P. Duhamel, J.F. Cardoso, S. Mayrargue, Subspace methods for the blind identification of multichannel FIR filters, IEEE Trans. Signal Process. 43 (2) (1995) 516–525.

[18] A.J. Paulraj, D.A. Gore, R.U. Nabar, H. Bolcskei, An overview of MIMO communications—a key to gigabit wireless, Proc. IEEE 92 (2) (2004) 198–218.

[19] J. Rissanen, Modeling by shortest data description, Automatica 14 (1978) 465–471.

[20] G. Schwartz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.

[21] S.K. Sin, C.H. Sen, A comparison of deconvolution techniques for the ultrasonic nondestructive evaluation of materials, IEEE Trans. Image Process. 1 (1) (1992) 3–10.

[22] P. Stoica, R.L. Moses, Introduction to Spectral Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1997.

[23] L. Tong, G. Xu, B. Hassibi, T. Kailath, Blind channel identification based on second-order statistics: a frequency domain approach, IEEE Trans. Inform. Theory 41 (1) (1995) 329–334.

[24] L. Tong, G. Xu, T. Kailath, A new approach to blind identification and equalization of multipath channels, in: Proceedings of the Asilomar Conference on Signals, Systems, and Computers, 1991, pp. 856–860.

[25] L. Tong, G. Xu, T. Kailath, Blind identification and equalization based on second-order statistics: a time domain approach, IEEE Trans. Inform. Theory 40 (2) (1994) 340–349.

[26] L. Vielva, D. Erdogmus, J.C. Príncipe, Underdetermined blind source separation using a probabilistic source sparsity model, in: Proceedings of the Second International Conference on Independent Component Analysis and Blind Signal Separation (ICA), San Diego, CA, USA, 2001, pp. 675–679.

[27] L. Vielva, Y. Pereiro, D. Erdogmus, J.C. Príncipe, Inversion techniques for underdetermined BSS in an arbitrary number of dimensions, in: Proceedings of the Fourth International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Nara, Japan, 2003, pp. 131–136.

[28] L. Vielva, I. Santamaría, C. Pantaleón, J. Ibáñez, D. Erdogmus, J.C. Príncipe, Estimation of the mixing matrix for underdetermined blind source separation using spectral estimation techniques, in: Proceedings of the XI European Signal Processing Conference (EUSIPCO), 2002, pp. 557–560.

[29] Y. Wang, J. Xuan, R. Srikanchana, J. Zhang, Z. Szabo, Z. Bhujwalla, P. Choyke, K. Li, Computed simultaneous imaging of multiple biomarkers, in: Proceedings of the XIII IEEE International Workshop on Neural Networks for Signal Processing (NNSP), Toulouse, France, 2003, pp. 269–278.

[30] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust. Speech Signal Process. ASSP-33 (2) (1985) 387–392.

[31] M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, Neural Comput. 13 (2001) 863–882.

David Luengo was born in Santander, Spain, in 1974. He received the Radiocommunication Bachelor Engineer degree and the Telecommunication Engineer Degree from the University of Cantabria, Spain, in 1994 and 1998, respectively. From 1998 to 2003 he was a Research Associate at the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Spain. In 2002, he spent a visiting period at the Coordinated Science Laboratory (CSL), University of Illinois. In 2003 he joined the Universidad Carlos III at Madrid, where he is currently an Assistant Professor. His current research interests include applications of chaos theory to signal processing and communications, machine learning algorithms, and blind source separation.

Ignacio Santamaría received the Telecommunication Engineer Degree and the Ph.D. in electrical engineering from the Polytechnic University of Madrid, Spain, in 1991 and 1995, respectively. In 1992 he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Spain, where he is currently an Associate Professor. He held visiting positions in 2000 and 2004 at the Computational NeuroEngineering Laboratory (CNEL), University of Florida. He has authored and coauthored more than 70 publications in refereed journals and international conference papers. His current research interests include nonlinear modelling techniques, adaptive systems, machine learning algorithms and their application to digital communication systems.

Luis Vielva was born in Santander, Spain, in 1966. He received his Licenciado degree in Physics and his Ph.D. in Physics from the University of Cantabria, Spain in 1989 and 1997, respectively. In 1989 he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Spain, where he is currently an Associate Professor. In 2001, he spent a visiting period at the Computational NeuroEngineering Laboratory (CNEL), University of Florida. Dr. Vielva has more than 50 publications in refereed journals and international conference papers. His current research interests include blind source separation and bioinformatics.