INVERSION TECHNIQUES FOR UNDERDETERMINED BSS IN AN ARBITRARY NUMBER OF DIMENSIONS

L. Vielva, Y. Pereiro

GTAS, DICOM
University of Cantabria
Santander, Spain

D. Erdoğmuş, J. C. Príncipe

CNEL
University of Florida
Gainesville, FL 32611, USA

ABSTRACT

The underdetermined Blind Source Separation (BSS) problem consist of estimating n sources from the measurements provided by $m < n$ sensors. In the noise-free linear model, the measurements are a linear combination of the sources, so that the mixing process is represented by a rectangular mixing matrix of m rows and n columns. The solution process can be decomposed in two stages: first estimate the mixing matrix from the measurements, and then estimate the "best" solution compatible with the underdetermined set of linear equations. Most of the results presented for the underdetermined BSS problem are based on geometric ideas valid for the $m = 2$ scenario. In this paper we extend these ideas to higher dimensions, and develop techniques to both estimate the mixing matrix and to invert the underdetermined linear problem that are valid for an arbitrary number of sources and measurements, provided $1 < m < n$.

1. INTRODUCTION

Blind source separation (BSS) is concerned with the problem of estimating n source signals from the measurements provided by m sensors. The measurements are generated as a mixture of the original sources, and "blind" adjective indicates that neither the sources nor the mixing process are known. Many models of the mixing process can be considered, such as instantaneous, delayed or convolved sources; noisy or noise-free mixing; and the relative number of sensors and sources. In this paper we focus on the noise-free linear model, in which n sources are linearly combined, through an unknown mixing matrix, in order to provide m measurements.

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

where $\mathbf{s} \in \mathbb{R}^n$ is the source random vector, $\mathbf{x} \in \mathbb{R}^m$ is the measurement random vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the unknown mixing matrix. The square case, with as many sensors as sources ($m = n$), has been extensively studied in the literature [1, 2]. The algorithms designed for this case usually exploit the assumed independence of the sources, and the problem reduces to estimate $\mathbf{A}$, since once the mixing matrix is known, the sources can be readily obtained by applying the inverse of $\mathbf{A}$ to the measurements, as (1) indicates. The overdetermined case, with more measurements than sources ($m > n$) can also be readily solved. Since now there are more linear equations than sources, the problem will not have in general an exact solution. There exist however a canonical procedure to find the solution that minimizes the L_2-norm of the error: estimate $\mathbf{A}$ and apply its pseudo-inverse [3] to the measurements in order to estimate the sources.

In this paper we are interested in the underdetermined case, in which the number of available sensors is smaller than the number of sources to estimate ($m < n$). In this case the solution of the BSS problem can also be formulated as a two-stage procedure: first estimate the mixing matrix $\mathbf{A}$ and then estimate the sources [4, 5]. The pseudo-inverse provides the solution that minimizes the L_2-norm. Observe that there is a clear difference between the solution provided by the pseudo-inverse in the overdetermined and underdetermined cases: in the first it provides the estimation of the sources that minimizes the error; in the second it provides the compatible solution with minimum norm. Nonetheless, in the underdetermined BSS case we are faced with an inescapable fact: since we have less sensors than sources, some information is lost and the separation procedure can not be perfect in general. This loss of information can be readily understood from a geometrical point of view. The mixing process consists on a linear transform of the sources from the linear space $\mathbb{R}^n$ into the lower-dimension linear space $\mathbb{R}^m$, so that a kind of projection has been performed, and the separation process consists on estimating the source vectors from the projections.

A well-known fact of the underdetermined BSS problem is that the performance of the separation algorithms improve, both for the estimation of the mixing matrix [6] and for the estimation of the sources [7], as the sparsity of the sources is higher. When the sources are not sparse in the original domain, a suitable linear transform, like Discrete Cosine Transform (DCT), short-time Fourier Transform (STFT), and wavelets [8], can be performed so that the coefficients that represent the signal in the new domain are indeed sparse.

To parametrically model sources with different degrees of sparsity, the following model for the source densities is used

$$p_{S_j}(s_j) = p_j\, \delta(s_j) + (1 - p_j) f_{S_j}(s_j), \quad j = 1, \ldots, n, \tag{2}$$

where s_j is the j-th source, p_j is the sparsity factor for s_j, and $f_{S_j}(s_j)$ is the PDF when the source j—that is as-

This work has been partially supported by Spanish Ministry of Science and Technology under project TIC2001-0751-C04-03.

sumed to be zero-mean—is active. The performance of this two-stage procedure strongly depends on the sparsity of the sources, both for the estimation of the mixing matrix [6] and for the estimation of the sources [7]: the higher the sparsity factor—the lower the probability of sources being active simultaneously—the better the estimation of mixing matrix and the recovery of the sources.

Most of the results on underdetermined BSS [7, 8] consider the case with two sensors ($m = 2$), in which the mixing matrix can be obtained, from a geometrical point of view [9], by finding the maxima of a unidimensional probability density function (PDF). However, the direct extension of this method to scenarios with more than two sensors requires finding the maxima of a multidimensional PDF [10], that, in addition to be computationally more complex, requires a number of samples that depends exponentially on the number of dimensions. In the usual $m = 2$ underdetermined scenario it is also customary to improve the results provided by pseudo-inverse when the sparsity of the sources is high by relying on heuristic criteria that have a clear geometrical interpretation on the basis of regions defined by the columns of the mixing matrix [8]. However, the extension of this criteria to higher dimensions it is also not obvious.

In this paper, we extend our previous work on underdetermined BSS [4] to deal with an arbitrary number of sensors (more than one) and an arbitrary number of sources. We present extensions for the two stages of the separation process: the estimation of the mixing matrix and the estimation of the sources. For the first stage, the basic procedure is to parametrically describe the mixing matrix in generalized spherical coordinates, to project onto the plane associated to each one of the coordinated angles, to estimate the maxima of the $m - 1$ unidimensional PDFs, and to select from all the possible combinations of angles those that really correspond to the columns of the mixing matrix. For the second stage, the generalization covers two aspects: on the one hand, for the heuristic approach, we extend the concept of angle bisector, that is appropriate for every combination of two vectors in the plane for the $m = 2$ scenario, to an m-dimensional Delaunay tessellation that is appropriate for an arbitrary high number of dimensions. On the other hand, we propose maximum likelihood (ML) and maximum a posteriori (MAP) estimators valid for an arbitrary number of dimensions.

The organization of the paper is as follows: In Section 2, we formulate the problem of estimating the mixing matrix as the problem of finding the maxima of an $(m - 1)$-dimensional PDF. In Section 2.1, we introduce the projection procedure that reduces the peak estimation problem from a multidimensional PDF to $m - 1$ decoupled unidimensional PDFs, and show how to elucidate the spurious combinations of peaks from those that are true maxima of the $(m - 1)$-dimensional PDF. In Section 3 we derive both heuristic and probabilistic estimators to perform the inversion of the matrix in an arbitrary number of dimensions. In Section 4, we validate the proposed method with a series of Montecarlo simulations. In Section 5 we present the conclusions of this work.

2. ESTIMATION OF THE MIXING MATRIX

Equation (1) can be interpreted from a geometrical point of view as the projection of the source vectors $\mathbf{s}$ from $\mathbb{R}^n$ into the vector space $\mathbb{R}^m$ of the measurement vectors $\mathbf{x}$. If we denote by $\mathbf{a}_j$ the j-th column of the mixing matrix, so that $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n]$, (1) can be rewritten as

$$\mathbf{x} = \sum_{j=1}^{n} s_j \mathbf{a}_j = s_1\mathbf{a}_1 + \cdots + s_n\mathbf{a}_n, \tag{3}$$

that explicitly shows that the measurement vector is a linear combination of the columns of the mixing matrix. According to this interpretation, if at a given time only the j-th source is non-zero, the measurement vector will be collinear with $\mathbf{a}_j$. When more than one source is active at the same time, the measurement will be a linear combination of the corresponding columns of the mixing matrix. For higher sparsity factors, the measurements are more concentrated along the directions of the columns of the mixing matrix [4].

The first step in our recovery procedure is to convert all the points of the m-dimensional vector space of the measurements and the columns of the mixing matrix from a Cartesian representation to a spherical coordinate system, where every point $\mathbf{x}$ of Cartesian coordinates $(x_1, \ldots, x_m)$ is represented by its modulus r and by $m - 1$ angles θ_i as

$$\begin{aligned}
x_1 &= r\cos\theta_{m-1}\cos\theta_{m-2}\cdots\cos\theta_3\cos\theta_2\cos\theta_1,\\
x_2 &= r\cos\theta_{m-1}\cos\theta_{m-2}\cdots\cos\theta_3\cos\theta_2\sin\theta_1,\\
x_3 &= r\cos\theta_{m-1}\cos\theta_{m-2}\cdots\cos\theta_3\sin\theta_2,\\
&\vdots\\
x_{m-1} &= r\cos\theta_{m-1}\sin\theta_{m-2},\\
x_m &= r\sin\theta_{m-1}.
\end{aligned}$$

According to this definition, the angles can be determined from the rectangular coordinates as

$$\theta_i = \arctan\frac{x_{i+1}}{\sqrt{\sum_{l=1}^{i} x_l^2}}, \quad i = 1, \ldots, m-1. \tag{4}$$

If we apply (4) to the measurements of an scenario with three sensors ($m = 3$) and four sources ($n = 4$) of sparsity factor 0.5, and represent an histogram taking as independent variables the $m-1$ angles, we obtain the results shown in Figure 1.

2.1. Dimension Reduction by projection

In Figure 1 it can be observed that the $(m-1)$-dimensional PDF is composed of a set of n peaks that, even for an sparsity factor of 0.5, are quite narrow. In Figure 2, a top view of the $(m - 1)$- dimensional PDF is shown. The black spots correspond to the locations of the maxima from Figure 1.

Since we are interested in determining only the position of the peaks, and not the shape of the PDF, all the information we are looking for can be extracted from the $m - 1$ projections onto the unidimensional vector spaces corresponding to conserving only one spherical coordinate

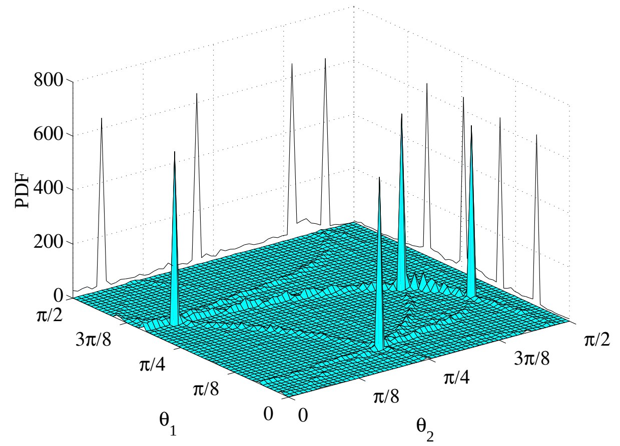

Figure 1: Histogram of angles for the measurements of a scenario with three sensors ($m = 3$) and four sources ($n = 4$) of sparsity factor 0.5. The $(m-1)$-unidimensional projections onto the plane of angle θ_i, $i = 1, \ldots, m-1$ are also shown.

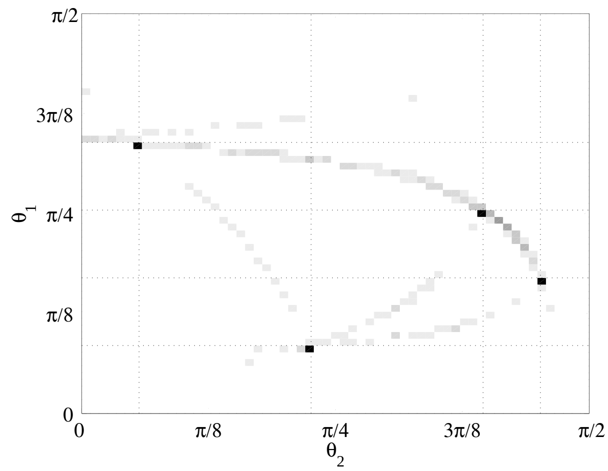

Figure 2: Top view of the $(m-1)$-dimensional PDF corresponding to the spherical angles of the measurements. The black spots correspond to the locations of the maxima from Figure 1.

and making zero all the other angles. These projections are shown in Figure 1 for the case of three sensors and four sources, which we are using as an example. They can be considered as the set of $m-1$ unidimensional PDFs of the $m-1$ spherical angles that are shown as projections in Figure 1.

To each of these $m-1$ unidimensional PDFs of the angles that parameterize the measurements, a method has to be applied to find up to n maxima, whose locations correspond to the estimates $\hat{\theta}_{ij}$, $i = 1, \ldots, m-1$, $j = 1, \ldots, n$. A number of methods could be applied, from the simpler one of calculating the histogram and finding the maxima, to the use of nonparametric estimation by means of Parzen windowing [6], or to the use of spectral estimation techniques suitable for the estimation of sinusoids in noise [11].

Once the estimations of the individual spherical angles are obtained, it is necesary to reconstruct the position of the maxima of the multidimensional PDF from the unidimensional projections. From these projections, all the combinations of angles could be constructed, as it is shown with dotted lines in Figure 2, and a method has to be implemented that allows to distinguish the correct combinations from the spurious solutions. We define a small area around each combination of angles, that constitutes a tentative solution, and count how many measurements fall into that area. The correct combinations will have a high number of occurrences, but a point falling into the region associated to a spurious combination will be an improbable event. Since the number of combinations of angles is n^{m-1}, the procedure to elucidate which are the correct combinations of angles is to construct an $(m-1)$-dimensional count array of length n in each of the dimensions and find the maxima for each intersection of the $m-1$ dimensions.

3. INVERSION OF THE LINEAR PROBLEM

The problem of estimating $\mathbf{s}$ from equation (1) when the mixing matrix $\mathbf{A}$ —which is assumed to be full rank— and $\mathbf{x}$ are known depends of the relation between m and n. If $m = n$ the problem is trivial, because the solution is given by $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$. In the overdetermined case ($m > n$), the pseudo inverse [12] $\mathbf{A}^{+}$ provides the solution $\mathbf{s} = \mathbf{A}^{+}\mathbf{x}$ that minimizes the L_2 norm of the residue, $||\mathbf{x} - \mathbf{A}\mathbf{s}||$. In the underdetermined case ($m < n$) the problem (1) has an infinite number of solutions, so it is necessary to impose some aditional criterion to select one solution vector $\mathbf{s}$. One possible criterion of general applicability could be to impose some norm L_p of the solution to be a minimum. Specifically, the solution provided by the pseudo inverse is the one that minimices the L_2 norm of the solution $||\mathbf{s}||$, and with no additional knowledge of the statistics of the sources could be the canonical option to choose [3]. As we will show next, if the signals admit a sparse representation, it is possible to design better inversion strategies.

3.1. Heuristic approaches

If, for a given sample of the source vector $\mathbf{s}$, only the j-th component is not null, the measurement $\mathbf{x}$ will be collinear with the j-th column of the matrix $\mathbf{A}$ and the components of the source vector will be

$$s_k = \frac{\mathbf{a}_j^T \mathbf{x}}{\mathbf{a}_j^T \mathbf{a}_j} \delta_j^k, \quad k = 1, \ldots, n, \tag{5}$$

where the superscript T denotes transpose, and δ_j^k is the Kronecker delta. In a real situation, even with highly sparse sources, the signals will rarely be exactly zero, but at each sample there will be some probability of one of the sources being significatively bigger than the others. To estimate which one that component is, we could choose the one that

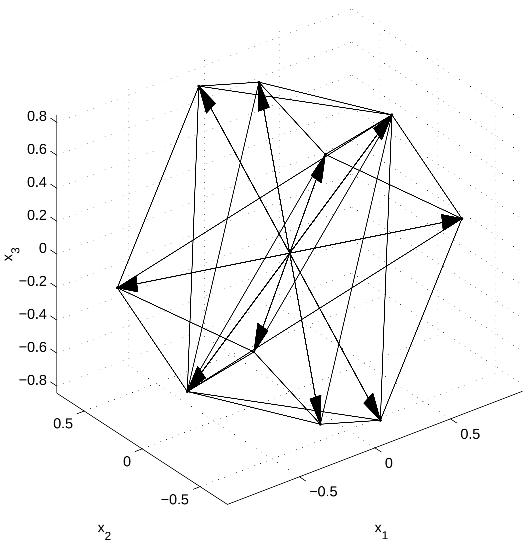

Figure 3: Division of the three-dimensional space into disjoint regions by a Delaunay tessellation for a scenario with $m = 3$ measurements and $n = 5$ sources.

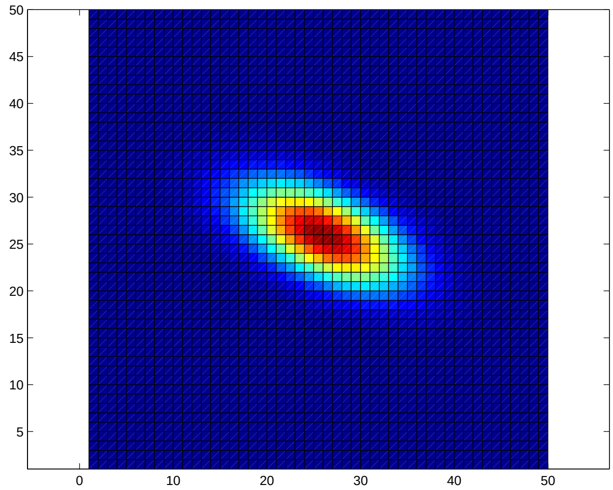

Figure 4: Bidimensional PDF of a reduced mixing matrix for an scenario with $m = 2$ measurements and $n = 3$ sources.

maximizes the normalized projection on to the directions of each column of $\mathbf{A}$.

Another family of methods could be based on building at each time step a reduced square matrix $\mathbf{A}_r \in \mathbb{R}^{m\times m}$ using m vectors of $\mathbb{R}^m$, chosen between the n column vectors $\mathbf{a}_j$ according to some optimization criterion. The resulting source vector $\mathbf{s}$ will have $n - m$ zeros corresponding to the non-selected columns, and the other components will be given by $\mathbf{A}_r^{-1}\mathbf{x}$. There are many ways of selecting the appropiate columns of the reduced matrix. In [8] a method of this family is proposed for the $m = 2$ case. The criterion it uses is to divide $\mathbb{R}^2$ into the sectors defined by the column vectors $\mathbf{a}_j$ and to choose at each sample those two vectos that surround the measurement $\mathbf{x}$.

When the number of measurements is greater than two, we are faced with the problem of dividing an m-dimensional space into regions delimitated by m vectors. A generalization of the division based on sectors, that has been successfully applied for the scenario with $m = 2$ [8], could be to use a Delaunay tessellation. In figure 3 we show a tessellation for a scenario with three measurements and five sources. Each vector corresponds to a column of the mixing matrix or its opposite. As it can be readily observed, the Delaunay tessellation divides the three dimensional measurement space into regions so that an unique set of three vectors can be used to build a reduced inversion matrix depending on the region associated to each individual measurement.

3.2. ML and MAP estimation

According to (3), if at any given time, we knew that at most m components of the signal are non zero, the problem would not be underdetermined any more and we could invert it (provided that we know which are the non zero components). The first step is to build all the possible reduced matrices, taking m from the n columns of the mixing matrix $\mathbf{A}$, and calculate the PDF associated to the linear combinations of sources through each of those reduced matrices. In this way we obtain $\binom{n}{m}$ m-dimensional PDFs that allow us to evaluate the likelihood of a given measurement in the m-dimensional space. In figure 4 we show an example for the PDF associated to the combination of two columns of the mixing matrix. It can be observed that the shape is an ellipsoid, which main axes could be computed by means of the eigenvectors $\mathbf{A}\mathbf{A}^T$.

The ML approach to estimate the n sources would consist on associating a given measurement to the reduced matrix that maximizes the likelihood, and invert $\mathbf{s}$ according to the inverse of the choosen reduced mixing matrix. If the sparsity factor of the sources is known beforehand, we can calculate the a priori probabilities of the different combinations of sources, so that the MAP estimator could be formulated [7].

This ML and MAP estimation can be considered as a classification problem: given a measurement, which is the most probable mixing matrix that produced it? If we estimate that the most probable is a reduced mixing matrix, we apply its inverse. If we estimate that the most probable is the complete mixing matrix $\mathbf{A}$, we apply its pseudo-inverse. Figures 5 and 6 illustrate this classification problem for a scenario with three sources and two measurements, for sparsity factors of $p = 0.5$ and $p = 0.8$, respectivelly. Since we have $m = 2$ and $n = 3$, a total of 3 bidimensional reduced matrices is available, that, in addition to the class corresponding to the complete mixing matrix $\mathbf{A}$ amounts to four different regions in the bidimensional measurement space. In both figures, the dark grey region that is divided into six different subregions is associated with the pseudo inverse. The other three regions, each one with two opposited subregions, correspond to each reduced mixing matrix. The directions of the columns of the mixing matrix mark the the discontinuities between regions near the origin. By

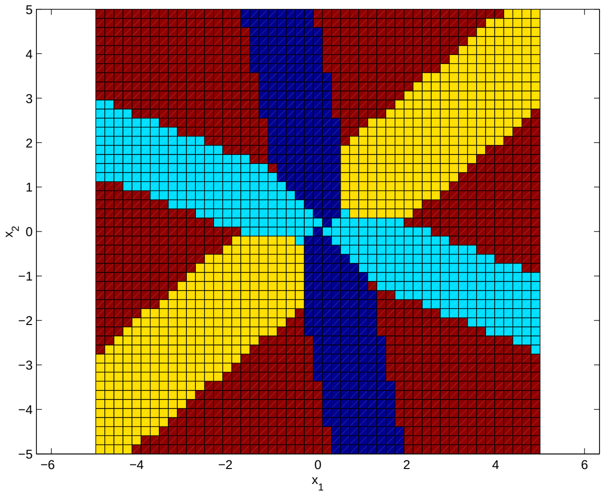

Figure 5: Classification regions of the bidimensional measurement space for a scenario with $m = 2$ and $n = 3$ and sparsity factor $p = 0.5$.

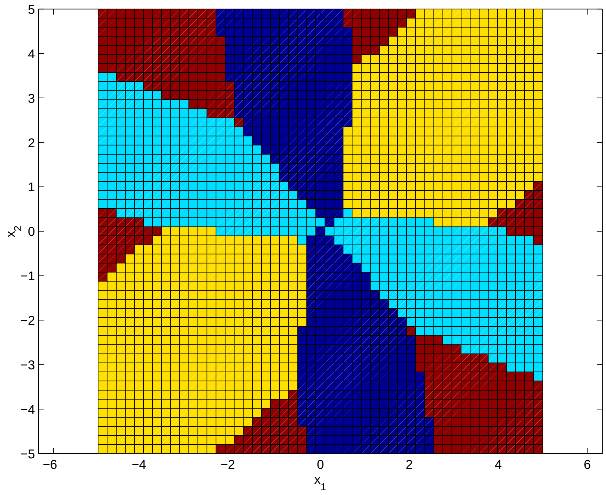

Figure 6: Classification regions of the bidimensional measurement space for a scenario with $m = 2$ and $n = 3$ and sparsity factor $p = 0.8$.

comparing these figures, some insight can be obtained: on the one hand, as the sparsity factor grows, the number of measurements that should be inverted by the pseudo inverse decreases (since is less probable that more than m sources are active at the same time). On the other hand, as the sparsity factor grows, the boundaries of the regions tend from being dictated by the bisector (the main eigenvector of $\mathbf{AA}^T$) to being dictated by the directions of the columns of the mixing matrix (which is a justification to the heuristic approach suggested above).

4. NUMERICAL RESULTS

We have performed simulations to validate both stages of the inversion process: matrix estimation and inversion of the underdetermined linear problem.

To characterize the performance of our matrix estimation method, Montecarlo simulations have been performed to estimate the mixing matrix from scenarios with different numbers of sources and sensors. In all the cases, the source realizations have been generated according to the model in (2), using as $f_{S_j}(s_j)$, $j = 1, \ldots, m$, Gaussian densities with zero mean and unit variance. The simulations have been performed as follows: for each scenario, twenty thousand samples of sources with sparsity factors from 0.05 to 0.95 have been produced. For each scenario and sparsity factor, four hundred mixing matrices have been randomly generated, the spherical angles have been estimated from the unidimensional projected PDFs, and the criterion to select the correct combination of angles has been applied. The different scenarios considered are those associated with a number of sensors ranging from two to five, and a number of sources ranging from one to ten. As the figure of merit we have selected the number of errors in the estimation of the angles (defining a tolerance on the basis of the bin length used on the histograms). We define the mean error rate as the mean number of errors for all the mixing matrices divided by the total number of angles to estimate. Figure 7 shows the results from scenarios with five sensors ($m = 5$) and a number of sources from six to twelve ($6 \leq n \leq 12$) for all the sparsity factors considered. It can be observed that the number of errors grows with the number of sources (more peaks have to be estimated from the same data, and the mean distance between peaks decreases), and diminish with the sparsity factor (the measurements tend to be more concentrated along the columns of the mixing matrix, reducing the spreading that confuses the estimation).

To characterize the performance of the inversion criteria, a scenario with three measurements and five sources has been simulated by using one hundred random matrices. Figure 8 shows the signal to noise ratio (SNR) on the estimation of the sources as a function of the sparsity factor for three different criteria: pseudo-inverse, ML, and MAP estimators. It can be observed that nothing outperforms the pseudo-inverse when the sparsity factor is very low (since the most probable situation is that all the sources are active at the same time), but, as the sparsity factor grows, the ML and MAP estimators both outperform the pseudo inverse.

5. CONCLUSIONS

In this paper we have presented methods for underdetermined BSS problems that allow to both estimate the mixing matrix and invert the underdetermined linear problem in an arbitrary number of dimensions. For the first stage of estimating the mixing matrix, the approach is based on parameterizing both the measurements and the columns of the mixing matrix in spherical coordinates and on estimating the peaks of the multidimensional PDF associated with the angles of the measurements. Since the estimation of multidimensional PDFs is a complex problem, we propose to project onto as many unidimensional PDFs as the num-

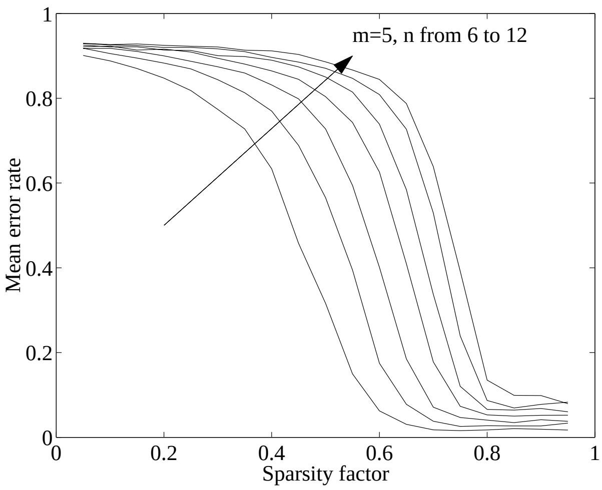

Figure 7: Mean error rate for scenarios with a fixed number of five sensors and a number of sources ranging from six to twelve, as a function of the sparsity factor of the sources.

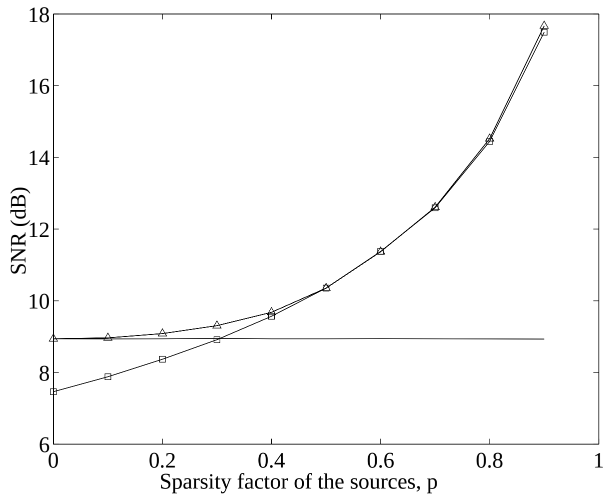

Figure 8: Signal to noise ratio for the estimation of five sources with three measurements as a function of the sparsity factor of the sources. Pseudo-inverse (solid line), ML estimator (□), and MAP estimator (△).

ber of spherical angles (the number of sensors minus one). For the second stage of inverting the underdetermined linear problem, we have presented both a heuristic approach in which the classification is performed by means of a m-dimensional Delaunay tessellation, and obtained ML and MAP estimators. The Montecarlo simulations have shown that our method provides excellent results for an arbitrary number of sources and sensors provided that the sparsity factor is high enough (around 0.75). The intuitive result that the performance improves with the number of measurements and the sparsity factor, and degrades with the number of sources has also been corroborated.

6. REFERENCES

[1] A. Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.

[2] S. Haykin, Ed., *Unsupervised Adaptive Filtering, Vol I: Blind Source Separation*, John Wiley & Sons, New York, 2000.

[3] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd edition, 1996.

[4] L. Vielva, D. Erdoğmuş, and J. C. Príncipe, "Underdetermined blind source separation in a time-varying environment," in *Proceedings ICASSP-02 (IEEE International Conference on Acoustics, Speech and Signal Processing)"*, Orlando, FL, May 2002, pp. 3049–3052.

[5] M. Zibulevsky, B. Pearlmutter, P. Bofill, and P. Kisilev, *Independent Components Analysis: Principles and Practice*, chapter Blind source separation by sparse decomposition in a signal dictionary, Cambridge University Press, 2000.

[6] D. Erdoğmuş, L. Vielva, and J. C. Príncipe, "Nonparametric estimation and tracking of the mixing matrix for underdetermined blind source separation," in *Proceedings of ICA-2001, Independent Component Analysis*, San Diego. CA, 2001, pp. 189–193.

[7] L. Vielva, D. Erdoğmuş, and J. C. Príncipe, "Underdetermined blind source separation using a probabilistic source sparsity model," in *Proceedings of ICA-2001, Independent Component Analysis*, San Diego. CA, Dec. 2001, pp. 675–679.

[8] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[9] C. G. Puntonet, A. Prieto, C. Jutten, M. Rodriguez-Alvarez, and J.Ortega, "Separation of sources: A geometry-based procedure for reconstruction of n-valued signals," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 46, no. 3, pp. 267–284, 1995.

[10] A. Jung, F. Theis, C. Puntonet, and E. Lang, "Fastgeo - a histogram based approach to linear geometric ICA," in *Independent Component Analysis*, San Diego. CA, Dec. 2001, pp. 349–354.

[11] L. Vielva, I. Santamaría, C. Pantaleón, J. Ibáñez, D. Erdoğmuş, and J. C. Príncipe, "Estimation of the mixing matrix for underdetermined blind source separation using spectral estimation techniques," in *Proceedings Eusipco-2002 (XI European Signal Processing Conference)*, Toulouse, France, September 3–6 2002, vol. I, pp. 557–560.

[12] Lloyd N. Trefethen and David Bau, *Numerical Linear Algebra*, pp. 77–85, SIAM, 1997.