Geometrical Blind Separation of Sources Based in a Discrete-Lattice of the Space of Observations

M. Rodríguez-Álvarez, C. G. Puntonet, I. Rojas, F. Rojas, A. Prieto


Abstract. The techniques of Blind Separation of Sources (BSS) are used in many Signal Processing applications in which the data sampled by sensors are a mixture of signals from different sources, and the goal is to obtain an estimation of the sources from the mixtures. This work shows a new method for blind separation of sources, based on geometrical considerations concerning the observation space. This new method is applied to a mixture of two sources and it obtains the coefficients of the unknown mixture matrix $A$ and separates the unknown sources $S_o$. Following an introduction, we present a brief abstract of previous work by other authors, the principles of the method and a description of the algorithm, together with some simulations.

Keywords. Blind Separation of Sources, Mixture Matrix, Linear Mixture of Sources, Space of Observations.

1 Introduction

The separation of independent source signals from mixed observed data, henceforth called sensor signals, is a fundamental and challenging signal processing problem. In many practical situations, one or more desired signals need to be recovered blindly knowing only the observed sensor signals. The term Blind refers to the lack of a priori information of the source signals and propagation model from the sources to the sensors. In many practical situations, observations may be modeled as linear mixtures of a number of source signals, i.e. a linear multi-input multi-output system. A typical example is speech recordings made in an acoustic environment in the presence of background noise and/or competing speakers. Other examples include Electro Encephalographics signals, passive sonar applications and cross-talk in data communications.

When $p$ different signals propagating through a real medium have to be captured by sensors, these sensors are sensitive to all sources $s_n(t)$ and thus the signal, $e_k(t)$, observed at the output of sensor $k$, is a mixture of source signals. The solution of the problem of separation consists of retrieving the unknown sources $s_n(t)$ from just the observations $e_k(t)$, eliminating the effect introduced by the medium. To achieve this it is necessary to apply the following hypotheses:

1. The sources $s_n(t)$ and the mixture matrix $A$ are unknown.
2. The number of sensors is equal to the number of sources, $p$.
3. The observed signals $e_k(t)$, satisfy:
\[ e_k(t) = \sum_{i=1}^{p} a_{ki} \cdot s_{io}(t), \ k = 1, \ldots, p, \ a_{ki} \in \mathbb{R} \]  
\[ (1) \]

In matrix notation equation (1) can be written:

\[ E(t) = A \cdot S_o(t); \ E, S_o \in \mathbb{R}^p, \ A \in \mathbb{R}^{p \times p} \]  
\[ (2) \]

Hypothesis 1 establishes the conditions for a Blind Separation of Sources. Equation (1) shows an instantaneous linear mixture. \( E(t) \) and \( S_o(t) \) denote column vectors with components \( e_k(t) \) and \( s_{io}(t) \) respectively, which are included in expression (2). \( A \) is a \( p \times p \) matrix (called mixture matrix), which we assume to be regular. This matrix \( A \) is introduced in order to model the effect of the real medium. To resolve the problem of separation of sources a matrix \( W \) is defined, such that:

\[ S(t) = W^{-1}(t) \cdot E(t) = W^{-1}(t) \cdot A \cdot S_o(t) \]  
\[ (3) \]

The procedure is intended to obtain \( W^{-1} \) such that \( W^{-1} \cdot A = D \cdot P \), that is, a diagonal matrix \( D \) modified by a permutation matrix \( P \).

# 2 Previous Work

A great diversity of estimation methods for ICA have been proposed to solve the problem of Blind Separation of Sources, most of which use some kind of statistical analysis. Initially, Jutten et al. [1] proposed a solution based on a neural network (Herault-Jutten’s network). Other authors have developed algorithms based on higher-order statistics [2], the contrast function concept [3], Maximum Likelihood (ML) techniques [4], neural networks for the separation of non-stationary signals [5], the entropy concept [6] or the geometric structure of the signal spaces [7], [8], [9]. The two methods most widely used in practice seem to be the fixed-point algorithm [10] and the maximum likelihood stochastic gradient algorithm [11]. The first alternative was originally derived from objective functions motivated by projection pursuit [12] and therefore its connection to the estimation of the ICA data model has been rather indirect. Maximum likelihood estimation is, in contrast, the mainstream method of statistical estimation.

In our previous work ([7], [8], [9]) we have proposed a geometrical procedure to separate digital or analog signals mixed linearly or non-linearly from independent sources.

Another closely related subject is the Adaptive Noise Cancellation (ANC) problem [13][14][15]. Widrow et al first addressed the Adaptive Noise Cancellation problem which then became one of the fundamental signal processing principles. Strube [16] suggests a method for speaker separation recorded by two microphones using Adaptive Noise Cancellation (ANC) techniques as a solution to the cocktail-party problem.

Recently, blind source separation by Independent Component Analysis (ICA) has received much attention because of its potential applications in signal processing such
as in speech recognition systems, telecommunications and medical signal processing. The goal of ICA is to extract from observed sensor signals underlying source signals which are as statistically independent as possible. With linear ICA it is assumed that the observed sensor signals are unknown linear mixtures of unobservable independent source signals. In contrast to correlation-based transformations such as Principal Component Analysis (PCA), ICA not only decorrelates the signals (2nd-order statistics) but also reduces higher-order statistical dependencies, attempting to make the signals as independent as possible.

In parallel to blind source separation studies, unsupervised learning rules based on information-theory were proposed in [17]. The goal was to maximize the mutual information between the inputs and outputs of a neural network, later called the Infomax-principle. This approach is related to the principle of redundancy reduction suggested in [18] as a coding strategy in neurons. Each neuron should encode features that are as statistically independent as possible from other neurons over a natural ensemble of inputs; decorrelation as a strategy for visual processing was explored by Atick [19]. In [6] and [20] the Infomax-principle was generalized to nonlinear processing units to cast the blind source separation problem into an information-theoretic framework and to demonstrate the separation and deconvolution of mixed sources. Their adaptive methods are more plausible from a neural processing perspective than the cumulant-based cost functions proposed in [2].

In the sequel many different approaches have been attempted by numerous researchers using neural networks, artificial learning, higher order statistics, the geometry of the signal spaces, minimum mutual information, beam forming and adaptive noise cancellation, each claiming various degrees of success. Despite the diversity of the approaches, the fundamental idea of the source signals being statistically independent remains the single most important assumption in most of these schemes. The independence assumption is used as the separation criterion in the majority of the schemes mentioned, in other words, separation is deemed successful if the output signals satisfy the independence criterion.

3 Principles of the new Method

For \( p = 2 \) and \((s_{o1}, s_{o2}) \in \mathbb{R}^+\), with bounded values \( s_{o1} \in [0, s_{o1}^M] \) and \( s_{o2} \in [0, s_{o2}^M] \), and for two random uniform noises, the observed signals \((e_{1}(t), e_{2}(t))\) form a parallelogram in the \((e_{1}, e_{2})\) space. In this special case we have demonstrated [8] that, through a matrix transformation, the coefficients of the \(W\) matrix coincide with the slopes of the parallelogram.

Nevertheless, if \((s_{o1}, s_{o2}) \in \mathbb{R}^+\), then the parallelogram is not strictly positive. In this case, it is possible to translate the parallelogram by detecting translation vector \(T\) and to translate the space of observations with this vector \(T\). This translation is now equivalent to considering the signals \((s_{o1}, s_{o2}) \in \mathbb{R}^+\). As shown in Fig. 1, the linear application represented for the mixture matrix \(A\) transforms the original form in the space of sources \((s_{o1}, s_{o2})\), into a parallelogram in the space of observations \((e_{1}, e_{2})\).
From Fig. 2 and it can be seen the space of observations \((e_1, e_2)\), that, for random uniform noises as sources, the parallelogram is geometrically bounded for the segments between the points \(P_1 = (p_{11}, p_{12})\), \(P_2 = (p_{21}, p_{22})\), \(P_3 = (p_{31}, p_{32})\) and \(P_4 = (p_{41}, p_{42})\). The slopes of these segments give the coefficients of the mixture matrix \(W\), as shown in Fig. 1. In order to obtain these segments, it is necessary to know the coordinates of the points \(P_1 = (p_{11}, p_{12})\), \(P_2 = (p_{21}, p_{22})\), \(P_3 = (p_{31}, p_{32})\) y \(P_4 = (p_{41}, p_{42})\).

In other case, for example for speech signals, the form of the figure in the space of observations is very different, and it can be seen in Fig. 3.

Than this method proceeds as follows:

1.- First of all, the algorithm computes the kurtosis of each component of the sensor signals and also the correlation coefficients between all observations. This is to detect whether the underlying source signal distributions correspond to uniform or non-uniform distributions. If the kurtosis of all signal components is positive, the algorithm looks for high density regions of the sensor signal distribution. With uniformly distributed signals, the algorithm estimates the bounding box of the parallelogram representing the space of observations.

2.- The algorithm subdivides the space of observations \((e_1, e_2)\) into a regular lattice of cells with \(N\)-rows and \(M\)-columns (lattice of \(N\) by \(M\)) as shown in Fig. 4. Then, the algorithm computes the number of cells in the lattice in which the number of points inside it is greater than a given threshold TH.

3.- The distribution of sensor signals within each of these cells then is replaced by a prototype sensor signal vector. In this way these cells are bookmarked with a prototype vector whose position within the cell represents the corresponding distribution of observations lying inside (see Fig. 4). The prototype vector mostly
does not point towards the centre of the cell because its position is weighted by the
density of points \((e_1, e_2)\) in this cell. This step greatly reduces the complexity of the
algorithm, because the greatest number of points that the procedure needs to compute
is \(N * M\) (i.e. if the space of observations has \(10^5\) points \((e_1, e_2)\) and \(N = M = 20\)
then the procedure only requires \(N * M = 400\) points).

4.- To further reduce the number of points, the next step of the algorithm finds
those points which either form the border of the hyperparallelepiped or mark the high
density regions of the sensor signal distribution in the space of observations. To do
this, the algorithm looks for cells that have a neighbourhood (for a cell, the
neighbourhood is formed by the cells that are above, below, to the left and to the right
of it) in which there is no cell containing the endpoint of a prototype observation
vector (such cells have fewer points than the threshold TH). Then these cells without
a complete neighbourhood form the border of the distribution encompassing NR data
points in the space of observations.

5.- The algorithm then calculates the coordinates of the points \(P_1 = (p_{11}, p_{12})\) and
\(P_2 = (p_{21}, p_{22})\). The space of observations has been reduced in step 4 to NR data
points which, in two dimensions, represent pairs of coordinates \((e_1, e_2)\). In this
reduced set of NR data points, there exist data points \(P_1\) and \(P_2\) with largest Euclidean
distance (see equation (4)) between them in the space of observations:

\[
\begin{align*}
d (P_i, P_j) &= \max_i d (P_i, P_j) \quad \forall \quad i, j \in \{1, 2, \ldots, NR\} \\
\end{align*}
\]
6.- Once points $P_1$ and $P_2$ are obtained, the algorithm calculates the equation of the straight line $R_1$ which passes through points $P_1$ and $P_2$ (see equation (5)):

$$A e_1 + B e_2 + C = 0$$

where

- $A = (P_{22} - P_{12})$
- $B = (P_{11} - P_{21})$
- $C = (P_{21} \cdot P_{12}) - (P_{22} \cdot P_{11})$. 

---

**Fig. 3.** Space of observations for the linear mixture of two speech signals

**Fig. 4.** Lattice of the space of observations in cells
7.- Next the algorithm estimates the coordinates of the points \( P_3 = (p_{31}, p_{32}) \) and \( P_4 = (p_{41}, p_{42}) \) as follows: the straight line \( R_1 \) divides the space of observations \((e_1, e_2)\) into two subspaces with \( R_1 \) the border between them. Data points which lie within one of these subspaces yield a nonzero result in equation (5). For example, data point \( s \) lying above the straight line \( R_1 \) yield a negative result in equation (5). There is then one data point \( P_3 = (p_{31}, p_{32}) \) which provides the most negative value of all possible outcomes of equation (5), hence which also represents the point with the greatest Euclidean distance from the straight line \( R_1 \) in the subspace above \( R_1 \). In the same way, points in the other subspace, below the straight line \( R_1 \), yield a positive result in equation (5). Again, there is one point \( P_4 = (p_{41}, p_{42}) \) that provides the most positive value of all possible results from equation (5), and which is also the point with greatest Euclidean distance from the straight line \( R_1 \) in the subspace below \( R_1 \). In both cases, the algorithm calculates the Euclidean distance from a generic point \( P_i = (p_{i1}, p_{i2}) \) in the space of observations \((e_1, e_2)\) to the straight line \( R_1 \) by equation (6):

\[
d(P_i, R_1) = \frac{AP_{i1} + BP_{i2} + C}{\sqrt{A^2 + B^2}}
\]

where \( A, B \) and \( C \) are the coefficients of the straight line \( R_1 \) given in equation (5).

8.- Once the characteristic points of the parallelogram have been obtained, the algorithm computes either the slopes of the segments \((P_1 P_3 \text{ and } P_1 P_4)\) or, equivalently \((P_2 P_4 \text{ and } P_3 P_2)\) or the slopes of the diagonals \((P_1 P_2 \text{ and } P_3 P_4)\) in order to obtain the coefficients of the matrix \( W \) as in Eq. (7) (see Fig. 5).

\[
\begin{bmatrix}
\frac{a_{12}}{a_{11}} \\
\frac{a_{12}}{a_{22}}
\end{bmatrix} = \begin{bmatrix}
p_{12} - p_{13} \\
p_{31} - p_{11}
\end{bmatrix} \begin{bmatrix}
a_{11} \\
a_{12}
\end{bmatrix} = \begin{bmatrix}
p_{43} - p_{33} \\
p_{41} - p_{31}
\end{bmatrix}
\]

or

\[
\begin{bmatrix}
\frac{a_{12}}{a_{11}} \\
\frac{a_{12}}{a_{22}}
\end{bmatrix} = \begin{bmatrix}
p_{12} - p_{13} \\
p_{31} - p_{11}
\end{bmatrix} \begin{bmatrix}
a_{11} \\
a_{12}
\end{bmatrix} = \begin{bmatrix}
p_{23} - p_{22} \\
p_{41} - p_{42}
\end{bmatrix}
\]

9.- Using the coefficients of the calculated mixture matrix \( W \), the procedure computes the inverse matrix and reconstructs the signals \((s_i(t))\) (see equation (3)).
Fig. 5. Obtention of the straight lines of the space of observations

4 Simulations and Results

4.1 Simulation 1.

The original signals are two uniform noises. The original mixture matrix (A) and the matrix (W) obtained for the procedure are:

\[
A = \begin{pmatrix}
1 & 0.5 \\
-0.5 & 1
\end{pmatrix} ;
\quad W = \begin{pmatrix}
1 & 0.499 \\
-0.501 & 1
\end{pmatrix}
\]

The procedure uses a lattice of 20 by 20 cells to divide the space of observations (N = 20, M = 20). The total number of points for each signal is 10000. The threshold TH is set at 15 points. The Crosstalk figures obtained for each signal after reconstruction of the estimated signals \(s_i(t)\) were:

- Crosstalk 1 = - 61.50 dB
- Crosstalk 2 = - 59.09 dB

4.2 Simulation 2.

The original signals are two words. The original mixture matrix (A) and the matrix (W) obtained for the procedure are:

\[
A = \begin{pmatrix}
1 & 0.5 \\
0.5 & 1
\end{pmatrix} ;
\quad W = \begin{pmatrix}
1 & 0.52 \\
0.52 & 1
\end{pmatrix}
\]

The procedure uses a lattice of 15 by 15 cells to divide the space of observations (N = 15, M = 15). The total number of points for each signal is 10000. The threshold
TH is set at 25 points. The Crosstalk figures obtained for each signal after reconstruction of the estimated signals $s_i(t)$ were:

- Crosstalk 1 = -19.95 dB
- Crosstalk 2 = -20.13 dB

5 Conclusions

In this work we have shown a new geometry-based method for blind separation of two sources. In comparison with other methods this new one has a low computational cost and it is very intuitive in terms of computer application. Furthermore, this method could be used to detect the perimeter or outlines in simple two-dimensional figures.

6 Future Research

In the future we will intend to implement this method for more than two signals (in general for a $p$-dimensional space with $p$ mixed signals). The main idea to resolve with this new method of Blind Source Separation is to project two by two in different planes ($e_i, e_j$) the space of observations and then, to apply our algorithm.

Acknowledgments

This work has been supported by the Spanish CICYT Projects TIC2001-2845 “PROBIOCOM” (Procedures for Biomedical and Communications Source Separation), TIC2000-1348 (Hybrid Procedures for Parallel Optimisation in Clusters), and DPI2001-3219 (Design and Implementation of new Hybrid Neuro-Fuzzy Methodologies in Parallel Computing Systems: application to model and control of complex real-time systems).

References