Icassp'89. pp. 2109-2112

SOURCE SEPARATION USING HIGHER ORDER MOMENTS

Jean-Francois Cardoso

Ecole Nat. Sup. des Telecommunications - Dept SIGNAL
46 rue Barrault, 75634 PARIS CEDEX 13, FRANCE.
and
CNRS-URA 820, GRECO-TDSI.

ABSTRACT

This communication presents a simple algebraic method for the extraction of independent components in multidimensional data. Since statistical independence is a much stronger property than uncorrelation, it is possible, using higher-order moments, to identify source signatures in array data without any a-priori model for propagation or reception, that is, without directional vector parametrization, provided that the emitting sources be independent with different probability distributions. We propose such a "blind" identification procedure. Source signatures are directly identified as covariance eigenvectors after data have been orthonormalized and non linearily weighted. Potential applications to Array Processing are illustrated by a simulation consisting in a simultaneous range-bearing estimation with a passive array.

INTRODUCTION

For a lot of reasons (of various kinds), the most common Signal Processing methods deal with second-order statistics, expressed in terms of covariance matrices. It is well known that Gaussian stochastic processes are exhaustively described by their second-order statistics. Nonetheless, when the Gaussian assumption is not valid, some information is lost by retaining only second-order statistics.

This information, which can be expressed in terms of cumulants [1], is sometimes necessary, for instance in the problem of identifying a non-minimal phase process [2,3]. In the recent years, research on higher-order statistics has mainly focused on Time Series Analysis, often using more specific description tools such as bispectra [4,5]. In contrast, less attention was given to applications in Multidimensional Signal Processing.

In Array Processing, recently proposed methods for High Resolution are also based on structural analysis of the second-order covariance. Impressive performances may be achieved by giving a priori information about signal propagation from sources to sensors in terms of "directional" or "steering" vectors. For a superimposition of narrowband plane waves impinging on a uniform linear array of identical sensors, these steering vectors show linear phase and, as a consequence, the problem of source location is analog to a harmonic retrieval problem. Most of the work on using higher order information in Array Processing has been done within this framework [6,7,8]. However, actual physical settings are often such that source signatures (directional vectors) depart from the assumed model. As expected, model-based methods are very sensitive to such discrepancies. Multipath, unknown antenna deformation are among the common causes of severe performance degradation.

It is the purpose of this communication to present a simple algebraic method allowing source identification when NO a priori information about the propagation and the reception is available. The key requirement is that the observed data consist in a linear superimposition of statistically independent components. It may seem strange that such a blind identification procedure be possible, but it should be recalled that statistical independence between sources is a much stronger requirement than mere uncorrelation. The question of blind separation of multidimensional components by taking advantage of statistical independence has already been adressed in recent litterature. A non-linear adaptive procedure has been proposed in [9,10] while a direct solution using explicitely cumulants was given for the case of two sources and two sensors in [11]. In contrast, we propose here a simple algebraic method to separate an arbitrary number of sources, given measurements from a larger number of sensors.

THE SOURCE SEPARATION PROBLEM

We are given an arbitrary array of N sensors providing samples in time and space of some random field. The situation we are interested in is the following. There are several sources located at different points in space, numbered with index i. The source i emits at time t a random message $\alpha_i(t)$. Influence of source i on the sensor array is described by a fixed vector X_i called the "source signature". At time t, the observed data vector $X(t)$ is the linear superimposition of the contributions from each source:

$$(M) \begin{cases} X(t) = \sum_{sources} \alpha_i(t)\, X_i \\ \alpha_i(t) : \textit{real or complex valued stochastic variable} \\ X_i : \textit{unknown deterministic vector} \end{cases}$$

The preceeding model (M) is quite general and underlies most of the narrow-band beamforming and bearing estimation methods. In these methods, however, source signatures are usually assumed to be of known structure, parameterized by angular position. We do not want to introduce such an a priori knowledge in our problem, but we assume in place that

(H) The source messages α_i are zero-mean stationnary *independent* processes.

Because independence is a strong statistical property, it is generally possible to solve the problem:

Given the model (M) and hypothesis (H), separate the sources, that is, extract signatures X_i and messages $\alpha_i(t)$ from the observations $X(t)$.

In the following, we omit explicit dependancy on time t. We will also assume to be in the general case where the source signatures X_i are linearily independent vectors. A point to be noted, when examining the model equation (M), is that any identification procedure can be successful only up to a scale factor. This is obvious since multiplication of a source message α_i by a constant factor combined with a division of the corresponding signature X_i by the same factor leaves the observation X unchanged. So we are free, without any loss of generality, to normalize either the source amplitudes or the signature norms, In the following we choose the normalization:

(1) $$E(|\alpha_i|^2)=1 \quad \text{for } \textit{all sources}$$

SECOND-ORDER IDENTIFICATION

Statistical second-order information contained in the data is expressed by the covariance matrix:

(2) $$R_X = E(XX^T)$$

where X^T denotes the conjugate transpose of vector X. Expression of the covariance in terms of the model is immediate since the $\alpha_i's$ are independent, zero mean and normalized, yielding:

(3) $$R_X = \sum_{sources} X_i X_i^T$$

For simplicity we assume in the following that there are exactly N sources so that R_X is full rank (since in the general case, signatures are linearily independent). This is not an actual restriction: we could introduce simple modifications (consisting in working in the "signal subspace") which would leave essentially unchanged the method to be exposed but this is postponed to a following section in order to keep notations as simple as possible.

Starting from an estimate of the covariance it is impossible to solve equation 3 for unique $X_i's$ without incorporating more information. This degeneracy can be explicited as follows. The covariance being symmetric, it can be factorized into:

(4) $$R_X = CC^T$$

where C is a (non necessarily symmetric) full rank matrix. Many covariance factorizations exist but, to our our purpose, we do not need to specify one. By direct substitution it is easily checked that if $(X_i)_{i=1,N}$ is a set of vectors verifying equation 3 then another solution is $(CUC^{-T}X_i)_{i=1,N}$ where U is any unitary $(UU^T = I)$ matrix. Hence, second-order information alone allows the problem to be solved only up to a unitary transform. If a model is at hand for signatures X_i (for instance angular parametrization of X_i as a directional vector in a source bearing estimation problem) then this degeneracy can be overcome and identification can be completed. If no such model is available, we are facing a "blind" identification problem and more information is to be extracted from the data themselves: higher order moments are needed.

FOURTH-ORDER BLIND IDENTIFICATION

Our method for blind identification using fourth-order moments operates in two steps: orthonormalization and quadratic weighting. Orthonormalization is nothing new and, in our case, may be seen as a preprocessing intended to extract all second-order structure from the data. We use a covariance factorization as in equation 4 to form a new set of data according to:

(5) $$Y = C^{-1} X$$

We have

$$R_Y = E(YY^T) = C^{-1} R_X C^{-T} = I$$

where I denotes the identity matrix, so that second-order information is no longer present in the new data set Y. The linear model for X is turned into a linear model for Y:

(6) $$Y = \sum_{i=1}^{N} \alpha_i Y_i \qquad \textit{with} \quad Y_i = C^{-1} X_i$$

From 3 and 5 it appears that:

(7) $$\sum_{i=1}^{N} Y_i Y_i^T = I$$

which means that the $Y_i's$ form an orthonormal set of vectors, justifying the term "orthonormalization" for this first step.

The second step consists in forming a quadratically weighted covariance defined by:

(8) $$\tilde{R}_Y = E(|Y|^2 YY^T)$$

which may also be seen as the usual covariance of the random variable $|Y|Y$. The structure of this weighted covariance matrix is easily derived. The $Y_i's$ forming an orthonormal basis, we have $|Y|^2 = \sum_{k=1}^{N} |\alpha_k|^2$ so that:

$$|Y|^2 YY^T = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} |\alpha_k|^2 \alpha_i \alpha_j^* Y_i Y_j^T$$

Let us consider the three cases:

- If $i \neq j$ then, for any k, k differs from either i or j. If, for instance, $k \neq i$ then $E(\alpha_i \alpha_j^* |\alpha_k|^2) = E(\alpha_i) \cdot E(\alpha_j^* |\alpha_k|^2) = 0$
- If $i = j = k$ then $E(\alpha_i \alpha_j^* |\alpha_k|^2) = E(|\alpha_i|^4)$
- If $i = j \neq k$ then $E(\alpha_i \alpha_j^* |\alpha_k|^2) = E(|\alpha_i|^2) \cdot E(|\alpha_k|^2) = 1$

Finally, we get for the weighted covariance the expression:

(9) $$\tilde{R}_Y = \sum_{i=1}^{N} (\ \mu_i + N - 1\ )\ Y_i Y_i^T$$

where we denote the fourth-order moment: $\mu_i = E(|\alpha_i|^4)$.

This last equation gives a solution to our problem because the $Y_i's$, being orthonormal, appear as the eigenvectors of the weighted covariance. Hence blind identification of superimposed independent components is achieved through the following fourth-order blind identification (FOBI) algorithm:

doublebox,expand,tab (%),delim ; c s l|l. FOBI : FOURTH-ORDER BLIND IDENTIFICATION = Form the data covariance % $R_X = E(XX^T)$ Factorize the covariance % $R_X = CC^T$ Orthonormalize the data % $Y = C^{-1}X$ Form the weighted covariance % $\tilde{R}_Y = E(|Y|^2 YY^T)$ Extract eigenvectors % $\tilde{R}_Y = \sum_{i=1}^{N} (\mu_i + N - 1)\, Y_i Y_i^T$ Extract the messages % $\alpha_i = Y_i^T Y$ Or identify the signatures % $X_i = CY_i$

RESTRICTIONS, EXTENSIONS AND OTHER COMMENTS

Singular Covariance

If there are less sources than sensors, slight modifications are to be introduced as the R_X covariance is no longer full rank and its square root C is no longer invertible. If there are M < N sources, then the null subspace of R_X has dimensionality M-N. Its orthogonal subspace, called the "signal subspace" is of dimension M and is generated by the signatures X_i $i = 1, M$. This subspace can be determined from the knowledge of the covariance. Since it contains all the signatures, projecting the observed data on it preserves the information but changes the apparent dimensionality of the observations to value M. In this subspace the covariance is not singular and the FOBI algorithm can operate exactly as previously described. The only change is that the total amount of computation is decreased by dimensionality reduction.

Degeneracy

There is an obvious restriction to the previous method. A difficulty arises if some eigenvalues of the weighted covariance are equal, so that equation 9 does not uniquely determine the Y_i. In our case, this condition is equivalent to two sources having identical (or close enough) normalized fourth-order moments. However, under accidental degeneracy, FOBI fails only partially: only those sources having identical fourth-order normalized moments remain unseparable while the others signatures can be identified, and corresponding messages separated. In addition, there is a "built-in" warning in the method: eigenvalues are extracted together with eigenvectors so that the detection of two close eigenvalues is an indication that corresponding signatures (and only them) cannot be reliably identified. Incidentally, this shows that we can afford the presence of one Gaussian source, but no more than one, if we aim at complete separation, as all normalized complex Gaussian variable are such that $\mu_i = 2$.

Extension to Higher Order Moments.

Degeneracy due to identical kurtosis can be overcome by extending our scheme to higher moments. Introduction of a quadratically weighted covariance as in definition 8 may be seen as a "trick" to make fourth-order moments appear within the covariance structure. If a different weighting is applied, moments of order higher than 4 affect the eigenvalues of the weighted covariance. As an example, if we consider:

$$\tilde{\tilde{R}}_Y = E\left[|Y|^4 YY^T \right] \qquad (10)$$

it comes easily, by reasoning as in the preceeding section, that:

$$\tilde{\tilde{R}}_Y = \sum_{i=1}^{N} \left(\lambda_i + (2N-3)\,\mu_i + c \right)\, Y_i Y_i^T \qquad (11)$$

where we have set $\lambda_i = E\left(|\alpha_i|^6\right)$ and $c = (N-1)(N-2) + \sum_i \mu_i$ and where it is also assumed that third-order moments of all the distributions are null. Hence moments of order 6 appear in the eigenvalues of the new weighted covariance and, if different from one source to another, allow discrimination between sources with (possibly) equal fourth-order moments.

More generally, let us consider an arbitrary weighting function g and define:

$$R^g_Y = E\left[g(|Y|^2)\, YY^T \right] \qquad (12)$$

In addition let us assume that the source distributions are symmetric, so that odd-order moments are null. By expanding the g function in Taylor series, it comes, following the usual line:

$$R^g_Y = \sum_{i=1}^{N} g_i\, Y_i Y_i^T \qquad (13)$$

where g_i is some combination of even-order moments of the sources according to:

$$g_i = E\left[|\alpha_i|^2 \, g\left[\sum_j |\alpha_j|^2 \right] \right] \qquad (14)$$

Hence, for symmetric sources distribution, the $Y_i's$ remain eigenvectors of any weighted covariance, and we can expect that, if source probability distributions are all different, there exist weighting functions g such that the weighted covariance R^g_Y eigenvalues are not degenerate. Again, should this degeneracy appear while running the method, it would be readily detectable by eigenvalues inspection.

PRELIMINARY EXPERIMENTS

It is quite uneasy to compare FOBI with others methods since basic assumptions and evaluation criteria are necessarily different. Nonetheless we show here simulations intended to demonstrate its ability to deal with source location problems. We run, on the same set set of data, FOBI and the MUSIC algorithm [12]. We consider a passive listening linear array of 8 identical sensors. Distance between sensors is 1 m and wavelength is 2 m. Two sources are present: the first one is Gaussian, located 100 m away from the array at a 5 degrees bearing, the second one is a binary source located 200 m away at 10 degrees. Due to finite distance, phase of directional vectors are not purely linear but contains a small quadratic term (in Fresnel approximation) whose amplitude is directly related to source range. As a consequence, the MUSIC algorithm is unadjusted because its localisation function is computed assuming a linear phase directional vector. If sources were located at infinite distance, this noise-free MUSIC simulation would easily success in resolving the two peaks Due to near-field phase distortion, the sources are hardly resolved, as shown in figure a.

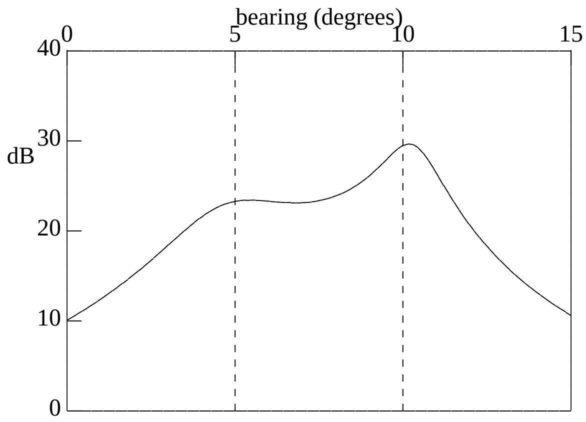

figure a : MUSIC localisation function
in a near-field situation

Note that we did not introduce any noise so that this poor performance is only due to MUSIC ignoring source ranges.

In contrast, the blind algorithm yields estimates of the actual directional vector. From this FOBI estimates, bearing and range can then be extracted. We used a simple unoptimized range-bearing extraction procedure: for each estimated signature, we performed a linear regression on the phase derivative and determined the bearing from the origin of the regression line while range was determined from the regression slope. In figure b, each point is a bearing-range estimate for statistics accumulated on 400 samples. The figure shows 16 estimates to give an idea of the estimation variance, (not very realistic, though, as no noise is present in this simple simulation).

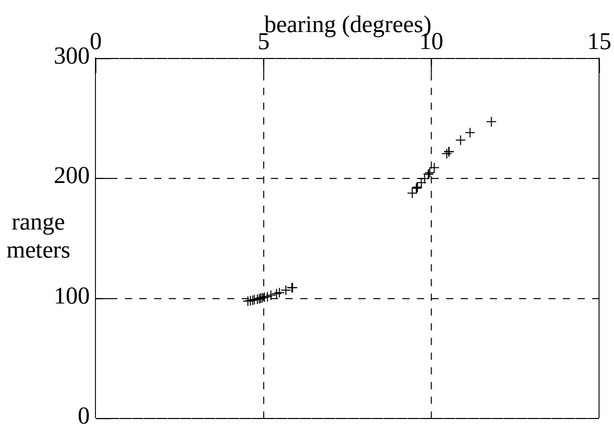

figure b : bearing-range estimates from FOBI signatures
16 experiments with 400 accumulated samples

CONCLUSION

When a multidimensional signal consists in a linear superimposition of independent components, blind (model-free) identification of these components is possible. To this purpose, we have proposed a straightforward method operating in two steps: orthonormalization and distortion (weighting). This is not an iterative algorithm: independent components are readily identified as eigenvectors of a modified covariance so that only standard numerical procedures are to be used. The basic FOBI algorithm relies on fourth-order moments, but may as well be tuned for exploitation of moments of higher order. As its main limitation, our algorithm is not able to discriminate components with identical probability distribution. A simple but original application to array processing was also given: a simultaneous range-bearing estimation in passive listening.

REFERENCES

[1] D.R. Brillinger, "Time Series Data Analysis and Theory", Holden Day, 1981.

[2] J.K. Tugnait, "Identification of Linear Stochastic Systems via Second- and Fourth-Order Cumulant Matching", IEEE Trans. on I.T., Vol. IT-33, No. 3, pp.393-407, May 1987.

[3] A. Benveniste, M. Goursat, G.Ruget, "Robust Identification of a Nonminimum Phase System: Blind Adjustement of a Linear Equalizer", IEEE Trans. on A.C., Vol. AC-25, No. 3, pp.385-399, June 1980.

[4] A.W. Lohmann, B. Wirnitzer, "Triple Correlations", Proc. IEEE, Vol. 72, No. 7, July 1984.

[5] C.L. Nikias, M.R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework", Proc. IEEE, Vol. 75, No. 7, July 1987.

[6] R. Pan and C.L. Nikias, "Harmonic Decomposition Methods in Cumulant Domains", Proc. ICASSP-88,pp. 2356-2359, 1988.

[7] A. Swami and J.M. Mendel, "Cumulant-Based Approach to the Harmonic Retrieval Problem", Proc. ICASSP-88,pp. 2264-2267, 1988.

[8] M. Amengual, F. Vallverdu and G. Vazquez, "On the Use of Higher Order Information in SVD Based Methods", Signal Processing IV: Theo. and Appl. Elsevier publishers,pp. 431-434, 1988.

[9] C. Jutten and J. Herault, "Independent Components Analysis Versus Principal Components Analysis", Signal Processing IV: Theo. and Appl. Elsevier publishers, pp. 643-646, 1988.

[10] J. Herault, C. Jutten and B. Hans, "Detection de Grandeurs Primitives dans un Message Composite", 10th GRETSI Nice, pp. 1017-1022, 1985.

[11] J.L. Lacoume, P.Ruiz, "Source Identification: A Solution Based on the Cumulants", Proc. of the 4-th ASSP Workshop on Spectral Estimation and Modeling,pp. 199-203, Aug. 1988.

[12] R.O. Schmidt "Multiple Emitter Loaction and Signal Parameter Estimation", Proc. RADC Spectral Estimation Workshop, pp. 243-258, 1979.