BLIND IDENTIFICATION OF UNDER-DETERMINED MIXTURES BASED ON THE CHARACTERISTIC FUNCTION

Pierre Comon, Myriam Rajih

I3S, Algorithmes-Euclide-B, BP 121, F-06903 Sophia-Antipolis, France
{comon,rajih}@i3s.unice.fr

ABSTRACT

Linear Mixtures of independent random variables (the so-called sources) are sometimes referred to as Under-Determined Mixtures (UDM) when the number of sources exceeds the dimension of the observation space. The algorithms proposed are able to identify algebraically a UDM using the second characteristic function of the observations. With only two sensors, the first algorithm only needs a SVD. With a larger number of sensors, the second algorithm executes an ALS. The joint use of statistics of different orders is possible, and a LS solution can be computed.

1. INTRODUCTION

This paper is devoted to *Underdetermined* Linear Mixtures (UDM), that is, mixtures of independent random sources where the number of sources, N, always exceeds the number of sensors, P. In other words, underdetermined mixtures do not enjoy sparsity properties such as disjoint source spectra, or sources non permanently present (this property is often exploited in Speech applications [15]).

Moreover, we are only interested in Blind Identification, and not in Source Extraction. These two problems are closely related when the number of sources does not exceed the number of sensors. In fact, the linear mixture can then be linearly inverted, and looking for its inverse is an equivalent problem (see [12] and earlier references therein). Techniques that have been utilized in this framework, such as second order pre-whitening, or deflation, are not applicable for UDM.

Identifiability of linear mixtures received on a single sensor requires source distributions to have an indecomposable characteristic function (c.f.) [17] [13]; for instance in digital communications, BPSK sources are indecomposable but QPSK are not. This condition can be deflated for underdetermined mixtures received on 2 sensors [19] [6]. In contrast, for over-determined mixtures, the only pathological distributions are Gaussian [10] [7] [13]. In the sequel, it is assumed that an under-determined mixture is available on more than one sensor, viz $1 < P < N$. In addition, it is not assumed that spectral or multi-spectral differences can be exploited as in [11] for instance, and the time dimension is merely ignored.

Blind source extraction from underdetermined mixtures is a difficult problem since these mixtures cannot be linearly inverted [4]. On the other hand, Blind Identification (BI) of the mixture matrix can be performed without extracting the sources (at least in a first stage), as in [3] [4] [8] [18] [1]. More precisely, the methods proposed in [3] [4] [8] [11] only use the data Fourth Order statistics, whereas in [18] or [6], the information contained in the second c.f. of observations is exploited. We extend this kind of approach by using additional equations, which makes the solution more stable. Contrary to cumulant based approaches such as [1] or [4], for a given number of sensors, the number of sources is theoretically not limited, which constitutes the main motivation in using the c.f. This advantage has not been fully exploited in [20], where the mixture is assumed square.

2. ASSUMPTIONS AND NOTATION

In accordance with the remarks made in introduction, we assume the observation model below:

$$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} + \boldsymbol{w} \tag{1}$$

where array variables are distinguished from scalars by bold faces, $\boldsymbol{x}$ and $\boldsymbol{s}$ are random vectors of size P and N respectively, $\boldsymbol{A}$ is a $P \times N$ full rank matrix, and $\boldsymbol{w}$ accounts for modeling errors and additive noise. From now on, its presence is just ignored in the remaining, except when running computer experiments. The entries s_n of vector $\boldsymbol{s}$ are assumed to be non Gaussian and statistically independent.

For simplicity, we shall restrict our attention in this paper to real variables and mixture. As pointed out in [6], the immersion of the complex framework in a real framework of larger size introduces some additional constraints, which make the problem more difficult, but at the same time allow a better stability of the solution. Most of the reasoning developed in this paper applies to the complex case, up to some complication in the notation.

We also assume the following hypotheses:

- **H1** the columns of $\boldsymbol{A}$ are pairwise linearly independent.
- **H2** source distributions are unknown and non Gaussian
- **H3** the number N of sources is known
- **H4** the moments of the sources are unknown, but finite up to some order larger than N

Under **H1**, **H2**, and **H3**, $\boldsymbol{A}$ can be shown to be essentially unique [13, pp.311-313].

Two practical algorithms are subsequently described. The first is a significant improvement of the approach described in [18] and [6], and the second is able to solve the derivative matching problem for $N > 2$ with the help of an ALS algorithm. Both are based on the core functional equation below, which is a direct consequence of source independence:

$$\Psi_x(\boldsymbol{u}) = \sum_{n=1}^{N} \psi_n(\sum_{p=1}^{P} A_{pn} u_p) \tag{2}$$

where $\Psi_x(\boldsymbol{u})$ denotes the second c.f. of $\boldsymbol{x}$ defined as $\Psi_x(\boldsymbol{u}) = \log \mathrm{E}\{\exp(\boldsymbol{u}^\intercal \boldsymbol{x})\}$, and where $\psi_n(v)$ denotes the second c.f. of source s_n: $\psi_n(v) = \log \mathrm{E}\{\exp(v s_n)\}$. This core equation can be used in an open neighborhood Ω of the origin, where Ψ_x does not vanish.

3. ALGORITHM ALGECAF: AN ALGEBRAIC SOLUTION

It is easy to verify that any two derivatives of (2) can be combined in order to cancel the nth term of the sum. More precisely, for any triplet of indices, define the differential operator:

$$D_{n,i,j} \stackrel{\text{def}}{=} A_{in} \frac{\partial \Psi_x}{\partial u_j} - A_{jn} \frac{\partial \Psi_x}{\partial u_i}$$

In other words, $D_{n,i,j}\Psi_(\boldsymbol{u})$ does not depend on ψ_n, for any values of (i, j). Thus, by applying such an operator N times for different n's and for arbitrary pairs (i_n, j_n), one eventually gets zero. In order to be able to estimate $\boldsymbol{A}$, it is interesting to fix the pair (i, j), which leads to:

$$\Big\{ \prod_{n=1}^{N} D_{n,i,j} \Big\} \psi_{\boldsymbol{x}}(\boldsymbol{u}) = \sum_{k=0}^{N} q_k[i,j] \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u})}{\partial u_j^{N-k} \partial u_i^k} = 0, \forall \boldsymbol{u} \in \Omega \tag{3}$$

where $q_k[i, j]$ are known functions of the (yet unknown) entries of $\boldsymbol{A}$. In order to obtain the exact relation between vector $\boldsymbol{q}[i, j]$ and rows i and j of $\boldsymbol{A}$, it suffices to plug equation (2) into (3), which yields:

$$\sum_{n=1}^{N} \left[\sum_{k=0}^{N} q_k[i,j] A_{jn}^{N-k} A_{in}^k \right] \psi_n^{(N)}(\sum_p A_{pn} u_p) = 0 \tag{4}$$

where $\psi_n^{(N)}$ denotes the Nth derivative of ψ_n. Since this holds true for any $\boldsymbol{u} \in \Omega$, one can deduce that

$$\sum_{k=0}^{N} q_k[i,j] A_{jn}^{N-k} A_{in}^k, \ \forall n \tag{5}$$

This shows that the N ratios A_{in}/A_{jn} can be obtained as the N roots in the projective space (*i.e.*including infinity) of a polynomial of degree N, once $\boldsymbol{q}$ has been obtained.

Now, imposing (3) to be satisfied on a grid of K values $\{\boldsymbol{u}[1], \ldots, \boldsymbol{u}[K]\} \in \Omega$, one can build the over-determined linear system $\boldsymbol{H}[N]\,\boldsymbol{q} = \boldsymbol{0}$, where $\boldsymbol{H}[N]$ is the matrix of Nth order derivatives given below:

$$\begin{pmatrix} \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[1])}{\partial u_j^N} & \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[1])}{\partial u_j^{N-1} \partial u_i} & \cdots & \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[1])}{\partial u_i^N} \\ \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[2])}{\partial u_j^N} & \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[2])}{\partial u_j^{N-1} \partial u_i} & \cdots & \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[2])}{\partial u_i^N} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[K])}{\partial u_j^N} & \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[K])}{\partial u_j^{N-1} \partial u_i} & \cdots & \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u}[K])}{\partial u_i^N} \end{pmatrix}$$

Because $\boldsymbol{A}$ can be estimated only up to a scale factor, it is entirely identified by this procedure if it contains only two rows ($P = 2$). If there are more than 2 sensors, this algorithm can be adapted to the price of an important increase in complexity, as shown in [6].

Our contribution here is different: we show that one can improve on the stability of this solution by adding extraneous equations. In fact, expression (3) is still null if we take further derivatives:

$$\frac{\partial}{\partial u_\ell} \sum_{k=0}^{N} q_k[i,j] \frac{\partial^N \psi_{\boldsymbol{x}}(\boldsymbol{u})}{\partial u_j^{N-k} \partial u_i^k} = 0, \ \forall \boldsymbol{u} \in \Omega \tag{6}$$

For instance, for $\boldsymbol{u} = \boldsymbol{0}$, $P = 2$, and $N = 3$, this yields the two fourth-order cumulant equations used in [4].

An even more interesting results is that (3) and (6) involve the same unknown q_k, so that they can be combined to build a single larger over-determined system. Indeed, denote $\boldsymbol{H}[N+1, 0]$ and $\boldsymbol{H}[N+1, 1]$ the two $K \times N$ matrices built from (6) when $\ell \in \{i, j\}$. Then $\boldsymbol{q}[i, j]$ satisfies the following linear system:

$$\begin{bmatrix} \boldsymbol{H}[N] \\ \boldsymbol{H}[N+1,0] \\ \boldsymbol{H}[N+1,1] \end{bmatrix} \cdot \boldsymbol{q}[i,j] = \boldsymbol{0}$$

Example: To make it clear, in order to identify a 2×3 mixture, one wishes to estimate a vector $\boldsymbol{q}$ of dimension $N + 1 = 4$. To do this, one can either build a linear system with 3rd order derivatives taken at (at least) 3 different points of Ω, or the two types of 4th order derivatives taken at (at least) 2 different points of Ω. But one can also build a system combining both, including then both 3rd and 4th order derivatives, possibly taken at a single point of Ω (the linear system needs in fact at least $N = 3$ rows in order to have a null space of dimension at most 1).

4. ALGORITHM ALESCAF: AN ALTERNATE LS SOLUTION

As already pointed out, the ALGECAF algorithm is very attractive for the Blind Identification of $2 \times N$ mixtures, but more complicated to implement for $P > 2$ [6]. Therefore, there is a great interest in looking for other ways of exploiting the c.f. From (2), one can easily obtain that

$$\frac{\partial^3 \Psi_x(\boldsymbol{u})}{\partial u_i \partial u_j \partial u_p} = \sum_{n=1}^{N} A_{in} A_{jn} A_{pn} \, \psi_n^{(3)}(\sum_q A_{qn} u_q) \tag{7}$$

Again, take this equation on K points $\boldsymbol{u}[k] \in \Omega$. Then, storing the left hand side of (7) in a family of symmetric matrices $T_{ij}[p,k]$, and denoting $D_{kn} = \psi_n^{(3)}(\sum_q A_{qn} u_q[k])$, (7) can be arranged in compact form as

$$\boldsymbol{T}[p,k] = \boldsymbol{A}\,\mathrm{Diag}\{\boldsymbol{A}(p,:)\}\,\mathrm{Diag}\{\boldsymbol{D}(k,:)\}\,\boldsymbol{A}^{\mathsf{T}}, \qquad (8)$$

with $1 \leq p \leq P$, $1 \leq k \leq K$, and where $\mathrm{Diag}\{\boldsymbol{v}\}$ denotes the diagonal matrix whose entries are those of vector $\boldsymbol{v}$. Expression (8) is a four way PARAFAC model and can be solved using an ALS algorithm described in the appendix 7.1. This procedure constitutes algorithm ALESCAF (Alternate Least Squares Identification based on the Characteristic Function) and is able to compute $\boldsymbol{A}$ and $\boldsymbol{D}$ from symmetric matrices $\boldsymbol{T}[p,k]$ (the implicit dependence of $\boldsymbol{D}$ on $\boldsymbol{A}$ is ignored). Uniqueness of $\boldsymbol{A}$ is achieved as Kruskal's inequality is verified [14]:

$$3r_k(\boldsymbol{A}) + r_k(\boldsymbol{D}) \geq 2\,\mathrm{rank}\{\boldsymbol{T}\} + 3 \qquad (9)$$

where $r_k(\boldsymbol{A})$ is the Kruskal rank [14] of $\boldsymbol{A}$. Results are expected to be better when increasing the order of the statistics as we move away from the Parafac limit, but this need to be verified by simulations.

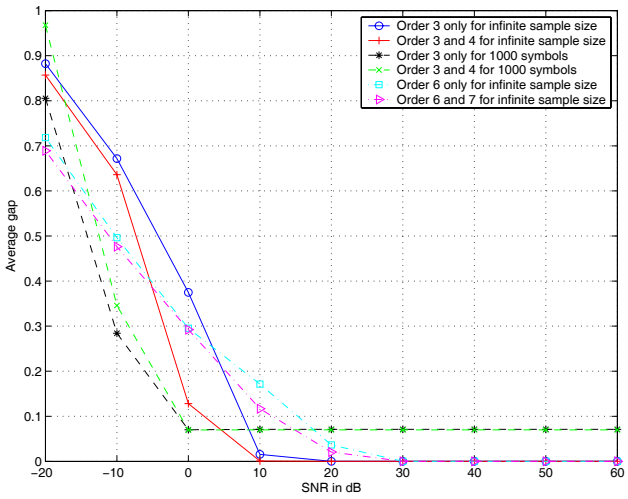

Fig. 1: Gap between estimated and actual mixing matrix for $(P,N) = (2,6)$, with algorithms ALGECAF with use of 6th and 7th derivatives, and for $(P,N) = (2,3)$ with use of 3rd and 4th derivatives. Average gap values over 21 independent trials are plotted.

5. COMPUTER RESULTS

Estimates of matrices $\boldsymbol{H}[\cdot]$ are computed in the following manner. First, all derivatives of $\Psi_x(\boldsymbol{u})$ of required order are formally expressed as a function of moments $\mu(\boldsymbol{n},\boldsymbol{u}[k]) = \mathrm{E}\{\boldsymbol{x}^{\boldsymbol{n}} e^{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{u}[k]}\}$. Then sample moments $\hat{\mu}(\boldsymbol{n},\boldsymbol{u}[k]) = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{x}[m]^{\boldsymbol{n}} e^{\boldsymbol{x}[m]^{\mathsf{T}}\boldsymbol{u}[k]}\}$ are computed, yielding eventually estimates of entries of $\boldsymbol{H}$ (here $\boldsymbol{x}^{\boldsymbol{n}}$ stands for $\Pi_p x_p^{n_p}$). The number of sensors is taken to be $P = 2$, and the number of sources ranges from $N = 3$ to $N = 6$. Sources are BPSK, that is, they take their values in $\{-1, 1\}$ with equal probabilities.

Two types of results are reported for ALGECAF. First, the influence of the noise alone is analyzed. For this purpose, a block of data of length 2^N is generated with exactly all possible combinations of $\{-1, 1\}$; in this manner, sources are always seen as perfectly independent. Independent realizations of a Gaussian noise are added, with various noise level (SNR). Gaps averaged over 21 trials are reported in figure 1 with the label "infinite sample size". Second, other experiments are reported where source blocks are also randomly generated; therefore sources are seen as statistically independent only for large block lengths. As reported with the curves labeled "1000 samples" in figure 1, one can observe a plateau for high SNR's.

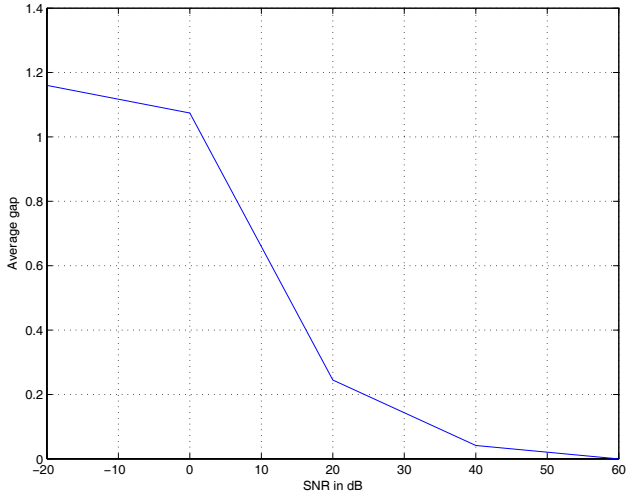

Fig. 2: Gap between estimated and actual mixing matrix for $(P,N) = (2,3)$, with algorithm ALESCAF. Average gap values over 21 independent trials are plotted.

In figure 2 we report the influence of noise on ALESCAF algorithm applied on "infinite" blocks of data. We start with SNR=60dB, check for convergence, and use the value of the corresponding loading matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ to initialize the next ALESCAF algorithm for SNR=50dB and so on. By doing so, one expects to access ultimate performances, *i.e.* in actual situations, performances will be poorer.

6. CONCLUDING REMARKS

Our contribution was three-fold: (i) we have demonstrated that it was possible to derive an algebraic solution to the $2 \times N$ Blind Identification problem by simultaneously using derivatives of different orders, and that it improves the stability of the solution, (ii) we have made the connection with cumulant-based approaches, and proved that improvement (i) also applies to the joint use of cumulants of different orders, larger than or equal to N; (iii) we showed that an ALS

algorithm of PARAFAC type could be utilized to identify a $P \times N$ mixture, and that only third-order derivatives of the c.f. are necessary, although higher orders can also be used.

Future works include: (a) the proof of identifiability when using only derivatives of given orders of the c.f.; (b) improvement of the convergence of ALS algorithms, slow for topological reasons (likely because of a lack of closure [5]); (c) in order to account for a possibly different variance in estimates of moments of different orders, a weighting can be rather easily introduced, and may improve on asymptotic performance; (d) take into account part of the symmetry in the Parafac algorithm, for instance as outlined in appendix 7.2.

7. APPENDIX

7.1 Minimizing the gap between both sides of (8) consists of minimizing

$$\Upsilon = \sum_{p,k} ||\boldsymbol{T}[p,k] - \boldsymbol{A}\,\mathrm{Diag}\{\boldsymbol{C}(p,:)\}\,\mathrm{Diag}\{\boldsymbol{D}(k,:)\}\,\boldsymbol{B}^{\mathsf{T}}||^2 \tag{10}$$

with respect to matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$, if the symmetry constraint is relaxed. The set of matrices $\boldsymbol{T}[p,k]$ defined in (8) can be stored in a tensor T_{ijpk}. Then, this problem can be solved with the help of the Harshman's PARAFAC algorithm [14] [16], originally developed for 3rd order tensors, and improved by Bro [2].

7.2 Define the family of diagonal matrices $\boldsymbol{\Lambda}[p,k] = \mathrm{Diag}\{\boldsymbol{A}(p,:)\}\,\mathrm{Diag}\{\boldsymbol{D}(k,:)\}$. Ignoring the dependence of $\boldsymbol{\Lambda}$ on $\boldsymbol{A}$, one can try to impose the symmetry $\boldsymbol{A} = \boldsymbol{B}$ in (10); but things are more complicated because the optimization criterion (10) is not quadratic anymore in the unknown rectangular matrix $\boldsymbol{B}$. Let $\boldsymbol{\lambda}[k] = \mathbf{diag}\boldsymbol{\Lambda}[k]$. Two writings are derived in order to obtain stationary values with respect to the rectangular matrix and to the diagonal one:

$$\Upsilon = \sum_k ||\boldsymbol{T}[k] - \boldsymbol{B}\boldsymbol{\Lambda}[k]\boldsymbol{B}^{\dagger}||^2 \tag{11}$$

and, with $\boldsymbol{t}[k] = \mathbf{vec}\boldsymbol{T}[k]$ and an appropriate definition of $\mathcal{B}$:

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \mathcal{B}\,\boldsymbol{\lambda}[k]||^2 \tag{12}$$

Some manipulations would show that the stationary values $\boldsymbol{\lambda}[k]$ are given by

$$\boldsymbol{\lambda}[k] = \{\mathcal{B}^{\dagger}\,\mathcal{B}\}^{-1}\mathcal{B}^{\dagger}\,\boldsymbol{t}[k] \tag{13}$$

Last, the stationary value of each column $\boldsymbol{b}[\ell]$ of matrix $\boldsymbol{B}$ is the dominant eigenvector of the Hermitian matrix

$$\boldsymbol{P}[\ell] = \frac{1}{2}\sum_k \lambda_\ell[k]\{\tilde{\boldsymbol{T}}[k;\ell]^{\dagger} + \tilde{\boldsymbol{T}}[k;\ell]\} \tag{14}$$

where $\tilde{\boldsymbol{T}}[k;\ell] \stackrel{\mathrm{def}}{=} \boldsymbol{T}[k] - \sum_{n\neq\ell}\lambda_n[k]\boldsymbol{b}[n]\boldsymbol{b}[n]^{\dagger}$. A LS solution is computed when matrices involved are singular. This turns out to have similarities with a work of Yeredor [21], developed for diagonalizing a set of square matrices by an invertible transform, i.e. applicable only for $\mathrm{rank}\{\boldsymbol{T}\} \leq P$; however, the algorithm theoretically also works for square tensors of rank larger than P.

8. REFERENCES

[1] L. ALBERA, A. FERREOL, P. COMON, P. CHEVALIER, "Blind identification of overcomplete mixtures of sources (BIOME)", *Lin. Algebra Appl.*, to appear, 2004.

[2] R. BRO, C. A. ANDERSSON, "Improving the speed of multiway algorithms. Part II: Compression", *Chemometrics and Intelligent Laboratory Systems*, vol. 42, no. 1-2, pp. 105–113, 1998.

[3] J. F. CARDOSO, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors", in *Proc. ICASSP*, Toronto, 1991, pp. 3109–3112.

[4] P. COMON, "Blind identification and source separation in 2x3 under-determined mixtures", *IEEE Trans. Signal Processing*, pp. 11–22, Jan. 2004.

[5] P. COMON, "Canonical tensor decompositions", Research Report RR-2004-17, I3S, June 17 2004.

[6] P. COMON, M. RAJIH, "Blind identification of complex under-determined mixtures", in *ICA Conference*, Granada, Spain, Sept 22–24 2004.

[7] G. DARMOIS, "Analyse générale des liaisons stochastiques", *Rev. Inst. Internat. Stoch.*, vol. 21, pp. 2–8, 1953.

[8] L. de LATHAUWER, P. COMON, B. de MOOR, "ICA algorithms for 3 sources and 2 sensors", in *6th Workshop on HOS*, Caesarea, Israel, June 14–16 1999, pp. 116–120.

[9] J. ERIKSSON, V. KOIVUNEN, "Identifiability and separability of linear ICA models revisited", in *4th Int. Symp. Indep. Comp. Ana. (ICA'03)*, Nara, Japan, 2003.

[10] W. FELLER, *An Introduction to Probability Theory and its Applications*, vol. II, Wiley, 1966.

[11] A. FERREOL, L. ALBERA, P. CHEVALIER, "Fourth order blind identification of underdetermined mixtures", in *ICASSP'03*, Hong Kong, April 6-10 2003, pp. 41–44.

[12] A. HYVÄRINEN, J. KARHUNEN, E. OJA, *Independent Component Analysis*, Wiley, 2001.

[13] A. M. KAGAN, Y. V. LINNIK, C. R. RAO, *Characterization Problems in Mathematical Statistics*, Probability and Mathematical Statistics. Wiley, New York, 1973.

[14] J. B. KRUSKAL, "Three-way arrays: Rank and uniqueness of trilinear decompositions", *Linear Algebra and Applications*, vol. 18, pp. 95–138, 1977.

[15] T. W. LEE, M. S. LEWICKI et al., "Blind source separation of more sources than mixtures using overcomplete representations", *IEEE Sig. Proc. Letters*, pp. 87–90, Apr. 1999.

[16] S. LEURGANS, R. T. ROSS, R. B. ABEL, "A decomposition for three-way arrays", *SIAM Jour. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, Oct. 1993.

[17] E. LUKACS, *Characteristic functions*, Griffin, 1970.

[18] A. TALEB, "An algorithm for blind identification of N independent signals with 2 sensors", in *ISSPA'01*, Kuala-Lumpur, 2001.

[19] A. TALEB, C. JUTTEN, "On underdetermined source separation", in *ICASSP99*, Phoenix, Arizona, March 15-19 1999.

[20] A. YEREDOR, "Blind source separation via the second characteristic function", *Signal Processing*, vol. 80, pp. 897–902, 2000.

[21] A. YEREDOR, "Non-orthogonal joint diagoinalization in the LS sense with application in blind source separation", *IEEE Trans. Sig. Proc.*, vol. 50, no. 7, pp. 1545–1553, 2002.

[0]This work has been supported in part by the Pascal Network IST-2002-506778, and by the CNRS ACI2M project 032324.