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 [4] [3] [13]. 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.) [16] [14]; 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 [18] [7]. In contrast for over-determined mixtures, the only pathological distributions are Gaussian [11] [8] [14]. 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 [12] 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 [5]. 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 [2] [5] [9] [17] [1]. More precisely, the methods proposed in [2] [5] [9] [12] only use the data FO statistics, whereas in [17] or [7], 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 [5], for a given number of sensors, the number of sources is theoretically not limited, which constitutes the main motivation in using the c.f.

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 Vector, Matrix, or Tensor 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 [7], 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 [14, pp.311-313].

Two practical algorithms are subsequently described. The first is a significant improvement of the approach described in [17] and [7], and the second turns out to have similarities with a work of Yeredor [19], developed for diagonalizing a set of square matrices by an invertible transform, i.e. applicable only for $r_s(\boldsymbol{T}) \leq P$; however, the algorithm described here also works for square tensors of rank $r_s(\boldsymbol{T}) \geq P$. 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 $\Psi_x(\boldsymbol{u}) = \log \mathrm{E}\{\exp(\boldsymbol{u}^{\mathsf{T}}\boldsymbol{x})\}$, and where $\psi_n(v)$ denotes the second c.f. of source s_n: $\psi_n(v) = \log \mathrm{E}\{\exp(vs_n)\}$. 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 [7].

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 [5].

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 AICF algorithm is very attractive for the Blind Identification of $2 \times N$ mixtures, but more

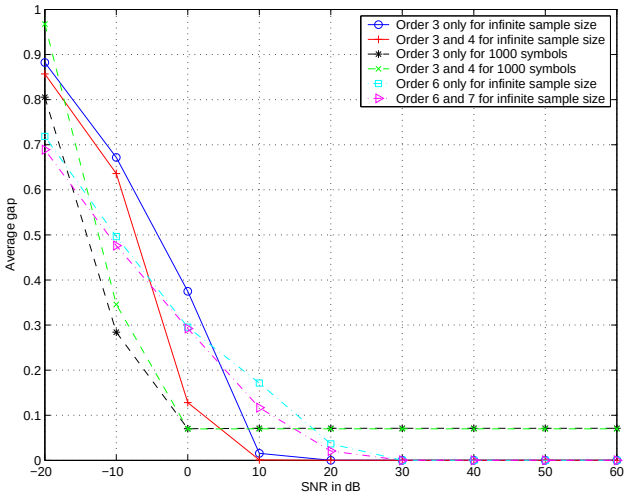

Fig. 1. Gap between estimated and actual mixing matrix for $(P, N) = (2, 6)$, with algorithms ALGECAF with use of 6th and 7th derivatives, and $(P, N) = (2, 3)$ with use of 3rd and 4th derivatives. Average gap values over 21 independent trials are plotted.

complicated to implement for $P > 2$. Therefore, there is a great interest in looking for other ways of exploiting the c.f. From (2), one can easily obtain that

$$\frac{\partial^2 \Psi_x(\boldsymbol{u})}{\partial u_i \partial u_j} = \sum_{n=1}^{N} A_{in} A_{jn}\, \psi_n^{(2)}(\sum_p A_{pn} u_p) \qquad (7)$$

Again, one can take this equation on K points $\boldsymbol{u}[k]$ on a grid of Ω. With obvious notations, the result can be arranged in compact form as

$$\boldsymbol{T}[k] = \boldsymbol{A}\, \boldsymbol{\Lambda}[k]\, \boldsymbol{A}^{\intercal},\ 1 \leq k \leq K \qquad (8)$$

We describe in appendix two ALS algorithms able to compute $\boldsymbol{A}$ and diagonal matrices $\boldsymbol{\Lambda}[k]$ from symmetric matrices $\boldsymbol{T}[k]$. This procedure constitutes algorithm ALESCAF (Alternate Least Squares Identification based on the Characteristic Function).

Now in some cases, in particular in the presence of strong additive Gaussian noise of unknown covariance, it may be attractive to use only statistics of order greater than 2. It turns out that this is also possible with quite minor modifications, as reported in [6]. Matrices $\boldsymbol{\Lambda}[k]$ then depend on the kth column of $\boldsymbol{A}$, but this is ignored, hence the suboptimality of the algorithm.

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 $\mu(\boldsymbol{n}, \boldsymbol{u}[k]) = \mathrm{E}\{\boldsymbol{x}^{\boldsymbol{n}} e^{\boldsymbol{x}^{\intercal} \boldsymbol{u}[k]}\}$. Then sample moments $\hat{\mu}(\boldsymbol{n}, \boldsymbol{u}[k]) = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{x}[m]^{\boldsymbol{n}} e^{\boldsymbol{x}[m]^{\intercal} \boldsymbol{u}[k]}\}$ 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. 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.

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 second-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 [6]); (c) in order to account for a possibly different variance ion estimates of moments of different orders, a weighting can be rather easily introduced, and may improve on asymptotic performance.

7. APPENDIX

Asymmetric ALESCAF Given a set of (possibly rectangular) matrices $\boldsymbol{T}[k]$ (typically tensor slabs), the algorithm aims at minimizing

$$\Upsilon = \sum_k ||\boldsymbol{T}[k] - \boldsymbol{B}\, \boldsymbol{\Lambda}[k]\, \boldsymbol{C}^{\dagger}||^2 \qquad (9)$$

with respect to matrices $\boldsymbol{B}$ and $\boldsymbol{C}$, where matrices $\boldsymbol{\Lambda}[k]$ are diagonal. In the present framework however, even if matrices $\boldsymbol{T}[k]$ are all square symmetric, it might still be of interest

to run asymmetric iterations. This criterion can alternatively be written in the form of a distance between vectors as

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \sum_n \boldsymbol{\lambda}_n[k]\, \boldsymbol{c}[n]^* \otimes \boldsymbol{b}[n]||^2$$

where $\boldsymbol{t}[k] = \mathbf{vec}(\boldsymbol{T}[k])$, $\boldsymbol{b} = \mathbf{vec}(\boldsymbol{B})$, and $\boldsymbol{c} = \mathbf{vec}(\boldsymbol{C})$. By introducing a $P^2 \times N$ matrix $\mathcal{M}$, whose nth column is $\boldsymbol{c}[n]^* \otimes \boldsymbol{b}[n]$, it is possible to obtain a more compact expression:

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \mathcal{M}\, \boldsymbol{\lambda}[k]||^2 \tag{10}$$

Stationary values of $\boldsymbol{B}$ and $\boldsymbol{C}$ are given by

$$\boldsymbol{B} = \{\sum_k \boldsymbol{T}[k]\boldsymbol{C}\boldsymbol{\Lambda}[k]\}\{\sum_\ell \boldsymbol{\Lambda}[\ell]\boldsymbol{C}^\dagger\boldsymbol{C}\boldsymbol{\Lambda}[\ell]\}^{-1} \tag{11}$$

$$\boldsymbol{C} = \{\sum_k \boldsymbol{T}[k]^\dagger\boldsymbol{B}\boldsymbol{\Lambda}[k]\}\{\sum_\ell \boldsymbol{\Lambda}[\ell]\boldsymbol{B}^\dagger\boldsymbol{B}\boldsymbol{\Lambda}[\ell]\}^{-1} \tag{12}$$

whereas stationary values of the diagonal of $\boldsymbol{\Lambda}[k]$ are given by the vectors

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

The ALS algorithm ALESCAF1 consists of executing alternatively (13), (11), and (12). When matrices involved in a system solution are singular, a LS solution is computed.

Symmetric ALESCAF In the symmetric case, things are more complicated because the optimization criterion is not quadratic anymore in the unknown rectangular matrix. Again, two writings are derived in order to obtain stationary values with respect to the rectangular matrix and to the diagonal one:

$$\Upsilon = \sum_k ||\boldsymbol{T}[k] - \boldsymbol{B}\boldsymbol{\Lambda}[k]\boldsymbol{B}^\dagger||^2 \tag{14}$$

and, with an appropriate definition of $\mathcal{B}$ similar as above:

$$\Upsilon = \sum_k ||\boldsymbol{t}[k] - \mathcal{B}\, \boldsymbol{\lambda}[k]||^2 \tag{15}$$

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

$$\boldsymbol{\lambda}[k] = \{\mathcal{B}^\dagger\, \mathcal{B}\}^{-1}\mathcal{B}^\dagger\, \boldsymbol{t}[k] \tag{16}$$

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

$$\boldsymbol{P}[\ell] = \frac{1}{2}\sum_k \lambda_\ell[k]\{\tilde{\boldsymbol{T}}[k;\ell]^\dagger + \tilde{\boldsymbol{T}}[k;\ell]\} \tag{17}$$

where $\tilde{\boldsymbol{T}}[k;\ell] \stackrel{\text{def}}{=} \boldsymbol{T}[k] - \sum_{n\neq\ell}\lambda_n[k]\boldsymbol{b}[n]\boldsymbol{b}[n]^\dagger$. The ALS algorithm ALESCAF2 consists of executing successively (16) and the calculation of the dominant eigenvector of the K matrices (17). As before, a LS solution is computed when matrices involved are singular.

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] 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.

[3] J. F. CARDOSO, "High-order contrasts for independent component analysis", *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan. 1999.

[4] P. COMON, "Independent Component Analysis, a new concept ?", *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

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

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

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

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

[9] 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.

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

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

[12] 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.

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

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

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

[17] A. TALEB, "An algorithm for the blind identification of n independent signals with 2 sensors", in *Sixth International Symposium on Signal Processing and its Applications (ISSPA'01)*. 2001, IEEE.

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

[19] 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.