Blind Identification of Noisy Under-Determined Mixtures Based on Multiple Order Derivatives of the Characteristic Function

Myriam Rajih and Pierre Comon

I3S, 2000 route des Lucioles, B.P.121, F-06903 Sophia-Antipolis, France
phone: +33 4 9294 2793/2717, fax: +33 4 9294 2896, {rajih,comon}@i3s.unice.fr
www.i3s.unice.fr

Abstract—Based on Multi-Linear algebra tools, solutions to the Blind Identification of Under-Determined Mixtures (UDM) are developed. In [7], we proposed algorithm ALESCAF (Alternating LEast Squares identification based on the Cha-racteristic Function), that uses the derivatives of the second characteristic function (c.f.) of the observations, without any need of sparsity assumption on sources, but assuming their statistical independence. In [7], only one order derivative was considered. In this paper, new versions of ALESCAF are proposed, that take advantage of the joint use of derivatives of different orders. Computer simulations demonstrate that new versions of ALESCAF accelerate the convergence. Convergence of ALESCAF algorithms is accelerated by using ELS (Enhanced Line Search), proposed in [17].

I. INTRODUCTION

Linear Mixtures of independent random variables (the so-called sources) are referred to as Under-Determined Mixtures (UDM) when the number of sources, \( N \), exceeds the number of sensors, \( P \). In other words, UDM do not enjoy sparsity properties such as disjoint source spectra, or sources non permanently present. The latter property is often exploited in Speech applications; see [15] among others. The exploitation of sparsity is a promising technique, but is not applicable in the present framework.

This paper aims at blindly identify the mixture matrix without extracting the sources (at least in a first stage). This has been achieved in [2] [6] [8] [9], by the only use of the data Fourth Order statistics. Here, we exploit the information contained in the second c.f. of the observations.

The c.f. has been already utilized in [19] to blindly separate sources under the assumption that there are at most as many sources as sensors (i.e. over-determined mixtures). However, the advantage that the c.f. may allow to identify linear mixtures where the number of sources exceeds the number of sensors has neither been noticed nor exploited therein; hence the originality of the present contribution.

In [7], we proposed an algebraic solution ALGECAF (AL-GEmbraic identification based on the ChAracteristic Function), proven to be very attractive for the blind identification of \( 2 \times N \) mixtures. However, ALGECAF turns out to be complicated to implement for \( P > 2 \). Hence, the importance of ALESCAF, which reduces the Blind Identification problem to a tensor decomposition and accelerated the convergence [17]. Here, faster convergence is obtained by simultaneously using derivatives of different orders. In [16], the authors presented a method that accelerates the convergence given the distributions of the sources. Here, the source distributions are unknown, and we considered noisy data.

Some recent works dealt with the Blind Identification of an UDM using tensor decomposition [14]. The number of sources is limited in [14], whereas the algorithms proposed here do not impose any bound on the number of sources, at least in theory. Practically, the complexity of the solution increases as the number of sources increases.

The paper is organized as follows: Multi-linear algebra tools are presented in section II. Section IV presents the ALESCAF algorithms along with an example. Finally, computer results are given in section V.

II. PARAFAC TOOLS

PARAFAC can be seen as a generalization of the two-way factor analysis to multi-way data. It was first introduced by Harshman in 1970 [10] based on the principle of parallel Proportional Profiles (PP) proposed by Cattell in 1944 [4]. The PP principle states that if two (or more) different two-way models are described by the same set of loading vectors and only proportions or weights change from one model to the other, then those loading vectors lead to a new model which is unambiguous with respect to (w.r.t) rotation [1]. In other words, suppose that the matrix \( X_1 \) can be modeled as:

\[
X_1 = a_1b_1^Tc_{11} + a_2b_2^Tc_{12} + \ldots + a_Fb_F^Tc_{1F}
\]

\( a_f \) and \( b_f \) \( (1 \leq f \leq F) \) being the columns of matrices \( A \) and \( B \). And suppose that another matrix \( X_2 \) can be modeled using the same set of loading vectors only in different proportions described by \( c_{ij} \):

\[
X_2 = a_1b_1^Tc_{21} + a_2b_2^Tc_{22} + \ldots + a_Fb_F^Tc_{2F}
\]

Then, we can build a combined model:

\[
X_k = A \text{Diag}\{C(k,:)}B^T, k = 1, 2
\]

which can be alternatively written as :

\[
X_{ijk} = \sum_f A_{ij}B_{jf}C_{kf}.
\]

The trilinear model is also known as CANDECOMP for CANonical DECOMPosition introduced by Caroll and Chang in 1970 [3]. The 3-way PARAFAC model is very popular in psychometrics and chemometrics where it was first used along with its extension to higher orders [10] [3] [18]. It also finds applications in the signal processing area [13] [5] [7]. While the 2-way model suffers a rotational indeterminacy that yields an infinite set of solutions, the PARAFAC model enjoys a uniqueness property under simple conditions summarized in the Kruskal theorem [12], hence its importance.
Many algorithms propose a solution to fit the PARAFAC model, one of which is the Alternating Least Square (ALS) algorithm. The convergence of ALS was found to be very slow in some cases. Enhanced Line Search (ELS) [17] is the preferred solution to cope with the problem of slow convergence, and will be used in the ALESCAF algorithms.

III. ASSUMPTIONS AND NOTATION

We assume the observation model below:

$$x = A s + w$$  \hspace{1cm} (4)

where array variables are distinguished from scalars by bold faces, \(x\) and \(s\) are real random vectors of size \(P \times N\) respectively, \(A\) is a real \(P \times N\) full rank matrix, and \(w\) 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 \(s\) are assumed to be non Gaussian and statistically independent.

We also assume the following hypotheses:

\(\textbf{H1}\) the columns of \(A\) are pairwise linearly independent.

\(\textbf{H2}\) source distributions are unknown and non Gaussian.

\(\textbf{H3}\) the number \(N\) of sources is known.

\(\textbf{H4}\) the moments of the sources are unknown, but finite up to some order larger than \(N\).

Under \(\textbf{H1}, \textbf{H2},\) and \(\textbf{H3}\), \(A\) can be shown to be essentially unique [11].

The algorithms proposed in the following are based on the core functional equation below, which is a direct consequence of source independence:

$$\Psi_x(u) = \sum_{n=1}^{N} \psi_n(\sum_{p=1}^{P} A_{pn} u_p)$$  \hspace{1cm} (5)

where \(\Psi_x(u)\) denotes the second c.f. of \(x\) defined as \(\Psi_x(u) = \log E\{\exp(j u' x)\}\), and where \(\psi_n(v)\) denotes the second c.f. of source \(s_n\): \(\psi_n(v) = E\{\exp(j u_n v)\}\), where the dotless \(j\) denotes the square root of \(-1\). This core equation can be used in an open neighborhood \(\Omega\) of the origin where \(\Psi_x\) does not vanish, which always exists.

IV. ALGORITHM ALESCAF

In this section, we assume the hypothesis:

\(\textbf{H4a}\) the source second characteristic functions \(\psi_n\) admit finite derivatives up to order 4, at every point of some grid \(G\) of K values \(\{u[1], ..., u[K]\} \in \Omega\)

A. One Order Derivatives

For notation simplicity, suppose that \(P \geq 3\), the case with 2 sensors will be discussed in IV-C. And take the second order derivatives of (5):

$$\frac{\partial^2 \Psi_x(u)}{\partial u_i \partial u_j} = \sum_{n=1}^{N} A_{in} A_{jn} \psi_n^{(2)}(\sum_{q} A_{qn} u_q)$$  \hspace{1cm} (6)

with \(1 \leq i, j \leq P\). Take this equation on \(K\) points \(u[k] \in G \subset \Omega\). Then, storing the left hand side of (6) in a family of symmetric matrices \(T[i,j][k]\), and denoting \(D_{kn} = \psi_n^{(2)}(\sum_{q} A_{qn} u_q[k])\), (6) can be arranged in compact form as:

$$T[k] = A \text{Diag}\{D(k,:)\} A^\top$$  \hspace{1cm} (7)

with \(1 \leq k \leq K\), where \(\text{Diag}\{\cdot\}\) denotes the diagonal matrix whose entries are those of vector \(\cdot\), and where \(D(k,:)\) denotes the \(k\)th row of \(D\). Expression (7) can be alternatively written as:

$$T_{ijk} = \sum_{n=1}^{N} A_{in} A_{jn} D_{kn}$$  \hspace{1cm} (8)

where \(T\) is a \(P \times P \times K\) tensor. (7) and (8) equivalently represent a 3-way PARAFAC model, and can be efficiently solved by using the ELS algorithm described in [17].

As the symmetry constraint is relaxed in the ELS algorithm, minimizing the gap between both sides of (8) consists of minimizing:

$$\gamma = \sum_{ijk} ||T_{ijk} - \sum_{n=1}^{N} A_{in} B_{jn} D_{kn}||^2$$

This procedure constitutes algorithm ALESCAF(2), and is able to compute \(A\) and \(D\) from tensor \(T\), where the implicit dependence of \(D\) on \(A\) is ignored.

B. Multiple Order Derivatives

The previous algorithm only uses one (the second) order derivatives of the second c.f. of the observations to build tensor \(T\), and proved to have some limitations in terms of identifiability and convergence speed. It can be made faster by adding extraneous terms to tensor \(T\), that correspond to other order derivatives of (5). In fact, take the \(P\) further derivatives of (6):

$$\frac{\partial^p \Psi_x(u)}{\partial u_i \partial u_j} = \sum_{n=1}^{N} A_{in} A_{jn} \psi_n^{(p)}(\sum_{q} A_{qn} u_q)$$  \hspace{1cm} (9)

with \(1 \leq p \leq P\). Denote by \(T[2]\) the \(P \times P \times K\) tensor previously defined in (8), where the number 2 stands for the order of the derivatives used. Again, take equation (9) on \(K\) points \(u[k] \in G \subset \Omega\). Note that the latter \(K\) points can be different from those of section (IV-A). For simplicity, we keep the same \(K\) points \(u[k]\).

Then, (9) leads to \(P\) new \(P \times P \times K\) tensors \(T[3, p] : T[3, p]_{i,j,k} = A \text{Diag}\{D_{p}(k,:)\} A^\top\)

where \((D_{p})_{kn} = A_{pn} \psi_n^{(p)}(\sum_{q} A_{qn} u_q[k])\). By putting \(T[2]\) and the \(P\) tensors \(T[3, p]\) next to each other in the third mode, we obtain a bigger tensor of size \(P \times P \times (P+1)K\).

This rearrangement is visualized in figure (1).

![Building new P x P x (P+1)K tensor from tensors T[2] and T[3, p], 1 ≤ p ≤ P.](image)

This constitutes algorithm ALESCAF(2,3), that simultaneously uses the second and third order derivatives of the second c.f. of the observations, and accelerates the convergence as will be shown in the simulation section.
C. Example 1

As mentioned in IV-A, let’s take a more concrete example and suppose that $P = 2$, $N = 3$. In this case, we need to take the third order derivatives of the second c.f. of the observations, as the second order derivatives yield a 3-way PARAFAC model that does not achieve Kruskal uniqueness condition [12]:

$$2 r_k(A) + r_k(D) \geq 2 \text{rank}\{T[2]\} + 3 \quad (10)$$

Then, we obtain a 4-way PARAFAC model defined by:

$$T_{ijk} [3] = \sum_{n=1}^{N} A_{in} A_{jn} A_{kn} D_{kn}$$

where $T[3]$ is a $P \times P \times P \times K$ tensor, and $D_{kn} = \psi_{kn}^{(3)}(\sum_{q} A_{qn} u_q[k])$. Now, uniqueness is achieved as (10) is verified: $3P + N = 9 \geq 2N + 3 = 9$.

Previous tensor $T[3]$ only uses the third order derivatives. We can accelerate the convergence by simultaneously using third and fourth derivatives of the second c.f. of the observations. Then, entries of the new $P \times P \times P \times (P + 1)K$ tensor $T$ are defined as:

$$T_{l:p:k} = \left( \frac{\partial^3 \psi_x(u[k])}{\partial u_1^3 \partial u_p} \right) \left( \frac{\partial^3 \psi_y(u[k])}{\partial u_2^3 \partial u_p} \right) \left( \frac{\partial^3 \psi_z(u[k])}{\partial u_3^3 \partial u_p} \right) , \quad 1 \leq k \leq K$$

$$T_{l:p:k} = \left( \frac{\partial^3 \psi_x(u[k])}{\partial u_1^3 \partial u_2 \partial u_p} \right) \left( \frac{\partial^3 \psi_y(u[k])}{\partial u_2^3 \partial u_2 \partial u_p} \right) \left( \frac{\partial^3 \psi_z(u[k])}{\partial u_3^3 \partial u_2 \partial u_p} \right) , \quad K + 1 \leq k \leq 2K$$

$$T_{l:p:k} = \left( \frac{\partial^4 \psi_x(u[k])}{\partial u_1^4 \partial u_2 \partial u_2 \partial u_p} \right) \left( \frac{\partial^4 \psi_y(u[k])}{\partial u_2^4 \partial u_2 \partial u_2 \partial u_p} \right) \left( \frac{\partial^4 \psi_z(u[k])}{\partial u_3^4 \partial u_2 \partial u_2 \partial u_p} \right) , \quad 2K + 1 \leq k \leq 3K$$

with $1 \leq p \leq P$.

As mentioned in the introduction, the number of sources $N$ is theoretically unbounded. However, as $N$ increases, the complexity of ALESCAF increases because the order of the derivatives and the tensor order need to be increased. In fact, if we keep $P = 2$ and increase $N$ by one ($N=4$), uniqueness is no longer achieved with ALESCAF(3) (10). One solution is to use the 4th order derivatives of (5) and build the corresponding 5-way tensor $T[4]$ of size $P \times P \times P \times P \times K$ defined as:

$$T_{l:p:k} = \left( \frac{\partial^4 \psi_x(u[k])}{\partial u_1^4 \partial u_2 \partial u_2 \partial u_p} \right) \left( \frac{\partial^4 \psi_y(u[k])}{\partial u_2^4 \partial u_2 \partial u_2 \partial u_p} \right) \left( \frac{\partial^4 \psi_z(u[k])}{\partial u_3^4 \partial u_2 \partial u_2 \partial u_p} \right) , \quad 1 \leq l, p \leq P, \text{ and } 1 \leq k \leq K.$$

V. COMPUTER RESULTS

We analyze the influence of the joint use of several order derivatives of the second c.f. of the observations on the Blind Identification of an UDM. For this purpose, we consider the linear model of expression (4), with $P = 2$ and $N = 3$. The channel matrix $A$ is:

$$A = \begin{pmatrix}
1 & \cos(\theta) & 0 \\
0 & \sin(\theta) & 1
\end{pmatrix}$$

In figures 2 and 3, sources are BPSK, that is, they take their values in $\{-1, 1\}$ with equal probabilities. An “infinite block” of data is generated by taking all the $2^P$ possible combinations of $\{-1, 1\}$; in this manner, sources are always seen as perfectly independent. Figure 2 and 3 report the performances of ALESCAF(3), ALESCAF(3, 4), and ALESCAF(3, 4, 5) in terms of the gap and the error $\gamma$ respectively.

Figures 4 and 5 report the same results for 4PAM sources, that is, they take their values in $\{-3, -1, 1, 3\}$ with equal probabilities.

Figure 6 reports the influence of noise on ALESCAF algorithms. Independent realizations of a Gaussian noise are added, with various noise levels (SNR), and the median gap over 21 independent trials is plotted. We start with $\text{SNR}=30\text{dB}$, check for convergence, and use the value of the corresponding loading matrices $A$ and $D$ to initialize
the next ALESCAF algorithms for SNR=25dB, and so on. By doing so, one expects to access ultimate performances, i.e. in actual situations, performances will be poorer.

VI. CONCLUDING REMARKS

We proposed new versions of ALESCAF, that simultaneously use several order derivatives of the second c.f. of the observations, and demonstrated through simulations that they accelerated the convergence. We also demonstrated that ALESCAF algorithms succeed on blindly identifying the mixing matrix in the presence of noise, and that the more orders utilized, the more accurate is the solution.

Performances with finite blocks of data also need to be investigated.

REFERENCES


