ABSTRACT
This paper is devoted to under-determined linear mixtures of independent random variables (i.e. with more inputs than outputs). Blind identifiability of general under-determined mixtures is first discussed, and the maximum number of sources is given, depending on the hypotheses assumed. Then an algorithm proposed by Taleb, essentially usable for 2-dimensional mixtures, is extended to the complex field. A procedure is proposed in order to avoid the enormous increase in complexity. Computer simulations demonstrate the ability of the algorithm to identify mixtures of N QPSK sources received on 1 or 2 sensors.

1. INTRODUCTION
Consider N statistically independent zero-mean random variables, which will be referred to as sources, \( s_1, \ldots, s_N \), and \( P \) linear mixtures defined as:

\[
x = A s
\]

where \( x \) is the column vector formed of \( x_p = \sum A_{pn} s_n \), \( A \) denotes the \( P \times N \) mixing matrix, and \( s \) the source column vector. When \( P \geq N \), the mixture is said to be over-determined, whereas in the case we are interested in, namely \( P < N \), the mixture is referred to as under-determined.

There exists now a large literature on Over-Determined Mixtures (ODM), and we shall not even review basic pointers because of lack of space. On the other hand, much less attention has been drawn on Under-Determined Mixtures (UDM). Under particular hypotheses, UDM can be sometimes deflated to ODM, for instance with the help of sparse decompositions in overcomplete bases [12]. On the contrary, we shall focus our attention to UDM that cannot be deflated. In the Statistics community, the first basic theorems can be traced back to the fifties, and can be found in [10] for instance. The Blind Identification of UDM’s can be viewed as a problem of Factor Analysis, in which the number of factors exceeds the dimension [5]; as such, it has been addressed in the seventies, but under restricting assumptions [11][13]; the decomposition is then known as PARAFAC. In the Signal Processing community, the problem has been addressed only ten years ago [3] [2]. Several approaches are possible, under various assumptions [6] [5] [16] [15] [7]. This will be briefly surveyed in the next section.

2. IDENTIFIABILITY AND SEPARABILITY
We are interested in the blind identification of the mixing matrix \( A \); if the solution is unique, then \( A \) is identifiable. But we may also want to uniquely determine source distributions [14] [8]. It turns out that a unique solution for \( A \) indeed does not always yield a unique set of source distributions. Under hypothesis H1 for instance, this holds true only for over-determined mixtures. Uniqueness should be understood throughout this paper up to a permutation among the sources, and up to a scale factor; because of this inherent indeterminacy, we shall rather talk about essentially unique solutions.

It is then useful to introduce the following hypotheses:

- **H1**: the columns of \( A \) are pairwise linearly independent.
- **H2**: source distributions are unknown and non Gaussian.
- **H3**: the number \( N \) of sources is known.
- **H4**: the characteristic function of \( x \) does not vanish.
- **H5**: for a given order \( r > 2 \), all source marginal cumulants of order \( r \) are unknown but finite, and it is known that at most one of them is null.
- **H6**: source cumulants are all known up to some order \( r \).
- **H7**: source distributions are known, discrete, and indecomposable.

Assumption H1 is not restrictive; in fact, if two columns \( i \) and \( j \) of \( A \) are proportional, then we can add sources \( s_i \) and \( s_j \) to form a new source, still independent from the others, and model (1) holds with merely \( N - 1 \) sources instead of \( N \).

Assumptions H1 and H2 together yield the unicity of \( A \) if it is known to be invertible [10, pp.89-90]. But this cannot be the case when \( N > P \). So consider now 6 distinct instances of the problem, each leading to identifiability properties of various strengths.

- **P1**: under H1, H2, and H3, \( A \) can be shown to be essentially unique [10, pp.311-313].
P2 under H1, H2, and H4. A and the N source distributions \( p_{s_n} \) are essentially unique, provided [10, pp.470-471]:

\[
N \leq P(P + 1)/2
\]  

(2)

P3 under assumptions H1, H3, and H6. A and the N source distributions \( p_{s_n} \) are essentially unique, provided the condition below holds true [14]:

\[
N \leq \left( \frac{P + r}{r + 1} \right)
\]  

(3)

P4 under assumptions H1, H3, and H5 for some even order \( r > 3 \), and if \( P > 1 \), A is essentially unique provided the sufficient condition below is satisfied [4]:

\[
2N \leq r(P - 1) + 2
\]  

(4)

If in addition sources are complex and non circular at order \( r \), then the bound in the right hand side can be made larger [6].

P5 if sources and mixture are real, then under assumptions H1, H3, and H5, A is essentially unique the number of sources is small enough to verify [11][13]:

\[
2N \leq r(P - 1) + 1
\]  

(5)

P6 under H1 and H7, then A is essentially unique, and for any finite P, there is no upper bound on N, except for rare ambiguous mixtures [9]; see also references quoted in [6].

If \((N, P) = (3, 2)\), then (2) holds true, as well as (3) and (4) for \( r = 4 \), but not (5). This is why PARAFAC methods are considered to be restrictive, even if they are recognized to be useful for large \( P \).

3. SINGLE COMPLEX SENSOR

Our concern is essentially to solve the Blind Identification problem P1, and we suppose that the corresponding identifiability conditions are verified. Taleb proposed in [15] an algorithm for the blind identification of real mixtures of N independent real source signals received on 2 sensors. The algorithm is based on a partial differential equation verified by the joint second characteristic function of the 2 sensors, \( \psi_x(u, v) \). A complex mixture of N complex sources received on \( P \) sensors can be viewed as a particular real mixture of \( 2N \) real sources received on \( 2P \) sensors. The latter real sources are independent provided the original complex sources have independent real and imaginary parts. We briefly review in the subsection below the algorithm allowing to identify \( 1 \times N \) complex mixtures, or equivalently \( 2 \times 2N \) real mixtures.

Real \( 2 \times N \) mixture.

Restricting our attention to the case of \( P = 2 \) sensors, (1) can be rewritten as:

\[
\begin{align*}
    x_1 &= a_1 s_1 + a_2 s_2 + \ldots + a_N s_N \\
    x_2 &= b_1 s_1 + b_2 s_2 + \ldots + b_N s_N
\end{align*}
\]  

(6)

The joint second characteristic function of \( x_1 \) and \( x_2 \) can be written as:

\[
\psi_x(u, v) = \log E[\exp(iux_1 + ivx_2)], \quad (u, v) \in \Omega
\]  

(7)

where \( \Omega \) is the largest subset of \( \mathbb{R}^2 \) containing the origin and where the characteristic function of the pair \((x_1, x_2)\) does not vanish. As the sources are independent, we can write:

\[
\psi_x(u, v) = \sum_{n=1}^{N} \psi_{s_n}(a_n u + b_n v)
\]  

(8)

It is assumed that \( \psi_{s_n} \) has derivatives up to the \( N \)th order, that is, \( E[|s_n|^k] \) exists and is finite for all \( k \leq N \). Then define the differential operator \( D_n \) as:

\[
D_n = b_n \frac{\partial}{\partial u} - a_n \frac{\partial}{\partial v}
\]  

(9)

When \( D_n \) is applied to (8), its nth term is canceled. By applying successively operators \( D_n \), \( n = 1, \ldots, N \), all the terms are eventually canceled and we get:

\[
\left\{ \prod_{n=1}^{N} D_n \right\} \psi_x(u, v) = 0 \Rightarrow \sum_{j=0}^{N} q_j \frac{\partial^N \psi_x(u, v)}{\partial u^{N-j} \partial v^j} = 0
\]  

(10)

where \( q_n \) are known functions of the \( a_n \)'s and the \( b_n \)'s. Using (8) hence yields:

\[
\sum_{n=1}^{N} \sum_{j=0}^{N} q_j a_n^{N-j} b_j \psi_{s_n}^{(N)}(a_n u + b_n v)
\]  

(11)

This states the relation between \( q_n \) and \((a_n, b_n)\) [6][15]:

\[
\sum_{j=0}^{N} q_j a_n^{N-j} b_j = 0, \quad \forall n = 1, \ldots, N
\]  

(12)

It is thus possible to estimate \( \hat{q} = [q_0, q_1, \ldots, q_N]^T \) from (10). In order to do this, start by selecting \( K \) points \((u_k, v_k) \in \Omega\), and estimate for each of these points all the \( N \)th order derivatives; there are \( N + 1 \) of them. This allows to form a \( K \times (N + 1) \) matrix \( H \) defined as:

\[
H = \begin{pmatrix}
    \frac{\partial^N \psi_x(u_1, v_1)}{\partial u^K} & \frac{\partial^N \psi_x(u_1, v_1)}{\partial v^K} & \cdots & \frac{\partial^N \psi_x(u_1, v_1)}{\partial v^N}
    \\
    \frac{\partial^N \psi_x(u_2, v_2)}{\partial u^K} & \frac{\partial^N \psi_x(u_2, v_2)}{\partial v^K} & \cdots & \frac{\partial^N \psi_x(u_2, v_2)}{\partial v^N}
    \\
    \vdots & \vdots & \ddots & \vdots
    \\
    \frac{\partial^N \psi_x(u_K, v_K)}{\partial u^K} & \frac{\partial^N \psi_x(u_K, v_K)}{\partial v^K} & \cdots & \frac{\partial^N \psi_x(u_K, v_K)}{\partial v^N}
\end{pmatrix}
\]
Now, to solve $Hq = 0$, it suffices to compute the right singular vector $\mathbf{q}_1$ associated with the smallest singular value of $H$. Once $\mathbf{q}_1$ has been computed, it can be seen from (12) that $(\hat{a}_n, \hat{b}_n)$ can be readily obtained as roots of the homogeneous polynomial $F(x, y) = \sum_{j=0}^{N} q_j x^{N-j} y^j$, as already suggested in [6] [15]. Note that $K$ should be large enough to ensure that the null space of $H$ is one-dimensional, and to avoid finding complex solutions for $a_n$ and $b_n$ when they are actually real.

The previous algorithm is valid for real sources and mixture. In the next sections, an algorithm is described that can be used to identify complex mixtures.

### 4. TWO COMPLEX SENSORS

When sources and mixture are complex, one may separate real and imaginary parts of the sources and mixture are complex, which is satisfied for numerous basic modulations, as QPSK. Then define the two characteristic functions of $(\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2)$ can be written, for $(u_1, v_1, u_2, v_2) \in \Omega$:

$$
\psi_{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2} = \sum_{n=1}^{N} \psi_n (\hat{a}_nu_1 + \hat{b}_nu_2 + \hat{b}_nu_2) + \psi_n (\hat{a}_nu_1 + \hat{b}_nu_2 - \hat{b}_nu_2)
$$

Define differential operator $D_n$ as:

$$
D_n = -(\hat{b}_nu_1 + \hat{b}_nu_2) \psi_{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2} = \sum_{n=0}^{N} \sum_{k=0}^{n} d_{nk} \psi_n (\hat{a}_nu_1 + \hat{b}_nu_2 + \hat{b}_nu_2)
$$

By applying $D_n$, we remove the $n$th term of the sum in (14). When applying all the $D_n$’s, $n = 1, \ldots, N$, we obtain:

$$
\sum_{n=0}^{N} \sum_{k=0}^{n} d_{nk} \psi_n \frac{\partial^{N-n} \psi_{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2}}{\partial u_1^{N-n} \partial u_2^k} = 0
$$

By replacing $\psi_{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2}$ by its expression (14) we get:

$$
\sum_{n=0}^{N} \sum_{k=0}^{n} d_{nk} \psi_n \frac{\partial^{N-n} \psi_{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2}}{\partial u_1^{N-n} \partial u_2^k} = 0
$$

which implies:

$$
\sum_{n=0}^{N} \sum_{k=0}^{n} d_{nk} \psi_n (\hat{a}_nu_1 + \hat{b}_nu_2 + \hat{b}_nu_2) = 0
$$

As a consequence, by computing directly the joint second characteristic function of $\{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2\}$, we end up with an over-determined system of 8 homogeneous equations of the form: $\sum_{n=0}^{N} \sum_{k=0}^{n} d_{nk} \psi_n \frac{\partial^{N-n} \psi_{\hat{x}_1, \hat{x}_1, \hat{x}_2, \hat{x}_2}}{\partial u_1^{N-n} \partial u_2^k} = 0$, that we are hardly able to solve. General Gröbner bases approaches are indeed rather impracticable for such systems in 12 variables of degree larger than 3; special purpose algebraic approaches are being investigated, and will be the subject of a future paper. We propose in the next section another approach to the problem.

### 5. A SOLUTION WITH REDUCED COMPLEXITY

The idea consists of solving three $2 \times 2N$ real identification problems instead of a single $2 \times N$ complex one. These are:

$$
\begin{bmatrix}
\begin{pmatrix}
\psi_{\hat{x}_1} \\
\psi_{\hat{x}_1}
\end{pmatrix}
\end{bmatrix} = A_1 s' , \quad \begin{bmatrix}
\begin{pmatrix}
\psi_{\hat{x}_2} \\
\psi_{\hat{x}_2}
\end{pmatrix}
\end{bmatrix} = A_2 s' , \quad \begin{bmatrix}
\begin{pmatrix}
\psi_{\hat{x}_1} \\
\psi_{\hat{x}_2}
\end{pmatrix}
\end{bmatrix} = A_R s',
$$

where $s'$ is the $2N \times 1$ real source vector defined as $s'_n = \hat{s}_n$ and $s'_{N+n} = \bar{\hat{s}}_n$, for $1 \leq n \leq N$. With the algorithm described in section 3, we obtain $2 \times 2N$ matrices $A_1$, $A_2$, and $A_R$, which are estimates of $A_1 A_1 P_1$, $A_2 A_2 P_2$, $A_R A_R P_R$, respectively, where the $A$’s are diagonal regular matrices and $P$’s are permutations. Matrices $A_k$ contain $N$ pairs of orthogonal vectors. First delete in each pair the vector of the form $(-q, r)$ and retain the other of the form $(r, q)$, in order to get $2 \times N$ matrices. Then define the two
1 × N vectors below, by taking the ratios of their first and second lines:

\[ R_a(k) = \frac{\hat{A}_1(1, k)}{\hat{A}_1(2, k)}, \quad R_b(\ell) = \frac{\hat{A}_2(1, \ell)}{\hat{A}_2(2, \ell)} \]

and the 1 × 2N vector \( \rho(m) = \frac{\hat{A}_n(1, m)}{\hat{A}_n(2, m)} \). A careful inspection of the theoretical values of these terms shows that \( R_a(k) \in \{a_n/\hat{a}_n, 1 \leq n \leq N\} \), \( R_b(\ell) \in \{b_n/\hat{b}_n, 1 \leq n \leq N\} \), \( \rho(m) \in \{\hat{a}_n/\hat{b}_n, \hat{a}_n/ib_n, 1 \leq n \leq N\} \). This reveals a simple algorithm to recover \( A \) up to the inherent indeterminacies of the original complex identification problem. In fact, form the 1 × N² ratio vector \( R_{ab}(k, \ell) = R_a(k)/R_b(\ell) \), and the 1 × (N² − N) ratio vector \( \rho_{ab}(j, m) = \rho(j)/\rho(m) \), \( j \neq m \). Next detect the N best index quadruplets that minimize the distance \( |R_{ab}(k, \ell) - \rho_{ab}(j, m)| \). This distance indeed becomes null when \( \hat{a}_k b_{\ell b} = \hat{a}_j b_{\ell m} \). Thus, it can be seen that this association procedure has fixed the permutation and scale ambiguities, and that the columns \( [\hat{a}_k, \hat{a}_j, \hat{b}_k, \hat{b}_j] \) and \( [1, 1/R_a(k), 1/\rho(j), 1/w] \) are proportional, where \( w \) is taken to be either \( \rho(m) R_a(k) \), or \( \rho(j) R_b(\ell) \), or their average. Note that this holds true as long as \( \hat{a}_k \) is not null. Should \( \hat{a}_k = 0 \), the same procedure can be run by setting another of the 4 entries to 1 (the four are not allowed to be all zero). We skip the programming details.

As a conclusion, we have obtained an estimate of each column of \( A \) up to \( N \times N \) permutation and scale ambiguities, as requested.

6. COMPUTER RESULTS

Sources that have been generated are i.i.d QPSK, and have therefore independent real and imaginary parts. Their whiteness has little influence on the results, but the correlation between real and imaginary parts obviously does, if any, up to some extent.

There are at least two ways to implement the algorithms previously described, depending on the estimator of derivatives \( \partial^{k+\ell} \psi/\partial u^k \partial v^\ell \). The first approach consists of estimating the characteristic function \( \psi(x) \) of the observation over a set \( (u, v) \in \Omega \) containing the origin. We merely utilized in figure 1 the sample estimate:

\[ \hat{\psi}_x(u, v) = \log \left[ \frac{1}{T} \sum_{t=1}^{T} \exp[iv(u^* x_1(t) + v^* x_2(t))] \right] \]

where vectors \( x(t) \) are realizations of the random variable \( x \). The successive derivatives can then be obtained from the values of \( \hat{\psi}_x(u, v) \) over a grid included in \( \Omega \) by central finite differences.

The second approach is theoretically more accurate, but turned out to be more sensitive to deviations of the sample mean from zero. It consists of computing the theoretical expressions of the requested derivatives of \( \psi_x \) as a function of successive derivatives of \( \phi_x \). This rather cumbersome calculation has been carried out by Maple once for all. Then, one can eventually replace the terms involved in the latter expressions by their sample estimates. As an example in the real case, \( \frac{\partial}{\partial u^k} \hat{\psi} = \frac{\partial}{\partial u^k} \hat{\phi} = \frac{1}{2} \), and a sample estimate of \( \partial^{k+\ell} \hat{\phi}/\partial u^k \partial v^\ell \) is:

\[ i^{k+\ell} \frac{1}{T} \sum_{t=1}^{T} x_1(t)^k x_2(t)^\ell \exp \{i(x_1(t) u + x_2(t) v)\} \]

This implementation has been chosen for obtaining figures 2 and 3.

6.1. 1 × 3 complex mixture of 3 complex sources

In this subsection, the mixture is received on a single sensor and is taken to be:

\[ A = [ 1 \cos \pi/6 + i \sin \pi/6 \cos \pi/3 + i \sin \pi/3 ] \]
The SNR is varied between -20dB and 60dB. The gap criterion measures the Frobenius norm between the estimated mixture and the actual one, \( A \), for the most favorable scale and permutation \( 3 \times 3 \) ambiguities, as in [6] for instance. Figure 1 reports the median of the gap values obtained with 41 independent trials, plus or minus the standard deviation. This gap has a maximal value of 6 for \( 2 \times 3 \) mixtures, because it is insensitive to scale. A value below 0.3 can be considered as quite good in the present framework.

6.2. \( 2 \times 3 \) complex mixture of 3 complex sources

Now the mixture has two rows. The first is the same as in the previous subsection, and the second is:

\[
\begin{bmatrix}
\cos 2\pi/5 + i\sin 2\pi/5 & \cos \pi/5 + i\sin \pi/5 & 1
\end{bmatrix}
\]

The performance criterion used in this section is that described in [1] for \( P \times N \) mixtures, and is suboptimal. In fact, computing the optimal scale and permutation ambiguities would be too computationally costly. The performance obtained is reported in figure 3 as a function of SNR, and becomes acceptable above 20dB.

7. REFERENCES