ON UNDERDETERMINED SOURCE SEPARATION

Anisse Taleb, Christian Jutten

LIS - INPG, 46 Avenue Félix Viallet
38000 Grenoble FRANCE
taleb@tirf.inpg.fr

ABSTRACT

This paper discusses some theoritical results on underdetermined source separation *i.e.* when the mixing matrix is degenerate, espcially when there is more sources than observations. In this case, we show that the sources can be restored up to an arbitrary additive random vector. In the particular case of discrete sources, very relevant for digital communications, we show that this vector is certain.

1. INTRODUCTION AND PROBLEM FORMULATION

Let an array of m sensors, each one receiving an unknown linear instantaneous combination of n unknown sources. Source separation consists in recovering the sources by using only sensor outputs. It relies on the main assumption of source spatial statistical independence.

Most source separation algorithms assume that the number of sources is less than the number of sensors $n \le m$. Despite its high interest, the case where the number of sources is greater than the number of sensors $n > m$ (called underdetermined source separation) has only been treated in particular cases.

In [2], Cao *et al.* gave necessary and sufficient conditions for the existence of a separating matrix, which can separate the sources into several groups. In [10], the use of high order statistics, under some conditions, can estimate the distinct angles of arrivals for more sources than sensors. In [3], it is shown how more sources than sensors can be identified by using only the fourth order cumulants. Belouchrani *et al.* [1] use the Expectation-Maximization algorithm for the separation of more sources than sensors, assuming that the sources are k-valued. A competitive learning algorithm has been used by Pajunen [9] to solve a similar problem, in which the sources were assumed to be binary. Finally, Comon *et al.* [4] adress the problem by forming virtual measurements in order to increase the observation vector dimension. These supplementary observations are nonlinear functions of the truly observed sensor outputs.

Let $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_n(t)]^T$ denote the $n \times 1$ unknown stationary source vector, components of which are assumed to be statistically independent. Let

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{v}(t) \tag{1}$$

be the output of the array of m sensors ($m < n$). $\boldsymbol{A}$ is an $m \times n$ full rank unknown matrix modeling the channel transfer function. $\boldsymbol{v}(t)$ is a stationary white additive Gaussian noise with covariance matrix $\boldsymbol{\Gamma} = E[\boldsymbol{v}\boldsymbol{v}^T]$. It is assumed that $\boldsymbol{s}(t)$ and $\boldsymbol{v}(t)$ are independent.

One can write $\boldsymbol{v}(t) = \boldsymbol{\Gamma}^{1/2}\boldsymbol{w}(t)$, where $\boldsymbol{w}(t)$ is Gaussian with unit variance and independent components. The observations can then be rewritten as :

$$\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{\Gamma}^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}(t) \\ \boldsymbol{w}(t) \end{bmatrix} = \boldsymbol{A}'\boldsymbol{s}'(t) \tag{2}$$

Equation (2) shows that the additive noise can be merged with the source signals to form a new source vector. Thus we will assume the model (1) without noise :

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) = \sum_{i=1}^{n} \boldsymbol{a}_i s_i(t) \tag{3}$$

The $\boldsymbol{a}_i$'s are the column vectors of matrix $\boldsymbol{A}$. They are assumed to be pairwise linearly independent. In fact, if it exists two proportional columns, for example $\boldsymbol{a}_k = \alpha \boldsymbol{a}_p$, we can then merge s_k and s_p to form a new source $s_k(t) + \alpha s_p(t)$ which will be statistically independent from the others.

Contrary to [2] and to classical source separation, it is restrictive to design a matrix that uses $\boldsymbol{x}(t)$ as input, and provides as output $\boldsymbol{y}(t)$ whose components are independent. This is in general impossible, and when it is possible, under some assumptions on the mixing matrix $\boldsymbol{A}$, sources can only be separated into several complementary groups [2].

2. UNIQUENESS OF SOLUTIONS

In underdetermined source separation, there are two distinct but very dependent issues : *identifiability* and *separability*. The former discusses the uniqueness of the rectangular mixing matrix $\boldsymbol{A}$ based on the problem assumptions *i.e.* statistical independence of the sources. Separability for its part is concerned by retrieving the sources, given the observation vector $\boldsymbol{x}(t)$ and given the matrix $\boldsymbol{A}$. The two problems can be summarized by the following uniqueness equation[1] :

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} = \boldsymbol{B}\boldsymbol{y} \tag{4}$$

which formalizes the following question: *Does there exist a couple* $(\boldsymbol{y}, \boldsymbol{B})$ *satisfying the same assumptions (***A1***,***A2***) than* $(\boldsymbol{s}, \boldsymbol{A})$ *and giving the same observation vector* $\boldsymbol{x}$ *?*

- **A1:** components of source vector ($\boldsymbol{s}$ or $\boldsymbol{y}$) are independent.
- **A2:** mixing matrix ($\boldsymbol{A}$ or $\boldsymbol{B}$) is full rank and its column vectors pairwise linearly independent

Here the Darmois-Skitovic theorem [6, 11] (see also [7] for easier accessibility) is of no help since neither $\boldsymbol{A}$ nor $\boldsymbol{B}$ are invertible. We will rather extend, for any m, a more general theorem proved by Darmois for $m = 2$ in the same paper [6].

Recall that the characteristic function $\varphi_x(u)$ of a random variable x always exists and is continuous. According to (4), the characteristic function of $\boldsymbol{x}$ writes as:

$$\begin{aligned}\varphi_{\boldsymbol{x}}(\boldsymbol{u}) &= E\Big[\exp(j\boldsymbol{u}^T\boldsymbol{x})\Big] \\ &= \prod_{i=1}^{n} \varphi_{y_i}(\boldsymbol{b}_i^T\boldsymbol{u}) = \prod_{i=1}^{n} \varphi_{s_i}(\boldsymbol{a}_i^T\boldsymbol{u}) \quad \forall \boldsymbol{u} \in \mathbb{R}^m\end{aligned} \tag{5}$$

Since $\varphi_{\boldsymbol{x}}(\boldsymbol{0}) = 1$ and $\varphi_{\boldsymbol{x}}$ is continuous, then there exists a small neigbourhood $\mathcal{U}$ of $\boldsymbol{0}$ in which $\varphi_{\boldsymbol{x}}$ does not vanish. Denoting log the principal branch of the logarithm in the right half plane and

$$\psi_{s_i} = \log\varphi_{s_i} \quad \text{and} \quad \psi_{y_i} = \log\varphi_{y_i} \quad 1 \le i \le n \tag{6}$$

the second characteristic functions, then for $\boldsymbol{u} \in \mathcal{U}$ we have :

$$\sum_{i=1}^{n} \psi_{y_i}(\boldsymbol{b}_i^T\boldsymbol{u}) = \sum_{i=1}^{n} \psi_{s_i}(\boldsymbol{a}_i^T\boldsymbol{u}) \tag{7}$$

[1] dependence over time will be omitted since the model is assumed to be instantaneous

According to the pairwise linear dependence between the columns of $\boldsymbol{A}$ and $\boldsymbol{B}$, we can consider three cases:

- Suppose that the columns of $\boldsymbol{B}$ are pairwise linearly independent with the columns of $\boldsymbol{A}$, *i.e.*

$$\forall i, j \text{ if} \quad \alpha_{ij}\boldsymbol{a}_i + \beta_{ij}\boldsymbol{b}_j = 0 \quad \text{then} \quad \alpha_{ij} = \beta_{ij} = 0 \tag{8}$$

In (7), replacing $\boldsymbol{u}$ by $\boldsymbol{u} + \boldsymbol{\delta}_1$ such that $\boldsymbol{b}_n^T\boldsymbol{\delta}_1 = 0$ and substracting (7) from the obtained equation, gives:

$$\sum_{i=1}^{n-1} \Delta_{\epsilon_{i1}}\psi_{y_i}(\boldsymbol{b}_i^T\boldsymbol{u}) = \sum_{i=1}^{n} \Delta_{\xi_{i1}}\psi_{s_i}(\boldsymbol{a}_i^T\boldsymbol{u}) \tag{9}$$

where Δ denotes the finite difference operator,

$$\Delta_h f(x) = f(x+h) - f(x),$$

and

$$\epsilon_{ik} = \boldsymbol{b}_i^T\boldsymbol{\delta}_k \tag{10}$$

$$\xi_{ik} = \boldsymbol{a}_i^T\boldsymbol{\delta}_k \tag{11}$$

In (9), the function ψ_{y_n} disappeared (other differences are non null due to pairwise linear independence). By induction, we can successively eliminate all ψ_{y_i}, and all ψ_{s_i} except for $i = l$ and get:

$$\Delta_{\xi_{l,2n-1}}\Delta_{\xi_{l,2n-2}}\cdots\Delta_{\xi_{l1}}\psi_{s_l}(\boldsymbol{a}_l^T\boldsymbol{u}) = 0, \;\; 1 \le l \le n$$

This shows that each second characteristic function must be a polynomial of degree at most $2n - 1$ (one could get better estimates of the degree of these polynomials but it is without great interest here). The theorem of Marcinkiewicz [8] shows that these polynomial can not be of degree greater than 2: $\boldsymbol{x}$ and $\boldsymbol{y}$ are then Gaussian.

- If there exists a column $\boldsymbol{b}_l = \alpha\boldsymbol{a}_k$, using the same reasonning, we deduce that the functions ψ_{y_i} $(i \ne l)$, ψ_{s_i} $(i \ne k)$ and $\psi_{y_l}(\alpha u) - \psi_{s_k}(u)$ are second order polynomials. Hence, y_i $(i \ne l)$ and s_i $(i \ne k)$ are Gaussian, but nothing can be said about y_l and s_k. Similar conclusions can be derived if p columns of $\boldsymbol{B}$ are proportional to p columns of $\boldsymbol{A}$.

- Now, suppose that all columns of $\boldsymbol{B}$ are proportional to the columns of $\boldsymbol{A}$, or without loss of generality[2] $\boldsymbol{B} = \boldsymbol{A}$, then (7) becomes :

$$\sum_{i=1}^{n} (\psi_{y_i} - \psi_{s_i})(\boldsymbol{a}_i^T\boldsymbol{u}) = 0 \tag{12}$$

Using successive finite differences, we get that each function $\psi_{y_i} - \psi_{s_i}$ $(1 \le i \le n)$ is a polynomial of degree at most $n - 1$.

In other words, the above discussion summerizes as:

[2] In the general case $\boldsymbol{B} = \boldsymbol{ADP}$, where $\boldsymbol{D}$ is a diagonal matrix, and $\boldsymbol{P}$ is a permutation matrix.

- If all sources are non Gaussian, the matrix $\boldsymbol{A}$ is unique up to a right multiplication by a diagonal and a permutation matrix.

- If there exists p non Gaussian sources, p columns of $\boldsymbol{B}$ are necessarily proportional to p columns of $\boldsymbol{A}$.

2.1. Non Gaussian sources

In the following, sources are supposed to be non Gaussian. According to above results, (4) becomes:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} = \boldsymbol{A}\boldsymbol{y} \tag{13}$$

Here the indeterminacies on $\boldsymbol{B}$ are absorbed by $\boldsymbol{y}$ without modifying the independence of its components. When $m = n$ it is obvious that $\boldsymbol{y} = \boldsymbol{x}$, the couple $(\boldsymbol{y}, \boldsymbol{B})$ is unique (up to the usual indeterminacies).

However, when $m < n$, $\boldsymbol{y}$ is not unique. Algebraically, general solutions are given by:

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{z} \tag{14}$$

where $\boldsymbol{z}$ is an arbitrary vector belonging to $\mathrm{Ker}(\boldsymbol{A})$. In fact, $\boldsymbol{z}$ can not be arbitrary since its statistical dependence over $\boldsymbol{y}$ will influence the statistical dependence between the components of $\boldsymbol{s}$.

One may first notice that if $\boldsymbol{z}$ is independent from $\boldsymbol{y}$, then it must be a certain vector. Indeed, by (14) the characteristic function of $\boldsymbol{y}$ writes as:

$$\prod_{i=1}^{n} \varphi_{y_i}(u_i) = \varphi_{\boldsymbol{z}}(\boldsymbol{u}) \prod_{i=1}^{n} \varphi_{s_i}(u_i) \tag{15}$$

which shows that $\varphi_{\boldsymbol{z}}(\boldsymbol{u})$ factorizes as the product of marginal characteristic functions, hence $\boldsymbol{z}$ has independent components. Moreover, $\sum_{k=1}^{n} a_{ik} z_k = 0$ for $0 \le i \le m$. Since a certain random variable is closed under decomposition (Cramér [5])[3], we deduce that each z_k is certain.

In general, the characterization of $\boldsymbol{z}$ remains an open problem. However, in some particular situations, the problem becomes tractable especially for discrete sources, which are the main support of digital communications.

2.2. Discrete sources

Suppose each source s_i takes its values in a finite discrete set $\mathcal{C}_i = \{s_i^1, s_i^2, \cdots, s_i^{k_i}\}$. With this *a priori* information about the source distribution, one seeks for a solution $\boldsymbol{y}$ of (13) which is discrete, *i.e.* each y_i takes its values in $\mathcal{C}'_i = \{y_i^1, y_i^2, \cdots, y_i^{k'_i}\}$. Referring to the discussion on equation (12), we have:

$$\varphi_{y_i}(u) = \varphi_{s_i}(u) \exp(P_i(u)) \tag{16}$$

where P_i is a polynomial of degree $n-1$. Replacing φ_{y_i} and φ_{y_i} by their values leads to:

$$\sum_{l=1}^{k'_i} p(y_i^l) \exp(juy_i^l) = \{\sum_{l=1}^{k_i} p(s_i^l) \exp(jus_i^l)\} \exp(P_i(u)) \tag{17}$$

Clearly, P_i must be a polynomial of the first order. In fact, the analytic extension of φ_{y_i} is an entire function whose order[4] equals 1, while the right hand side of (17) is of order $n-1$. The general form of this polynomial is $P_i(z) = az + b$. Since $\varphi_{y_i}(0) = \varphi_{s_i}(0) = 1$ then b is necessarily null. Moreover, for any real variable v, $\varphi_{y_i}(jv) = \varphi_{s_i}(jv)\exp(jav)$, then a is necessarily a purely imaginary number. Let $P_i(u) = jut_i$ where t_i is an arbitrary real, (17) becomes:

$$\sum_{l=1}^{k'_i} p(y_i^l) \exp(juy_i^l) = \sum_{l=1}^{k_i} p(s_i^l) \exp(ju(s_i^l + t_i)) \tag{18}$$

Since the representation of a function by a trigonometric polynomial is unique, necessarily $k_i = k'_i$. Moreover, there exists a permutation σ of $\{1, 2, \cdots, k_i\}$ such that:

$$y_i^l = t_i + s_i^{\sigma(l)} \quad \text{and} \quad p(y_i^l) = p(s_i^{\sigma(l)}) \qquad 1 \le l \le k_i,\ 1 \le i \le n \tag{19}$$

According to (14), and since $\boldsymbol{s}$ and $\boldsymbol{y}$ are discrete, then $\boldsymbol{z}$ is discrete too. Moreover:

$$\sum_{l} p(y_i^l) \exp(juy_i^l) = \sum_{l,r} p(s_i^l, z_i^r) \exp(ju(s_i^l + z_i^r) \tag{20}$$

and by using (19):

$$\sum_{l} p(s_i^l) \exp(ju(s_i^l + t_i)) = \sum_{l,r} p(s_i^l, z_i^r) \exp(ju(s_i^l + z_i^r)) \tag{21}$$

the same argument on uniqueness of trigonometric polynomial representation leads to:

$$\forall l, r: \quad s_i^l + z_i^r \in \mathcal{C}_i + t_i = \{s_i^1 + t_i, s_i^2 + t_i, \cdots, s_i^{k_i} + t_i\}$$

[3]A simple way to see this consists in computing the variance (when it exists) of $\sum_{k=1}^{n} a_{ik} z_k = 0$, taking into account that $\boldsymbol{z}$ has independent components.

[4]The definition of the order of an entire function may be found in: W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1966.

which obviously shows that:

$$\forall r : z_i^r - t_i \in \mathcal{C}_i - s_i^l \quad , \forall l$$

thus:

$$\forall r : z_i^r - t_i \in \cap_{l=1}^{k_i} \{\mathcal{C}_i - s_i^l\}$$

This finite intersection of sets contains the unique element 0. Thus $\forall r : z_i^r = t_i$, and $\boldsymbol{z} = \boldsymbol{t}$ is certain.

2.3. Mixed discrete and continuous sources

Suppose there exists $n-m$ discrete sources and m continuous sources. As seen in the previous section, the $n-m$ discrete sources can be restored up to a certain vector $\boldsymbol{t} \in \mathbb{R}^{n-m}$. According to (14), $\boldsymbol{z} \in \mathrm{Ker}(\boldsymbol{A})$ whose dimension is $n-m$, and whose $n-m$ components are certain and equal to the components of $\boldsymbol{t}$. Obviously, this shows that $\boldsymbol{z}$ is a certain vector. Each source can then be restored up to an additive constant.

This result can be extended to the case where we have $p \geq n-m$ discrete sources.

3. CONCLUSION

In this paper, we adress the problem of underdetermined source separation and prove that: non Gaussian sources can only be restored up to an arbitrary additive random vector, for discrete sources this vector is certain. This approach could also be generalized if the sources number is unknown.

This study proves that source separation is not impossible even if there are more sources than observations. It remains now to design algorithms validating the theory.

4. REFERENCES

[1] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation by the expectation-maximization technique. In *NOLTA 95*, pages 49–53, Las Vegas, USA, 1995.

[2] X. Cao and R. Liu. General approach to blind source separation. *IEEE Trans. S.P.*, 44(3):562–571, March 1996.

[3] J-F. Cardoso. Super-symetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors. In *ICASSP 91*, pages 3109–3112, Toronto (Canada), May 1991.

[4] P. Comon and O. Grellier. Nonlinear inversion of underdetermined mixtures. In *ICA 99*, Aussois (France), January 1999. submitted.

[5] H. Cramér. *Random variables and probability distributions.* Cambridge, 1937.

[6] G. Darmois. Analyse générale des liaisons stochastiques. *Rev. Inst. Internat. Stat.*, 21:2–8, 1953.

[7] A.M. Kagan, Y.V. Linnik, and C.R. Rao. *Characterization problems in mathematical statistics.* Wiley, 1973.

[8] J. Marcinkiewicz. Sur une propriété de la loi de gauss. *Math. Ziet.*, 44:612–618, 1938.

[9] P. Pajunen. A competitive learning algorithm for separating binary sources. In *ESANN 97*, pages 255–260, Bruges (Belgium), April 1997.

[10] S. Shamsunder and Giannakis G. Modeling of non-gaussian array data using cumulants: doa estimation of more sources with less sensors. *Signal Processing*, 30(3):279–297, February 1993.

[11] V. P. Skitovic. Linear forms of independent random variables and the normal distribution law. *Izvestiya Akademii Nauk SSSR. Seriya Matematiceskaya*, 18:185–200, 1954. in Russian.