Blind Separation of Instantaneous Mixtures of Dependent Sources

Marc Castella, Pierre Comon

Blind separation of instantaneous mixtures of dependent sources

Marc Castella[1] and Pierre Comon[2]

[1] GET/INT, UMR 5157, Département CITI, 9 rue Charles Fourier, 91011 Évry Cedex, France
marc.castella@int-evry.fr,
[2] CNRS, I3S, UMR 6070, BP.121, Sophia-Antipolis cedex, France
pcomon@i3s.unice.fr

Abstract. This paper deals with the problem of Blind Source Separation. Contrary to the vast majority of works, we do not assume the statistical independence betwwen sources and explicitly consider that they are dependent. We introduce three particular models of dependent sources and show that their cumulants have interesting properties. Based on these properties, we investigate the behaviour of classical Blind Source Separation algorithms when applied to these sources: depending on the source vector, the separation may be sucessful or some additionnal indeterminacies can be identified.

1 Introduction

Independent Component Analysis is now a well recognized concept, which has fruitfully spread out to a wide panel of scientific areas and applications. Contrary to other frameworks where techniques take advantage of a strong information on the diversity, for instance through the knowledge of the array manifold in antenna array processing, the core assumption is much milder and reduces to the statistical mutual independence between the inputs.

However, this assumption is not mandatory in Blind Source Separation. For instance, in the case of static mixtures, sources can be separated if they are only decorrelated when their nonstationarity or their color can be exploited. Other properties such as the fact that sources belong to a finite alphabet can alternatively be utilized [1,2] and do not require statistical independence.

Inspired from [3,4], we investigate the case of dependent sources, without assuming nonstationarity nor color. To our knowledge, only few references have tackled this issue [5,6].

2 Mixture model and notations

We consider a set of N source signals $(s_i(n))_{n\in\mathbb{Z}}, i = 1, \ldots, N$. The dependence on time of the signals will not be made explicit in the paper. The sources are mixed, yielding a P-dimensional observation vector $\mathbf{x} = (\mathbf{x}(n))_{n\in\mathbb{Z}}$ according to the model:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{s} = (s_1, \ldots, s_N)^\top$, $\mathbf{x} = (x_1, \ldots, x_P)^\top$ and $\mathbf{A}$ is a $P \times N$ matrix called the mixing matrix. We assume that $\mathbf{A}$ is left-invertible.

Source separation consists of finding a $P \times N$ separating matrix $\mathbf{B}$ such that its output $\mathbf{y} = \mathbf{Bx}$ corresponds to the original sources. When only the observations are used for this, the problem is referred to as the blind source separation (BSS) problem. Introducing the $N \times N$ global matrix $\mathbf{G} \triangleq \mathbf{BA}$, the BSS is problem is solved if $\mathbf{G}$ is a so-called trivial matrix, i.e. the product of a diagonal matrix with a permutation: these are well known ambiguities of BSS.

In this paper, we will study separation criteria as functions of $\mathbf{G}$. Source separation sometimes proceeds iteratively, extracting one source at a time (e.g. deflation approach). In this case, we will write $y = \mathbf{bx} = \mathbf{gs}$ where $\mathbf{b}$ and $\mathbf{g} = \mathbf{bA}$ respectively correspond to a row of $\mathbf{B}$ and $\mathbf{G}$ and y denotes the only output of the separating algorithm. In this case, the separation criteria are considered as functions of $\mathbf{g}$. Finally, we denote by $\mathrm{E}\{.\}$ the expectation operator and by $\mathrm{Cum}\{.\}$ the cumulant of a set of random variables. $\mathrm{Cum}_4\{y\}$ is equivalent to $\mathrm{Cum}\{y, y, y, y\}$ and, for complex variables, $\mathrm{Cum}_{2,2}\{y\}$ stands for $\mathrm{Cum}\{y, y, y^*, y^*\}$.

3 Examples and properties of dependent sources

We introduce in this section different cases of vector sources that are dependent and that will be considered in this paper.

3.1 Three dependent sources

Binary phase shift keying (BPSK) signals have specificities that will allow us to obtain source vectors with desired properties. By definition, BPSK sources take values $s = +1$ or $s = -1$ with equal probability 1/2. We define the following source vector:

A1. $\mathbf{s} \triangleq (s_1, s_2, s_3)^\top$ where s_1 is BPSK; s_2 is real-valued non Gaussian, independent of s_1 and satisfies $\mathrm{E}\{s_2\} = \mathrm{E}\{s_2^3\} = 0$ and $s_3 = s_1 s_2$.

Interestingly, the following lemma holds true:

Lemma 1. *The sources s_1, s_2, s_3 defined by A1 are obviously mutually dependent. Nevertheless they are decorrelated and their fourth-order cross-cumulants vanish, that is:*

$$\mathrm{Cum}\{s_i, s_j\} = 0 \text{ except if } i = j, \tag{2}$$

$$\mathrm{Cum}\{s_i, s_j, s_k, s_l\} = 0 \text{ except if } i = j = k = l. \tag{3}$$

Proof. Using the definition of s_1, s_2 and their independence, one can easily check that $\mathrm{E}\{s_1\} = \mathrm{E}\{s_2\} = \mathrm{E}\{s_3\} = 0$. For these centered random variables, it is known that cumulants can be expressed in terms of moments:

$$\mathrm{Cum}\{s_i, s_j\} = \mathrm{E}\{s_i s_j\} \tag{4}$$

$$\begin{aligned}\mathrm{Cum}\{s_i, s_j, s_k, s_l\} = {} & \mathrm{E}\{s_i s_j s_k s_l\} - \mathrm{E}\{s_i s_j\}\mathrm{E}\{s_k s_l\} \\ & - \mathrm{E}\{s_i s_k\}\mathrm{E}\{s_j s_l\} - \mathrm{E}\{s_i s_l\}\mathrm{E}\{s_j s_k\}\end{aligned} \tag{5}$$

Using again the definition of s_1, s_2 and their independence, it is then easy to check all cases of equations (4) and (5) and to verify that these 4th order cross-cumulants are indeed null. On the other hand, the third order cross-cumulant reads:

$$\mathrm{Cum}\left\{s_1, s_2, s_3\right\} = \mathrm{E}\left\{s_1 s_2 s_3\right\} = \mathrm{E}\left\{s_1^2 s_2^2\right\} = \mathrm{E}\left\{s_1^2\right\} \mathrm{E}\left\{s_2^2\right\} > 0 \tag{6}$$

and this proves that s_1, s_2, s_3 are mutually dependent. $\square$

Depending on s_2, more can be proved about the source vector defined by A1. For example, if the probability density function of s_2 is symmetric, then s_1 and s_3 are independent. On the contrary s_2 and s_3 are generally not independent.

An even more specific case is obtained when s_2 is itself BPSK. In this case, one can check that the sources (s_1, s_2, s_3) are pairwise independent, although not mutually independent.

3.2 Pairwise independent sources

We now investigate further the case of pairwise independent sources and introduce the following source vector:

A2. $\mathbf{s} = (s_1, s_2, s_3, s_4)^\intercal$ where s_1, s_2 and s_3 are independent BPSK and $s_4 = s_1 s_2 s_3$.

This case has been considered in [3], where it has been shown that

$$\forall i \in \{1, \ldots, 4\}, \mathrm{Cum}\left\{s_i, s_i, s_i, s_i\right\} = -2 \ , \quad \mathrm{Cum}\left\{s_1, s_2, s_3, s_4\right\} = 1 \tag{7}$$

and all other cross-cumulants vanish. The latter cumulant value shows that the sources are mutually dependent; although it can be shown that they are pairwise independent. It should be clear that pairwise independence is not equivalent to mutual independence but in an ICA context, it is relevant to recall the following proposition, which is a direct consequence of Darmois' theorem [7, p.294]:

Proposition 1. *Let* $\mathbf{s}$ *be a random vector with mutually independent components, and* $\mathbf{x} = \mathbf{G}\mathbf{s}$. *Then the mutual independence of the entries of* $\mathbf{x}$ *is equivalent to their pairwise independence.*

Based on this proposition, the ICA algorithm in [7] searches for an output vector with pairwise independent component. Let us stress that this holds only if the source vector has mutually independent components: pairwise independence is indeed not sufficient to ensure identifiability as we will see in Section 4.2.

3.3 Complex valued sources

We consider quaternary phase shift keying (QPSK) sources which by definition take their values in $\{1, \imath, -1, -\imath\}$ with equal probability $1/4$ and we define the following source vector:

A3. $\mathbf{s} = (s_1, s_2, s_3, s_4)^\intercal$ where s_1, s_2 and s_3 are mutually independent QPSK and $s_4 = s_1 s_2 s_3 s_4$.

Based on the Equations (4) and (5) which hold for the above centered sources, one can check the following proposition:

Lemma 2. *The sources in A3 are dependent and satisfy* $\mathrm{Cum}\{s_1, s_2, s_3, s_4^*\} = 1$. *However, they are second-order decorrelated and all their 4th order circular cross-cumulants (i.e. with as many conjugates as non-conjugates) vanish, that is:*

$$\mathrm{Cum}\left\{s_i, s_j^*\right\} = 0 \textit{ and } \mathrm{Cum}\left\{s_i, s_j\right\} = 0 \textit{ except if } i = j, \tag{8}$$

$$\mathrm{Cum}\left\{s_i, s_j, s_k^*, s_l^*\right\} = 0 \textit{ except if } i = j = k = l \tag{9}$$

4 ICA algorithms and dependent sources

4.1 Independence is not necessarily required

The sources given by A1 provide us with a specific example of dependent sources that are sucessfully separated by several ICA methods:

Proposition 2. *Let* $y = \mathbf{g}\mathbf{s}$ *where the vector of sources is defined by A1. Then, the function*

$$\mathbf{g} \mapsto |\mathrm{Cum}_4\{y\}|^\alpha, \ \alpha \geq 1 \tag{10}$$

defines a MISO contrast function, that is, its maximization over the set of unit norm vectors ($\|\mathbf{g}\|^2 = 1$) *leads to a vector* $\mathbf{g}$ *with only one non-zero component.*

Proof. The above proposition follows straightforwardly from Lemma 1 since the proof of the validity of the above contrast functions only relies on the property in Equation (3). □

Considering again the argument in the proof, one should easily notice that the above proposition can be generalized to the case of sources which satisfy:

A4. $\mathbf{s} = (s_1, \ldots, s_{3K})$ where: $\forall i \in \{0, \ldots, K-1\}$, s_{3i+1} is BPSK; s_{3i+2} is non Gaussian and satisfies $\mathrm{E}\{s_{3i+2}\} = \mathrm{E}\left\{s_{3i+2}^3\right\} = 0$; $s_{3i+3} = s_{3i+1}s_{3i+2}$; and the random variables $\{s_{3i+1}, s_{3i+2} \ ; \ i = 0, \ldots, K-1\}$ are mutually independent.

In addition, the above result can be generalized to MIMO (multiple input/multiple output) contrast functions as defined in [7] [2]:

Proposition 3. *Let* $\mathbf{y} = \mathbf{G}\mathbf{s}$ *where the vector of sources is defined by A1. Then the function:*

$$\mathbf{G} \mapsto \sum_{i=1}^{N} |\mathrm{Cum}_4\{y_i\}|^\alpha, \ \alpha \geq 1 \tag{11}$$

is a MIMO contrast, that is, its maximization over the group of orthogonal matrices leads to a solution $\mathbf{G}$ *which is a trivial matrix (permutation, scaling).*

Many classical algorithms for BSS or ICA first whiten the data: it is known that when doing so, they constrain matrix $\mathbf{G}$ to be orthogonal. In particular so does the algorithm proposed in [7], which relies on the contrast function in (11). It justifies that this algorithm successfully separates the sources A1. Actually, any algorithm relying on a prewhitening and associated with a contrast function based on the vanishing of the fourth-order cross cumulants (e.g. JADE) is able to separate sources such as A1.

4.2 Pairwise independence is not sufficient

We now consider the pairwise independent sources given by A2 and show that pairwise independence is not sufficient to ensure identifiability of the ICA model. We first have the following preliminary result:

Lemma 3. *Let $y = \mathbf{g}\mathbf{s}$ where the vector of sources is defined by A2. Assume that the vector (s_1, s_2, s_3) takes all 2^3 possible values. If the signal y has values in $\{-1, +1\}$, then $\mathbf{g} = (g_1, g_2, g_3, g_4)$ is either one of solutions below:*

$$\begin{cases} \exists i \in \{1, \dots, 4\} \quad g_i = \pm 1, & \text{and: } \forall j \neq i, g_j = 0 \\ \exists i \in \{1, \dots, 4\} \quad g_i = \pm 1/2, & \text{and: } \forall j \neq i, g_j = -g_i \end{cases} \tag{12}$$

Proof. If $y = \mathbf{g}\mathbf{s}$, using the fact that $s_i^2 = 1$ for $i = 1, \dots, 4$, we have with the particular sources given by A2:

$$y^2 = g_1^2 + g_2^2 + g_3^2 + g_4^2 + 2\Big[\big(g_1 g_2 + g_3 g_4\big) s_1 s_2 + \big(g_1 g_3 + g_2 g_4\big) s_1 s_3 + \big(g_2 g_3 + g_1 g_4\big) s_2 s_3\Big]$$

Since (s_1, s_2, s_3) take all possible values in $\{-1, 1\}^3$, we deduce from $y^2 = 1$ that the following equations necessarily hold:

$$\begin{cases} g_1^2 + g_2^2 + g_3^2 + g_4^2 = 1 \\ g_1 g_2 + g_3 g_4 = g_1 g_3 + g_2 g_4 = g_2 g_3 + g_1 g_4 = 0 \end{cases} \tag{13}$$

First observe that values given in (12) indeed satisfy (13). Yet, if a polynomial system of N equations of degree d in N variables admits a finite number of solutions[3], then there can be at most d^N distinct solutions. Hence, we have found them all in (12), since (12) provides us with 16 solutions for (g_1, g_2, g_3, g_4). $\square$

Using the above result, we are now able to specify the output of classical ICA algorithms when applied to a mixture of sources which satisfy A2.

Constant modulus and contrasts based on fourth order cumulants The constant modulus (CM) criterion is one of the most known criteria for Blind Source Separation. In the real valued case, it simplifies to:

$$J_{\mathrm{CM}}(\mathbf{g}) \triangleq \mathrm{E}\left\{ \left(y^2 - 1\right)^2 \right\} \qquad \text{with: } y = \mathbf{g}\mathbf{s} \tag{14}$$

Proposition 4. *For the sources given by A2, the minimization of the constant modulus criterion with respect to $\mathbf{g}$ leads to either one of the solutions given by Equation* (12).

Proof. We know that the minimum value of the constant modulus criterion is zero and that this value can be reached (for $\mathbf{g}$ having one entry being ± 1 and other entries zero). Moreover, the vanishing of the constant modulus criterion implies that $y^2 - 1 = 0$ almost surely and one can then apply Lemma 3. $\square$

[3] One can show that the number of solutions of (13) is indeed finite.

A connection can now be established with the fourth-order autocumulant if we impose the following constraint:

$$\mathrm{E}\left\{y^2\right\} = 1 \qquad (\text{or equivalently } \|g\| = 1 \text{ since } y = \mathbf{gs}) \tag{15}$$

Because of the scaling ambiguity of Blind Source Separation, the above normalization can be freely imposed. Under (15), we have $\mathrm{Cum}_4\{y\} = \mathrm{E}\left\{\left(y^2 - 1\right)^2\right\} - 2$ and minimizing $J_{\mathrm{CM}}(\mathbf{g})$ thus amounts to maximizing $-\mathrm{Cum}_4\{y\}$. Unfortunately, since $\mathrm{Cum}_4\{y\}$ may be positive or negative, no simple relation between $|\mathrm{Cum}_4\{y\}|$ and $J_{\mathrm{CM}}(\mathbf{g})$ can be deduced from the above equation. However, we have the following proposition:

Proposition 5. *Let $y = \mathbf{gs}$ where the vector of sources is defined by A2. Then, under the constraint* (15) *($\|\mathbf{g}\| = 1$), we have:*

(i) *The maximization of $\mathbf{g} \mapsto -\mathrm{Cum}_4\{y\}$ leads to either one of the solutions given by Equation* (12).
(ii) *$|\mathrm{Cum}_4\{y\}| \leq 2$ and the equality $|\mathrm{Cum}_4\{y\}| = 2$ holds true if and only if $\mathbf{g}$ is one of the solutions given in Equation* (12).

Proof. Part (i) follows from the arguments given above. In addition, using multilinearity of the cumulants and (7), we have for $y = \mathbf{gs}$:

$$\mathrm{Cum}_4\{y\} = -2\left(g_1^4 + g_2^4 + g_3^4 + g_4^4\right) + 24\left(g_1 g_2 g_3 g_4\right) \tag{16}$$

The result then follows straightfowardly from the study of the polynomial function in Equation (16). Indeed, optimizing (16) leads to the following Lagrangian:

$$\mathcal{L} = -2\sum_{i=1}^{4} g_i^4 + 24\prod_{i=1}^{4} g_i - \lambda\left(\sum_{i=1}^{4} g_i^2 - 1\right) \tag{17}$$

After solving the polynomial system which cancels the Jacobian of the above expresssion, one can check that all solutions are such that $|\mathrm{Cum}_4\{y\}| \leq 2$. Details are omitted for reasons of space. Part (ii) of the proposition easily follows. □

Similarly to the previous section, the above proposition can be generalized to MIMO contrast functions. In particular, this explains why, for a particular set of mixing matrices such as that studied in [3], the pairwise maximization algorithm of [7] still succeeded: a separation has luckily been obtained for the considered mixing matrices and initialization point of the algorithm, but it actually would not succeed in separating BPSK dependent sources for general mixing matrices.

Let us stress also that the results in this section are specific to the contrast functions given by (10) or (11). In particular, these results do no apply to algorithms based on other contrast functions such as JADE, contrary to the results in Sections 4.1 and 4.3.

4.3 Complex case

The output given by separation algorithms in case of complex valued signals may differ from the previous results which have been proved for real valued signals only. Indeed, complex valued Blind Source Separation does not always sum up to an obvious generalization of the real valued case [8]. We illustrate it in our context and show that, quite surprisingly, blind separation of the sources given by A3 can be achieved up to classical inderterminations of ICA. This is in contrast with the result in Equation (12) where additionnal indeterminacies appeared. First, we have:

Lemma 4. *Let $y = \mathbf{g}\mathbf{s}$ where the vector of sources is defined by A3. Assume that the vector (s_1, s_2, s_3) takes all 4^3 possible values. If the signal y is such that its values satisfy $|y|^2 = 1$, then $\mathbf{g} = (g_1, g_2, g_3, g_4)$ satisfies:*

$$\exists i \in \{1, \ldots, 4\} \quad |g_i| = 1, \textit{and: } \forall j \neq i, g_j = 0 \tag{18}$$

Proof. If $y = \mathbf{g}\mathbf{s}$, using the fact that $|s_i|^2 = 1$ for $i = 1, \ldots, 4$, we have with the particular sources given by A3:

$$|y|^2 = \sum_{i=1}^{4} |g_i|^2 + \sum_{i \neq j} g_i g_j^* s_i s_j^* \tag{19}$$

Since (s_1, s_2, s_3) take all possible values in $\{1, \imath, -1, -\imath\}^3$, we deduce from $|y|^2 = 1$ that the following equations necessarily hold:

$$\begin{cases} |g_1|^2 + |g_2|^2 + |g_3|^2 + |g_4|^2 = 1 \\ g_1 g_2^* = g_1 g_3^* = g_1 g_4^* = g_2 g_3^* = g_2 g_4 * = g_3 g_4^* = 0 \end{cases} \tag{20}$$

Solving for the polynomial system in the variables $|g_1|$, $|g_2|$, $|g_3|$ and $|g_4|$, we obtain that the solutions are the ones given in Equation (18). □

Constant modulus and fourth-order cumulant based contrasts In contrast with Propositions 4 and 5 we have the following result:

Proposition 6. *Let $y = \mathbf{g}\mathbf{s}$ where the sources satisfy A3. Then, the functions:*

$$\mathbf{g} \mapsto -\mathrm{E}\left\{ \left| |y|^2 - 1 \right|^2 \right\} \qquad \textit{and:} \tag{21}$$

$$\mathbf{g} \mapsto |\mathrm{Cum}_{2,2}\{y\}| \textit{ under constraint } \mathrm{E}\left\{|y|^2\right\} = 1 \tag{22}$$

are contrast functions, that is, their maximization leads to $\mathbf{g}$ *satisfying* (18).

Proof. The validity of the first contrast function is obtained with the same arguments as in the proof of Proposition 4: we have $|y|^2 \overset{m.s.}{=} 0$, which yields (20) via (19). In the case of independent sources, the proof of the validity of the second contrast involves only cumulants with equal number of conjugate and non conjugate variables: invoking Lemma 2, one can see that the same proof still holds here. □

Note that the same arguments can be applied to ICA methods such as the pairwise algorithm in [2] or JADE [9]. Figure 1 illustrates our result.

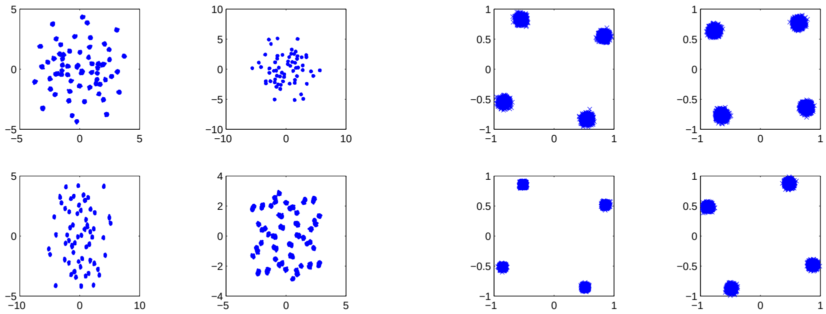

Fig. 1. Typical observed separation result of the sources A3 with the algorithm JADE (left: sensors, right: separation result)

References

1. Li, T.H.: Finite-alphabet information and multivariate blind deconvolution and identification of linear systems. IEEE Trans. on Information Theory **49**(1) (January 2003) 330–337
2. Comon, P.: Contrasts, independent component analysis, and blind deconvolution. Int. Journal Adapt. Control Sig. Proc. **18**(3) (April 2004) 225–243 special issue on Signal Separation: http:// www3.interscience.wiley.com/cgi-bin/jhome/4508. Preprint: I3S Research Report RR-2003-06.
3. Comon, P.: Blind identification and source separation in 2×3 under-determined mixtures. IEEE Trans. Signal Processing **52**(1) (January 2004) 11–22
4. Comon, P., Grellier, O.: Non-linear inversion of underdetermined mixtures. In: Proc. of ICA'99, Aussois, France (January 1999) 461–465
5. Hyvärinen, A., Shimizu, S.: A quasi-stochastic gradient algorithm for variance-dependent component analysis. In: Proc. International Conference on Artificial Neural Networks (ICANN2006), Athens, Greece (2006) 211–220
6. Cardoso, J.F.: Multidimensional independent component analysis. In: Proc. ICASSP '98. Seattle. (1998)
7. Comon, P.: Independent component analysis, a new concept. Signal Processing **36**(3) (April 1994) 287–314
8. Eriksson, J., Koivunen, V.: Complex random vectors and ICA models: Identifiability, uniqueness and separability. IEEE Trans. on Information Theory **52**(3) (March 2006) 1017–1029
9. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non gaussian signals. IEE Proc. F **140**(6) (1993) 362–370