A NEW CUMULANT-BASED CRITERION FOR ITERATIVE BLIND SOURCE SEPARATION

Rubén Martín-Clemente, José I. Acha
Área de Teoría de la Señal y Comunicaciones, University of Sevilla
Avda. de los Descubrimientos s/n., 41092-Sevilla, Spain
e-mail: ruben@us.es, acha@viento.us.es

ABSTRACT

This contribution introduces a new iterative algorithm for Blind Source Separation based on the cancellation of some second order derivatives of the output cross-cumulants. It is proven that the estimation of the separating matrix can be achieved by solving linear equations repeatedly. A key feature is thus that the algorithm requires a low computational effort. Simulations show the effectiveness of the approach.

1 INTRODUCTION

In Blind Source Separation (BSS), a vector of recorded signals $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ is assumed to be produced from a vector of unknown sources $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ using the mixture model:

$$\mathbf{x}(t) = A\,\mathbf{s}(t) \quad (1)$$

where $A = (a_{ij})$ is an unknown $N \times N$ matrix, called the mixing matrix. Our aim is to find a separating $N \times N$ matrix $B = (b_{ij})$ so that

$$\mathbf{y}(t) = B\,\mathbf{x}(t) = G\,\mathbf{s}(t) \quad (2)$$

is an estimate of the source signals – possibly up to permutations and scaling –. Source separation is thus achieved when matrix $G \overset{def}{=} B\,A = (g_{ij})$ has one and only one non-zero coefficient per row and column. For future use, note that such a matrix G verifies the simple property

($P.1$) Let i, j, n be indices from 1 to N, if $i \neq j$ then, for all n, one of the coefficients g_{in}, g_{jn} is zero at separation.

The following assumptions hold throughout: ($A.1$) Matrix A is invertible, ($A.2$) the sources are zero-mean, unit-variance, stationary random processes, ($A.3$) the random variables $s_1(t), \ldots, s_N(t)$ *are statistically independent* at each time t, ($A.4$) there is *at most one* gaussian distributed source. In practice, assumption ($A.3$) is plausible when the sources have different origins. Asumption ($A.4$) ensures the identifiability of the mixing matrix [1, 3].

There are several methods for solving the BSS problem. On the one hand, Information-Theoretic models have usually attracted a great deal of attention. Nevertheless, one must estimate the marginal probability density functions of the sources in order to use these criteria. On the other hand, cumulant-based methods are "universal" since they do not require any a priori knowledge of the source distributions. In addition, higher-order cumulants of additive Gaussian noise are null and thus cumulant-based methods are also blind to the noise structure. In practice, though, cumulant-based methods are usually restricted to the separation of small number of sources due to computational reasons. One possible solution is to use a reduced set of cumulants in order to reduce the computational cost.

In this paper, we present a new set of necessary and sufficient cumulant-based conditions for BSS. The separating matrix is estimated by solving linear equations repeatedly, leading to a computationally simple algorithm. The paper is organized as follows: Section 2 explains the separation principles and algorithms. Section 3 compares our approach with other existing methods. Section 4 presents experiments and Section 5 is devoted to conclusions.

2 CUMULANT-BASED EQUATIONS FOR BSS

2.1 What role do second-order statistics play in BSS?

One observation noted by many contributors[2, 3] is that solving the BSS problem is simpler when $\mathbf{x}(t)$ has uncorrelated unit-variance components. Uncorrelatedness means that

$$E\{\mathbf{x}(t)\,\mathbf{x}^T(t)\} = I \quad (3)$$

where I is the identity matrix. Since $E\{\mathbf{x}(t)\,\mathbf{x}^T(t)\} = A\,A^T$, we then have $A\,A^T = I$. Thus, the mixing matrix A is orthogonal and, consequently, matrices B and G must be orthogonal as well. This is convenient because an $N \times N$ orthogonal matrix only contains $N(N-1)/2$ independent parameters.

The point is that one can always linearly transform the observed data so that $\mathbf{x}(t)$ has uncorrelated components. This is usually called 'whitening' or 'sphering'. Whitening is done via the eigenvalue decomposition of the spatial covariance matrix of the observed data (see *Appendix*).

2.2 Differentiating the Output Cumulants

Second-order information, however, is not sufficient for blind separation of the sources [1]. In order to explote the crucial independence assumption (A.3), higher-order information must be used. This is the point of this Section. From now on, note that we will drop time index t for convenience of presentation.

Let $c_{31}(y_i, y_j)$ be the fourth-order cumulant $cum(y_i, y_i, y_i, y_j)$. This cumulant can be expanded as follows:

$$c_{31}(y_i, y_j) = \sum_{n=1}^{N} g_{in}^3\, g_{jn} \kappa_n \tag{4}$$

where κ_n is the kurtosis of the n-th source. Invoking property (P.1) of matrix G, it is a simple matter to check that the true separating matrices are roots of the equations $c_{31}(y_i, y_j) = 0$ $(i \neq j)$. In practice, though, these equations have two drawbacks: *i.)* spurious roots are possible, depending on the source statistics and *ii.)* the equations depend on as-high-as-fourth order powers of the coefficients of B through the relation $G = B\,A$.

In the following, we discuss a simple idea that circumvents both disadvantages: the key is that it is unnecessary to have a cubic term g_{in}^3 in (4); rather, in view of the above discussion and (P.1), this high power may be considered even as a nuisance and should be reduced. Our solution is to differentiate the cumulant: let

$$\Delta_{ij} \triangleq \frac{1}{6} \frac{\partial^2 c_{31}(y_i, y_j)}{\partial b_{ij}^2} = \sum_{n=1}^{N} g_{in}\, g_{jn}\, a_{jn}^2\, \kappa_n, \quad (i \neq j) \tag{5}$$

Using $cum(x_l, x_m, x_j, x_j) = \sum_p a_{lp} a_{mp} a_{jp}^2 \kappa_p$ (which is a consequence of the linearity property of the cumulants), it is possible to express (5) as:

$$\Delta_{ij} = \sum_{n=1}^{N} \sum_{m=1}^{N} b_{in} b_{jm} cum(x_n, x_m, x_j, x_j), \tag{6}$$

Then, by assuming that the signals in $\mathbf{x}(t)$ are uncorrelated and have unit-variance, we can state that the set of equations $\Delta_{ij} = 0$ for all $i \neq j$ provide us with *necessary* and *sufficient* conditions for source separation.

Proof.–

First, observe that (4) can be written compactly in a matrix form:

$$\Delta_{ij} = \mathbf{g}_i^T \Gamma_j \mathbf{g}_j \tag{7}$$

where $\mathbf{g}_k^T$ is the k-th row of G and Γ_j is the diagonal matrix whose (l,l)-entry is equal to $a_{jl}^2 \kappa_l$. If at most one source is gaussian (i.e., its kurtosis equals zero), we can also assume without loss of generality that each diagonal element of Γ_j is different (i.e. Γ_j has no repeated eigenvalues) since A (and thus Γ_j) can be properly adjusted by multipliying the observations $\mathbf{x}(t)$ by any orthogonal matrix.

Since both A and B are orthogonal matrices, it follows that $\mathbf{g}_i^T \mathbf{g}_j = \delta_{ij}$, where δ_{ij} stands for the Kronecker delta. Now, we prove our main result:

(*Necessity*). If B is a separating matrix, then each vector $\mathbf{g}_i$ is a different canonical vector. Therefore, it follows from (7) that $\Delta_{ij} = 0$ for all i, j $(i \neq j)$.

(*Sufficiency*). Let us assume that $\Delta_{ij} = 0$ for all $i \neq j$. The vector $\Gamma_j \mathbf{g}_j$ can be expressed as a linear combination of the vectors $\mathbf{g}_i$ since the set $\{\ \mathbf{g}_1,\ \mathbf{g}_2\ ,\ \ldots\ ,\ \mathbf{g}_n\ \}$ forms an orthonormal basis of the space. Therefore:

$$\Gamma_j \mathbf{g}_j = \Delta_{jj} \mathbf{g}_j + \sum_{i \neq j} \Delta_{ij} \mathbf{g}_i \tag{8}$$

where $\Delta_{ij} = \mathbf{g}_i^T \Gamma_j \mathbf{g}_j$ as was defined before and $\Delta_{jj} = \mathbf{g}_j^T \Gamma_j \mathbf{g}_j$. Since, $\Delta_{ij} = 0$ for all $i \neq j$, we obtain that $\Gamma_j \mathbf{g}_j = \Delta_{jj} \mathbf{g}_j$, which implies that each $\mathbf{g}_i$ is an eigenvector of a diagonal matrix, i.e., a canonical vector. As a consequence, B is a separating matrix.

2.3 Solving the Equations

2.3.1 Blind Separation of Two Sources

In the two-source case, a very simple solution can be obtained: by assuming a preliminary whitening of the observations, it is easy to show that the mixing matrix is of the form

$$A = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \tag{9}$$

The problem is thus reduced to determining the rotation angle α. Combining Δ_{12} and Δ_{21}, as defined in (6), one obtains the second-degree equation:

$$c_{31}^x + \{c_{22}^x - c_{40}^x\}\epsilon - c_{31}^x \epsilon^2 = 0 \tag{10}$$

where $c_{pq}^x = cum(\underbrace{x_1, \ldots, x_1}_{p\,times}, \underbrace{x_2, \ldots, x_2}_{q\,times})$.

Solving (10) gives the pair of solutions

$$\epsilon_1 = \tan(\alpha) \quad \epsilon_2 = -\cot(\alpha) \tag{11}$$

Finally, we can contruct two separating matrices as follows:

$$B = \frac{1}{\sqrt{1 + \epsilon_p^2}} \begin{bmatrix} 1 & \epsilon_p \\ -\epsilon_p & 1 \end{bmatrix} \tag{12}$$

where the value of p can be either 1 or 2.

2.3.2 *More than Two Sources*

In the general N-source case, the equations $\Delta_{ij} = 0$ $(i \neq j)$, are best solved by means of the following fixed-point-like algorithm[8]: let

$$P_i \stackrel{def}{=} E\{\mathbf{x}\,\mathbf{x}^T\, x_i^2\} - (1 + \lambda\, \kappa_i^x)\, I \qquad (13)$$

where $\mathbf{x}$ is the observation vector with unit-variance uncorrelated components, κ_i^x is the kurtosis of the i-th observation x_i and λ is an adjustable parameter whose optimal values will be given below. In addition, let $\mathbf{e}_i$ be the i-th canonical vector, *i.e.*, the i-th column of the identity matrix. The proposed algorithm proceeds as follows:

(Step 1) Whiten the observations (see *Appendix*).

(Step 2) For $i = 1$ to N do

1.1 Solve the system of linear equations $P_i\, \mathbf{v}_i = \mathbf{e}_i$.

1.2 Let $\mathbf{b}_i = \mathbf{v}_i / ||\mathbf{v}_i||_2$ be the i-th row of the separating matrix B.

(Step 3) Orthogonalize the rows of B.

(Step 4) Set $\mathbf{y} = B\mathbf{x}$. If $\mathbf{y} = \mathbf{x}$, stop. Otherwise, replace $\mathbf{x}$ with $\mathbf{y}$ and go back to Step 1.

Some comments are in order:

- First of all, Step 2 projects B after every computation on the space of orthogonal matrices (*i.e.*, the Stiefel manifold). In principle, one could take any method for orthogonalizing a set of vectors and apply it on the columns of B (for example, use the well-known Gram-Schmidt algorithm[4]).
- Secondly, in our experiments, the matrices P_i are calculated by the linearity properties of cumulants[10] from the matrices used in the previous iteration. This has the effect of a considerable reduction of the computational burden.
- Finally, stability analysis[8] suggests the choice $\lambda \geq 1$. Simulations show that $\lambda \in [1, 1.5]$ is a good tradeoff between convergence speed and stability.

3 LINKS OF THE COST FUNCTION WITH OTHER METHODS

Note that (6) can be written in a matrix form as:

$$\Delta_{ij} = \mathbf{b}_i^T\, C_j\, \mathbf{b}_j \qquad (14)$$

where $\mathbf{b}_k = [b_{k1}, \ldots, b_{kN}]^T$ and C_j is the $N \times N$ matrix whose (n, m)-entry equals $cum(x_n, x_m, x_j, x_j)$. Hence, the same argument that led to eqn. (8) now gives the identity:

$$C_j\, \mathbf{b}_j = \Delta_{jj}\mathbf{b}_j + \sum_{i \neq j} \Delta_{ij}\mathbf{b}_i = \Delta_{jj}\mathbf{b}_j \qquad (15)$$

where $\Delta_{jj} = \mathbf{b}_j^T C_j \mathbf{b}_j$, which means that $\mathbf{b}_j$ is an *eigenvector* of the cumulant matrix C_j. Interestingly, following a different approach, Cardoso and Souloumiac minimize the sum of squares of non-diagonal entries of the matrices $B^T C_j B$ in JADE [2]. Clearly, (14) comes from a *sufficient* subset of those equations which are solved by JADE. Consequently: *i.)* a significant computational cost saving is expected and *ii.)* since JADE satisfies more statistical equations, it may also be numerically more robust. The experiments are in accordance with these intuitions.

4 COMPUTER SIMULATIONS

We present several simulation results to show the performance of the algorithm.

Experiment 1. A "toy" experiment in which three simple sources are perfectly separated. It is useful in order to visualize the results. See Figures 1 and 2.

Experiment 2. In order to compare with existing algorithms, we used a mixture of ten sources: five of them were uniform variables (sub-gaussian) and the other five were generated as the cube of normal variables (super-gaussian) –5000 samples each–. The mixing matrix was randomly chosen in each experiment. We compare our algorithm with standard implementations of JADE[2] and FastIca[5], obtained from their authors via Internet [6, 7]. The results of twenty independent experiments were averaged. After the separation, the *mean signal to noise ratio* obtained by our algorithm was equal to 14.40 dB, whereas JADE obtained 15.78 dB and FastIca obtained 14.42 dB. On the other hand, the number of floating point operations required by JADE (Fastica) was 5.28 (1.92) times the number of operations required by our algorithm.

Experiment 3. *Thirty* independent sources were mixed by a randomly chosen mixing matrix and then separated by our algorithm, which used 5000 samples of each observation. The parameter λ was taken equal to 1.5. The experiment was repeated twenty times and the averaged Moreau and Macchi's Index [9], which is defined as:

$$p = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{|g_{ij}|^2}{\max_k |g_{ik}|^2} - 1 + \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{|g_{ij}|^2}{\max_k |g_{kj}|^2} - 1 \qquad (16)$$

is presented in Figure 3. The lower the value of the index, the better the performance.

Classical cumulant-based algorithms *are not* usually able to handle a so great number of sources in practice.

5 CONCLUSIONS

We have presented a new cumulant-based algorithm for BSS. The benefits of our approach are: *i.)* Simple im-

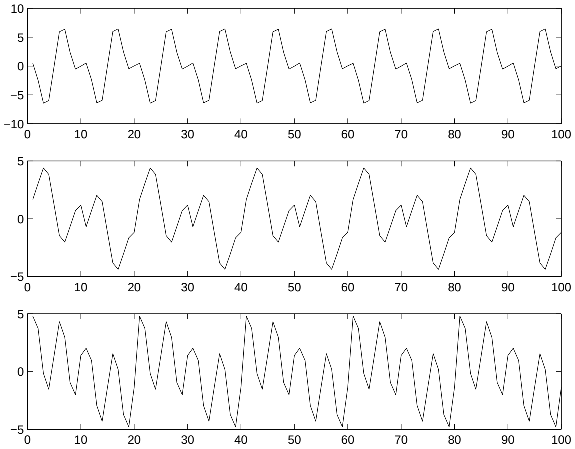

Figure 1: Original (correlated) observations.

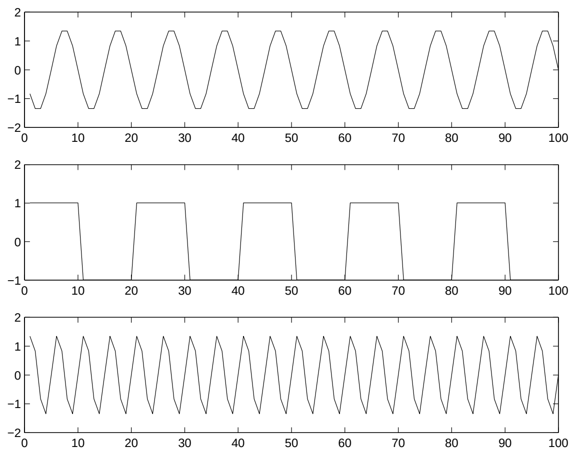

Figure 2: Estimated Sources.

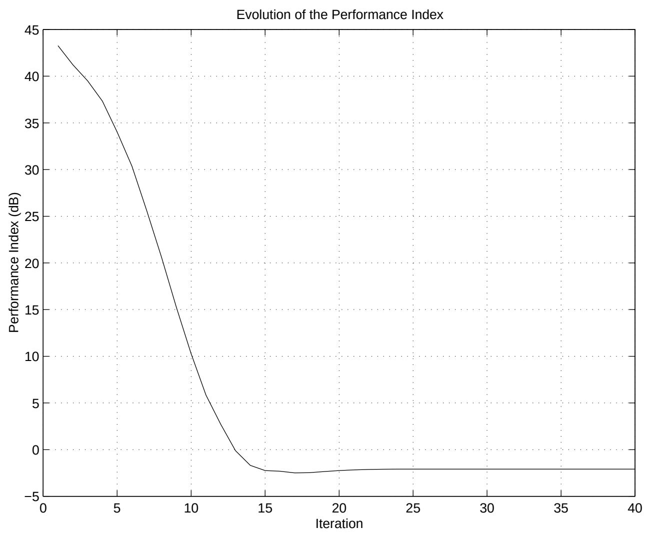

Figure 3: Evolution of the Separation Index in Experiment 3.

plementation, *ii.*) Fast Convergence. Simulations show that it can handle a great number of sources in practice. Due to the lack of space, calculations were ommited and will be addressed in a forthcoming paper [8].

APPENDIX (WHITENING THE OBSERVATIONS)

Let the spatial covariance matrix of the measurements be $R = E[\mathbf{x}\,\mathbf{x}^T]$. It can be expressed in terms of its own eigenvalue decomposition as $R = V\,D\,V^T$, where matrix V is formed by assembling the eigenvectors of R and D is a diagonal matrix whose main diagonal contains the eigenvalues of R. As the eigenvectors form an orthonormal basis, it follows that $V^T\,V = I$, the identity matrix. Then, let the whitening matrix be $W = D^{-1/2}\,V^T$: it is straightforward to check that replacing $\mathbf{x}$ with $W\,\mathbf{x}$, we obtain a vector with uncorrelated unit-variance components, as desired.

References

[1] X.-R Cao and R.-W.Liu, " General approach to blind source separation " *IEEE Trans. on Signal Proc.*, Vol.44, pp.562-571, 1996

[2] J-F. Cardoso and A. Souloumiac, " Blind Beamforming for non-Gaussian Signals ", *Proceedings of the Inst. Elect. Eng.*, Vol.140 (F6), pp.362-370, 1993.

[3] P. Comon, " Independent Component Analysis - A New Concept ? ", *Signal Processing* , Vol. 36, No. 3, pp. 287-314, 1994.

[4] G. Golub and C. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1996.

[5] A. Hyvarinen, E. Oja, " A Fast Fixed-Point Algorithm for Independent Component Analysis", *Neural Computation.*, Vol. 6, pp. 1483-1492, 1997.

[6] http://sig.enst.fr:80/~cardoso/stuff.html

[7] http://www.cis.hut.fi:80/~aapo

[8] R. Martín-Clemente and J.I.Acha, " Equations for the Blind Separation of Sources ", submitted to *Signal Processing* (*Elsevier*).

[9] E. Moreau and O. Macchi, " High-Order Contrasts for Self Adaptive Source Separation", *Internat. J. Adaptive Control Signal Process.*, Vol.10, No. 1, pp.19-46, 1996

[10] C. Nikias and A. Petropulu *Higher-Order Spectra Analysis*, Prentice-Hall, 1993.