Fourth-Order Cumulant Based Blind Source Separation

A. K. Nandi and V. Zarzoso

Department of Electronic and Electrical Engineering, University of Strathclyde,
204 George Street, Glasgow G1 1XW, U.K.

Abstract

Several methods based on higher-order statistics are currently used to solve the Blind Source Separation problem. In the present paper an approach from Mansour and Jutten is considered, and simplified via the application of the Singular Value Decomposition, which decorrelates the observed signals and leads to a straightforward solution based on some results developed by Comon.

1 Introduction

The problem of Blind Source Separation (BSS) consists in the reconstruction of a set of unknown source signals from the observation of another set of linear mixtures of the sources, where the mixture pattern is also unknown. The only assumptions are that the sources must be statistically independent (but not necessarily identically distributed) and zero-mean. More general identifiability conditions are established in [6]. Several approaches have already been taken to solve this problem and most of them resort to higher-order statistics (HOS). Cardoso, in [4], employs higher-order moments, but his method does not deal with identically distributed signals. The separation is improved using higher-order cumulants (HOC), as Comon carries out in [2], [3] and [5]. In the former two, a closed formula for two-source case is developed, which gives a direct solution when the observed signals come from an orthogonal transformation applied on a normalized set of sources such that they have the same power. In [5] the concept of Independent Component Analysis (ICA) is mathematically introduced and used in the BSS problem. The separation is achieved via the maximization of a contrast function formed by pairwise 4th-order marginal cumulants. Finally, in [1] Mansour and Jutten use a 4th-order cross-cumulant as a cost function for the case of two source and observed signals. This function is minimized and, when the resultant outputs are also constrained to be uncorrelated, the spurious solutions are eliminated. This last method, although looks simple, has some important drawbacks. In the present paper these drawbacks are highlighted and a modification, based on the Singular Value Decomposition (SVD) and results from [2] and [3], is found to overcome them, making the method much more robust and computationally simple.

IEEE Signal Processing Letters, Vol. 3, No. 12, December 1996, pp. 312-314.

2 Mansour-Jutten (MJ) approach

2.1 Model equations

Consider the noiseless BSS problem, which in the case of two source and two observed signals becomes:

$$E = M \cdot X; \qquad \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \tag{1}$$

where $\{\, e_i(t),\ i = 1,\ 2\,\}$ are the instantaneous linear mixtures (only data available), M is the mixture matrix (non-singular) and $\{\, x_i(t),\ i = 1,\ 2\,\}$ the source signals, assumed zero-mean and statistically independent. The separation is attained by estimating a 2×2 matrix W such that $G = W \cdot M = P \cdot D$, where P is a permutation matrix and D a diagonal matrix. If we call $\{\, s_i(t),\ i = 1,\ 2\,\}$ the outputs of the W matrix then

$$\begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \cdot \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \tag{2}$$

Now the statistical parameters of $s_1(t)$ and $s_2(t)$ can easily be written as a function of the elements of G and the statistical properties of the sources. In particular, if P_i represents the power and β_i the kurtosis (4th-order cumulant) of the source signals:

$$\mathrm{E}[s_1 \cdot s_2] = g_{11}\, g_{21}\, P_1 + g_{12}\, g_{22}\, P_2 \tag{3}$$

$$\mathrm{Cum}_{22}(s_1,\, s_2) = g_{11}^2\, g_{21}^2\, \beta_1 + g_{12}^2\, g_{22}^2\, \beta_2. \tag{4}$$

2.2 Solutions

In [1] the cost function $J = \mathrm{Cum}_{22}(s_1,\, s_2)$ is considered. From (4) the following set of solutions is found for $J = 0$:

a) If the source signals have the same sign of kurtosis ($\mathrm{sign}(\beta_1) = \mathrm{sign}(\beta_2)$): $g_{11} = g_{22} = 0$, $g_{12} = g_{21} = 0$, $g_{11} = g_{12} = 0$, and $g_{21} = g_{22} = 0$. The first two solutions achieve the separation, i.e., yield a matrix G with either a diagonal form or its permutation. The last two solutions are *trivial* and it is easy to get rid of them by imposing $[\, w_{11},\, w_{12}\,] \neq [\,0,\, 0\,]$ to avoid the penultimate one and $[\, w_{21},\, w_{22}\,] \neq [\,0,\, 0\,]$ to avoid the last one. Thus, it is sufficient to constrain $w_{11} \neq 0$, $w_{22} \neq 0$.

b) If the source signals have different sign of kurtosis ($\mathrm{sign}(\beta_1) \neq \mathrm{sign}(\beta_2)$): the solutions above still exist, but also *spurious* solutions in the form $g_{12} = p \cdot g_{11}$ and $g_{22} = q \cdot g_{21}$, with $(p \cdot q)^2 = \left|\frac{\beta_1}{\beta_2}\right|$. If the output signals are forced to be uncorrelated then, from (3), the first four solutions also satisfy this additional constraint. However, now the spurious solutions may only appear if $\left(\frac{P_1}{P_2}\right)^2 = \left|\frac{\beta_1}{\beta_2}\right|$, $\mathrm{sign}(\beta_1) \neq \mathrm{sign}(\beta_2)$, what is very unlikely to happen.

In short, by solving the system

$$\begin{cases} \mathrm{E}[s_1 \cdot s_2] = 0 \\ \mathrm{Cum}_{22}(s_1,\, s_2) = 0 \end{cases} \tag{5}$$

with $w_{11} \neq 0$, $w_{22} \neq 0$, the solutions found for the rest of the elements of W are such that the separation of the sources is achieved, i.e., the source signals appear in S up to, perhaps, a scale factor and a change in order.

2.3 Drawbacks

For the sake of simplicity, let us assume that $w_{11} = w_{22} = 1$, as done in [1]. Writing (5) as a function of the unknown variables $(w_{12},\, w_{21})$ and the statistical properties of the known data signals, we get:

$$\begin{cases} w_{21}\,\mu_{20} + w_{12}\,\mu_{02} + (1 + w_{12}w_{21})\,\mu_{11} = 0 \\ (1 + 4\,w_{12}\,w_{21} + w_{12}^2\,w_{21}^2)\,k_{22} + 2\,(1 + w_{12}\,w_{21})\,(w_{21}\,k_{31} + w_{12}\,k_{13}) + w_{21}^2\,k_{40} + w_{12}^2\,k_{04} = 0 \end{cases} \tag{6}$$

where $\mu_{ij} = E\,[e_1^i \cdot e_2^j]$ represents the moment of order $(i+j)$ of $e_1(t)$ and $e_2(t)$, and $k_{ij} = \mathrm{Cum}_{ij}(e_1,\, e_2)$ stands for the $(i+j)$th-order cumulant of the same pair of signals. To solve this system, a numerical method (see [7] for some of them) has to be used in the most general case. Some of them have been tested in different contexts. At the end, a method based on the minimization of the sum of squares of the functions above was chosen, as it gave the best results. However, two main difficulties crop up in this direct implementation:

- *Computational load*: one has to solve a system like (6) every time two signals are to be separated.
- *Errors in the solutions*: sometimes the numerical algorithm employed comes to rest in a non-desired local minimum which does not give the source signals properly separated. As a result, a more complex method to compute absolute minima (such as "simulated annealing", for instance) is needed in order to attain a good performance of this method.

Therefore, some kind of improvement is wanted.

3 Modification

3.1 Application of the SVD to the problem

Let us compute the SVD of the data signals: $E = U \cdot \Sigma \cdot V^T$, where U and V are the left and right singular vectors of E, respectively, and Σ is a diagonal matrix formed by the singular values of E. Now, let E' be equal to V^T. The MJ method will be applied to this last matrix: $E' = V^T = (U \cdot \Sigma)^{-1} \cdot E = (U \cdot \Sigma)^{-1} \cdot M \cdot X$.

The source signals are statistically independent and, in particular, uncorrelated, so $cov\{X\} = C_X^2 = diag(P_1,\ P_2)$. Let X' be a normalized version of X such that $cov\{X'\} = I,$ i.e., $X = C_X \cdot X'$. It is obvious that X' is an equivalent representation of the source signals, since it contains in its rows scaled versions of them (unit power). Then, the matrix E' can be written as:

$$E' = (U \cdot \Sigma)^{-1} \cdot M \cdot C_X \cdot X' = M' \cdot X' \tag{7}$$

which is an equivalent problem to that we had at the beginning. As $E' \cdot E'^T = I$ (because $V^T \cdot V = I$) and $cov\{X'\} = I,\ M'$ must necessarily be an orthogonal matrix. Now we are looking for a matrix W such that $S = W \cdot E'$ contains the source signals. The system that has to be solved is given by (6). If W was orthogonal, as the signals in E' are uncorrelated and have the same power, the signals in S would always be uncorrelated and the first equation of the system would not need to be solved any more. Hence, let us assume that W takes the form of a Givens rotation matrix:

$$W = c \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix}, \qquad c = \frac{1}{\sqrt{1+\theta^2}}. \tag{8}$$

In that case, the second equation of the system becomes a simple polynomial in θ equated to zero:

$$k_{22}^{s}/c^4 = k_{22}^{e'}\,\theta^4 + 2\,(k_{31}^{e'} - k_{13}^{e'})\,\theta^3 + (k_{40}^{e'} - 4\,k_{22}^{e'} + k_{04}^{e'})\,\theta^2 + 2\,(k_{13}^{e'} - k_{31}^{e'})\,\theta + k_{22}^{e'} = 0 \tag{9}$$

where $k_{ij}^{s} = \mathrm{Cum}_{ij}(s_1,\ s_2)$ and $k_{ij}^{e'} = \mathrm{Cum}_{ij}(e'_1,\ e'_2)$. Extracting the roots of this 4th-order polynomial the separation is obtained. But what guarantees that there exist real solutions of (9), i.e., that the solution can be derived from a matrix like (8)?

3.2 Solution for orthogonal transformation and normalized input signals

The answer to the above question is given by Comon, for instance in [3]. When M' is orthogonal it is proved there that the solutions of (9) are given by two double roots θ_0 and $-1/\theta_0$, with $\theta_0 \in]-1,\,1]$:

$$\theta_0 = -\frac{\rho}{2} + \mathrm{sign}(\rho)\sqrt{\frac{\rho^2}{4}+1}, \qquad \rho = \frac{k_{31}^{e'} - k_{13}^{e'}}{k_{22}^{e'}}. \tag{10}$$

These solutions also make the rest of the 4th-order cross-cumulants of the outputs signals vanish. In addition, if the signals in E' have the same power and are uncorrelated, the transformation (8) yields uncorrelated outputs, as desired. Note that in [3] the first condition is not explicitly stated (see section 2.3 in that paper) but it is absolutely necessary to get uncorrelated outputs with this type of transformation. This condition can also be extended to the signals contained in X', due to the orthogonality of M'. In short, the comprehensive search for the solution of (6), as done in the MJ approach above, need not be carried out any longer.

4 Simulation results

For the direct implementation of the MJ method the Simplex algorithm [7] is employed to look for the minima of the sum of squares of the functions of (6). For the proposed modification, after computing the SVD of the input signals matrix, the solutions are directly extracted from (10). To compare the performance of the methods, the parameter P_k measures the ratio of the sum of the absolute values of the 4th-order marginal cumulants to the sum of the absolute values of all the 4th-order cumulants of the obtained signals, and the parameter P_e, defined in [5], is a measurement of the distance between G and the identity matrix I, modulo a multiplicative factor of the form $P \cdot D$, where P is a permutation and D a diagonal invertible matrix. In a good separation: $P_k \approx 1, \;\; P_e \approx 0$. The parameters $Time$ and $Flops$ do not take into account the computation of the cumulants which is necessary in both cases, but only the main body of each algorithm. In the first simulation two white uniformly distributed signals are used as sources, both with zero mean and unit power. The mixture matrix M is formed by uniformly distributed elements in the range $[-10, \; 10]$. The experiment is repeated 100 times with different signals and mixture matrix each time. From these simulations both the mean value and the standard deviation (figures between parentheses in the table) are calculated for each parameter and each method. The results obtained are summarized in *Table 1*. Although both methods yield similar parameter P_k, this is not the case in the rest of the parameters, for which the modification produces much better results. In order to consider the case of source signals with different sign of kurtosis, the experiment is repeated with a white zero-mean unit-power exponentially distributed signal employed as a second source. The results also appear in *Table 1*. The same comment as before is applicable here. However, it is worth pointing out that in this situation 3 solutions from the direct implementation of the MJ method did not give a proper G matrix, and therefore an acceptable separation, while the proposed modification offered good separation in all cases.

5 Conclusion

It has been seen that by applying the SVD to the input matrix we can get rid of the exhaustive search which is necessary in the direct implementation of the MJ method. Moreover, this modification turns out to be more robust, more accurate and much faster than the former. On the other hand, the SVD could be considered as a preprocessing to decorrelate the observed signals, which simplifies further operations to achieve higher-order independence. In fact, once the decorrelation is attained the separation is given directly by Comon in [2] and [3]. Therefore, although the methods developed by Mansour-Jutten and by Comon seem to be different approaches to solve the BSS problem, when a decorrelation is applied to the former it degenerates into the latter.

	Uniform - uniform		Uniform - exponential	
Parameter	MJ Method	Modification	MJ Method	Modification
P_k	0.9813 (0.0208)	0.9689 (0.0604)	0.9381 (0.0690)	0.9659 (0.0703)
P_e	0.1974 (0.7374)	0.0189 (0.0807)	0.4680 (0.8672)	0.0219 (0.0768)
$Time$ (s)	4.0799 (0.9202)	0.0083 (0.0059)	6.4606 (3.5701)	0.0092 (0.0180)
$Flops$	7405 (1478)	27 (0)	11865 (6185)	27 (0)

Table 1: Results obtained from the simulations. On the left-hand side, both sources are white zero-mean uniformly distributed signals. On the right-hand side, the second source is replaced by a white zero-mean process with exponential distribution.

6 References

1. A. Mansour and C. Jutten, "Fourth-Order Criteria for Blind Sources Separation", *IEEE Transactions on Signal Processing*, Vol. 43, No. 8, August 1995, pp. 2022-2025.

2. P. Comon, "Separation of Stochastic Processes", *Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 1989, pp. 174-179.

3. P. Comon, "Separation of Sources using Higher-Order Cumulants", *SPIE Vol. 1152 Advanced Algorithms and Architectures for Signal Processing IV*, 1989, pp. 170-181.

4. J.-F. Cardoso, "Source Separation using Higher-Order Moments", *Proceedings ICASSP*, Glasgow, Scotland, May 1989, pp. 2109-2112.

5. P. Comon, "Independent Component Analysis, A new concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 287-314.

6. L. Tong, S. Yu, Y. Inouye and R. Liu, "A Necessary and Sufficient Condition of Blind Identification", *International Signal Processing Workshop on Higher-Order Statistics*, Chamrousse, France, July 1991, pp. 261-264.

7. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, 1988.