THE POTENTIAL OF DECORRELATION IN BLIND SEPARATION OF SOURCES BASED ON CUMULANTS†

Vicente ZARZOSO and Asoke K. NANDI

Signal Processing Division, Department of Electronic and Electrical Engineering
University of Strathclyde, 204 George Street, Glasgow G1 1XW, U.K.
*Phone: +44-141-552 4400 * Fax: +44-141-552 2487 * E-mail: vicente@spd.eee.strath.ac.uk*

ABSTRACT

Several methods based on higher-order statistics are currently used to solve the problem of blind separation of sources. The present contribution considers an approach from Mansour and Jutten for the case of two noiseless observations of two sources. This method is simplified via the application of the Singular Value Decomposition, which decorrelates the observed signals and makes higher-order independence very simple to obtain, from some results developed by Comon. The connection discovered between the two procedures, in turn, provides an interesting interpretation of this last method: the decorrelation process transforms the original model into a new one made up of equivalent sources which are much easier to extract.

1 INTRODUCTION

The problem addressed here is the *Blind Source Separation (BSS)*. It consists in the reconstruction of a set of mutually independent signals — so-called *sources* — from the sole observation of another set of instantaneous linear mixtures. Both the source signals and the transformation from them to the observations are unknown, and hence the term *blind*. Only two assumptions are usually made: the independence of the sources and the linearity of the transformation. Many situations may be modelled from the BSS standpoint. For instance, the extraction of signals impinging on an array of sensors. Other fields of interest include direct estimation of arrival (DOA), blind deconvolution, identification of multichannel ARMA processes when the input is not observed, electrocardiogram extraction, etc.

Mainly due to its wide range of applications, this area has been extensively studied in the last decade, although an optimum solution still remains to be found. Jutten and Herault [9] implemented a neural network which was able to separate sources with non-Gaussian and even probability density function (pdf). At that time, this solution lacked a theoretical explanation, and indeed its convergence was not wholly understood. In [5] the algorithm was interpreted in statistical terms, and so were the conditions for its convergence. It was found that the neural network implicitly searches for the zeros of the cross-cumulants of the outputs. Cardoso [1] attained the separation through the Eigenvalue Decomposition (EVD) of a weighted covariance matrix. His algorithm only deals with different non-normal symmetric pdfs. The extension to any set of non-normal pdfs is carried out by the same author in [2], where the 4th-order moment tensor is employed. Comon [3], [4], found a closed-form analytical solution for the noiseless two-source two-sensor case when the transformation is represented by an orthogonal matrix. In [6] the concept of *Independent Component Analysis (ICA)* is mathematically introduced. The ICA of a random vector consists in searching for a linear transformation that minimizes the statistical dependence between its components, as opposed to 2nd-order techniques, such as *Principal Component Analysis (PCA)*, which only aim at uncorrelation. From this point of view, a successful algorithm for BSS is developed. The rationale behind it is basically the diagonalization of the 4th-order cumulant tensor of the observations. Another form of ICA is unfolded in [10] via a higher-order extension of the Singular Value Decomposition (HOSVD). More recently, other non-cumulant-based approaches have been suggested. In [8] a maximum likelihood approach is adopted, in which the separating matrix must maximize the likelihood of the given observations.

1.1 General Cumulant-Based Approach

The cumulants can efficiently measure statistical independence, and hence the potential of the methods for BSS that rely on them. Consider the noiseless model:

$$Y = M \cdot X \tag{1}$$

where $Y = [y_1(t), \ldots, y_p(t)]^T$ contains the p observations, M is the $p \times q$ transfer matrix, and $X = [x_1(t), \ldots, x_q(t)]^T$ represents the q source signals. The general procedure is to look for a linear transformation W such that applied on the observations results in an independent set of signals, i.e. the elements of that matrix must make the output cross-cumulants vanish. It

†This work is partially supported by the Defence Research Agency of the UK.

In *Proceedings ECSAP '97*, Prague, Czech Republic, 24-27 June 1997, pp. 293-296.

is guaranteed in [14] that, provided there exists at most one Gaussian source, the separation can be achieved from the output statistics alone leaving the sources waveforms undistorted. This means that an indetermination of the form $P \cdot D$ is allowed, with P being a $(q \times q)$-permutation and D a regular diagonal matrix with the same dimensions. Therefore, W achieves a solution in the waveform-preserving sense if $W \cdot M = P \cdot D$, so that the output signals contain the sources up to a rearrangement and a scale factor.

In the present paper, which may be considered as an extension and update of [12], a cumulant-based method proposed by Mansour and Jutten in [11] for the two-source two-sensor noiseless case is reviewed. It extracts the sources by solving a system of non-linear equations involving 2nd- and 4th-order cross-cumulants of the observations, so that uncorrelation and 4th-order independence are imposed in the same step. In order to avoid an exhaustive search, the SVD is first used to decorrelate and normalize the observed signals. From this new set of observations, higher-order independence is obtained by using certain results unfolded by Comon [3], [4], which supply the separating matrix in an analytical manner and thus save the use of iterative numerical algorithms.

2 MANSOUR-JUTTEN (MJ) METHOD

2.1 Model Equations

Consider the noiseless BSS problem (1), which in the case of two sources and two observations becomes:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \tag{2}$$

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 invertible one (this is the condition for waveform-preserving BSS commented above). If we call $\{ s_i(t),\ i = 1,\ 2 \}$ the outputs of the W matrix:

$$\begin{aligned} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} &= \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \cdot \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \\ &= \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \end{aligned} \tag{3}$$

The statistical parameters of the outputs can easily be written as a function of the elements of the global matrix G and the statistical properties of the sources. If P_i represents the power and β_i the kurtosis (4th-order marginal cumulant) of the source signals, then:

$$E\left[s_1 \cdot s_2\right] = g_{11}\, g_{21}\, P_1 + g_{12}\, g_{22}\, P_2 \tag{4}$$

$$Cum_{22}(s_1,\, s_2) = g_{11}^2\, g_{21}^2\, \beta_1 + g_{12}^2\, g_{22}^2\, \beta_2. \tag{5}$$

2.2 Solutions

In [11], Mansour and Jutten consider the cost function $J = Cum_{22}(s_1,\, s_2)$. From (5) the following set of solutions is found for $J = 0$. When the source signals have the same sign of kurtosis ($sign(\beta_1) = sign(\beta_2)$), both desired and trivial solutions appear. The latter set can be avoided through certain constraints made on the elements of W (in particular, it is sufficient to impose $w_{11} \neq 0$, $w_{22} \neq 0$). On the other hand, if the source signals turn out to have different sign of kurtosis ($sign(\beta_1) \neq sign(\beta_2)$), the solutions above still exist, but also *spurious* solutions which do not achieve a proper extraction of the sources. Nevertheless, if we also force uncorrelation at the outputs, it can be shown from (4) that the spurious solutions do not appear any more.

In short, by solving the system

$$\left\{ \begin{array}{c} E\left[s_1 \cdot s_2\right] = 0 \\ Cum_{22}(s_1,\, s_2) = 0 \end{array} \right. \tag{6}$$

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 attained, i.e. the source signals can be found in S.

Assuming hereafter that $w_{11} = w_{22} = 1$, as done in [11], the system (6) may be written as a function of the unknown variables $(w_{12},\, w_{21})$ and the statistical properties of the known data signals:

$$\left\{ \begin{array}{l} w_{21}\, \mu_{20} + w_{12}\, \mu_{02} + (1 + w_{12} w_{21})\, \mu_{11} = 0 \\ (1 + 4\, w_{12}\, w_{21} + w_{12}^2\, w_{21}^2)\, k_{22} + 2\, (1 + w_{12}\, w_{21}) \\ \qquad (w_{21}\, k_{31} + w_{12}\, k_{13}) + w_{21}^2\, k_{40} + w_{12}^2\, k_{04} = 0 \end{array} \right. \tag{7}$$

where $\mu_{ij} = E\left[y_1^i \cdot y_2^j\right]$ represents the moment of order $(i+j)$ of $y_1(t)$ and $y_2(t)$, and $k_{ij} = Cum_{ij}(y_1,\, y_2)$ stands for the $(i+j)$th-order cumulant of the same pair of signals. To solve this system, a numerical method (see [13] for some of them) has to be used in the most general case. As a result, there is a computational burden inherent to the procedure. Furthermore, due to the numerical nature of the algorithm to be used, errors in the solution do not make the results very accurate.

3 MODIFICATION

3.1 Application of the SVD to the Problem

Let us compute the SVD [7] of the data signals: $Y = U \cdot \Sigma \cdot V^T$, and let Y' be equal to V^T. The MJ method will be applied to this last matrix, which may be seen as a first estimate of the sources: $Y' = V^T = (U \cdot \Sigma)^{-1} \cdot Y = (U \cdot \Sigma)^{-1} \cdot M \cdot X$.

Since the source signals are statistically independent their covariance matrix exhibits a diagonal structure: $cov\{X\} = C_X^2 = diag(P_1,\ P_2)$. Let X' be a normalized version of X such that $cov\{X'\} = I$, i.e. $X' = C_X^{-1} \cdot X$. It is obvious that X' is an equivalent representation of the source signals, since it contains in its rows just scaled versions of them (unit power). Then, the matrix Y' can be stated as:

$$Y' = (U \cdot \Sigma)^{-1} \cdot M \cdot C_X \cdot X' = M' \cdot X' \tag{8}$$

which may be regarded as an equivalent problem to the one we had at the beginning. In addition, as $Y' \cdot Y'^T = I$

(because $V^T \cdot V = I$) and $cov\{X'\} = I$, M' must necessarily be an orthogonal matrix. Thus, by applying the SVD, the original observed signals have been transformed into another set of signals which are uncorrelated and have unit power. Accordingly, the SVD is a numerically reliable way to carry out the PCA of the inputs [6], which provides uncorrelation but does not guarantee a complete extraction of the sources. Nevertheless, as far as BSS is concerned, the new problem is equivalent to the original one, for X and X' are related by the waveform-preserving equivalence, so further processing is still possible. Specifically, the MJ method will be used on the new set of signals Y', instead of Y, to achieve higher-order independence through the second equation of the system (6). These further operations are a means of implementing the ICA on the observations, which will improve the first estimation of the sources obtained by the PCA [6].

3.2 Solutions for the New Model

Then, along the same lines of section 1.1, a matrix W is sought, such that $S = W \cdot Y'$ contains the source signals. The system to be solved is given by (7). As the signals in Y' are spatially white and equal-power, if W was orthogonal then the signals in S would always be uncorrelated and the first equation of the system would not need to be solved any more. Thus, let us assume that W takes the form of a Givens rotation matrix:

$$W = c \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix}, \qquad c = \frac{1}{\sqrt{1+\theta^2}}. \tag{9}$$

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

$$\begin{aligned} & k^s_{22}/c^4 = \\ & \quad = k^{y'}_{22}\,\theta^4 + 2\,(k^{y'}_{31} - k^{y'}_{13})\,\theta^3 + (k^{y'}_{40} - 4\,k^{y'}_{22} + k^{y'}_{04})\,\theta^2 + \\ & \quad + 2\,(k^{y'}_{13} - k^{y'}_{31})\,\theta + k^{y'}_{22} = 0 \end{aligned} \tag{10}$$

where $k^s_{ij} = Cum_{ij}(s_1,\, s_2)$ and $k^{y'}_{ij} = Cum_{ij}(y'_1,\, y'_2)$. It is proved by Comon [3], [4] that, when the transfer matrix is orthogonal (as it is M' in our case) the solutions of this polynomial are given by two double roots θ_0 and $-1/\theta_0$, with $\theta_0 \in]-1,\, 1]$:

$$\theta_0 = -\frac{\rho}{2} + sign(\rho)\sqrt{\frac{\rho^2}{4} + 1}, \qquad \rho = \frac{k^{y'}_{31} - k^{y'}_{13}}{k^{y'}_{22}}. \tag{11}$$

These solutions also make the rest of the 4th-order cross-cumulants of the output signals vanish, i.e. the outputs are independent up to that order. In addition, owing to the fact that the signals in Y' have the same power and are uncorrelated, the transformation (9) yields uncorrelated outputs and so the first equation of the system (7) remains fulfilled. Note that the former condition is strictly necessary if uncorrelation is to be preserved. Therefore, (7) has been solved, but in a two-step procedure, forcing uncorrelation in the first step and 4th-order independence in the second one. The discussion in the previous section guarantees us that the signals obtained through the solution of such a system supply the desired sources. In short, the goal has been achieved, but sparing the comprehensive search that was necessary in the direct implementation of the MJ method.

4 SIMULATION RESULTS

The MATLAB V4.2c environment running on a UNIX Sun 4/40 SPARCstation IPC was employed in the simulations. For the direct implementation of the MJ method the SIMPLEX algorithm [13] is used to search for the minima of the sum of squares of the functions of (7). For the proposed modification, after computing the SVD of the input signals matrix (through the MATLAB built-in function *svd()*), the solutions are directly extracted from (11). 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 [6], 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. A Monte Carlo approach has been chosen for the simulations. In the first one, two white uniformly distributed signals are used as sources, both with zero mean, unit power and composed of 5000 samples. 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 the separations, both the mean value and the standard deviation (figures in parentheses in the table) are calculated for each parameter and each method. Table 1 summarizes the results obtained. Although both procedures yield similar parameter P_k, this is not the case in the rest of the parameters, for which the modification produces much better results. 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, with equal number of samples. The results also appear in table 1. The same comment as before is applicable here. Nevertheless, in this situation 3 solutions from the direct implementation of the MJ method did not give a proper G matrix, and thus an acceptable separation, while the proposed modification offered good separation in all cases.

5 CONCLUSIONS

In the direct implementation of the MJ method, 2nd- and 4th-order independence are sought at the same time. However, if uncorrelation is obtained in the first place,

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.

	Unif. - Unif.		Unif. - Exp.	
Parameter	MJ	Mod.	MJ	Mod.
P_k	0.98 (0.02)	0.97 (0.06)	0.94 (0.07)	0.97 (0.07)
P_e	0.20 (0.74)	0.02 (0.08)	0.47 (0.87)	0.02 (0.08)
Time (s)	4.08 (0.92)	0.01 (0.01)	6.46 (3.57)	0.01 (0.02)
Flops	7405 (1478)	27 (0)	11865 (6185)	27 (0)

the numerical search is no longer necessary. From the simulation results, this two-step procedure turns out to be more robust, more accurate and much faster than the former. Furthermore, the MJ method has been found to have difficulties when the sign of sources kurtosis is different, surely due to the system sensitivity to the initial approximation in that particular situation. On the other hand, the SVD could be regarded as a preprocessing to whiten and normalize the observed signals, which simplifies new operations to achieve higher-order independence. In point of fact, once uncorrelation is attained the separation is given analytically by Comon in [3] and [4]. As a result, although the methods developed by Mansour-Jutten and by Comon seem to be different approaches to BSS, when a decorrelation is applied to the former it degenerates into the latter. In spite of all the apparent advantages, one must keep in mind that the MJ method is essentially oriented to real-time applications, in the form of gradient-based algorithms, and hence the SVD limits the use of the idea in an adaptive fashion. Nevertheless, other ways of adaptively getting uncorrelated observations from the original ones exist, such as that suggested by Comon in [3]. At the same time, Comon's method could be looked at from another point of view. Indeed, the prewhitening strategy transforms the original problem into an equivalent one (in the waveform-preserving sense) for which the source signals can be simply retrieved. In turn, this standpoint highlights the paramount importance of a decorrelation as a previuos step when facing a problem of blind separation of sources.

All the above is applicable to the simplified scenario composed of two noiseless observations of two sources. The introduction of noise in the BSS model would certainly disturb the MJ system, for which exact solutions might well not exist any more, although this remains to be studied. The application of the SVD, on the other hand, would allow to work on the signal subspace and thus still obtain good results, as long as noise and sources were mutually independent. This would hold true even when increasing the number of sources and sensors. In order to achieve higher-order independence from the transformed set of observations, no closed-form solution has been found yet for the case of more than two sources and two sensors. In [3], [4] and [6], for instance, iterative approaches based on pairwise independence are proposed, which, although not optimum, provide a reasonable performance, at the expense of increased computational cost.

References

[1] J.-F. Cardoso, "Source Separation using Higher-Order Moments", *Proc. ICASSP*, Glasgow, Scotland, May 1989, pp. 2109–2112.

[2] J.-F. Cardoso, "Blind Identification of Independent Signals", *Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 1989.

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

[4] P. Comon, "Separation of Sources using Higher-Order Cumulants", *SPIE Vol. 1152 Advanced Alg. and Archit. for Signal Processing IV*, 1989, pp. 170–181.

[5] P. Comon, C. Jutten and J. Herault, "Blind Separation of Sources, Part II: Problems Statement", *Signal Processing*, Vol. 24, No. 1, Nov. 1991, pp. 11–20.

[6] P. Comon, "Independent Component Analysis, A new concept?", *Signal Processing*, Vol. 36, No. 3, Apr. 1994, pp. 287–314.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 2nd Edition, 1989.

[8] F. Harroy and J.-L. Lacoume, "Maximum likelihood estimators and Cramer-Rao bounds in source separation", *Signal Processing*, Vol. 55, Dec. 1996, pp. 167–177.

[9] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", *Signal Processing*, Vol. 24, No. 1, Nov. 1991, pp. 1–10.

[10] L. D. Lathauwer *et al.*, "Blind Source Separation by Higher-Order Singular Value Decomposition", in: M. Holt *et al.* (Eds.), *Proc. EUSIPCO-94*, Edinburgh, Scotland, U.K., Sept. 1994, pp. 175–178.

[11] A. Mansour and C. Jutten, "Fourth-Order Criteria for Blind Sources Separation", *IEEE Trans. on Signal Processing*, Vol. 43, No. 8, Aug. 1995, pp. 2022–2025.

[12] A. K. Nandi and V. Zarzoso, "Fourth-Order Cumulant Based Blind Source Separation", *IEEE Signal Processing Letters*, Vol. 3, No. 12, Dec. 1996, pp. 312–314.

[13] W. H. Press *et al.*, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, 2nd Edition, 1992.

[14] L. Tong *et al.*, "A Necessary and Sufficient Condition of Blind Identification", *International Signal Processing Workshop on HOS*, Chamrousse, France, July 1991, pp. 261–264.