BLIND SOURCE SEPARATION BY SIMULTANEOUS THIRD-ORDER TENSOR DIAGONALIZATION*

Lieven De Lathauwer, Bart De Moor, Joos Vandewalle
K.U.Leuven - E.E. Dept.- ESAT - SISTA
Kard. Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium
tel: 32/16/321805 fax: 32/16/321986
e-mail: Lieven.DeLathauwer@esat.kuleuven.ac.be.

ABSTRACT

We develop a technique for Blind Source Separation based on simultaneous diagonalization of (linear combinations of) third-order tensor "slices" of the fourth-order cumulant. It will be shown that, in a Jacobi-type iteration scheme, the computation of an elementary rotation can be reformulated in terms of a simultaneous matrix diagonalization.

1 INTRODUCTION

This paper deals with an alternative technique to solve the classical problem of *Blind Source Separation* or *Independent Component Analysis* (ICA). Denote the basic linear statistical model as:

$$Y = \mathbf{M}X + N \tag{1}$$

in which the observed vector Y, the source vector X and the noise vector N are zero-mean random vectors with values in $\mathbb{R}$ or $\mathbb{C}$. Assume that the transfer matrix $\mathbf{M}$ has linearly independent columns. The components of X are mutually statistically independent, as well as statistically independent from the noise components. The goal of ICA consists of the estimation of the transfer matrix (or "mixing matrix") $\mathbf{M}$ and the corresponding realizations of X, given only realizations of Y.

Without a priori knowledge the ICA-problem cannot be solved using only second-order statistics. Usually the second-order statistics of the observation vector Y are used to prewhiten the data; the covariance matrix can be diagonalized by e.g. Eigenvalue Decomposition (EVD). In the prewhitening stage, the column space of the transfer matrix (signal subspace) can be determined, which allows to reduce the dimensionality of the problem in a more-sensors-than-sources set-up by simple projection. However the transfer matrix itself remains unidentified up to an orthogonal (unitary) factor $\mathbf{Q}$. In the second step $\mathbf{Q}$ is then estimated from higher-order cumulants of the standardized data. Several algorithms have been presented in literature. Among the most well-known algebraic approaches are the one by Comon [6] and the JADE-algorithm (Joint Approximate Diagonalization Estimation) by Cardoso and Souloumiac [2].

In the former approach $\mathbf{Q}$ is estimated as the best diagonalizer, in least-squares sense, of the standardized cumulant tensor $\mathcal{C}$. Its computation takes the form of a Jacobi-type iteration, where each elementary rotation is basically obtained by rooting a polynomial of degree four (for the diagonalization of a fourth-order cumulant). This technique is further analyzed in [8]: it is shown that the calculation can also take the form of a tensorial Power Algorithm, based on only sums and products.

In the JADE-algorithm $\mathbf{Q}$ is estimated as the solution of a simultaneous EVD. The matrices to be decomposed span the range of $\mathcal{C}$, considered as a matrix-to-matrix mapping (e.g. the cumulant slices $\mathbf{C}^{(k,l)}$, obtained by fixing the indices k and l in c_{ijkl}). The simultaneous EVD can be computed by Jacobi-type iteration. Each iteration step gives rise to the best rank-1 approximation of a real symmetric (2×2)-matrix (real-valued data) or (3×3)-matrix (complex data).

Less well-known is the technique proposed in [7], where it is shown that, at the expense of numerical accuracy, the computational cost of the preceding methods can be lowered to the complexity of a matrix SVD. $\mathbf{Q}$ is estimated as the singular matrix of a "matrix unfolding" of $\mathcal{C}$, the columns of which are obtained by varying only one index (say i) in c_{ijkl}.

The aim of this paper is to fill the gap in the framework discussed above. Instead of examining the properties of $\mathcal{C}$ considered as a fourth-order tensor, or as a set of matrix- or vector-slices, we will investigate how the properties of the "third-order tensor slices" can be exploited.

The paper is organized as follows. In Section 2 the solution concept of simultaneous third-order tensor di-

* This research was partially supported by the Belgian Program on Interuniversity Attraction Poles (IUAP-17, IUAP-50), the European Community Research program ESPRIT, Basic Research Working Group nr. 6620 (ATHOS) and the Flemish Institute for Support of Scientific-Technological Research in Industry (I.W.T.) Lieven De Lathauwer is a Research Assistant supported by the I.W.T. Bart De Moor is a Research Associate of the National Fund for Scientific research (N.F.W.O.)

agonalization is developed, in terms of real-valued data. Explicit expressions for the computation of an elementary Jacobi-rotation are given in Section 3 (real and complex case). In Section 4 performance properties are discussed. Section 5 contains simulation results. Section 6 is a summarizing conclusion.

2 SOLUTION CONCEPT

Denote the whitened ICA-model as:

$$Y' = \mathbf{Q}^t X + N' \tag{2}$$

where the standardized observation vector Y', the source vector X and the noise vector N' are zero-mean random vectors in $\mathbb{R}^N$. We denote the fourth-order cumulant tensor of Y' by $\mathcal{C}$.

We associate to $\mathcal{C}$ a linear transformation of $\mathbb{R}^N$ to the vector space of real third-order tensors $\mathbb{R}^{N\times N\times N}$ in the following way:

$$\mathcal{V}' = \mathcal{C}(V) \quad \Longleftrightarrow \quad v'_{ijk} = \sum_l \mathcal{C}_{ijkl} v_l \tag{3}$$

for all index values. This linear mapping has a special structure. Neglecting the noise term in Eq. (2), the singular values are given by $\operatorname{sign}(\kappa_n)$ $(1 \leqslant n \leqslant N)$, where κ_n symbolizes the kurtosis of the nth source. The corresponding right singular vectors are the columns of $\mathbf{Q}$, denoted by Q_n $(1 \leqslant n \leqslant N)$. The corresponding "left singular tensors" $\mathcal{Q}_n$ $(1 \leqslant n \leqslant N)$ are given by:

$$(\mathcal{Q}_n)_{ijk} = \operatorname{sign}(\kappa_n)\,(Q_n)_i (Q_n)_j (Q_n)_k \tag{4}$$

So the SVD of the linear mapping reveals the standardized transfer matrix. Like in other ICA-algorithms we claim for identifiability reasons that at most one source is non-kurtic. A crucial remark is that all the third-order tensors in the range space of $\mathcal{C}$ can be written as a linear combination of the left singular tensors, such that they can be diagonalized by $\mathbf{Q}$.

When noise is present and/or when the statistics of Y are only available with limited accuracy, the derivation above is only approximately valid. We propose to estimate $\mathbf{Q}$ as the orthogonal matrix that simultaneously diagonalizes as far as possible (in least-squares sense) a set of third-order tensors that form a basis for the range of $\mathcal{C}$. Formally, if we denote the set to be diagonalized as $\{\mathcal{T}^{(l)}\}_{(1\leqslant l\leqslant N)}$, $\mathbf{Q}$ is estimated as the orthogonal matrix $\mathbf{U}$ that maximizes the following criterion function f:

$$f(\mathbf{U}) = \sum_l (\sum_n |{t'}^{(l)}_{nnn}|^2) \tag{5}$$

where $\mathcal{T}'^{(l)}$ denotes the tensor $\mathcal{T}^{(l)}$, after multiplication with $\mathbf{U}$:

$$t'_{nnn} = \sum_p \sum_q \sum_r u_{np} u_{nq} u_{nr} t^{(l)}_{pqr} \tag{6}$$

An orthogonal basis for the range of the linear mapping can be obtained from its SVD, together with a first estimate of $\mathbf{Q}$. It is also possible to find an ordinary basis by simple transformation under $\mathcal{C}$ of N linearly independent vectors. E.g. transformation of the canonical unit vectors corresponds to choosing the third-order tensor slices $\mathcal{C}^{(l)}$, obtained by fixing the index l in c_{ijkl}.

The algebraic technique of simultaneous third-order tensor diagonalization can be linked to the statistical concept of contrast optimization, established in [6]. Without going into details, the related contrast function takes the form of

$$c(\mathbf{U}) = \sum_{n,l}^{N} \operatorname{cum}(z_n, z_n, z_n, z_l)^2 \tag{7}$$

where the stochastic vector Z, with components z_n $(1 \leqslant n \leqslant N)$, is defined according to $Z = \mathbf{U}Y'$.

The best simultaneous diagonalizer $\mathbf{Q}$ will be computed by Jacobi-iteration. In each step the set of N $(2\times 2\times 2)$-tensors associated with the elementary rotation, will be diagonalized as far as possible.

3 ELEMENTARY ROTATIONS

The determination of an elementary rotation is the actual core of the algorithm. Although this result might seem contra-intuitive, it turns out that the simultaneous diagonalization of N $(2\times 2\times 2)$-tensors is mathematically equivalent to the simultaneous diagonalization of N symmetric (Hermitean) (2×2)-matrices, constructed from the third-order tensors. Simultaneous diagonalization of a set of matrices is discussed in [1, 2, 9]. The translation of the tensor problem to the matrix problem depends on the context.

3.1 Real-valued Data

First we introduce some new notations. We use the standard representation of Givens rotations:

$$\mathbf{U} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \tag{8}$$

Further we denote by $\{\tilde{\mathcal{C}}^{(l)}\}_{(1\leqslant l\leqslant N)}$ the set of $(2\times 2\times 2)$-tensors to be diagonalized. We also construct the real symmetric (2×2)-matrix $\mathbf{B}$, defined by:

$$b_{11} = 9a_2/4 + 3a_3/2 + a_1/4 \tag{9}$$

$$b_{12} = 3a_4/2 \qquad b_{22} = a_1 \tag{10}$$

in which a_1, a_2, a_3, a_4 are given by:

$$a_1 = \sum_l (\tilde{c}^{(l)}_{111})^2 + (\tilde{c}^{(l)}_{222})^2 \tag{11}$$

$$a_2 = \sum_l (\tilde{c}^{(l)}_{112})^2 + (\tilde{c}^{(l)}_{122})^2 \tag{12}$$

$$a_3 = \sum_l \tilde{c}^{(l)}_{111}\tilde{c}^{(l)}_{122} + \tilde{c}^{(l)}_{112}\tilde{c}^{(l)}_{222} \tag{13}$$

$$a_4 = \sum_l \tilde{c}^{(l)}_{122}\tilde{c}^{(l)}_{222} - \tilde{c}^{(l)}_{111}\tilde{c}^{(l)}_{112} \tag{14}$$

Using these notations, it is only a matter of tedious calculations to show that the criterion (5) takes the form

$$f(\mathbf{U}) = W^T \cdot \mathbf{B} \cdot W \tag{15}$$

in which $W \stackrel{\text{def}}{=} (\sin(2\alpha) \;\; \cos(2\alpha))^T$. Hence the optimal rotation can be found by computing the dominant eigenvector of $\mathbf{B}$ and normalizing it to unit-length. The sign of this vector can be fixed by restricting $\mathbf{U}$ to the set of inner rotations ($\alpha \in [-\pi/4, +\pi/4]$).

The core of the JADE-algorithm is also in the form of Eq. (15). When JADE aims at diagonalizing the cumulant slices $\mathbf{C}^{(k,l)}$, the cost of computing the equivalent of $\mathbf{B}$ is roughly the same as well.

3.2 Complex-valued Data: Case 1

The derivation in Section 2 can be repeated for complex data. Let us define the fourth-order cumulant $\mathcal{C}$ as

$$c_{ijkl} = \text{cum}(y'_i, {y'_j}^*, {y'_k}^*, y'_l) \tag{16}$$

and adjust the notation of Eqs.(2,4,6,7). In addition, we will represent an elementary complex Givens rotation as

$$\mathbf{U} = \begin{pmatrix} \cos\alpha & -e^{j\theta}\sin\alpha \\ e^{-j\theta}\sin\alpha & \cos\alpha \end{pmatrix} \tag{17}$$

Now consider the real symmetric (3×3)-matrix $\mathbf{B}$, defined by (*Re* and *Im* denote the real resp. imaginary part of a complex number):

$$b_{11} = a_1 \qquad b_{23} = \text{Im}(a_{10} - a_9)/2 \tag{18}$$
$$b_{12} = \text{Im}(a_2 + a_3 + a_4 + a_5)/2 \tag{19}$$
$$b_{13} = \text{Re}(a_4^* + a_5 - a_2 - a_3^*)/2 \tag{20}$$
$$b_{22} = a_{11} - a_{12} \qquad b_{33} = a_{11} + a_{12} \tag{21}$$

in which the auxiliary variables are given by:

$$a_1 = \sum_l |\tilde{c}_{111}^{(l)}|^2 + |\tilde{c}_{222}^{(l)}|^2 \quad a_3 = \sum_l (\tilde{c}_{111}^{(l)})^* \tilde{c}_{211}^{(l)} \tag{22}$$
$$a_2 = \sum_l \tilde{c}_{111}^{(l)} (\tilde{c}_{112}^{(l)} + \tilde{c}_{121}^{(l)})^* \tag{23}$$
$$a_4 = \sum_l (\tilde{c}_{222}^{(l)})^* \tilde{c}_{112}^{(l)} \tag{24}$$
$$a_5 = \sum_l \tilde{c}_{222}^{(l)} (\tilde{c}_{212}^{(l)} + \tilde{c}_{221}^{(l)})^* \tag{25}$$
$$a_6 = \sum_l \tilde{c}_{111}^{(l)} (\tilde{c}_{212}^{(l)} + \tilde{c}_{221}^{(l)})^* + \tilde{c}_{222}^{(l)} (\tilde{c}_{112}^{(l)} + \tilde{c}_{121}^{(l)})^* \tag{26}$$
$$a_7 = \sum_l |\tilde{c}_{112}^{(l)} + \tilde{c}_{121}^{(l)}|^2 + |\tilde{c}_{212}^{(l)} + \tilde{c}_{221}^{(l)}|^2 \tag{27}$$
$$a_8 = \sum_l |\tilde{c}_{211}^{(l)}|^2 + |\tilde{c}_{112}^{(l)}|^2 \tag{28}$$
$$a_9 = \sum_l \tilde{c}_{111}^{(l)} (\tilde{c}_{122}^{(l)})^* + \tilde{c}_{211}^{(l)} (\tilde{c}_{112}^{(l)} + \tilde{c}_{121}^{(l)})^* \tag{29}$$
$$a_{10} = \sum_l \tilde{c}_{222}^{(l)} (\tilde{c}_{211}^{(l)})^* + \tilde{c}_{122}^{(l)} (\tilde{c}_{212}^{(l)} + \tilde{c}_{221}^{(l)})^* \tag{30}$$
$$a_{11} = a_1/4 + \text{Re}(a_6)/2 + (a_7 + a_8)/4 \tag{31}$$
$$a_{12} = \text{Re}(a_9 + a_{10})/2 \tag{32}$$

Now the criterion function takes the form

$$f(\mathbf{U}) = W^T \cdot \mathbf{B} \cdot W \tag{33}$$

in which $W \stackrel{\text{def}}{=} (\cos(2\alpha) \;\; \sin(2\alpha)\sin\theta \;\; \sin(2\alpha)\cos\theta)^T$. Like in the real case, the optimal rotation can be found via the dominant eigenvector of $\mathbf{B}$.

3.3 Complex-valued Data: Case 2

With respect to the linear transformation in Eq. (3), it would be slightly more natural to define the fourth-order cumulant $\mathcal{C}$ by

$$c_{ijkl} = \text{cum}(y'_i, y'_j, y'_k, {y'_l}^*) \tag{34}$$

However, defined in this way, $\mathcal{C}$ is theoretically zero for circular random variables. On the other hand, an ICA-technique based on definition (34) may have a significantly higher performance than an algorithm based on definition (16) if the sources are substantially non-circular. The results of the preceding sections can be generalized when one starts from definition (34) as well. Detailed calculations are available on request.

4 PROPERTIES

4.1 Uniform Performance

In [4] it is proved that all ICA-techniques, based on optimisation of an orthogonal contrast function after prewhitening, exhibit the property of equivariance. This implies that, for a sufficiently low noise level, the quality of source separation is independent of the mixing matrix.

If we denote the estimated transfer matrix as $\hat{\mathbf{M}}$, the rejection of source q in the estimate of source p can be quantified by the Interference-to-Signal Ratio $\text{ISR}_{pq} = \text{E}(|\hat{\mathbf{M}}^\dagger \cdot \mathbf{M}|^2)$ ($\dagger$ symbolizes the Moore-Penrose inverse). For sufficiently long datasets, and for sufficiently low levels of the noise power σ_N^2, the performance is bounded as follows [3]:

$$\lim_{\sigma_N \to 0, T \to \infty} \frac{\text{ISR}_{pq} + \text{ISR}_{qp}}{2} \geqslant \frac{1}{4T} \tag{35}$$

in which T is the length of the dataset. The bound is actually reached by e.g. the technique presented in this paper (see Section 5).

4.2 Spatial Matched Filter

If the estimate $\hat{\mathbf{M}}$ is used to implement a Minimum Variance Distortionless Response filter (MVDR), the beamformer asymptotically corresponds to a spatial matched filter, when the noise is spatially white and Gaussian (the proof of this property is omitted for the sake of brevity). The resulting beamformer maximizes the power of each source estimate with respect to the sum of the power of the interferences and the noise.

5 SIMULATION RESULTS

We illustrate with two experiments, taken from [2]. Two mutually independent and temporally white signals, uniformly distributed on the unit circle, impinge on a linear $\lambda/2$ equispaced array of 10 unit-gain omnidirectional sensors in the far field of the emitters. The source powers are denoted as σ_1^2 and σ_2^2. In the simulations the lengths of the columns of $\hat{\mathbf{M}}$ are normalized in the sense that the source estimates have unit power. Hence, theoretically the elements of the transfer matrix are given by $m_{pq} = \sigma_q e^{2j\pi p\phi_q}$, where ϕ_q denotes the electrical angle of source q. In each experiment the datalength $T = 100$. All curves are obtained by averaging over 500 Monte Carlo simulations.

The results obtained by simultaneous tensor diagonalisation turn out to be nearly the same as the results of the JADE-algorithm. According to [5] the ICA-algorithm by Comon behaves similarly as well.

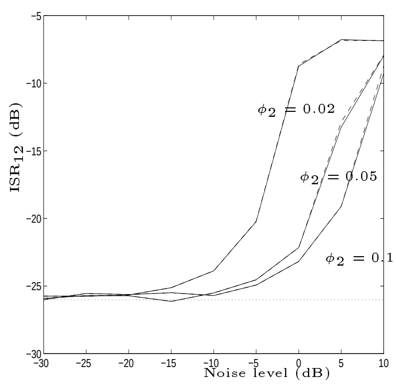

Figure 1: *Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0dB$) on the quality of separation. $\phi_1 = 0$. Solid: simultaneous third-order tensor diagonalization. Dashed: JADE-algorithm. Dotted: low-noise upper-bound of performance.*

6 CONCLUSIONS

We have shown that ICA can be realized by simultaneous diagonalization of third-order tensors, spanning the range of the standardized fourth-order cumulant. Computational requirements and performance are similar to those of the JADE-algorithm. The results in this paper can also be used in case the higher-order stage of ICA is based on the third-order cumulant itself (for sources with an unsymmetric probability distribution).

MATLAB-code is available by anonymous ftp from `ftp.esat.kuleuven.ac.be`.

References

[1] A. Bunse-Gerstner et al., *Numerical Methods for Simultaneous Diagonalization*, SIAM J. Matr. Anal. and Appl., 14 (4), Oct. 1993, pp. 927-949.

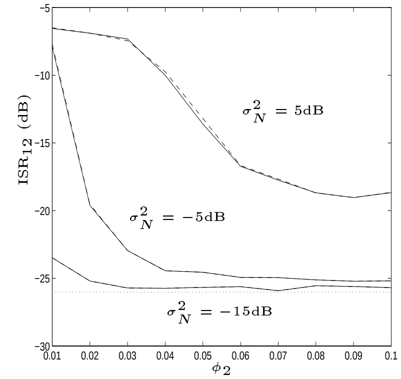

Figure 2: *Effect of the difference in DOA ($\phi_1 = 0$) on the quality of separation. $\sigma_1^2 = \sigma_2^2 = 0$ (dB). Solid: simultaneous third-order tensor diagonalization. Dashed: JADE-algorithm. Dotted: low-noise upper-bound of performance.*

[2] J.-F. Cardoso, A. Souloumiac, *Blind Beamforming for Non-Gaussian Signals*, IEE Proc.-F, 140 (6), 1994, pp. 362-370.

[3] J.-F. Cardoso, *On the Performance of Source Separation Algorithms*, Proc. EUSIPCO-94, Edinburgh, U.K., Sept. 13-16, 1994, Vol. 2, pp. 776-779.

[4] J.-F. Cardoso, *The Invariant Approach to Source Separation*, Proc. NOLTA'95, Las Vegas, USA, Dec. 10-14, 1995, Vol.1, pp. 55-60.

[5] P. Chevalier, *On the Performance of Higher Order Blind Sources Separation Methods*, Proc. IEEE SP / ATHOS Workshop on HOS, Girona, Spain, June 12-14, 1995, pp. 30-34.

[6] P. Comon, *Independent Component Analysis, A New Concept?* Sign. Proc., Special Iss. HOS, 36 (3), April 1994, pp. 287-314.

[7] L. De Lathauwer, B. De Moor, J. Vandewalle, *Blind Source Separation by Higher-Order Singular Value Decomposition*, Proc. EUSIPCO-94, Edinburgh, U.K., Sept. 13-16, 1994, Vol. 1, pp. 175-178.

[8] L. De Lathauwer, P. Comon, B. De Moor, J. Vandewalle, *Higher-Order Power Method - Application in Independent Component Analysis*, Proc. NOLTA'95, Las Vegas, USA, Dec. 10-14, 1995, Vol.1, pp. 91-96.

[9] A. Souloumiac, J.-F. Cardoso, *Givens Angles for Simultaneous Diagonalization*, SIAM J. Matr. Anal. and Appl., accepted for publication.