Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 1995-66

Independent Component Analysis based on Higher-Order Statistics Only [1]

Lieven De Lathauwer[2] Bart De Moor[3] Joos Vandewalle

December 1995

Accepted for:
8th IEEE SP Workshop
on Statistical Signal and Array Processing
Corfu, Greece
June 24-26, 1996

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/delathauwer/reports/ldl-95-66.ps.Z*

[2]ESAT - Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee), Belgium, tel 32/16/32 18 05, fax 32/16/32 19 86, email: *Lieven.DeLathauwer@esat.kuleuven.ac.be*. Lieven De Lathauwer is a research assistant with the I.W.T. (Flemish Institute for Support of Scientific-Technological Research in Industry). The research reported in this paper was partially supported by the Belgian Programme on Interuniversity Poles of Attraction (IUAP-17, IUAP-50), initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. It was also supported by the European Community Research program ESPRIT, Basic Research Working Group nr. 6620 (ATHOS). The scientific responsibility rests with its authors.

[3]Bart De Moor is a research associate with the N.F.W.O. (Belgian National Fund for Scientific Research).

Independent Component Analysis Based on Higher-Order Statistics Only*

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: 32/16/321085 fax: 32/16/321986
email: Lieven.DeLathauwer@esat.kuleuven.ac.be

Abstract

Most conventional techniques for Independent Component Analysis (ICA) resort to second-order statistics to decorrelate the observed data. The prewhitening step makes these algorithms sensitive to the presence of additive Gaussian noise. In this paper a higher-order-only technique is presented. The identification problem is approached in a (linear and multilinear) algebraic framework: our derivation starts with the observation that the solution can be obtained from the Canonical Decomposition (CANDECOMP) of a higher-order cumulant tensor. Next, it is demonstrated that the CANDECOMP components follow from the simultaneous diagonalization, by congruence transformation, of a set of matrices. A reformulation in terms of orthogonal unknowns leads to a simultaneous Schur decomposition, which is solved by a Givens-type iteration. The technique can be considered as the higher-order-only equivalent of the popular JADE-algorithm.

1 Introduction

The basic statistical model for *Independent Component Analysis* (ICA), or *Blind Source Separation*, is in this paper denoted 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 $\mathbb{C}$. The components of X are mutually statistically independent, as well as statistically independent from the noise components. The goal of ICA now consists of the estimation of the transfer matrix (or "mixture 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 for a whitening of the data. In this way the transfer matrix can be estimated up to an orthogonal factor $\mathbf{U}$. In the second step $\mathbf{U}$ is then obtained from higher-order cumulants of the standardized data. Several algorithms have been presented in literature. Among the most well-known approaches are the one by Comon [3] (further analyzed in [7]), where $\mathbf{U}$ is computed by a Jacobi-type diagonalization of the standardized cumulant tensor, and the JADE-algorithm (Joint Approximate Diagonalization Estimation) by Cardoso and Souloumiac [2], where $\mathbf{U}$ is found as the solution of a simultaneous eigenvalue decomposition.

In our paper the problem is solved using *only* the *higher-order* cumulant. This approach has the advantage that it is conceptually blind for the noise term N, when this term is Gaussian. For simplicity of notation, the exposition in this summary is restricted to fourth-order processing of real-valued data. The technique can be applied to cumulants of any order (higher than 2), as well as to complex data.

The paper is organized as follows. In the next section the relation between the columns of $\mathbf{M}$ and the fourth order observation cumulant is explicited. This relation takes the form of a tensorial decomposition of the cumulant in a sum of symmetric rank-1 tensors, and the uniqueness of this decomposition is discussed. In Section 3 the estimation of the transfer matrix from the cumulant model is presented as a simultaneous congruence transformation. In Section 4 it is

*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), the Flemish Institute for Support of Scientific-Technological Research in Industry (I.W.T.) and is part of a Concerted Action Project of the Flemish Community, entitled "Model-based Information Processing Systems". 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.) The scientific responsibility is assumed by the authors.

explained how the problem can be reformulated in terms of unknown orthogonal matrices. This leads to the simultaneous Schur decomposition of Section 5, for which a Givens-type computation scheme is derived. Section 6 contains a concluding discussion.

2 Canonical Decomposition

2.1 Model

As the name already suggests, the CANDECOMP of higher-order tensors is a basic concept in multilinear algebra. First we define the tensorial outer product of a set of vectors:

Definition 1 *The outer product of the vectors $U^{(1)} \in \mathbb{R}^{I_1}$, $U^{(2)} \in \mathbb{R}^{I_2}$, ..., $U^{(N)} \in \mathbb{R}^{I_N}$, denoted as $U^{(1)} \circ U^{(2)} \circ \ldots \circ U^{(N)}$, is an $(I_1 \times I_2 \times \ldots \times I_N)$-tensor $\mathcal{A}$, defined by the following element-wise equation:*

$$a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$$

In analogy with the vector/matrix case, the outer product leads to the definition of rank-1 tensors:

Definition 2 *An Nth-order tensor $\mathcal{A}$ has rank 1 when it equals the outer product of N vectors $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$.*

These elementary definitions allow to define the CANDECOMP:

Definition 3 *The Canonical Decomposition (CANDECOMP) of an Nth-order tensor $\mathcal{A}$ is the decomposition of $\mathcal{A}$ in a minimal sum of rank-1 components.*

The decomposition is also known as *Parallel Factors Model* (PARAFAC). It can be considered as the tensorial generalization of the diagonalization of matrices by equivalence transformation (unsymmetric case) or by congruence transformation (symmetric case). Despite the importance of CANDECOMP no robust general computation schemes have been proposed in the past.

2.2 Link with ICA

When the noise N is Gaussian, it does not contribute to the fourth-order cumulant of Y. This cumulant, denoted by $\mathcal{C}$, then shows the following structure:

$$\mathcal{C} = \sum_p \kappa_p M_p \circ M_p \circ M_p \circ M_p \tag{2}$$

where κ_p denotes the fourth-order cumulant of the pth source ($1 \leqslant p \leqslant P$) and M_p symbolizes the pth "steering vector" (i.e. the pth column of $\mathbf{M}$). Eq. (2) is clearly a symmetric CANDECOMP-model. The contribution of a non-Gaussian noise component, and the effect of other estimation errors when $\mathcal{C}$ is a finite sample cumulant, is considered as a perturbation of the equation.

2.3 Uniqueness

The uniqueness properties of CANDECOMP and its matrix counterparts are thoroughly different. Here we assume that the transfer matrix is square and regular, and that all the sources have non-vanishing kurtosis. It can be proved [9] that these conditions are sufficient to guarantee that decomposition (2) is unique up to the following trivial indeterminacies:

- permutation of the terms
- scaling of the steering vectors with a factor α_p, combined with inverse scaling (factor α_p^{-4}) of the coefficients κ_p.

(The interested reader is referred to the overview paper [4] for a discussion of some other uniqueness properties.) Note that in our setting different sources can have the same probability distribution, as long as they are mutually statistically independent in fourth order. The conditions can be weakened for the identification of at most one non-kurtic source. It is also possible to handle the "more-sensors-than-sources" case.

3 Simultaneous Congruence Transformation

We associate to $\mathcal{C}$ a linear matrix transformation in the following way:

$$\mathbf{B} = \mathcal{C}(\mathbf{A}) \quad \Longleftrightarrow \quad b_{ij} = \sum_{kl} c_{ijkl} a_{kl} \tag{3}$$

for all index values. From Eq. (2) follows that every matrix in the range space of $\mathcal{C}$ can be written as a linear combination of the "steering matrices" $M_p M_p^T$ ($1 \leqslant p \leqslant P$). In other words, the transfer matrix $\mathbf{M}$ diagonalizes every matrix in the range space of $\mathcal{C}$ by congruence transformation. Assume a basis for the range space is given by $\mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_P}$, then we have the following *simultaneous congruence transformation*:

$$\begin{aligned} \mathbf{T_1} &= \mathbf{M} \cdot \mathbf{D_1} \cdot \mathbf{M}^T \\ \mathbf{T_2} &= \mathbf{M} \cdot \mathbf{D_2} \cdot \mathbf{M}^T \\ &\vdots \\ \mathbf{T_P} &= \mathbf{M} \cdot \mathbf{D_P} \cdot \mathbf{M}^T \end{aligned} \tag{4}$$

where $\mathbf{D_1}, \mathbf{D_2}, \ldots, \mathbf{D_P}$ are diagonal. Remarkably, a similar set of equations arises in the blind separation of constant-modulus signals [11], which suggests a weird link between the constant-modulus property and non-Gaussianity. The computational technique of this paper differs from the one presented in [11], the latter being suboptimal.

Although two equations in (4) are generally sufficient to estimate the transfer matrix, we prefer to solve the complete set simultaneously, in order to exploit all the available information. This can be substantiated by numerical arguments [9].

The simultaneous solution of Eq. (4) is the higher-order-only equivalent of the *simultaneous eigenvalue decomposition* on which the ICA-algorithm by Cardoso and Souloumiac is based [2]. In the latter algorithm pre-whitening leads to a simultaneous matrix decomposition from which an orthogonal matrix has to be computed; in the current approach a general regular matrix has to be determined (up to the indeterminacies mentioned in Section 2.3), corresponding to the fact that there is no pre-whitening.

4 A new matrix representation

The fact that the unknown transfer matrix is basically an arbitrary regular matrix, makes it hard to deal with in a proper numerical way. Therefore we will represent the mixture matrix by a pair of *orthogonal* matrices, obtained from the QR-factorisation $\mathbf{M} = \mathbf{Q}^T \mathbf{R}'$ and the RQ-decomposition $\mathbf{M}^T = \mathbf{R}'' \mathbf{Z}^T$. The pair ($\mathbf{Q}$,$\mathbf{Z}$) is actually an equivalent representation of the transfer matrix, within the limits of identifiability. From the definition of $\mathbf{Q}$ and $\mathbf{Z}$ we have:

$$(\mathbf{Q} \cdot \mathbf{Z}) \cdot \mathbf{R}''^T = \mathbf{R}' \tag{5}$$

The orthogonal matrix $\mathbf{Q}\cdot\mathbf{Z}$ will be denoted as $\mathbf{V}$. The lower triangular part of Eq. (5) is a system of linear equations in the unknown coefficients of $\mathbf{R}''$:

$$\left[\begin{array}{cc} v_{P,P-1} & v_{PP} \end{array}\right] \left[\begin{array}{c} r''_{P-1,P-1} \\ r''_{P-1,P} \end{array}\right] = 0$$

$$\left[\begin{array}{ccc} v_{P-1,P-2} & v_{P-1,P-1} & v_{P-1,P} \\ v_{P,P-2} & v_{P,P-1} & v_{P,P} \end{array}\right] \left[\begin{array}{c} r''_{P-2,P-2} \\ r''_{P-2,P-1} \\ r''_{P-2,P} \end{array}\right] = \left[\begin{array}{cc} 0 & 0 \end{array}\right]^T \tag{6}$$

$$\vdots$$

Note that a scaling of the rows of $\mathbf{R}''$ does not affect this homogeneous set of equations, which is consistent with the fact that the steering vectors can only be determined up to a scalar multiple. By substitution of $\mathbf{R}''$ in Eq. (5) $\mathbf{R}'$ can be found as well.

5 Simultaneous Schur Decomposition

5.1 Principle

The notation of the simultaneous congruence transformation in terms of $\mathbf{Q}$ and $\mathbf{Z}$ leads to a set of matrix equations that we will denote as a *simultaneous Schur decomposition*:

$$\begin{aligned} \mathbf{Q} \cdot \mathbf{T_1} \cdot \mathbf{Z} &= \mathbf{R_1} = \mathbf{R}' \cdot \mathbf{D_1} \cdot \mathbf{R}'' \\ \mathbf{Q} \cdot \mathbf{T_2} \cdot \mathbf{Z} &= \mathbf{R_2} = \mathbf{R}' \cdot \mathbf{D_2} \cdot \mathbf{R}'' \\ &\vdots \\ \mathbf{Q} \cdot \mathbf{T_P} \cdot \mathbf{Z} &= \mathbf{R_P} = \mathbf{R}' \cdot \mathbf{D_P} \cdot \mathbf{R}'' \end{aligned} \tag{7}$$

From these equations the orthogonal matrices $\mathbf{Q}$ and $\mathbf{Z}$ have to be determined such that $\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_P}$ are "as upper triangular as possible" (in least-squares sense). The criterion function f to be minimized can be written as:

$$f(\mathbf{Q},\mathbf{Z}) = \|\mathbf{Q} \cdot \mathbf{T_1} \cdot \mathbf{Z}\|_{LF}^2 + \ldots + \|\mathbf{Q} \cdot \mathbf{T_P} \cdot \mathbf{Z}\|_{LF}^2 \tag{8}$$

in which $\|\mathbf{A}\|_{LF}$ denotes the below-diagonal Frobenius-norm of $\mathbf{A}$, i.e.

$$\|\mathbf{A}\|_{LF} = (\sum_{j<i} \sum_{i} a_{ij}^2)^{1/2} \tag{9}$$

It can be proved that the criterion f satisfies all the conditions for a higher-order-only contrast function [3, 5] that discriminates over the set of regular transfer matrices. Contrarily to classical approaches this contrast depends on two orthogonal matrices.

5.2 Solution by Givens-iteration

The core of our method is the computation of $\mathbf{Q}$ and $\mathbf{Z}$ from Eq. (7). The criterion function $f(\mathbf{Q},\mathbf{Z})$ is optimized by an iteration technique, in which $\mathbf{Q}$ and $\mathbf{Z}$ are determined as a sequence of elementary Givens rotations. Each elementary rotation makes the set $\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_P}$ simultaneously as upper triangular as possible.

First, the estimates of $\mathbf{Q}$ and $\mathbf{Z}$ are initialized as any orthogonal matrix, e.g. as the identity matrix: $\mathbf{Q}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{I}$. The estimates of $\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_P}$ are initialized accordingly: $\mathbf{R_1}^{(0)} = \mathbf{T_1}$, $\mathbf{R_2}^{(0)} = \mathbf{T_2}$, ..., $\mathbf{R_P}^{(0)} = \mathbf{T_P}$. In each iteration step k either $\mathbf{Q}^{(k)}$ or $\mathbf{Z}^{(k)}$ is updated. An update of $\mathbf{Q}^{(k)}$ takes the form of $\mathbf{Q}^{(k+1)} = \mathbf{G_{ij}} \cdot \mathbf{Q}^{(k)}$, in which $\mathbf{G_{ij}}$ denotes an elementary Givens rotation that affects rows i and j; at the same time $\mathbf{R_1}^{(k)}, \mathbf{R_2}^{(k)}, \ldots, \mathbf{R_P}^{(k)}$ are updated as $\mathbf{R_1}^{(k+1)} = \mathbf{G_{ij}^T} \cdot \mathbf{R_1}^{(k)}$, $\mathbf{R_2}^{(k+1)} = \mathbf{G_{ij}^T} \cdot \mathbf{R_2}^{(k)}$, ..., $\mathbf{R_P}^{(k+1)} = \mathbf{G_{ij}^T} \cdot \mathbf{R_P}^{(k)}$. $\mathbf{Z}^{(k)}$ is updated in a similar way, by working on the columns.

Let us focus on the updating of $\mathbf{Q}^{(k)}$ by multiplication with $\mathbf{G_{ij}}$. The Givens-rotation should be determined such that it minimizes the below-diagonal norm of all the ith and jth rows in $\mathbf{R_1}^{(k+1)}, \ldots, \mathbf{R_P}^{(k+1)}$. If we define

$$\mathbf{E_{ij}^{(k)}} = \left[\begin{array}{ccc} (\mathbf{R_1^{(k)}})_{i,\,i:j-1} & \cdots & (\mathbf{R_K^{(k)}})_{i,\,i:j-1} \\ (\mathbf{R_1^{(k)}})_{j,\,i:j-1} & \cdots & (\mathbf{R_P^{(k)}})_{j,\,i:j-1} \end{array}\right] \tag{10}$$

using MATLAB-notation, then $\mathbf{G_{ij}}$ should be determined such that the second row of $\mathbf{E}_{\mathbf{ij}}^{(k+1)} = \mathbf{G}_{\mathbf{ij}}^{T} \cdot \mathbf{E}_{\mathbf{ij}}^{(k)}$ has minimal Frobenius-norm. Hence $\mathbf{G_{ij}}$ can be obtained as the left singular matrix of $\mathbf{E}_{\mathbf{ij}}^{(k)}$: in this way the norm of the second row equals the smallest singular value of $\mathbf{E}_{\mathbf{ij}}^{(k)}$, which is the best one can do. In simulations the Givens-iteration shows a monotonous convergence to the global optimum of $f(\mathbf{Q}, \mathbf{Z})$.

6 Discussion and conclusions

We presented a new ICA-technique that resorts only to the higher-order cumulants of the observations. The transfer matrix estimate that is obtained shows exactly the same uniqueness properties as in the classical ICA-algorithms [2, 3], which also exploit second-order information in a prewhitening step. Higher-order-only Blind Source Separation has the advantage that it is asymptotically insensitive to additive Gaussian perturbations of the data. When dealing with finite sample cumulants, the accuracy of higher-order-only versus classical approaches is subject to a trade-off, caused by the fact that higher-order statistics are harder to estimate than second-order statistics [1, 10].

Our approach is based on the observation that the data cumulant can be expanded as a sum of rank-1 tensors. This implies that all the matrices in the range space of the cumulant tensor, considered as a super-symmetric matrix-to-matrix mapping, satisfy a simultaneous congruence transformation. For the numerical computation of this set of matrix equations we proposed a new representation of the transfer matrix: it turns out that any matrix, of which the columns are fixed up to multiplication with a scalar, can be represented by a pair of orthogonal matrices, obtained by QR- and RQ-factorisation. In this new format the simultaneous congruence transformation takes the form of a simultaneous Schur decomposition, that can be computed by a Givens-type iteration. The result can be considered as an approximate solution of *other* cumulant-based identification criterions, e.g. least-squares cumulant matching can be realized by means of a standard optimization routine, using the simultaneous Schur-solution as starting value [8].

The technique established in this paper is in fact the higher-order-only equivalent of the well-known combined second/higher-order ICA-algorithm by Cardoso and Souloumiac [2]. The concepts of this paper also lead to a higher-order-only equivalent ([9]) of the ICA-algorithm by Comon [3].

The technique can also be generalized for higher-order tensors without symmetry properties [6]. The unsymmetric version of the algorithm can be used for Factor Analysis of multiway datasets.

References

[1] C. Bourin and P. Bondon. Efficiency of high-order moment estimates. *Proc. IEEE SP / ATHOS Workshop on Higher-Order Statistics*: 186-190, Girona, Spain, June 12-14, 1995.

[2] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proceedings-F*, 140(6): 362-370, 1994.

[3] P. Comon. Independent component analysis, a new concept? *Signal Processing*, Special Issue *Higher Order Statistics*, 36(3): 287-314, April 1994.

[4] P. Comon and B. Mourrain. Decomposition of quantics in sums of powers of linear forms. Accepted for publication in *Signal Processing*.

[5] P. Comon, L. De Lathauwer and B. De Moor. A contrast-based independent component analysis without second-order moments. *Proc. SSAP-96*.

[6] L. De Lathauwer, B. De Moor and J. Vandewalle. Canonical decomposition of a fourth-order tensor. *IMA Conf. on Linear Algebra and Its Applications*, Manchester, U.K., July 10-12, 1995.

[7] L. De Lathauwer, P. Comon, B. De Moor and J. Vandewalle. Higher-order power method - Application in independent component analysis. *Proc. NOLTA'95*, 1: 91-96, Las Vegas, USA, December 10-14, 1995.

[8] L. De Lathauwer, B. De Moor and J. Vandewalle. A technique for higher-order-only blind source separation. *Proc. ICONIP'96*, Hong Kong, September 24-27, 1996 (to be published).

[9] L. De Lathauwer. *Signal Processing by Higher-Order Tensors*. Ph.D. thesis, K.U.Leuven, E.E.Dept.-ESAT, Belgium (in preparation).

[10] J. Fonollosa. Sample cumulants of stationary processes: asymptotic results. *IEEE Trans. on Signal Processing*, 43 (4): 967-977, April 1995.

[11] A.-J. van der Veen and A. Paulraj. An analytical constant modulus algorithm. Submitted to *IEEE Trans. Signal Processing*.