Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 1996-10

A Technique for Higher-Order-Only Blind Source Separation[1]

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

March 1996

Accepted for publication:
ICONIP'96
Sept. 24 - 27, 1996, Hong Kong
special invited session on
Blind Signal Processing
-
Adaptive and Neural Network Approaches

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/delathauwer/reports/report-96-10.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 supported by the European Community Research program ESPRIT, Basic Research Working Group nr. 6620 (ATHOS), the Belgian Institute for Scientific Research in Industry and Agriculture (I.W.O.N.L.), the Flemish Institute for Support of Scientific-Technological Research in Industry (I.W.T.), the Belgian Program on Interuniversity Attraction Poles (IUAP-17, IUAP-50), initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture and is part of a Concerted Action Project of the Flemish Community, entitled "Model-based Information Processing Systems". The scientific responsibility is assumed by its authors.

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

Abstract

We present a technique for Blind Source Separation that resorts only to the higher-order statistics of the data. In this way the algorithm is asymptotically insensitive to the effects of additive Gaussian noise. The identification of the transfer matrix is expressed as a simultaneous congruence transformation. By reformulating the problem as a simultaneous generalized Schur decomposition we restrict the unknowns to the set of orthogonal matrices. The solution can be obtained by Givens iteration, but this paper focuses on an alternative method based on bigradient optimization. This technique can be interpreted as a neural network method, in which the basic learning rule is Hebbian. Adaptivity is obtained by exponential data weighting. Cumulant matching takes the form of an extra least-squares descent algorithm.

A Technique for Higher-Order-Only Blind Source Separation*

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*— We present a technique for Blind Source Separation that resorts only to the higher-order statistics of the data. In this way the algorithm is asymptotically insensitive to the effects of additive Gaussian noise. The identification of the transfer matrix is expressed as a simultaneous congruence transformation. By reformulating the problem as a simultaneous generalized Schur decomposition we restrict the unknowns to the set of orthogonal matrices. The solution can be obtained by Givens iteration, but this paper focuses on an alternative method based on bigradient optimization. This technique can be interpreted as a neural network method, in which the basic learning rule is Hebbian. Adaptivity is obtained by exponential data weighting. Cumulant matching takes the form of an extra least-squares descent algorithm.**

1 Introduction

Denote the basic statistical model for *Blind Source Separation*, or *Independent Component Analysis* (ICA), 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 $\mathbf{M}$ and the corresponding realizations of X, given only realizations of Y.

Blind source separation is a very fundamental problem, with a large application area. The classical example is blind Space Division Multiple Access (SDMA) in the context of wireless communications, where one aims at the separation of messages (same time slot - same frequency bin) from different users at different locations, by using an array of antennas, even if this array is badly calibrated. Indeed, if the source signals are narrow-band, then the complex envelope representation of the transmission takes precisely the form of Eq. (1) (the components of X are the source signals and the components of Y are the signals captured by the antenna array) [13]. The same separation model arises in a very natural way in seismic prospection, psychometrics, astronomy, vibro-acoustics, biomedical signal processing, etc. - see e.g. [5, 12].

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 algebraic approaches are the one by Comon [4] (further analyzed in [7]), where $\mathbf{U}$ is computed by a Jacobi-type diagonalization of the standardized cumulant tensor, and the one by Cardoso [2], where $\mathbf{U}$ is found as the solution of a simultaneous eigenvalue decomposition. Solutions based on artificial neural networks (ANN) are discussed in e.g. [3].

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.

*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.)

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 tetrads, 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 the problem is reformulated as a simultaneous generalized Schur decomposition, for which we derive a bigradient optimization computation scheme. Section 5 discusses the issue of adaptivity, which is realized by means of an exponential forgetting function. In Section 6 the algorithm is interpreted as a neural network method. In Section 7 it is explained how the solution can be used for cumulant matching. Section 8 contains a concluding discussion.

2 A Decomposition in Tetrads

2.1 Model

When the noise N is Gaussian, it does not contribute to the fourth-order cumulant of Y. This cumulant, denoted by $\mathcal{C}^{(4)}$, has a very typical structure:

$$c^{(4)}_{ijkl} = \sum_{s}^{S} \kappa_s \, m_{is} m_{js} m_{ks} m_{ls} \tag{2}$$

where κ_s denotes the fourth-order cumulant of the sth source ($1 \leqslant s \leqslant S$). In tensorial terminology each term corresponds to a fourth-order rank-1 tensor, or tetrad, equal to the outer product of four times the same column of $\mathbf{M}$. (These columns will further be symbolized by M_s ($1 \leqslant s \leqslant S$) and denoted as "steering vectors"). The contribution of a non-Gaussian noise component, and the effect of other estimation errors when $\mathcal{C}^{(4)}$ is a finite sample cumulant, is considered as a perturbation of Eq. (2).

2.2 Uniqueness

For notational convenience we assume that the transfer matrix is square and regular, and that all the sources have non-vanishing kurtosis. It can be proved [6] 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_s, combined with inverse scaling (factor α_s^{-4}) of the coefficients κ_s.

Note that different sources can have the same probability distribution, as long as they are mutually statistically independent. 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 $\mathcal{C}^{(4)}$ a linear matrix transformation in the following way:

$$\mathbf{B} = \mathcal{C}^{(4)}(\mathbf{A}) \qquad \Longleftrightarrow \qquad b_{ij} = \sum_{kl} c_{ijkl} a_{kl} \tag{3}$$

for all index values. From Eq. (2) follows that every matrix in the range space of $\mathcal{C}^{(4)}$ can be written as a linear combination of the "steering matrices" $M_s M_s^T$ ($1 \leqslant s \leqslant S$). In other words, the transfer matrix $\mathbf{M}$ diagonalizes every matrix in the range space of $\mathcal{C}^{(4)}$ by congruence transformation. Assume that the range space is spanned by $\mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_P}$ ($P \geqslant S$), 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. This set of equations 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 2.2), corresponding to the fact that there is no pre-whitening.

4 Simultaneous Generalized Schur Decomposition

4.1 Principle

Theoretically the transfer matrix can be computed from 2 of the equations in (4). E.g. if we can assume that $\mathbf{D_2}$ does not contain any zero elements on the diagonal, then combination of the first two equations leads to the following eigenvalue decomposition:

$$\mathbf{T_1} \cdot \mathbf{T_2}^{-1} = \mathbf{M} \cdot \mathbf{D_1} \cdot \mathbf{D_2}^{-1} \cdot \mathbf{M}^{-1} \tag{5}$$

Intuitively however, it is clear that it is preferable to solve (4) simultaneously, in order to exploit all the available information. This can be substantiated by numerical arguments [9].

The fact that the unknown transfer matrix is basically an arbitrary regular matrix, makes it hard to deal with in a proper numerical way. We reformulate the problem in terms of orthogonal unknowns by introducing a QR-factorisation $\mathbf{M} = \mathbf{Q}^T \mathbf{R}'$ and an RQ-decomposition $\mathbf{M}^T = \mathbf{R}'' \mathbf{Z}^T$, which leads to a set of matrix equations that we will denote as a *simultaneous generalized Schur decomposition* (a set of two of the equations below is called "generalized Schur decomposition" in [11]):

$$\begin{aligned}
\mathbf{Q} \cdot \mathbf{T_1} \cdot \mathbf{Z} &= \mathbf{R_1} = \mathbf{R}' \cdot \mathbf{D_1} \cdot \mathbf{R}'' \\
\mathbf{Q} \cdot \mathbf{T_2} \cdot \mathbf{Z} &= \mathbf{R_2} = \mathbf{R}' \cdot \mathbf{D_2} \cdot \mathbf{R}'' \\
&\vdots \\
\mathbf{Q} \cdot \mathbf{T_P} \cdot \mathbf{Z} &= \mathbf{R_P} = \mathbf{R}' \cdot \mathbf{D_P} \cdot \mathbf{R}''
\end{aligned} \tag{6}$$

(The same trick can be found in [14], for the separation of constant-modulus signals.) 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 maximized can be written as:

$$f(\mathbf{Q}, \mathbf{Z}) = \|\mathtt{upp}(\mathbf{Q} \cdot \mathbf{T_1} \cdot \mathbf{Z})\|_F^2 + \ldots + \|\mathtt{upp}(\mathbf{Q} \cdot \mathbf{T_P} \cdot \mathbf{Z})\|_F^2 \tag{7}$$

in which $\|\mathbf{A}\|_F$ denotes the Frobenius-norm and $\mathtt{upp}(\mathbf{A})$ the upper triangular part of $\mathbf{A}$.

In the next section we will explain how the optimization can be realized, and in Section 4.3 we will describe how $\mathbf{M}$ can be derived, once $\mathbf{Q}$ and $\mathbf{Z}$ have been found.

4.2 Solution by bigradient optimization

The core of our method is the computation of $\mathbf{Q}$ and $\mathbf{Z}$ from Eq. (6). The maximization of the criterion function f with respect to $\mathbf{Q}$ and $\mathbf{Z}$, and the adaptive tracking of the solution (Section 5), can take various forms. In the companion paper [8] the solution is based on 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. For a continuous-time formulation of the problem a gradient flow is considered in [9]. Here we will present a bigradient optimization algorithm [15], which is more suitable for interpretation in terms of ANN's.

At each iteration step n a small correction term, in the direction of steepest ascent of f, is added to $\mathbf{Q}(t)$ and $\mathbf{Z}(t)$. This correction term does not take into account that $\mathbf{Q}(t)$ and $\mathbf{Z}(t)$ are orthogonal matrices; orthogonality is realized by simultaneous minimization of the criterion:

$$g(\mathbf{Q}, \mathbf{Z}) = 1/2 \, \|\mathbf{I} - \mathbf{Q}^T \mathbf{Q}\|_F^2 + 1/2 \, \|\mathbf{I} - \mathbf{Z}^T \mathbf{Z}\|_F^2 \tag{8}$$

in which $\mathbf{I}$ is the identity matrix. The updating rule then takes the form:

$$\mathbf{Q}(t+1) = \mathbf{Q}(t) + \mu \nabla_{\mathbf{Q}(t)} f - \gamma \nabla_{\mathbf{Q}(t)} g \tag{9}$$

$$\mathbf{Z}(t+1) = \mathbf{Z}(t) + \mu \nabla_{\mathbf{Z}(t)} f - \gamma \nabla_{\mathbf{Z}(t)} g \tag{10}$$

in which μ is a small and γ a relatively large $(0.5 \ldots 1)$ gain parameter, and in which the gradients are

given by:

$$\nabla_{\mathbf{Q}} f = 2 \sum_p \mathbf{M}_p \cdot \mathbf{Z} \cdot (\mathrm{upp}(\mathbf{R}_p))^T \qquad \nabla_{\mathbf{Z}} f = 2 \sum_p \mathbf{M}_p^T \cdot \mathbf{Q} \cdot \mathrm{upp}(\mathbf{R}_p) \tag{11}$$

$$\nabla_{\mathbf{Q}} g = -\mathbf{Q} \cdot (\mathbf{I} - \mathbf{Q}^T \mathbf{Q}) \qquad \nabla_{\mathbf{Z}} g = -\mathbf{Z} \cdot (\mathbf{I} - \mathbf{Z}^T \mathbf{Z}) \tag{12}$$

4.3 Reconstruction of the transfer matrix

It is not hard to show that the pair (**Q**,**Z**) is actually an alternative representation of the transfer matrix itself. From the definition of **Q** and **Z** we have:

$$(\mathbf{Q} \cdot \mathbf{Z}) \cdot \mathbf{R}''^T = \mathbf{R}' \tag{13}$$

The orthogonal matrix $\mathbf{Q} \cdot \mathbf{Z}$ will be denoted as **V**. The lower triangular part of Eq. (13) is a system of linear equations in the unknown coefficients of $\mathbf{R}''$:

$$\begin{bmatrix} v_{P,P-1} & v_{PP} \end{bmatrix} \begin{bmatrix} r''_{P-1,P-1} \\ r''_{P-1,P} \end{bmatrix} = 0$$

$$\begin{bmatrix} v_{P-1,P-2} & v_{P-1,P-1} & v_{P-1,P} \\ v_{P,P-2} & v_{P,P-1} & v_{P,P} \end{bmatrix} \begin{bmatrix} r''_{P-2,P-2} \\ r''_{P-2,P-1} \\ r''_{P-2,P} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{14}$$

$$\vdots$$

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. (13) $\mathbf{R}'$ can be found.

5 Adaptivity

When the data are slowly time-varying, the algorithm can be made adaptive by considering the weighted observations $Y(t), \alpha Y(t-1), \alpha^2 Y(t-2), \ldots$ instead of $Y(t), Y(t-1), Y(t-2), \ldots$ $(0 < \alpha \lesssim 1)$, in which the slowly decaying exponential function serves as a forgetting mechanism.

For long datasets a consistent estimate of the fourth-order moment is obtained as:

$$\hat{\mathcal{M}}^{(4)}(t) = (1-\alpha')(\tilde{\mathcal{M}}^{(4)}(t) + \alpha' \tilde{\mathcal{M}}^{(4)}(t-1) + \alpha'^2 \tilde{\mathcal{M}}^{(4)}(t-2) + \ldots) \tag{15}$$

in which $\alpha' = \alpha^4$ and $\tilde{\mathcal{M}}^{(4)}(t)$ is the tetrad corresponding to $Y(t)$: $\tilde{m}^{(4)}_{ijkl}(t) = y_i(t) y_j(t) y_k(t) y_l(t)$. It is clear that this estimate can be updated as:

$$\hat{\mathcal{M}}^{(4)}(t+1) = (1-\alpha')\tilde{\mathcal{M}}^{(4)}(t+1) + \alpha' \hat{\mathcal{M}}^{(4)}(t) \tag{16}$$

A consistent estimate of the covariance, $\hat{\mathbf{C}}^{(2)}(t)$, is found and updated in a similar way:

$$\hat{\mathbf{C}}^{(2)}(t+1) = (1-\alpha'')\tilde{\mathbf{C}}^{(2)}(t+1) + \alpha' \hat{\mathbf{C}}^{(2)}(t) \tag{17}$$

in which $\alpha'' = \alpha^2$ and $\tilde{\mathbf{C}}^{(2)}(t)$ equals $Y(t)Y(t)^T$. The fourth-order cumulant estimate at time $t+1$, $\hat{\mathcal{C}}^{(4)}(t+1)$, can be computed from its classical definition,or equivalently, by the following updating rule:

$$\begin{aligned} \hat{c}^{(4)}_{ijkl}(t+1) \;=\; & \alpha' \hat{c}^{(4)}_{ijkl}(t) + (6\alpha'' - 4\alpha' - 2)\, \tilde{m}^{(4)}_{ijkl}(t+1) + (\alpha' - \alpha'')\, (\tilde{c}^{(2)}_{ij}(t+1)\hat{c}^{(2)}_{kl}(t) + \hat{c}^{(2)}_{ij}(t)\tilde{c}^{(2)}_{kl}(t+1) \\ & + \tilde{c}^{(2)}_{ik}(t+1)\hat{c}^{(2)}_{jl}(t) + \hat{c}^{(2)}_{ik}(t)\tilde{c}^{(2)}_{jl}(t+1) + \tilde{c}^{(2)}_{il}(t+1)\hat{c}^{(2)}_{jk}(t) + \hat{c}^{(2)}_{il}(t)\tilde{c}^{(2)}_{jk}(t+1)) \end{aligned} \tag{18}$$

This expression can also be considered as the updating rule for $\mathbf{T_1}(t), \mathbf{T_2}(t), \ldots, \mathbf{T_P}(t)$, if these matrices are chosen as the cumulant slices of $\mathcal{C}^{(4)}(t)$ corresponding to fixed indices k and l.

For the Givens-solution scheme of [8] it is not necessary to start a complete iteration every time step, considering the fact that we are only dealing with a weak non-stationarity. Simulations show that e.g. one sweep of rotations per time step produces sufficiently accurate results.

6 Neural Network Architecture

The bigradient algorithm can be interpreted in terms of a two-layer ANN as follows. First of all, to deal with the multidimensional character of the data to be processed, the different layers of the ANN

should be considered as a stack with PS levels, each level containing S neurons. The ith input to the $((p-1)P+j)$th input level consists of the matrix element $(\mathbf{T_p}(t))_{ij}$. The weight vector of the nth neuron of the first layer is the nth column of $\mathbf{Q}(t)$, independent of the level. In this way the output of the nth neuron at the $((p-1)P+j)$th level of the first layer equals $(\mathbf{Q}(t)^T\mathbf{T}_p(t))_{ij}$. Summarized, the first layer computes $\mathbf{Q}(t)^T\mathbf{T_1}(t), \mathbf{Q}(t)^T\mathbf{T_2}(t), \ldots$ from the inputs $\mathbf{T_1}(t), \mathbf{T_2}(t), \ldots$ In a similar way the weights of the neurons in the second layer are the elements of $\mathbf{Z}(t)$; the output of the network are the matrices $\mathbf{R_1}(t), \mathbf{R_2}(t), \ldots$ The left part of Fig. 1 is a schematic representation of this ANN for the case where $P = S = 2$.

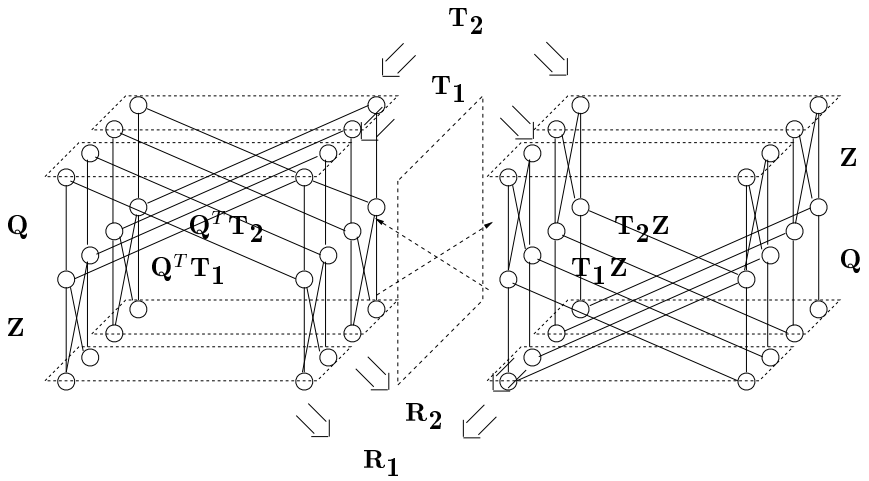

Figure 1: Schematic representation of an ANN for simultaneous generalized Schur decomposition.

It is easy to verify that the first correction term in the learning rule of Eq. (10) is in fact a Hebbian term: $\mathbf{Z}(t)$ is updated with a sum of products, each obtained by multiplication of the values at the in- and output of the connection. As far as the criterion f is concerned, the updating of $\mathbf{Z}(t)$ can be implemented as a simple, uniform, local learning rule. This doesn't hold true for $\mathbf{Q}(t)$. However by a mathematical trick the size of the network might be traded for simplicity of the learning rule. Fig. 1 shows the principle: in the right part of the figure the ordering of the multiplication with $\mathbf{Z}(t)$ and $\mathbf{Q}(t)$ is reversed with respect to the left part. In this way $\mathbf{Q}(t)$ can be updated by a Hebbian rule too. The new values $\mathbf{Q}(t+1)$ and $\mathbf{Z}(t+1)$ are immediately fed back to the weights for the first layer.

7 Least-squares Cumulant Matching

The least-squares criterion associated to the upper triangularity in Eq. (6) and the least-squares criterion associated to the fit in Eq. (2) are only asymptotically equivalent, although simulations show that they are quite close in practice. The simultaneous generalized Schur solution could be used as an interesting starting value for an additional least-squares descent algorithm that matches both sides of Eq. (2). For this extra optimization step one can use closed-form expressions for the gradient and Hessian of the criterion function. We stress the importance of the initial value, obtained by simultaneous generalized Schur decomposition: the cumulant matching criterion is a multivariate polynomial function, of degree 8 in the unknown transfer matrix entries, and shows a lot of irrelevant local optima.

8 Discussion and Conclusions

We presented a new technique for Blind Source Separation 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, 4], 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 tetrads. This implies that all the matrices in the range space of this 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 generalized Schur decomposition, that can be computed by bigradient optimization. The technique can be interpreted as an ANN method, in which the basic learning rule is Hebbian. Adaptivity can be realized by exponential data weighting. The result can be considered as an approximate solution of *other* cumulant-based identification criterions, e.g. least-squares cumulant matching.

The derivation 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]. Simultaneous congruence transformation and simultaneous eigenvalue decomposition play exactly the same role. However, without prewhitening *a pair* of orthogonal matrices has to be determined, instead of a single simultaneous eigenmatrix. The equivalent of simultaneous diagonalization turns out to be simultaneous *triangularization*. For a solution method based on Givens-iteration, like in [2], the reader is referred to [8]. The concepts of this paper also lead to a higher-order-only equivalent ([9]) of the ICA-algorithm by Comon [4].

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*, Girona, Spain, June 12-14, 1995, pp. 186-190.

[2] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, 1994.

[3] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. N.Y.: John Wiley, 1993.

[4] P. Comon, "Independent Component Analysis, a new concept?" *Signal Processing*, Special Issue *Higher Order Statistics*, vol. 36, no. 3, pp. 287-314, April 1994.

[5] P. Comon *et al.* (Ed.), *Proc. IEEE SP / ATHOS Workshop on Higher-Order Statistics*. Girona, Spain, June 12-14, 1995.

[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*, Las Vegas, USA, December 10-14, 1995, vol.1, pp. 91-96.

[8] L. De Lathauwer, B. De Moor and J. Vandewalle, "Independent Component Analysis based on higher-order statistics only," *Proc. SSAP-96, 8th IEEE SP Workshop on Stat. Signal and Array Processing*, Corfu, Greece, June 24-26, 1996.

[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*, vol. 43, no. 4, pp. 967-977, April 1995.

[11] G.H. Golub and C.F. Van Loan, *Matrix computations*. Johns Hopkins University Press, 1991.

[12] J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. of the International SP Workshop on Higher Order Statistics*. Chamrousse, France, July 10-12, 1991.

[13] K.S. Shanmugam, *Digital and Analog Communication Systems*. N.Y.: John Wiley & Sons, 1985.

[14] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 5, May 1996.

[15] L. Wang, J. Karhunen and E. Oja, "A bigradient optimization approach for robust PCA, MCA, and source separation," *Proc. IEEE Int. Conf. on Neural Networks*, Perth, Australia, Nov. 1995, pp. 1684-1689.