00-15

De Lathauwer L., De Moor B., Vandewalle J., ``ICA techniques for more sources than sensors", in *Proc. of the IEEE Signal Processing Workshop on Higher-Order Statistics (HOS'99)*, Caesarea, Israel, Jun. 1999, pp. 121-124., Lirias number: 178930.

ICA techniques for more sources than sensors

Lieven De Lathauwer, Bart De Moor, Joos Vandewalle
K.U.LEUVEN, E.E.Dept. (ESAT), SISTA/COSIC
Kard. Mercierlaan 94, B-3001 Leuven (Heverlee)
Belgium
Lieven.DeLathauwer@esat.kuleuven.ac.be

Abstract

In this paper we derive algorithms to identify the mixing matrix in the context of an Independent Component Analysis with more sources than sensors. First, by exploiting the fact that for complex-valued observations, depending on the type of complex symmetry, 2 different fourth-order cumulants are available, we develop a technique that can cope with $N(N+1)/2$ sources for only N sensors. Secondly, the technique presented in [2], based on a single cumulant, is modified to take both cumulants into account as well.

1. Introduction

The basic statistical model for *Independent Component Analysis* (ICA), or *Blind Source Separation* (BSS), is in this paper denoted as

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

in which the observed vector Y, the source vector X and the noise vector N are zero-mean random vectors with values in $\mathbb{C}$. 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 "mixing matrix" $\mathrm{M} \in \mathbb{C}^{N \times R}$ and the corresponding realizations of X, given only realizations of Y.

In this paper we focus on the situation in which there are more sources than sensors. Our approach is based on multilinear algebra. In this framework the ICA problem can be formulated as the decomposition of super-symmetric higher-order cumulant tensors in a minimal number of rank-1 terms, in which each rank-1 term consists of the contribution of one particular source signal to the overall linear combination. Higher-order tensors have the advantage, compared to matrices, that the number of rank-1 terms is not bounded by the dimension of the column nor the row space, which makes it a priori possible to address the ICA problem in a context where the number of sources exceeds the number of sensors. The rectangular size of the mixing matrix makes it of course impossible to estimate the source values by mere matrix inversion, but the source values may nevertheless be reconstructed by exploiting prior knowledge on the source distributions, like e.g. the finite alphabet property [4, 5].

The techniques developed in this paper start from the observation that for complex-valued non-Gaussian sources two types of fourth-order cumulants are available, even when the probability density is symmetric about the origin:

$$\begin{aligned} \mathcal{C}_Y^{(4,0)} &\stackrel{\text{def}}{=} \mathrm{Cum}\{Y, Y, Y, Y\}, \\ \mathcal{C}_Y^{(2,2)} &\stackrel{\text{def}}{=} \mathrm{Cum}\{Y, Y, Y^*, Y^*\}. \end{aligned}$$

Due to the multilinearity property, the observation cumulants are related to the source cumulants $\kappa_r^{(4,0)}$ and $\kappa_r^{(2,2)}$ $(1 \leqslant r \leqslant R)$ in the following way:

$$(\mathcal{C}_Y^{(4,0)})_{ijkl} = \sum_{r=1}^{R} \kappa_r^{(4,0)} m_{ir} m_{jr} m_{kr} m_{lr}, \tag{2}$$

$$(\mathcal{C}_Y^{(2,2)})_{ijkl} = \sum_{r=1}^{R} \kappa_r^{(2,2)} m_{ir} m_{jr} m_{kr}^* m_{lr}^*. \tag{3}$$

for all index values.

In Section 2, $(\mathcal{C}_Y^{(4,0)}, \mathcal{C}_Y^{(2,2)})$ will be considered as the tensor equivalent of a matrix pencil; by associating an Eigenvalue Decomposition (EVD) with it, a mixing matrix with $N(N+1)/2$ columns can be identified. Section 3 further examines the mathematical aspects of this technique. Next, Section 4 presents a modification of the powerful approach described in [1]. In this paper, a more-sources-than-sensors algorithm was derived by exploiting the algebraic structure of $\mathcal{C}_Y^{(2,2)}$; our modification consists in the fact that knowledge of $\mathcal{C}_Y^{(4,0)}$ can be incorporated without a substantial extra cost.

In [7] we address the special case of (2×3) ICA. Very efficient algorithms are developed here, which is important w.r.t. real-life applications.

2. Technique 1

Consider $\mathcal{C}^{(4,0)}$ and $\mathcal{C}^{(2,2)}$ as linear matrix-to-matrix mappings, according to

$$(\mathcal{C}(\mathbf{A}))_{ij} = \sum_{p,q=1}^{N} c_{ijpq} a_{pq},$$

and let $\mathbf{C}^{(4,0)}, \mathbf{C}^{(2,2)} \in \mathbb{C}^{N^2 \times N^2}$ be the corresponding matrix representations. In this format, Eqs. (2) and (3) can be expressed as follows:

$$\mathbf{C}_Y^{(4,0)} = \tilde{\mathbf{M}} \cdot \mathbf{C}_X^{(4,0)} \cdot \tilde{\mathbf{M}}^T, \quad (4)$$

$$\mathbf{C}_Y^{(2,2)} = \tilde{\mathbf{M}} \cdot \mathbf{C}_X^{(2,2)} \cdot \tilde{\mathbf{M}}^H, \quad (5)$$

in which $\mathbf{C}_X^{(4,0)}, \mathbf{C}_X^{(2,2)} \in \mathbb{C}^{N^2 \times N^2}$ are diagonal matrices, containing the source autocumulants, and in which $\tilde{\mathbf{M}} = \mathbf{M} \odot \mathbf{M}$, with $\odot$ the Kathri-Rao or column-wise Kronecker product. One can associate a generalized eigenvalue problem with the "tensor pencil" $(\mathcal{C}_Y^{(4,0)}, \mathcal{C}_Y^{(2,2)})$ as follows:

$$\mathbf{C}_Y^{(4,0)} (\mathbf{C}_Y^{(2,2)})^{\dagger *} = \tilde{\mathbf{M}} \cdot \left(\mathbf{C}_X^{(4,0)} (\mathbf{C}_X^{(2,2)})^{\dagger *} \right) \cdot \tilde{\mathbf{M}}^{\dagger *}. \quad (6)$$

This decomposition can be reduced to an ordinary EVD by multiplying each side of the equation with its complex conjugate:

$$\mathbf{C}_Y^{(4,0)} \cdot (\mathbf{C}_Y^{(2,2)})^{\dagger *} \cdot (\mathbf{C}_Y^{(4,0)})^{*} \cdot (\mathbf{C}_Y^{(2,2)})^{\dagger} = \tilde{\mathbf{M}} \cdot \mathbf{D} \cdot \tilde{\mathbf{M}}^{\dagger}, \quad (7)$$

in which $\mathbf{D} \stackrel{\text{def}}{=} \mathbf{C}_X^{(4,0)} \cdot (\mathbf{C}_X^{(2,2)})^{\dagger *} \cdot (\mathbf{C}_X^{(4,0)})^{*} \cdot (\mathbf{C}_X^{(2,2)})^{\dagger}$ is diagonal. If there are no mutually collinear pairs in the set $\{(|\kappa_r^{(4,0)}|, |\kappa_r^{(2,2)}|)\}_{1 \leqslant r \leqslant R}$ and if the matrices $\{M_r M_r^T\}_{1 \leqslant r \leqslant R}$ are linearly independent, then the EVD (7) is unique. Taking into account that the range of the mapping in (7) only contains symmetric matrices, we conclude that $R = N(N+1)/2$ factors can be identified. If the sample cumulants are perturbed by noise then the eigenvectors may not exactly correspond to rank-1 matrices. Best rank-1 approximations can then be used to initialize an optimization routine in which both sides of Eqs. (2) and (3) are matched.

For generic mixing matrices the multiplicity condition on the eigenvalues of Eq. (7) is not crucial. Actually one can allow a multiplicity S, bounded by $2S(S-1) \leqslant (N(N-1))^2$; then the rank-1 matrices in the eigenspace can still be identified in an essentially unique way. With this respect, we refer to [1]. One can use the same technique to deal with the possible ill-conditioning of the eigenvectors when eigenvalues are close.

3. A variant of the EVD

Eq. (6) is the unsymmetric version of the Tagaki EVD [9]. Let us have a closer look at this type of decomposition. In this section we use the following notation:

$$\mathbf{A} = \mathbf{Z} \cdot \mathbf{\Lambda} \cdot \mathbf{Z}^{-*}, \quad (8)$$

in which $\mathbf{Z}$ is regular and $\mathbf{\Lambda}$ diagonal. Eq. (8) e.g. originates from combining the equations

$$\mathbf{A_1} = \mathbf{Z} \cdot \mathbf{\Lambda_1} \cdot \mathbf{Z}^T, \quad (9)$$

$$\mathbf{A_2} = \mathbf{Z} \cdot \mathbf{\Lambda_2} \cdot \mathbf{Z}^H, \quad (10)$$

to a decomposition for $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{A_1} \cdot \mathbf{A_2}^{-*}$, as in the previous section.

First remark that

$$\mathbf{A} \cdot Z_k^* = \lambda_k Z_k \quad (11)$$

implies that, for arbitrary $r, \phi \in \mathbb{R}$, the "eigenvalue / eigenvector" pair $(re^{i2\phi}\lambda_k, re^{i\phi} Z_k)$ is also a solution of (11), and vice-versa ($r \neq 0$). Hence we can restrict ourselves to real-valued eigenvalues.

If $\mathbf{A}$ can be decomposed as in Eq. (8), then the components can be found from the ordinary EVD

$$\mathbf{A} \cdot \mathbf{A}^* = \mathbf{Z} \cdot \mathbf{\Lambda} \cdot \mathbf{\Lambda}^* \cdot \mathbf{Z}^{-1}. \quad (12)$$

On the other hand, for a given matrix $\mathbf{A}$, Eq. (12) might well admit complex conjugate eigenvalues. This corresponds to the fact that $\mathbf{A}$ cannot be decomposed as in Eq. (8); instead $\mathbf{\Lambda}$ should be replaced by a block-diagonal matrix consisting of antidiagonal (2×2)-blocks that contain the conjugate eigenvalue pairs. This, in turn, can be due to a similar structure of $\mathbf{\Lambda_2}$ in Eq. (10). The derivation of these assertions is technical and will be detailed in [8].

If the noise level is so high that complex eigenvalues do occur in Eq. (7), then one can still proceed in an approximate sense as follows. Instead of the EVD (12), let us consider the Schur decomposition

$$\mathbf{A} \cdot \mathbf{A}^* = \mathbf{Q} \cdot \tilde{\mathbf{R}} \cdot \mathbf{Q}^H, \quad (13)$$

in which $\mathbf{R}$ is upper triangular. From this unitary decomposition one can find a least-squares approximation of $\mathbf{A}\mathbf{A}^*$ that has real eigenvalues, by simply setting the diagonal of $\mathbf{R}$ equal to its real part; we denote the result by $\tilde{\mathbf{R}}$.

In terms of the QR-decomposition $\mathbf{Z} = \mathbf{Q}\mathbf{U}$, the remaining task in the estimation of $\mathbf{Z}$ is then the computation of the upper triangular eigenmatrix $\mathbf{U}$ of $\tilde{\mathbf{R}}$:

$$\tilde{\mathbf{R}} = \mathbf{U} \cdot \tilde{\mathbf{\Lambda}} \cdot \mathbf{U}^{-1}, \quad (14)$$

in which $\tilde{\mathbf{\Lambda}}$ is the diagonal part of $\tilde{\mathbf{R}}$. Complex conjugate eigenvalues of $\mathbf{A}\mathbf{A}^*$ yield eigenvalues of multiplicity 2 in Eq. (14). In the application under consideration, the indeterminacy in the eigenspace can be met by claiming that the corresponding columns of $\mathbf{Z}$ represent rank-1 matrices, as in the previous section.

4. Technique 2

In Section 2 the condition that $\tilde{\mathbf{M}}$ encodes rank-1 matrices, was only imposed after the computation of the EVD (7). However, by exploiting this condition one can extract from $\mathbf{C}_Y^{(2,2)}$, for a generic mixing matrix, R columns, bounded by $2R(R-1) \leqslant (N(N-1))^2$. We refer to [1]. In this section we will show that the scheme of [1] can easily be adapted to take knowledge of $\mathbf{C}_Y^{(4,0)}$ into account as well.

Assume that the range of $\mathbf{C}_Y^{(2,2)}$ is spanned by the matrices $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \ldots, \mathbf{B}^{(R)} \in \mathbb{C}^{N \times N}$. Because of Eq. (5), each matrix $\mathbf{B}^{(r)}$ can be expanded as

$$\mathbf{B}^{(r)} = \sum e_{rs} M_s M_s^T.$$

Along the lines of [1], it follows that the unknown matrix $\mathbf{E} \in \mathbb{C}^{R \times R}$ diagonalizes by congruence any matrix in the kernel of a linear matrix to fourth-order tensor mapping derived from $\{\mathbf{B}^{(r)}\}_{1 \leqslant r \leqslant R}$; this kernel is of dimension R. If we call the kernel matrices $\mathbf{K}_1^{(2,2)}, \mathbf{K}_2^{(2,2)}, \ldots, \mathbf{K}_R^{(2,2)}$, then we have that

$$\begin{aligned} \mathbf{E}^H \cdot \mathbf{K}_1^{(2,2)} \cdot \mathbf{E} &= \mathbf{D}_1^{(2,2)}, \\ \mathbf{E}^H \cdot \mathbf{K}_2^{(2,2)} \cdot \mathbf{E} &= \mathbf{D}_2^{(2,2)}, \\ &\vdots \\ \mathbf{E}^H \cdot \mathbf{K}_R^{(2,2)} \cdot \mathbf{E} &= \mathbf{D}_R^{(2,2)}, \end{aligned} \qquad (15)$$

in which $\mathbf{D}_1^{(2,2)}, \mathbf{D}_2^{(2,2)}, \ldots, \mathbf{D}_R^{(2,2)}$ are diagonal. Such a simultaneous congruence transformation also appeared in the derivation of an analytical constant modulus algorithm and was solved via a simultaneous generalized Schur decomposition by means of a QZ iteration or Jacobi iteration [10, 6].

$\mathbf{C}_Y^{(4,0)}$ can be addressed in a similar way. The resulting system of equations is

$$\begin{aligned} \mathbf{E}^T \cdot \mathbf{K}_1^{(4,0)} \cdot \mathbf{E} &= \mathbf{D}_1^{(4,0)}, \\ \mathbf{E}^T \cdot \mathbf{K}_2^{(4,0)} \cdot \mathbf{E} &= \mathbf{D}_2^{(4,0)}, \\ &\vdots \\ \mathbf{E}^T \cdot \mathbf{K}_R^{(4,0)} \cdot \mathbf{E} &= \mathbf{D}_R^{(4,0)}. \end{aligned} \qquad (16)$$

To solve Eqs. (15) and (16) simultaneously, modified QZ or Jacobi iterations can be devised [8].

By first computing the symmetric EVD of $\mathbf{C}_Y^{(2,2)}$, as in [1], the unknown matrix $\mathbf{E}$ can be restricted to the manifold of unitary matrices. (We assume that all the source cumulants have the same sign; if $\mathbf{C}_Y^{(2,2)}$ is indefinite, then one should work with J-unitary matrices, which leads to more complicated equations [8].) The joint diagonalization of (15) can then be realized by means of the Jacobi iteration derived in [2, 3]. We will now show that (16) can be cast in exactly the same scheme.

Let us denote the series of (2×2)-submatrices to be diagonalized by $\mathbf{A}_r^{(4,0)}$ $(1 \leqslant r \leqslant R)$ and let the unitary diagonalizer be given by

$$\mathbf{Q} = \begin{pmatrix} \cos\alpha & -\sin\alpha\, e^{j\phi} \\ \sin\alpha\, e^{-j\phi} & \cos\alpha \end{pmatrix}.$$

The off-diagonal entries of the transformed matrices

$$\mathbf{B}_r^{(4,0)} \stackrel{\text{def}}{=} \mathbf{Q}^T \cdot \mathbf{A}_r^{(4,0)} \cdot \mathbf{Q}$$

are then given by

$$(\mathbf{B}_r^{(4,0)})_{12} = \frac{1}{2} v^T \cdot g_r^{(4,0)},$$

in which

$$\begin{aligned} v^T &\stackrel{\text{def}}{=} (\cos 2\alpha, \sin 2\alpha \cos\phi, \sin 2\alpha \sin\phi), \\ g_r^{(4,0)} &\stackrel{\text{def}}{=} \begin{pmatrix} 2\,(\mathbf{A}_r^{(4,0)})_{12} \\ (\mathbf{A}_r^{(4,0)})_{22} - (\mathbf{A}_r^{(4,0)})_{11} \\ -j\,((\mathbf{A}_r^{(4,0)})_{22} + (\mathbf{A}_r^{(4,0)})_{11}) \end{pmatrix}. \end{aligned}$$

Consequently, the cost function

$$f^{(4,0)}(\mathbf{Q}) \stackrel{\text{def}}{=} \sum_r 2\,|(\mathbf{B}_r^{(4,0)})_{12}|^2$$

is given by

$$2f^{(4,0)}(\mathbf{Q}) = v^T \cdot \mathtt{Real}\{\mathbf{G}^{(4,0)} \cdot \mathbf{G}^{(4,0)H}\} \cdot v, \qquad (17)$$

in which

$$\mathbf{G}^{(4,0)} \stackrel{\text{def}}{=} \left(g_1^{(4,0)}, g_2^{(4,0)}, \ldots, g_R^{(4,0)}\right).$$

As far as (15) is concerned, [2] shows that the cost function

$$f^{(2,2)}(\mathbf{Q}) \stackrel{\text{def}}{=} \sum_r 2\,|(\mathbf{B}_r^{(2,2)})_{12}|^2$$

is given by

$$2f^{(2,2)}(\mathbf{Q}) = c - v^T \cdot \mathtt{Real}\{\mathbf{G}^{(2,2)} \cdot \mathbf{G}^{(2,2)H}\} \cdot v, \qquad (18)$$

in which c is a real constant and

$$\mathbf{G}^{(2,2)} \stackrel{\text{def}}{=} \left(g_1^{(2,2)}, g_2^{(2,2)}, \ldots, g_R^{(2,2)}\right),$$

with

$$g_r^{(2,2)} \stackrel{\text{def}}{=} \begin{pmatrix} (\mathbf{A}_r^{(2,2)})_{11} - (\mathbf{A}_r^{(2,2)})_{22} \\ 2\,\mathtt{Real}\{(\mathbf{A}_r^{(2,2)})_{12}\} \\ -2j\,\mathtt{Imag}\{(\mathbf{A}_r^{(2,2)})_{12}\} \end{pmatrix}.$$

From (17) and (18) it is clear that the joint cost function $f^{(4,0)} + f^{(2,2)}$ can be minimized by computing the smallest singular triplet of a real (3×3)-matrix; this can be done in a non-iterative way.

5. Conclusion

In this paper multilinear algebraic techniques for more-sources-than-sensors ICA were developed. A first technique was based on a tensor generalization of an unsymmetric variant of the Tagaki EVD. A second contribution was extending [2] to a scheme in which two observation cumulants are taken into account; this may increase the accuracy and robustness without seriously affecting the computational cost.

Acknowledgements

This research was partially supported by the Flemish Government (the Concerted Research Action GOA-MIPS, the FWO (Fund for Scientific Research - Flanders) project G.0240.99, the FWO Research Community Advanced Numerical Methods for Mathematical Modelling, the IWT Action Programme on Information Technology ITA/GBO/T23), the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs (the Interuniversity Poles of Attraction Programmes IUAP P4-02 and IUAP P4-24) and the European Commission (SCIENCE-ERNSI SC1-CT92-0779). Lieven De Lathauwer is a Post-Doctoral Research Assistant with the K.U.Leuven. Bart De Moor is a senior Research Associate with the F.W.O. and an Associate Professor with the K.U.Leuven. Joos Vandewalle is a Full Professor with the K.U.Leuven. The scientific responsibility is assumed by the authors.

References

[1] J.-F. Cardoso. Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors. *Proc. ICASSP-91*, 5: 3109-3112.

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

[3] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 17 (1): 161-164, 1996.

[4] P. Comon. Blind channel identification and extraction of more sources than sensors. *Proc. SPIE Conf.*, 2-13, San Diego, CA, U.S.A., July 19-24, 1998.

[5] P. Comon and O. Grellier. Non-linear inversion of underdetermined mixtures. *Proc. ICA'99*, 461-465, Aussois, France, Jan. 11-15, 1999.

[6] L. De Lathauwer. *Signal Processing based on Multilinear Algebra*. Ph.D. thesis, K.U.Leuven, E.E.Dept. (ESAT), Belgium, Sept. 1997.

[7] L. De Lathauwer, P. Comon, B. De Moor and J. Vandewalle. ICA algorithms for 3 sources and 2 sensors. *Proc. HOS'99*.

[8] L. De Lathauwer, B. De Moor and J. Vandewalle. ICA techniques for more sources than sensors. *Tech. Report*, ESAT/SISTA, K.U.Leuven (in preparation).

[9] R. Horn and C. Johnson. *Topics in Matrix Analysis*, Cambridge University Press, N.Y., 1991.

[10] A.-J. van der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. Signal Processing*, 44 (5): 1136-1155, 1996.