Handbook of Blind Source Separation, Independent Component Analysis and Applications

P. Comon and C. Jutten Eds



October 6, 2009

Glossary

$\mathbf{x}$	vector of components x_p, $1 \leq p \leq P$
s, x, y	sources, observations, separator outputs
N	number of sources
P	number of sensors
T	number of observed samples
$\star$	convolution
$\mathbf{A}$	matrix with components A_{ij}
$\mathbf{A}$, $\mathbf{B}$	mixing and separation matrices
$\mathbf{G}$, $\mathbf{W}$, $\mathbf{Q}$	global, whitening, and separating unitary matrices
$\check{g}$	Fourier transform of g
$\widehat{\mathbf{s}}$	estimate of quantity $\mathbf{s}$
$p_{\mathbf{x}}$	probability density of $\mathbf{x}$
ψ	joint score function
φ_i	marginal score function of source s_i
Φ	first characteristic function
Ψ	second characteristic function
$\mathbb{E}\,\mathbf{x}$, $\mathbb{E}\{\mathbf{x}\}$	mathematical expectation of $\mathbf{x}$
$I\{\mathbf{y}\}$ or $I(p_{\mathbf{y}})$	mutual information of $\mathbf{y}$
$K\{\mathbf{x};\mathbf{y}\}$ or $K(p_{\mathbf{x}};p_{\mathbf{y}})$	Kullback divergence between $p_{\mathbf{x}}$ and $p_{\mathbf{y}}$
$H\{\mathbf{x}\}$ or $H(p_{\mathbf{x}})$	Shannon entropy $\mathbf{x}$
$\mathcal{L}$	likelihood
$\mathcal{A}$, $\mathcal{B}$	mixing, and separating (non linear) operators
$\mathrm{cum}\{x_1, \ldots, x_P\}$	joint cumulant of variables $\{x_1, \ldots, x_P\}$
$\mathrm{cum}_R\{y\}$	marginal cumulant of order R of variable y
$\mathbf{Q}^{\mathrm{T}}$	transposition
$\mathbf{Q}^{\mathrm{H}}$	conjugate transposition
$\mathbf{Q}^{*}$	complex conjugation
$\mathbf{Q}^{\dagger}$	pseudo-inverse

Υ	contrast function
$\mathbb{R}$	real field
$\mathbb{C}$	complex field
$\widehat{\mathbf{A}}$	estimator of mixing matrix
diag $\mathbf{A}$	vector whose composents are the diagonal of matrix $\mathbf{A}$
Diag $\mathbf{a}$	diagonal matrix whos entries are those of vector $\mathbf{a}$
trace $\mathbf{A}$	trace of matrix $\mathbf{A}$
det $\mathbf{A}$	determinant of matrix $\mathbf{A}$
mean $\mathbf{a}$	arithmetic average of component of vector $\mathbf{a}$
$\check{s}(\nu)$	Fourier transform of process $s(t)$
$\otimes$	Kronecker product between matrices
$\otimes$	tensor product
$\bullet_j$	contraction over index j
krank$\{\mathbf{A}\}$	Kruskal's k-rank of matrix $\mathbf{A}$

Contents

Glossary **3**

1 Algebraic methods **7**
1.1 Introduction 7
1.1.1 Multilinear algebra 8
1.1.2 Higher-order statistics 8
1.1.3 Jacobi iteration 11
1.2 Independent component analysis 13
1.2.1 Algebraic formulation 13
1.2.2 Step 1: prewhitening 15
1.2.3 Step 2: fixing the rotational degrees of freedom using the higher-order cumulant 16
1.3 Diagonalization in least squares sense 17
1.3.1 Third-order real case 19
1.3.2 Third-order complex case 20
1.3.3 Fourth-order real case 21
1.3.4 Fourth-order complex case 22
1.4 Simultaneous diagonalization of matrix slices 22
1.4.1 Real case 25
1.4.2 Complex case 25
1.5 Simultaneous diagonalization of third-order tensor slices 26
1.6 Maximization of the tensor trace 26
Bibliography 28

Chapter 1

Algebraic Methods after Prewhitening

L. De Lathauwer

1.1 Introduction

In this chapter we discuss four prewhitening-based algebraic algorithms for Independent Component Analysis (ICA). The prewhitening takes into account the structure of the covariance matrix of the observed data. This is typically done by means of an Eigenvalue Decomposition (EVD) or a Singular Value Decomposition (SVD). The prewhitening only allows one to find the mixing matrix up to an orthogonal or unitary factor. In this chapter, the unknown factor is estimated from the higher-order cumulant tensor of the observed data. This involves orthogonal tensor decompositions that can be interpreted as higher-order generalizations of the EVD.

In the remainder of this Introduction, we introduce some basic material that is needed in the further discussion. Section 1.1.1 deals with some basic concepts of multilinear algebra. Section 1.1.2 deals with Higher-Order Statistics (HOS). Section 1.1.3 recalls the principle of a Jacobi iteration. The rest of the chapter is structured as follows. Section 1.2 introduces the decompositions of covariance matrix and cumulant tensor on which the methods are based, and explains the common structure of the algorithms. Four specific algorithms are subsequently discussed in Sections 1.3–1.6. The presentation is in terms of real-valued data. Wherever the generalization to the complex case is not straightforward, this will be explicitly stated.

Matlab code of the algorithms is available on the internet — see the corresponding entry in the Bibliography. Several variants of the algorithms can be thought of. In particular, it is possible to combine the decompositions of cumulant tensors of different order, as explained in [37].

1.1.1 Multilinear algebra

The outer product generalizes expressions of the type $\mathbf{a}\mathbf{b}^{\mathrm{T}}$, in which $\mathbf{a}$ and $\mathbf{b}$ are vectors.

DEFINITION 1.1 *The* outer product $\mathcal{A} \circ \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_P \times J_1 \times J_2 \times \ldots \times J_Q}$ *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_P}$ *and a tensor* $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_Q}$*, is defined by*

$$(\mathcal{A} \circ \mathcal{B})_{i_1 i_2 \ldots i_P j_1 j_2 \ldots j_Q} = a_{i_1 i_2 \ldots i_P} b_{j_1 j_2 \ldots j_Q},$$

for all values of the indices.

The entries of a Kth-order tensor $\mathcal{A}$, equal to the outer product of K vectors $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$, $\ldots$, $\mathbf{u}^{(K)}$, are thus given by $a_{i_1 i_2 \ldots i_K} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(K)}_{i_K}$.

The multiplication of a higher-order tensor with a matrix can be defined as follows.

DEFINITION 1.2 *The* mode-k product *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_K}$ *by a matrix* $\mathbf{U} \in \mathbb{R}^{J_k \times I_k}$*, denoted by* $\mathcal{A} \bullet_k \mathbf{U}$*, is an* $(I_1 \times I_2 \times \ldots \times I_{k-1} \times J_k \times I_{k+1} \ldots \times I_K)$*-tensor defined by*

$$(\mathcal{A} \bullet_k \mathbf{U})_{i_1 i_2 \ldots j_k \ldots i_K} = \sum_{i_k} a_{i_1 i_2 \ldots i_k \ldots i_K} u_{j_k i_k},$$

for all index values.

The mode-k product allows us to express the effect of a basis transformation in $\mathbb{R}^{I_k}$ on the tensor $\mathcal{A}$. In the literature sometimes the symbol $\times_k$ is used instead of $\bullet_k$.

By way of illustration, let us have a look at the matrix product $\mathbf{A} = \mathbf{U}^{(1)} \cdot \mathbf{B} \cdot \mathbf{U}^{(2)^{\mathrm{T}}}$, involving matrices $\mathbf{B} \in \mathbb{R}^{I_1 \times I_2}$, $\mathbf{U}^{(1)} \in \mathbb{R}^{J_1 \times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{J_2 \times I_2}$ and $\mathbf{A} \in \mathbb{R}^{J_1 \times J_2}$. Working with "generalized transposes" in the multilinear case (in which the fact that column vectors are transpose-free, would not have an inherent meaning), can be avoided by observing that the relationships of $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ (not $\mathbf{U}^{(2)^{\mathrm{T}}}$) with $\mathbf{B}$ are in fact completely similar. In the same way as $\mathbf{U}^{(1)}$ makes linear combinations of the rows of $\mathbf{B}$, $\mathbf{U}^{(2)}$ makes linear combinations of the columns. In the same way as the columns of $\mathbf{B}$ are multiplied by $\mathbf{U}^{(1)}$, its rows are multiplied by $\mathbf{U}^{(2)}$. In the same way as the columns of $\mathbf{U}^{(1)}$ are associated with the column space of $\mathbf{A}$, the columns of $\mathbf{U}^{(2)}$ are associated with the row space. This typical relationship is denoted by means of the $\bullet_k$-symbol: $\mathbf{A} = \mathbf{B} \bullet_1 \mathbf{U}^{(1)} \bullet_2 \mathbf{U}^{(2)}$. The symmetry is clear at the level of the individual entries: $a_{j_1 j_2} = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} u^{(1)}_{j_1 i_1} u^{(2)}_{j_2 i_2} b_{i_1 i_2}$, for all j_1, j_2. Note that $\mathbf{U}^{(1)} \cdot \mathbf{B} \cdot \mathbf{U}^{(2)^{\mathrm{H}}} = \mathbf{B} \bullet_1 \mathbf{U}^{(1)} \bullet_2 \mathbf{U}^{(2)^{*}}$.

1.1.2 Higher-order statistics

The basic HOS are higher-order moments and higher-order cumulants. Moment tensors of a real stochastic vector are defined as follows.

DEFINITION 1.3 *The Kth-order* moment *tensor $\mathcal{M}_{\mathbf{x}}^{(K)} \in \mathbb{R}^{I \times I \times \ldots \times I}$ of a real stochastic vector $\mathbf{x} \in \mathbb{R}^I$ is defined by the element-wise equation*

$$(\mathcal{M}_{\mathbf{x}}^{(K)})_{i_1 i_2 \ldots i_K} = \mathrm{mom}\{x_{i_1}, x_{i_2}, \ldots, x_{i_K}\} = \mathbb{E}\{x_{i_1} x_{i_2} \ldots x_{i_K}\}. \tag{1.1}$$

Moments are obtained from the coefficients of the first characteristic function of the random vector [36]. The first-order moment is equal to the mean. The second-order moment is the correlation matrix (following the definition adopted in e.g. [29], in which the mean is not subtracted).

On the other hand, cumulants of a real stochastic vector are defined as follows.

DEFINITION 1.4 *The Kth-order* cumulant *tensor $\mathcal{C}_{\mathbf{x}}^{(K)} \in \mathbb{R}^{I \times I \times \ldots \times I}$ of a real stochastic vector $\mathbf{x} \in \mathbb{R}^I$ is defined by the element-wise equation*

$$\begin{aligned}(\mathcal{C}_{\mathbf{x}}^{(K)})_{i_1 i_2 \ldots i_K} &= \mathrm{cum}\{x_{i_1}, x_{i_2}, \ldots, x_{i_K}\} \\ &= \sum (-1)^{L-1} (L-1)!\, \mathbb{E}\{\prod_{i \in \mathtt{A}_1} x_i\}\, \mathbb{E}\{\prod_{i \in \mathtt{A}_2} x_i\} \ldots \mathbb{E}\{\prod_{i \in \mathtt{A}_L} x_i\},\end{aligned} \tag{1.2}$$

where the summation involves all possible partitions $\{\mathtt{A}_1, \mathtt{A}_2, \ldots, \mathtt{A}_L\}$ $(1 \leqslant L \leqslant K)$ of the integers $\{i_1, i_2, \ldots, i_K\}$. For a real zero-mean stochastic vector $\mathbf{x}$ the cumulants up to order 4 are explicitly given by:

$$(\mathbf{c}_{\mathbf{x}})_i = \mathrm{cum}\{x_i\} = \mathbb{E}\{x_i\}, \tag{1.3}$$

$$(\mathbf{C}_{\mathbf{x}})_{i_1 i_2} = \mathrm{cum}\{x_{i_1}, x_{i_2}\} = \mathbb{E}\{x_{i_1} x_{i_2}\}, \tag{1.4}$$

$$(\mathcal{C}_{\mathbf{x}}^{(3)})_{i_1 i_2 i_3} = \mathrm{cum}\{x_{i_1}, x_{i_2}, x_{i_3}\} = \mathbb{E}\{x_{i_1} x_{i_2} x_{i_3}\}, \tag{1.5}$$

$$\begin{aligned}(\mathcal{C}_{\mathbf{x}}^{(4)})_{i_1 i_2 i_3 i_4} = \mathrm{cum}\{x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}\} &= \mathbb{E}\{x_{i_1} x_{i_2} x_{i_3} x_{i_4}\} \\ &\quad - \mathbb{E}\{x_{i_1} x_{i_2}\}\, \mathbb{E}\{x_{i_3} x_{i_4}\} \\ &\quad - \mathbb{E}\{x_{i_1} x_{i_3}\}\, \mathbb{E}\{x_{i_2} x_{i_4}\} \\ &\quad - \mathbb{E}\{x_{i_1} x_{i_4}\}\, \mathbb{E}\{x_{i_2} x_{i_3}\}.\end{aligned} \tag{1.6}$$

For every component x_i of $\mathbf{x}$ that has a nonzero mean, x_i has to be replaced in these formulas, except Eq. (1.3) and Eq. (1.2) when it applies to a first-order cumulant, by $x_i - \mathbb{E}\{x_i\}$.

In the same way as moments are obtained from the coefficients of the first characteristic function of the random vector, cumulants are obtained from its second characteristic function [36]. It turns out that, again, the first-order cumulant is the mean of the stochastic vector. The second-order cumulant is the covariance matrix. The interpretation of a cumulant of order higher than two is not straightforward, but the powerful properties listed below will demonstrate its importance. For the moment it suffices to state that cumulants of a set of random variables give an indication of their mutual statistical dependence (completely independent variables resulting in a zero cumulant), and that higher-order cumulants of a single random variable are some measure of its non-Gaussianity (cumulants of a Gaussian variable, for $K > 2$, being equal to zero).

The definitions above are given for real-valued random vectors. In the case of complex random vectors, one or more of the arguments of $\mathrm{mom}\{\cdot\}$ and $\mathrm{cum}\{\cdot\}$ may be complex conjugated. In this chapter we will in particular use the so-called "circular" covariance matrix $\mathbf{C}_{\mathbf{x}} = \mathrm{cum}\{\mathbf{x}, \mathbf{x}^*\}$ and cumulant $\mathcal{C}_{\mathbf{x}}^{(4)} = \mathrm{cum}\{\mathbf{x}, \mathbf{x}^*, \mathbf{x}^*, \mathbf{x}\}$.

At first sight higher-order moments, because of their straightforward definition, might seem more interesting quantities than higher-order cumulants. However, cumulants have a number of important properties, which they do not share with higher-order moments, such that in practice cumulants are more frequently used. We enumerate some of the most interesting properties [38, 39]:

1. *Symmetry:* real moments and cumulants are fully symmetric in their arguments, i.e.,

$$\begin{aligned}(\mathcal{M}_{\mathbf{x}}^{(K)})_{i_1 i_2 \ldots i_K} &= (\mathcal{M}_{\mathbf{x}}^{(K)})_{\mathtt{P}(i_1 i_2 \ldots i_K)}, \\ (\mathcal{C}_{\mathbf{x}}^{(K)})_{i_1 i_2 \ldots i_K} &= (\mathcal{C}_{\mathbf{x}}^{(K)})_{\mathtt{P}(i_1 i_2 \ldots i_K)},\end{aligned}$$

   in which $\mathtt{P}$ is an arbitrary permutation of the indices.

2. *Multilinearity:* if a real stochastic vector $\mathbf{x}$ is transformed into a stochastic vector $\tilde{\mathbf{x}}$ by a matrix multiplication $\tilde{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x}$, then we have:

$$\mathcal{M}_{\tilde{\mathbf{x}}}^{(K)} = \mathcal{M}_{\mathbf{x}}^{(K)} \bullet_1 \mathbf{A} \bullet_2 \mathbf{A} \ldots \bullet_K \mathbf{A}, \tag{1.7}$$

$$\mathcal{C}_{\tilde{\mathbf{x}}}^{(K)} = \mathcal{C}_{\mathbf{x}}^{(K)} \bullet_1 \mathbf{A} \bullet_2 \mathbf{A} \ldots \bullet_K \mathbf{A}. \tag{1.8}$$

   (In the complex case, some of the matrices $\mathbf{A}$ may be complex conjugated, depending on the complex conjugation pattern adopted in the definition of the moment or cumulant.) The characteristic way in which moments and cumulants change under basis transformations, is actually what makes that we can call them *tensors* [36]. Eqs. (1.7–1.8) are in fact more general: the matrix $\mathbf{A}$ does not have to represent a basis transformation; it does not even have to be square.

3. *Even distribution:* if a real random variable x has an even probability density function $p_x(x)$, i.e., $p_x(x)$ is symmetric about the origin, then the odd moments and cumulants of x vanish.

4. *Partitioning of independent variables:* if a subset of K stochastic variables x_1, x_2, …, x_K is independent of the other variables, then we have:

$$\mathrm{cum}\{x_1, x_2, \ldots, x_K\} = 0.$$

   This property does not hold in general for moments (e.g. for two mutually independent random variables x and y we have that $\mathrm{mom}\{x, x, y, y\} = \mathbb{E}\{x^2\} \cdot \mathbb{E}\{y^2\}$, which does not vanish, unless one of the variables is identically equal to zero). A consequence of the property is that a higher-order cumulant of a stochastic vector that has mutually independent components, is a diagonal tensor, i.e., only the entries of which all the indices

are equal can be different from zero. This very strong algebraic condition is the basis of all the ICA techniques that will be discussed in this chapter.

5. *Sum of independent variables:* if the stochastic variables x_1, x_2, ..., x_K are mutually independent from the stochastic variables y_1, y_2, ..., y_K, then we have:

$$\begin{aligned}&\mathrm{cum}\{x_1 + y_1, x_2 + y_2, \ldots, x_K + y_K\} = \\ &\mathrm{cum}\{x_1, x_2, \ldots, x_K\} + \mathrm{cum}\{y_1, y_2, \ldots, y_K\}.\end{aligned}$$

The cumulant tensor of a sum of independent random vectors is the sum of the individual cumulants. This property does not hold for moments either. As a matter of fact, it explains the term "cumulant". (One could expand $\mathrm{mom}\{x_1 + y_1, x_2 + y_2, \ldots, x_K + y_K\}$ as a sum over all possible x - y combinations, but the cross-terms, containing x- as well as y-entries, do not necessarily vanish, as for cumulants — see the previous property.)

6. *Non-Gaussianity:* if y is a Gaussian variable with the same mean and variance as a given stochastic variable x, then we have for $K \geqslant 3$:

$$\mathcal{C}_x^{(K)} = \mathcal{M}_x^{(K)} - \mathcal{M}_y^{(K)}.$$

As a consequence, higher-order cumulants of a Gaussian variable are zero. In combination with the multilinearity property, we observe that higher-order cumulants have the interesting property to be blind for additive Gaussian noise. Namely, if a stochastic variable x is corrupted by additive Gaussian noise b, i.e.,

$$\hat{x} = x + b,$$

then we nevertheless have that

$$\mathcal{C}_{\hat{x}}^{(K)} = \mathcal{C}_x^{(K)} + \mathcal{C}_b^{(K)} = \mathcal{C}_x^{(K)}.$$

Generally speaking, it becomes harder to estimate HOS from sample data as the order increases, i.e., longer datasets are required to obtain the same accuracy [2, 33]. Hence in practice the use of HOS is usually restricted to third- and fourth-order cumulants. For symmetric distributions fourth-order cumulants are used, since the third-order cumulants vanish, as mentioned in the third property.

1.1.3 Jacobi iteration

The ICA algorithms discussed in this chapter are of the Jacobi type. Therefore we repeat in this section the principle of a *Jacobi iteration*. We take the diagonalization of an Hermitean matrix as example. For a more extensive discussion, we refer to [30].

Given an Hermitean matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$, we want to find a unitary matrix $\mathbf{U} \in \mathbb{C}^{I \times I}$ such that

$$\mathbf{B} = \mathbf{U} \cdot \mathbf{A} \cdot \mathbf{U}^{\mathrm{H}} \tag{1.9}$$

is diagonal. In other words, we look for a unitary matrix $\mathbf{U}$ that minimizes the cost function

$$f(\mathbf{U}) = \mathrm{off}(\mathbf{B})^2 = \sum_{i \neq j} |b_{ij}|^2 . \tag{1.10}$$

Because $\mathbf{U}$ is unitary, this is equivalent to maximizing the objective function

$$g(\mathbf{U}) = \|\mathrm{diag}(\mathbf{B})\|^2 = \sum_{i} |b_{ii}|^2 . \tag{1.11}$$

Any unitary matrix can, up to multiplication by a diagonal matrix $\mathbf{D}$ of which the diagonal entries are unit-modulus, be written as a product of elementary Jacobi rotation matrices $\mathbf{J}(p, q, c, s)$, defined for $p < q$ by

$$\mathbf{J}(p,q,c,s) = \begin{array}{c} \begin{array}{ccccccc} & & p & & q & & \end{array} \\ \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & -s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \overline{s} & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ \\ p \\ \\ q \\ \\ \end{array} \tag{1.12}$$

with $(c, s) \in \mathbb{R} \times \mathbb{C}$ such that $c^2 + |s|^2 = 1$. The matrix $\mathbf{D}$ does not change the cost function and can therefore be left aside. The following explicit parameterizations of $\mathbf{J}$ will be used in Sections 1.3–1.6:

$$c = \cos\alpha \quad s = \sin\alpha \qquad \alpha \in [0, 2\pi) \quad \text{(real case)} \tag{1.13}$$

$$c = \tfrac{1}{\sqrt{1+\theta^2}} \quad s = -\tfrac{\theta}{\sqrt{1+\theta^2}} \qquad \theta \in \mathbb{R} \quad \text{(real case)} \tag{1.14}$$

$$c = \cos\alpha \quad s = \sin\alpha e^{i\phi} \qquad \alpha, \phi \in [0, 2\pi) \quad \text{(complex case)} \tag{1.15}$$

$$c = \tfrac{1}{\sqrt{1+|\theta|^2}} \quad s = -\tfrac{\theta}{\sqrt{1+|\theta|^2}} \qquad \theta \in \mathbb{C} \quad \text{(complex case).} \tag{1.16}$$

Due to the unitarity of $\mathbf{J}$, the only part of $\mathbf{J} \cdot \mathbf{A} \cdot \mathbf{J}^{\mathrm{H}}$ that affects the cost function, is the (2×2) submatrix at the intersection of rows and columns p and q. An explicit expression for the elementary rotation that exactly diagonalizes the (2×2) submatrix is easy to derive. It can be verified that the cost function is periodic in the rotation angle with period $\pi/2$. Hence, one can always choose an inner rotation, i.e., a rotation over an angle in the interval $(-\pi/4, \pi/4]$. In this way, implicit re-ordering of columns and rows, which may hamper convergence, is avoided. A Jacobi iteration consists of successively applying Jacobi rotations to $\mathbf{A}$, for different (p, q), until the cost function has been minimized. In the process, the unitary matrix $\mathbf{U}$ is gradually built up as the product of all elementary rotations.

We now turn our attention to the choice of (p, q). From the standpoint of maximizing the reduction of the cost function $f(\mathbf{J})$, it makes sense to choose

Table 1.1: Cyclic Jacobi for the diagonalization of an Hermitean matrix.

$\mathbf{U} = \mathbf{I}$
while $\text{off}(\mathbf{A}) > \epsilon$ do
    for $p = 1 : I - 1$ do
        for $q = p + 1 : I$ do
            Compute optimal Jacobi rotation $\mathbf{J}(p, q)$
            $\mathbf{A} \leftarrow \mathbf{J} \cdot \mathbf{A} \cdot \mathbf{J}^{\text{H}}$
            $\mathbf{U} \leftarrow \mathbf{U} \cdot \mathbf{J}^{\text{H}}$
        end for
    end for
end while

---

the pair (p, q) for which $|a_{pq}|^2$ is maximal. This is what is done in the *classical Jacobi algorithm*. Since

$$\text{off}(\mathbf{A})^2 \leqslant I(I-1)\,|a_{pq}|^2\,,$$

we have that

$$\text{off}(\mathbf{J} \cdot \mathbf{A} \cdot \mathbf{J}^{\text{H}})^2 \leqslant \left(1 - \frac{2}{I(I-1)}\right) \text{off}(\mathbf{A})^2, \tag{1.17}$$

such that the Jacobi algorithm for diagonalization of an Hermitean matrix converges at least linearly to a diagonal matrix. It can be proved that the asymptotic convergence rate is actually quadratic. However, finding the optimal (p, q) in each step is expensive. Therefore, one rather follows a fixed order when going through the different subproblems. This technique is called the *cyclic Jacobi algorithm*. Cyclic Jacobi also converges quadratically. The algorithm is outlined in Table 1.1.

1.2 Independent component analysis

1.2.1 Algebraic formulation

The basic linear ICA model is denoted by:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} + \mathbf{b} = \tilde{\mathbf{x}} + \mathbf{b}, \tag{1.18}$$

in which $\mathbf{x} \in \mathbb{R}^P(\mathbb{C}^P)$ is the *observation vector*, $\mathbf{s} \in \mathbb{R}^N(\mathbb{C}^N)$ is the *source vector* and $\mathbf{b} \in \mathbb{R}^P(\mathbb{C}^P)$ represents additive *noise*. $\tilde{\mathbf{x}} \in \mathbb{R}^P(\mathbb{C}^P)$ is the *signal part* of the observations. $\mathbf{A} \in \mathbb{R}^{P\times N}(\mathbb{C}^{P\times N})$ is the *mixing matrix*; its range is the *signal subspace*.

In this chapter we assume that the mixing vectors are linearly independent. This implies that the number of sensors P is not smaller than the number

of sources N. It is actually possible to estimate the mixing matrix in cases where $P < N$. For algebraic approaches to this so-called *underdetermined* ICA problem we refer to [13, 17, 18, 21, 22, 23, 24] and references therein.

In this chapter we focus on the estimation of the mixing matrix. Since this matrix is nonsingular, with at least as many rows as columns, the sources may afterwards be estimated by premultiplying the observations with the pseudo-inverse of the estimate $\widehat{\mathbf{A}}$:

$$\widehat{\mathbf{s}} = \widehat{\mathbf{A}}^{\dagger} \cdot \mathbf{x}.$$

This equation implements in fact a *Linear Constrained Minimum Variance* (LCMV) beamformer, which minimizes the mutual interference of the sources. One may also follow other beamforming strategies [42].

Using the properties discussed in Section 1.1.2, the model (1.18) leads in the real case to the following two key equations:

$$\mathbf{C}_{\mathbf{x}} = \mathbf{A} \cdot \mathbf{C}_{\mathbf{s}} \cdot \mathbf{A}^{\mathrm{T}} + \mathbf{C}_{\mathbf{b}} = \mathbf{C}_{\mathbf{s}} \bullet_1 \mathbf{A} \bullet_2 \mathbf{A} + \mathbf{C}_{\mathbf{b}}, \tag{1.19}$$

$$\mathcal{C}_{\mathbf{x}}^{(K)} = \mathcal{C}_{\mathbf{s}}^{(K)} \bullet_1 \mathbf{A} \bullet_2 \mathbf{A} \ldots \bullet_K \mathbf{A} + \mathcal{C}_{\mathbf{b}}^{(K)}. \tag{1.20}$$

(In the complex case, some of the matrices $\mathbf{A}$ may be complex conjugated, depending on the complex conjugation pattern adopted for the covariance matrix and cumulant tensor.) The crucial point in (1.19–1.20) is that $\mathbf{C}_{\mathbf{s}}$ and $\mathcal{C}_{\mathbf{s}}^{(K)}$ are diagonal, because of the statistical independence of the sources. Furthermore, $\mathcal{C}_{\mathbf{b}}^{(K)}$ vanishes if the noise is Gaussian.

Eqs. (1.19–1.20) can also be written as

$$\mathbf{C}_{\mathbf{x}} = \sum_{n=1}^{N} \sigma_{s_n}^2 \, \mathbf{a}_n \mathbf{a}_n^{\mathrm{T}} + \mathbf{C}_{\mathbf{b}} = \sum_{n=1}^{N} \sigma_{s_n}^2 \, \mathbf{a}_n \circ \mathbf{a}_n + \mathbf{C}_{\mathbf{b}}, \tag{1.21}$$

$$\mathcal{C}_{\mathbf{x}}^{(K)} = \sum_{n=1}^{N} c_{s_n}^{(K)} \, \mathbf{a}_n \circ \mathbf{a}_n \circ \ldots \circ \mathbf{a}_n + \mathcal{C}_{\mathbf{b}}^{(K)}, \tag{1.22}$$

in which $\sigma_{s_n}^2$ and $c_{s_n}^{(K)}$ denote the variance and the Kth-order cumulant of the nth source, respectively, and in which $\mathbf{a}_n$ denotes the nth column of $\mathbf{A}$. Let us for the moment make abstraction of the noise term. Eq. (1.21) is an expansion of the covariance of the observations in a sum of rank-1 terms, where each term consists of the contribution of one particular source. Eq. (1.22) is an analogous decomposition of the higher-order cumulant of the observations. The latter is known as a *Canonical Decomposition* (CANDECOMP) or *Parallel Factor Decomposition* (PARAFAC) of the tensor $\mathcal{C}_{\mathbf{x}}^{(K)}$ [7, 14, 17, 20, 27, 31, 35, 40].

It is well-known that Eq. (1.19)/(1.21) does not contain enough information to allow for the full estimation of the mixing matrix. This will be explained in detail in Section 1.2.2. However, the situation is very different at the tensor level. Under our conditions on the mixing matrix, and assuming that all source cumulants $c_{s_n}^{(K)}$ are nonzero, decomposition (1.20)/(1.22) is unique, up to trivial indeterminacies. This allows us to solve the ICA problem. The strategy behind the methods discussed in this chapter is as follows. First we extract as

much information as possible from Eq. (1.19)/(1.21). How this can be done, is explained in Section 1.2.2. The remaining free parameters are subsequently estimated using (1.20)/(1.22). This problem will in general terms be addressed in Section 1.2.3. Specific algorithms are discussed in Sections 1.3–1.6.

Other strategies are possible. As explained in the previous paragraph, the mixing matrix could in principle be estimated from Eq. (1.20)/(1.22) alone, without using second-order statistics. One could also combine the second-order and higher-order constraints by giving them each a finite weight. For such alternative approaches we refer to [13, 19, 22, 44] and references therein.

In general, it makes sense to extract much information from Eq. (1.19)/(1.21). First, second-order statistics can be estimated more reliably than higher-order statistics, as already mentioned at the end of Section 1.1.2 [2, 33]. Second, after prewhitening the remaining problem consists of the determination of an orthogonal (unitary) matrix. That is, after prewhitening, the matrix that has to be determined, is optimally conditioned. Prewhitening may be less appropriate when the data are subject to Gaussian noise with unknown color. In that case, the effect of the noise on the second-order statistics cannot be mitigated in the way explained in Section 1.2.2. An error that is introduced in a prewhitening step cannot be fully compensated in the step that follows [4, 28]. On the other hand, higher-order cumulants are theoretically insensitive to Gaussian noise, as explained in Section 1.1.2.

1.2.2 Step 1: prewhitening

Briefly, the goal of *prewhitening* is to transform the observation vector $\mathbf{x}$ into an other stochastic vector, $\mathbf{z}$, which has unit covariance. This involves the multiplication of $\mathbf{x}$ with the inverse of a square root of its covariance matrix $\mathbf{C_x}$. When $N < P$, a projection of $\mathbf{x}$ onto the signal subspace is carried out.

Let us discuss the problem in more detail. First, we observe that the covariance matrix $\mathbf{C_x}$ takes the following form (for the moment, the noise term in Eq. (1.18) is neglected, for clarity):

$$\mathbf{C_x} = \mathbf{A} \cdot \mathbf{C_s} \cdot \mathbf{A}^{\mathrm{T}}, \tag{1.23}$$

in which the covariance $\mathbf{C_s}$ of $\mathbf{s}$ is diagonal, since the source signals are uncorrelated. (Throughout this section, the transpose is replaced by the Hermitean transpose in the complex case.) Assuming that the source signals have unit variance (without loss of generality, as we may appropriately rescale the mixing vectors as well), we have:

$$\mathbf{C_x} = \mathbf{A} \cdot \mathbf{A}^{\mathrm{T}}.$$

A first observation is that the number of sources can be deduced from the rank of $\mathbf{C_x}$. Substitution of the SVD of the mixing matrix $\mathbf{A} = \mathbf{U} \cdot \mathbf{\Delta} \cdot \mathbf{Q}^{\mathrm{T}}$ shows that the EVD of the observed covariance allows us to estimate the components $\mathbf{U}$ and $\mathbf{\Delta}$ while the factor $\mathbf{Q}$ remains unknown:

$$\mathbf{C_x} = \mathbf{U} \cdot \mathbf{\Delta}^2 \cdot \mathbf{U}^{\mathrm{T}} = (\mathbf{U}\mathbf{\Delta}) \cdot (\mathbf{U}\mathbf{\Delta})^{\mathrm{T}}. \tag{1.24}$$

Hence the signal subspace can be estimated from the second-order statistics of the observations, but the actual mixing matrix remains unknown up to an orthogonal factor. Instead of working via the EVD of $\mathbf{C_x}$, one may also obtain $\mathbf{U}$ and $\mathbf{\Delta}$ from the SVD of the data matrix. EVD/SVD-based prewhitening amounts to a *Principal Component Analysis* (PCA) [32].

The effect of the additive noise term $\mathbf{b}$ can be neutralized by replacing $\mathbf{C_x}$ by the noise-free covariance $\mathbf{C_x} - \mathbf{C_b}$. In the case of spatially white noise (i.e., noise of which the components are mutually uncorrelated and have the same variance), $\mathbf{C_b}$ takes the form of $\sigma_b^2\,\mathbf{I}$, in which σ_b^2 is the variance of the noise on each data channel. In a more-sensors-than-sources setup, σ_b^2 can be estimated as the mean of the "noise eigenvalues", i.e., the smallest $P - N$ eigenvalues, of $\mathbf{C_x}$. The number of sources is estimated as the number of significant eigenvalues of $\mathbf{C_x}$; for a detailed procedure, we refer to [43].

After computation of $\mathbf{U}$ and $\mathbf{\Delta}$, a standardized random vector $\mathbf{z}$ can be defined as

$$\mathbf{z} = \mathbf{\Delta}^\dagger \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{x}. \tag{1.25}$$

This vector has unit covariance. It is related to the source vector in the following way:

$$\mathbf{z} = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{s} + \mathbf{\Delta}^\dagger \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{b}. \tag{1.26}$$

1.2.3 Step 2: fixing the rotational degrees of freedom using the higher-order cumulant

In this section we explain in general terms how the remaining degrees of freedom can be fixed, i.e., how the matrix $\mathbf{Q}$ that contains the right singular vectors of the mixing matrix can be estimated. As we have already exploited the information provided by the second-order statistics of the observations, we now resort to their HOS.

Assuming that the noise is Gaussian, we obtain from (1.26) that in the real case the higher-order cumulant of the standardized random vector $\mathbf{z}$ is given by

$$\mathcal{C}_{\mathbf{z}}^{(K)} = \mathcal{C}_{\mathbf{s}}^{(K)} \bullet_1 \mathbf{Q}^{\mathrm{T}} \bullet_2 \mathbf{Q}^{\mathrm{T}} \ldots \bullet_K \mathbf{Q}^{\mathrm{T}}. \tag{1.27}$$

(In the complex case, some of the matrices $\mathbf{Q}$ may be complex conjugated, depending on the complex conjugation pattern adopted in the definition of the cumulant.) Let us write $\mathbf{W} = \mathbf{Q}^{\mathrm{T}}$ for convenience. Then (1.27) can also be written as

$$\mathcal{C}_{\mathbf{z}}^{(K)} = \sum_{n=1}^{N} c_{s_n}^{(K)} \mathbf{w}_n \circ \mathbf{w}_n \circ \ldots \circ \mathbf{w}_n. \tag{1.28}$$

The standardized cumulant $\mathcal{C}_{\mathbf{z}}^{(K)}$ is related to the Kth-order cumulant of the observations by the multilinearity property:

$$\mathcal{C}_{\mathbf{z}}^{(K)} = \mathcal{C}_{\mathbf{x}}^{(K)} \bullet_1 (\mathbf{U} \cdot \mathbf{\Delta})^\dagger \bullet_2 (\mathbf{U} \cdot \mathbf{\Delta})^\dagger \ldots \bullet_K (\mathbf{U} \cdot \mathbf{\Delta})^\dagger. \tag{1.29}$$

We repeat that the source cumulant $\mathcal{C}_{\mathbf{s}}^{(K)}$ is theoretically a diagonal tensor, since the source signals are not only uncorrelated but also higher-order independent.

Hence Eq. (1.27) is in fact a symmetric EVD-like tensor decomposition. It can be shown that this decomposition is up to trivial indeterminacies unique if at most one diagonal entry of $\mathcal{C}_{\mathbf{s}}^{(K)}$ is equal to zero [19, pp. 127–128]. Note again that this is very different from the situation for matrices. In the matrix case we have for instance $\mathbf{I} = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{I} \cdot \mathbf{Q} = \mathbf{V}^{\mathrm{T}} \cdot \mathbf{I} \cdot \mathbf{V}$ for arbitrary orthogonal $\mathbf{V}$. This is actually what makes that we cannot find $\mathbf{Q}$ from the second-order statistics of the observations.

On the other hand, simply counting the degrees of freedom in Eq. (1.29) shows that in general a higher-order tensor cannot be diagonalized by means of orthogonal (unitary) transformations. In the real case for instance, the symmetric tensor $\mathcal{C}_{\mathbf{z}}^{(K)}$ contains $N(N+1)\dots(N+K-1)/K!$ independent entries, while the decomposition allows only $N(N-1)/2$ (orthogonal factor $\mathbf{Q}$) $+ N$ (diagonal of $\mathcal{C}_{\mathbf{s}}^{(K)}$) degrees of freedom. This means that if $\mathcal{C}_{\mathbf{z}}^{(K)}$ is not perfectly known (due to finite datalength, non-Gaussian additive noise, etc.), its estimate can in general not be fully diagonalized. This leaves room for different algorithms, which deal with the estimation error in a different manner. Four approaches will be discussed in the following sections. The general structure of prewhitening-based algorithms that use the higher-order cumulant is summarized in Table 1.2.

1.3 Diagonalization in least squares sense

The first reliable algebraic algorithm for ICA was derived by Comon [10]. This algorithm will be denoted as *COM2*. It works as follows. Since the higher-order cumulant cannot in general be diagonalized by means of orthogonal (unitary) transformations, we will make it as diagonal as possible in least squares sense. Denote

$$\mathcal{C}' = \mathcal{C}_{\mathbf{z}}^{(K)} \bullet_1 \mathbf{U} \bullet_2 \mathbf{U} \bullet_3 \dots \bullet_K \mathbf{U}. \tag{1.30}$$

(Some of the matrices $\mathbf{U}$ may be complex conjugated in the complex case.) We look for an orthogonal (unitary) matrix $\mathbf{U}$ that minimizes the cost function

$$f_{K,2}(\mathbf{U}) = \mathrm{off}(\mathcal{C}')^2 = \sum_{\neg(n_1=n_2=\dots=n_K)} \left| c'_{n_1 n_2 \dots n_K} \right|^2, \tag{1.31}$$

Because $\mathbf{U}$ is orthogonal (unitary), this is equivalent to maximizing the contrast function

$$g_{K,2}(\mathbf{U}) = \|\mathrm{diag}(\mathcal{C}')\|^2 = \sum_n \left| c'_{nn\dots n} \right|^2. \tag{1.32}$$

The idea is visualized in Fig. 1.1.

The optimal $\mathbf{U}$ is obtained by means of a higher-order generalization of the Jacobi iteration described in Section 1.1.3. Due to the orthogonality (unitarity) of $\mathbf{J}$ in (1.12), the only part of $\mathcal{C}_{\mathbf{z}}^{(K)}$ that affects the contrast function, is the $(2 \times 2 \times \dots \times 2)$-subtensor $\tilde{\mathcal{C}}$ formed by the entries of which all indices are equal to either p or q. Like $\mathcal{C}_{\mathbf{z}}^{(K)}$, $\tilde{\mathcal{C}}$ cannot in general be exactly diagonalized. Hence, contrary to the matrix case, we cannot say a priori which elementary

Table 1.2: Prewhitening-based ICA using the higher-order cumulant.

1. Prewhitening stage (PCA):
   - Estimate covariance $\mathbf{C}_{\mathbf{x}}$.
   - EVD: $\mathbf{C}_{\mathbf{x}} = \mathbf{U} \cdot \mathbf{\Delta}^2 \cdot \mathbf{U}^{\mathrm{T}}$
   - Standardize data: $\mathbf{z} = \mathbf{\Delta}^{\dagger} \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{x}$.

   (Section 1.2.2.)
2. Higher-order stage:
   - Estimate cumulant $\mathcal{C}_{\mathbf{z}}^{(K)}$.
   - (Approximate) diagonalization:

     $$\mathcal{C}_{\mathbf{z}}^{(K)} = \mathcal{C}_{\mathbf{s}}^{(K)} \bullet_1 \mathbf{Q}^{\mathrm{T}} \bullet_2 \mathbf{Q}^{\mathrm{T}} \ldots \bullet_K \mathbf{Q}^{\mathrm{T}},$$

     in which $\mathbf{Q}$ is orthogonal. A class of algebraic algorithms:
     - COM2 (Section 1.3).
     - JADE (Section 1.4).
     - STOTD (Section 1.5).
     - COM1 (Section 1.6).

   (Section 1.2.3.)
3. Mixing matrix: $\mathbf{A} = \mathbf{U} \cdot \mathbf{\Delta} \cdot \mathbf{Q}^{\mathrm{T}}$.
4. Source estimation by means of LCMV or other type of beamforming.

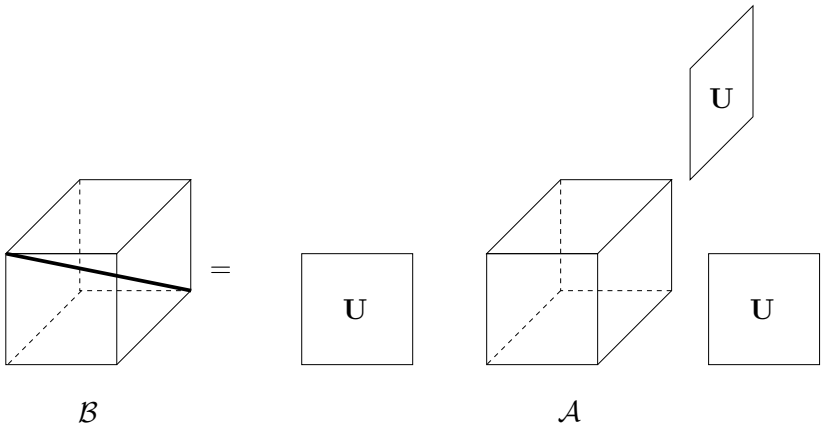


Figure 1.1: Visualization of the (approximate) diagonalization of a third-order tensor $\mathcal{A}$ by an orthogonal congruence transformation.

subtensor will most increase the contrast function. Therefore the rotation pairs (p, q) are swept in a cyclic way. The maximal diagonalization of elementary subtensors gradually minimizes the cost function, so we can expect convergence to at least a local minimum. Contrary to the matrix case, the cost function can have spurious local minima. The lack of higher-order equivalent of (1.17) makes it hard to prove convergence to the global optimum, or to formally study the convergence speed. In practice, the algorithm seems to converge reasonably fast and local optima do not seem to pose a problem.

We now explain for a number of specific cases how the optimal Jacobi rotation can be determined.

1.3.1 Third-order real case

We repeat that the $(2 \times 2 \times 2)$ symmetric tensor to be diagonalized is denoted by $\tilde{\mathcal{C}}$. We define $\mathcal{C} = \tilde{\mathcal{C}} \bullet_1 \mathbf{J} \bullet_2 \mathbf{J} \bullet_3 \mathbf{J}$, in which $\mathbf{J}$ is an elementary (2×2) Jacobi rotation matrix. The original derivation started from the parameterization (1.14) [11]. Here we use the parameterization (1.13).

We construct a real symmetric (2×2)-matrix $\mathbf{B}$, as follows:

$$\begin{aligned} b_{11} &= a_1, \\ b_{12} &= 3a_4/2, \\ b_{22} &= 9a_2/4 + 3a_3/2 + a_1/4, \end{aligned}$$

in which the auxiliary variables a_1, a_2, a_3, a_4 are given by

$$\begin{aligned} a_1 &= \tilde{c}_{111}^2 + \tilde{c}_{222}^2, \\ a_2 &= \tilde{c}_{112}^2 + \tilde{c}_{122}^2, \\ a_3 &= \tilde{c}_{111}\tilde{c}_{122} + \tilde{c}_{112}\tilde{c}_{222}, \\ a_4 &= \tilde{c}_{122}\tilde{c}_{222} - \tilde{c}_{111}\tilde{c}_{112}. \end{aligned}$$

It can be proved that

$$\sum_{i=1}^{2} c_{iii}^2 = \mathbf{q}^{\mathrm{T}} \cdot \mathbf{B} \cdot \mathbf{q},$$

in which $\mathbf{q} = (\cos(2\alpha) \;\; \sin(2\alpha))^{\mathrm{T}}$ [26]. Hence the optimal rotation can be found by computing the dominant eigenvector of $\mathbf{B}$ and normalizing it to unit length. It is clear that the function $\sum_{i=1}^{2} c_{iii}^2$ is periodic in the rotation angle, with period $\pi/2$, as the sign of $\mathbf{q}$ is of no importance. The sign of the dominant eigenvector can be chosen such that $\mathbf{J}$ is an inner rotation. The actual elements of the optimal inner Jacobi rotation can then be obtained from the entries of $\mathbf{q}$ by using the basic goniometric relations $\cos\alpha = \sqrt{1 + \cos(2\alpha)/2}$ and $\sin\alpha = \sin(2\alpha)/(2\cos\alpha)$.

1.3.2 Third-order complex case

In this section we assume that $\mathcal{C}_{\mathbf{z}}^{(3)}$ is defined by $(\mathcal{C}_{\mathbf{z}}^{(3)})_{ijk} = \mathrm{cum}\{z_i, z_j^*, z_k^*\}$. The $(2 \times 2 \times 2)$ tensor that has to be diagonalized, is denoted by $\tilde{\mathcal{C}}$. We define $\mathcal{C} = \tilde{\mathcal{C}} \bullet_1 \mathbf{J} \bullet_2 \mathbf{J}^* \bullet_3 \mathbf{J}^*$, in which $\mathbf{J}$ is an elementary (2×2) complex Jacobi rotation matrix. We use the parameterization (1.15).

We construct a real symmetric (3×3)-matrix $\mathbf{B}$, as follows:

$$\begin{aligned} b_{11} &= a_1, \\ b_{12} &= \mathrm{Im}(v_1) + \mathrm{Im}(v_2), \\ b_{13} &= \mathrm{Re}(v_1) - \mathrm{Re}(v_2), \\ b_{22} &= v_4 - \mathrm{Re}(v_3), \\ b_{23} &= \mathrm{Im}(v_3), \\ b_{33} &= v_4 + \mathrm{Re}(v_3), \end{aligned}$$

in which $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote, respectively, the real and complex part of a

complex number, and in which the auxiliary variables are given by

$$\begin{aligned}
a_1 &= |\tilde{c}_{111}|^2 + |\tilde{c}_{222}|^2, \\
a_2 &= |\tilde{c}_{112}|^2 + |\tilde{c}_{212}|^2, \\
a_3 &= |\tilde{c}_{211}|^2 + |\tilde{c}_{122}|^2, \\
a_4 &= \tilde{c}_{111}\tilde{c}^*_{112}, \\
a_5 &= \tilde{c}_{111}\tilde{c}^*_{211}, \\
a_6 &= \tilde{c}_{222}\tilde{c}^*_{122}, \\
a_7 &= \tilde{c}_{222}\tilde{c}^*_{212}, \\
a_8 &= \tilde{c}_{111}\tilde{c}^*_{212} + \tilde{c}_{222}\tilde{c}^*_{112}, \\
a_9 &= \tilde{c}^*_{111}\tilde{c}_{122} + \tilde{c}_{222}\tilde{c}^*_{211}, \\
a_{10} &= \tilde{c}^*_{211}\tilde{c}_{112} + \tilde{c}_{122}\tilde{c}^*_{212}, \\
v_1 &= a_7 - a_5/2, \\
v_2 &= a_4 - a_6/2, \\
v_3 &= a_9/2 + a_{10}, \\
v_4 &= (a_1 + a_3)/4 + a_2 + \mathrm{Re}(a_8).
\end{aligned}$$

It can be proved that

$$\sum_{i=1}^{2} |c_{iii}|^2 = \mathbf{q}^{\mathrm{T}} \cdot \mathbf{B} \cdot \mathbf{q},$$

in which $\mathbf{q} = (\cos(2\alpha) \;\; \sin(2\alpha)\sin\phi \;\; \sin(2\alpha)\cos\phi)^{\mathrm{T}}$ [26]. Like in the real case, the optimal rotation can be found by computing the dominant eigenvector of $\mathbf{B}$ and normalizing it to unit length. The actual elements of the optimal Jacobi rotation can be obtained from the entries of $\mathbf{q}$ by using the basic relations $\cos\alpha = \sqrt{1 + \cos(2\alpha)/2}$ and $\sin\alpha\, e^{i\phi} = (\sin(2\alpha)\,\cos\phi + i\,\sin(2\alpha)\,\sin\phi)/(2\cos\alpha)$. The sign of the dominant eigenvector can be chosen such that $\mathbf{J}$ is an inner rotation.

1.3.3 Fourth-order real case

The $(2 \times 2 \times 2 \times 2)$ symmetric tensor to be diagonalized, is denoted by $\tilde{\mathcal{C}}$. We define $\mathcal{C} = \tilde{\mathcal{C}} \bullet_1 \mathbf{J} \bullet_2 \mathbf{J} \bullet_3 \mathbf{J} \bullet_4 \mathbf{J}$, in which $\mathbf{J}$ is an elementary (2×2) Jacobi rotation matrix. We use the parameterization (1.14) and work as in [10].

It turns out that in the real case the contrast function $g_{4,2}$ can be expressed as a function of the auxiliary variable $\xi = \theta - 1/\theta$. Setting the derivative w.r.t. ξ equal to zero, yields:

$$\sum_{k=0}^{4} d_k\, \xi^k = 0, \tag{1.33}$$

in which

$$\begin{aligned} d_4 &= \tilde{c}_{1111}\tilde{c}_{1112} - \tilde{c}_{1222}\tilde{c}_{2222}, \\ d_3 &= u - 4(\tilde{c}_{1112}^2 + \tilde{c}_{1222}^2) - 3\tilde{c}_{1122}\,(\tilde{c}_{1111} + \tilde{c}_{2222}), \\ d_2 &= 3v, \\ d_1 &= 3u - 2\tilde{c}_{1111}\tilde{c}_{2222} - 32\tilde{c}_{1112}\tilde{c}_{1222} - 36\tilde{c}_{1122}^2, \\ d_0 &= -4(v + 4d_4), \end{aligned}$$

with

$$\begin{aligned} u &= \tilde{c}_{1111}^2 + \tilde{c}_{2222}^2, \\ v &= (\tilde{c}_{1111} + \tilde{c}_{2222} - 6\tilde{c}_{1122})(\tilde{c}_{1222} - \tilde{c}_{1112}). \end{aligned}$$

We first compute the real roots of Eq. (1.33). (Note that this can be done in a non-iterative way, since the polynomial is only of degree 4 [1].) After computation of these roots, the corresponding values of θ can be found from

$$\theta^2 - \xi\theta - 1 = 0. \tag{1.34}$$

This equation always shows precisely one solution in the interval $(-1, +1]$, corresponding to an inner rotation. If Eq. (1.33) has more than one real root, then the one corresponding to the maximum of the contrast function is selected.

1.3.4 Fourth-order complex case

In this section we assume that $\mathcal{C}_{\mathbf{z}}^{(4)}$ is defined by $(\mathcal{C}_{\mathbf{z}}^{(4)})_{ijkl} = \mathrm{cum}\{z_i, z_j^*, z_k^*, z_l\}$.

Starting from parameterization (1.16), we can work as in Section 1.3.3 [10]. It turns out that the contrast function $g_{4,2}$ can now be expressed as a function of the auxiliary variable $\xi = \theta - 1/\theta^*$. Differentiation of $g_{4,2}$ w.r.t. the modulus ρ and the argument ϕ of ξ yields a system of two polynomials of degree 3 and 4 in ρ, where the coefficients are polynomials in $e^{i\phi}$.

On the other hand, starting from (1.15), it was shown in [19] that the optimal Jacobi rotation can be found from the best rank-1 approximation of a real symmetric $(3\times3\times3\times3)$ tensor of which the entries are derived from $\tilde{\mathcal{C}}$. This best rank-1 approximation can be computed by means of the algorithms discussed in [25, 34, 35, 45].

1.4 Simultaneous diagonalization of matrix slices

Call a $(J_1 \times J_2)$ matrix slice of a $(J_1 \times J_2 \times \ldots \times J_K)$ tensor $\mathcal{A}$ a matrix obtained by varying the first two indices of $\mathcal{A}$ and picking one particular value for the other indices. We will use Matlab notation $(\mathcal{A})_{:,:,j_3,\ldots,j_K}$ to denote such a matrix

slice. From (1.28) we have that the slices $(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K}$ have the following structure:

$$\begin{aligned}(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K} &= \sum_{n_1=1}^{N} c_{s_{n_1}}^{(K)} w_{n_1 n_3} \ldots w_{n_1 n_K} \mathbf{w}_{n_1} \circ \mathbf{w}_{n_1} \\ &= \sum_{n_1=1}^{N} c_{s_{n_1}}^{(K)} w_{n_1 n_3} \ldots w_{n_1 n_K} \mathbf{w}_{n_1} \mathbf{w}_{n_1}^{\mathrm{T}} \\ &= \mathbf{W} \cdot \mathbf{D}_{n_3,\ldots,n_K} \cdot \mathbf{W}^{\mathrm{T}}, \end{aligned} \tag{1.35}$$

in which the $(N \times N)$ matrices $\mathbf{D}_{n_3,\ldots,n_K}$ are diagonal. (The transpose may be replaced by the Hermitean transpose in the complex case.) Eq. (1.35) is an EVD of $(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K}$. We see that the different slices have a common eigenmatrix, namely $\mathbf{W} = \mathbf{Q}^{\mathrm{T}}$.

In the presence of noise, the different matrix slices may not have an eigenmatrix in common, i.e., a matrix that simultaneously diagonalizes all slices may not exist. Denote

$$\mathcal{C}'_{:,:,n_3,\ldots,n_K} = \mathbf{U} \cdot (\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K} \cdot \mathbf{U}^{\mathrm{T}}. \tag{1.36}$$

(Again, the transpose may be replaced by the Hermitean transpose in the complex case.) Cardoso and Souloumiac proposed to determine $\mathbf{Q}$ as the orthogonal (unitary) matrix that makes the matrices $\mathcal{C}'_{:,:,n_3,\ldots,n_K}$ jointly as diagonal as possible in least squares sense [5]. Formally, an orthogonal (unitary) matrix $\mathbf{U}$ is determined that minimizes the cost function

$$f_{K,SMD}(\mathbf{U}) = \sum_{n_3 \ldots n_K} \mathrm{off}(\mathcal{C}'_{:,:,n_3,\ldots,n_K})^2 = \sum_{n_3 \ldots n_K} \sum_{n_1 \neq n_2} \left| c'_{n_1 n_2 \ldots n_K} \right|^2, \tag{1.37}$$

Because $\mathbf{U}$ is orthogonal (unitary), this is equivalent to maximizing the contrast function

$$g_{K,SMD}(\mathbf{U}) = \sum_{n_3 \ldots n_K} \|\mathrm{diag}(\mathcal{C}'_{:,:,n_3,\ldots,n_K})\|^2 = \sum_{n_1 n_3 \ldots n_K} \left| c'_{n_1 n_1 n_3 \ldots n_K} \right|^2. \tag{1.38}$$

The idea is visualized in Fig. 1.2.

The algorithm presented in [5] additionally exploits a second feature of the problem. Note that there are N^{K-2} matrix slices $(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K}$. On the other hand, (1.35) shows that, in the absence of noise, these matrices are all linear combinations of the N rank-1 matrices $\mathbf{w}_{n_1} \mathbf{w}_{n_1}^{\mathrm{T}}$. It makes sense to work with a reduced set of N matrices that best represents the original set $\{(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K}\}$. For the determination of the reduced set, let us consider the matrix slices $(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K}$ as vectors in the $(N \times N)$ matrix space. These vectors are the columns of the $(N^2 \times N^{K-2})$ matrix $\underline{\mathbf{C}}_{\mathbf{z}}^{(K)}$, defined by

$$\left(\underline{\mathbf{C}}_{\mathbf{z}}^{(K)}\right)_{(n_1-1)N+n_2,(n_3-1)N^{K-3}+\ldots+(n_{K-1}-1)N+n_K} = (\mathcal{C}_{\mathbf{z}}^{(K)})_{n_1,n_2,\ldots,n_K},$$

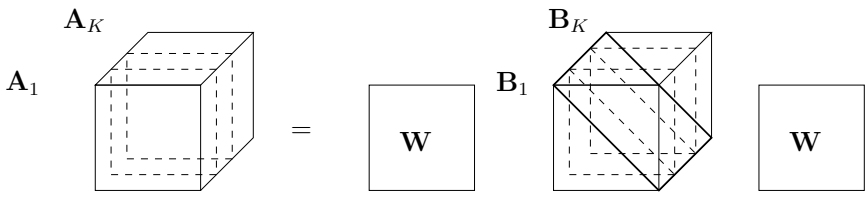


Figure 1.2: Visualization of the (approximate) simultaneous diagonalization of a set of matrices $\mathbf{A}_1, \ldots, \mathbf{A}_K$ by an orthogonal congruence transformation.

for all index entries. The dominant N-dimensional subspace of the column space of $\underline{\mathbf{C}}_{\mathbf{z}}^{(K)}$ is spanned by its N dominant left singular vectors. Denote these vectors by $\mathbf{e}_n$. Re-matricizing yields the matrices $\mathbf{E}_n$ that will be simultaneously diagonalized:

$$(\mathbf{E}_n)_{n_1,n_2} = (\mathbf{e}_n)_{(n_1-1)N+n_2}.$$

In the case of fourth-order cumulants, $\underline{\mathbf{C}}_{\mathbf{z}}^{(K)}$ is a symmetric (Hermitean) matrix of which the singular vectors can be seen as "eigenmatrices" of $\mathcal{C}_{\mathbf{z}}^{(4)}$. The algorithm in [5] was consequently called *JADE*, standing for *Joint Approximate Diagonalization of Eigenmatrices*.

Denote

$$\mathbf{C}_n^{'} = \mathbf{U} \cdot \mathbf{E}_n \cdot \mathbf{U}^{\mathrm{T}}. \tag{1.39}$$

(Again, the transpose may be replaced by the Hermitean transpose in the complex case.) JADE looks for an orthogonal (unitary) matrix $\mathbf{U}$ that minimizes the cost function

$$f_{JADE}(\mathbf{U}) = \sum_n \mathrm{off}(\mathbf{C}_n^{'})^2 = \sum_n \sum_{n_1 \neq n_2} \left| (\mathbf{C}_n^{'})_{n_1 n_2} \right|^2, \tag{1.40}$$

Because $\mathbf{U}$ is orthogonal (unitary), this is equivalent to maximizing the contrast function

$$g_{JADE}(\mathbf{U}) = \sum_n \|\mathrm{diag}(\mathbf{C}_n^{'})\|^2 = \sum_n \sum_{n_1} \left| (\mathbf{C}_n^{'})_{n_1 n_1} \right|^2. \tag{1.41}$$

The optimal $\mathbf{U}$ is obtained by means of a higher-order generalization of the Jacobi iteration described in Section 1.1.3. Due to the orthogonality (unitarity) of $\mathbf{J}$ in (1.12), the only part of $\mathbf{E}_n$ that affects the contrast function, is the (2×2) submatrix $\tilde{\mathbf{C}}_n$ at the intersection of rows and columns p and q. Similar comments as in Section 1.3 are in order. Rotation pairs (p, q) are swept in a cyclic way. Formal analysis of the convergence properties is difficult. Nevertheless, asymptotic quadratic convergence has been proven for the noise-free case [3]. In practice, the algorithm seems to converge reasonably fast and local optima do

not seem to pose a problem. Note that the Jacobi iteration can be initialized with the eigenmatrix of one of the matrices that have to be diagonalized.

In [41] it has been proved that the asymptotical accuracy (for infinitesimal errors in the cumulant estimate) of COM2 and JADE are the same. In [8, 9] several simulations were carried out to compare both methods in practical situations. The results seem to indicate that the methods have approximately the same accuracy provided that the number of sources N is not overestimated, while the performance of JADE degrades with respect to that of COM2 when N is overestimated.

We now explain how the optimal Jacobi rotation can be determined. It turns out that even in the complex case this rotation can be computed efficiently.

1.4.1 Real case

We repeat that the (2×2) symmetric matrices to be simultaneously diagonalized are denoted by $\tilde{\mathbf{C}}_n$. We define $\mathbf{C}_n = \mathbf{J}\cdot\tilde{\mathbf{C}}_n\cdot\mathbf{J}^{\mathrm{T}}$, in which $\mathbf{J}$ is an elementary (2×2) Jacobi rotation matrix. We use parameterization (1.13) and work as in [5, 6].

First define a $(2\times N)$ matrix $\mathbf{G}$ of which the columns $\mathbf{g}_n$ are given by

$$\mathbf{g}_n = \begin{pmatrix} (\tilde{\mathbf{C}}_n)_{11} - (\tilde{\mathbf{C}}_n)_{22} \\ -(\tilde{\mathbf{C}}_n)_{12} - (\tilde{\mathbf{C}}_n)_{21} \end{pmatrix}.$$

It can be proved that the objective function is given by

$$\sum_n (\mathbf{C}_n)_{11}^2 + (\mathbf{C}_n)_{22}^2 = \mathbf{q}^{\mathrm{T}}\cdot\mathbf{G}\mathbf{G}^{\mathrm{T}}\cdot\mathbf{q},$$

in which $\mathbf{q} = (\cos(2\alpha)\ \ \sin(2\alpha))^{\mathrm{T}}$. The vector $\mathbf{q}$ that maximizes this expression is the dominant eigenvector of $\mathbf{G}\mathbf{G}^{\mathrm{T}}$, or, equivalently, the dominant left singular vector of $\mathbf{G}$. The actual elements of the optimal Jacobi rotation can be obtained from the basic relations $\cos\alpha = \sqrt{1+\cos(2\alpha)/2}$ and $\sin\alpha = \sin(2\alpha)/(2\cos\alpha)$. The sign of the dominant eigenvector can be chosen such that $\mathbf{J}$ is an inner rotation.

1.4.2 Complex case

We repeat that the (2×2) Hermitean matrices to be simultaneously diagonalized are denoted by $\tilde{\mathbf{C}}_n$. We define $\mathbf{C}_n = \mathbf{J}\cdot\tilde{\mathbf{C}}_n\cdot\mathbf{J}^{\mathrm{H}}$, in which $\mathbf{J}$ is an elementary (2×2) Jacobi rotation matrix. We use parameterization (1.15) and work as in [5, 6].

First define a $(3\times N)$ matrix $\mathbf{G}$ of which the columns $\mathbf{g}_n$ are given by

$$\mathbf{g}_n = \begin{pmatrix} (\tilde{\mathbf{C}}_n)_{11} - (\tilde{\mathbf{C}}_n)_{22} \\ -(\tilde{\mathbf{C}}_n)_{12} - (\tilde{\mathbf{C}}_n)_{21} \\ i(-(\tilde{\mathbf{C}}_n)_{12} + (\tilde{\mathbf{C}}_n)_{21}) \end{pmatrix}.$$

It can be proved that

$$\sum_n |(\mathbf{C}_n)_{11}|^2 + |(\mathbf{C}_n)_{22}|^2 = \mathbf{q}^{\mathrm{T}} \cdot \mathrm{Re}(\mathbf{G}\mathbf{G}^{\mathrm{H}}) \cdot \mathbf{q},$$

in which $\mathbf{q} = (\cos(2\alpha) \;\; \sin(2\alpha)\sin\phi \;\; \sin(2\alpha)\cos\phi)^{\mathrm{T}}$. Hence the optimal rotation can be found via the dominant eigenvector of the real symmetric (3×3)-matrix $\mathrm{Re}(\mathbf{G}\mathbf{G}^{\mathrm{H}})$. The actual elements of the optimal Jacobi rotation can be obtained from the entries of $\mathbf{q}$ by using the basic relations $\cos\alpha = \sqrt{1 + \cos(2\alpha)/2}$ and $\sin\alpha\, e^{i\phi} = (\sin(2\alpha)\,\cos\phi + i\,\sin(2\alpha)\,\sin\phi)/(2\cos\alpha)$. The sign of the dominant eigenvector can be chosen such that $\mathbf{J}$ is an inner rotation.

1.5 Simultaneous diagonalization of third-order tensor slices

Instead of considering the tensor $\mathcal{C}_{\mathbf{z}}^{(K)}$ as a stack of N^{K-2} $(N \times N)$ matrices $(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,n_3,\ldots,n_K}$, we could also consider it as a stack of N^{K-3} $(N \times N \times N)$ tensors $(\mathcal{C}_{\mathbf{z}}^{(K)})_{:,:,:,n_4,\ldots,n_K}$ and simultaneously diagonalize these. This can be done using the expressions in Sections 1.3.1 and 1.3.2, taking into account that the cost function now consists of a sum of contributions associated with the different third-order tensor slices (i.e., the matrix $\mathbf{B}$ is now a sum of matrices associated with the different third-order tensor slices). We refer to [26] for more details. The technique is called *STOTD*, which stands for *Simultaneous Third-Order Tensor Diagonalization*.

Simulations in [26] indicate that the performance of JADE and STOTD are generally quite comparable, although in specific cases one method may perform better the other.

1.6 Maximization of the tensor trace

It is well-known that the trace of a matrix is invariant under orthogonal (unitary) congruence transformations. Again, the situation is very different for higher-order tensors. In the orthogonal (unitary) congruence transformation (1.30) the trace of the tensor, i.e., the sum of its diagonal entries, does in general change with $\mathbf{U}$. Comon realized that we can thus diagonalize the tensor by maximizing its trace, if we know that all source cumulants are positive [16]. If all cumulants are negative, then we can process $-\mathcal{C}_{\mathbf{z}}^{(K)}$ instead of $\mathcal{C}_{\mathbf{z}}^{(K)}$. The assumption that the cumulants of all sources have the same known sign is satisfied in many applications. For typical constellations used in telecommunication, the fourth-order cumulant is negative. Speech and audio signals typically have positive fourth-order cumulant. In this section we assume positive source cumulants.

Formally, we look for an orthogonal (unitary) matrix $\mathbf{U}$ that maximizes the contrast function

$$g_{K,1}(\mathbf{U}) = \sum_n c'_{nn\ldots n}. \tag{1.42}$$

The resulting algorithm will be denoted as *COM1*, where the "1" indicates that we maximize the 1-norm of the diagonal, instead of its 2-norm, like in COM2. The optimal $\mathbf{U}$ is again obtained by means of a Jacobi iteration. The fact that the diagonal entries are not squared, leads to simpler expressions than in COM2, such that the computation of an elementary rotation is easier.

Simulations indicate that the accuracy of COM1 is comparable to that of COM2 and JADE, provided (i) the number of sources is not over-estimated, (ii) the background noise is Gaussian, (iii) no source is Gaussian and (iv) the cumulants of the sources have the same sign [15]. COM1 seems to be less sensitive to non-Gaussian background noise than COM2 and JADE. When the number of sources is over-estimated, the convergence speed of the three methods decreases. COM2 is then the most reliable method whereas JADE becomes the slowest. On the other hand, the performance of COM1 degrades with respect to that of COM2 and JADE when one of the sources is Gaussian or quasi-Gaussian. In addition, COM1 is unable to separate sources with different cumulant signs.

We now explain how an optimal Jacobi rotation can be determined. Due to space limitations we only consider the fourth-order complex case. We assume that $\mathcal{C}_{\mathbf{z}}^{(4)}$ is defined by $(\mathcal{C}_{\mathbf{z}}^{(4)})_{ijkl} = \mathrm{cum}\{z_i, z_j^*, z_k^*, z_l\}$. Like in Section 1.3, the $(2 \times 2 \times 2 \times 2)$-subtensor that will be diagonalized, is denoted by $\tilde{\mathcal{C}}$. We define $\mathcal{C} = \tilde{\mathcal{C}} \bullet_1 \mathbf{J} \bullet_2 \mathbf{J}^* \bullet_3 \mathbf{J}^* \bullet_4 \mathbf{J}$, in which $\mathbf{J}$ is an elementary (2×2) Jacobi rotation matrix. A first approach was based on the parameterization (1.16) [16]. Here we use parameterization (1.15) and we work as in [12].

We construct a real symmetric (3×3)-matrix $\mathbf{B}$ as follows:

$$\begin{aligned}
b_{11} &= \tilde{c}_{1111} + \tilde{c}_{2222}, \\
b_{12} &= \mathrm{Im}(\tilde{c}_{1222} - \tilde{c}_{1211}), \\
b_{13} &= \mathrm{Re}(\tilde{c}_{1222} - \tilde{c}_{1211}), \\
b_{22} &= \frac{1}{2}(\tilde{c}_{1111} + \tilde{c}_{2222}) + 2\tilde{c}_{1212} - \mathrm{Re}(\tilde{c}_{1221}), \\
b_{23} &= \mathrm{Im}(\tilde{c}_{1221}), \\
b_{33} &= \frac{1}{2}(\tilde{c}_{1111} + \tilde{c}_{2222}) + 2\tilde{c}_{1212} + \mathrm{Re}(\tilde{c}_{1221}).
\end{aligned}$$

It can be shown that

$$\sum_{i=1}^{2} c_{iiii} = \mathbf{q}^{\mathrm{T}} \cdot \mathbf{B} \cdot \mathbf{q},$$

with $\mathbf{q} = (\cos(2\alpha) \;\; \sin(2\alpha)\sin\phi \;\; \sin(2\alpha)\cos\phi)^{\mathrm{T}}$. The optimal rotation can be found by computing the dominant eigenvector of $\mathbf{B}$ and scaling it to unit norm. The entries of the Jacobi rotation matrix are found from the entries of $\mathbf{q}$ by using the trigonometric relations $\cos\alpha = \sqrt{(1+\cos(2\alpha))/2}$ and $\sin\alpha\, e^{i\phi} = (\sin(2\alpha)\,\cos\phi + i\,\sin(2\alpha)\,\sin\phi)/(2\cos\alpha)$. The sign of the dominant eigenvector can be chosen such that $\mathbf{J}$ is a rotation over an angle in the interval $(-\pi/4, \pi/4]$.

Bibliography

[1] M. Abramowitz and I. E. Segun. *Handbook of Mathematical Functions*. Dover Publ., N.Y., 1968.

[2] C. Bourin and P. Bondon. Efficiency of high-order moment estimates. In *IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics*, pages 186–190, Girona, Spain, 1995.

[3] A. Bunse-Gerstner, R. Byers, and V. Mehrmann. Numerical methods for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 14(4):927–949, Oct. 1993.

[4] J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *Proc. VIIth European Signal Processing Conference (EUSIPCO-94)*, volume 2, pages 776–779, Edinburgh, Scotland, U.K., Sept. 1994.

[5] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proc.-F*, 140(6):362–370, 1994. [Matlab code available at http://www.tsi.enst.fr/∼cardoso/stuff.html].

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

[7] J. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35:283–319, 1970.

[8] P. Chevalier. Méthodes aveugles de filtrage d'antennes. *Revue d'Electronique et d'Electricité*, 3:48–58, Sept. 1995.

[9] P. Chevalier. On the performance of higher order separation methods. In *Proc. IEEE-ATHOS Workshop on Higher-Order Statistics*, pages 30–34, Begur, Spain, June 1995.

[10] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994. [Matlab code available at http://www.i3s.unice.fr/˜comon/].

[11] P. Comon. Tensor diagonalization, a useful tool in signal processing. In *10th IFAC Symposium on System Identification (IFAC-SYSID'94)*, volume 1, pages 77–82, Copenhagen, Denmark, July 1994.

[12] P. Comon. From source separation to blind equalization, contrast-based approaches. In *Int. Conf. on Image and Signal Processing (ICISP'01)*, Agadir, Morocco, May 2001.

[13] P. Comon. Blind identification and source separation in 2×3 underdetermined mixtures. *IEEE Trans. Signal Processing*, 52(1):11–22, Jan. 2004.

[14] P. Comon. Canonical tensor decompositions. Tech. rep. RR-2004-17, Lab. I3S, Sophia-Antipolis, France, June 2004.

[15] P. Comon, P. Chevalier, and V. Capdevielle. Performance of contrast-based blind source separation. In *SPAWC - IEEE Sig. Proc. Advances in Wireless Com.*, pages 345–348, Paris, France, Apr. 1997.

[16] P. Comon and E. Moreau. Improved contrast dedicated to blind separation in communications. In *Proc. ICASSP*, pages 3453–3456, Munich, Germany, Apr. 1997. [Matlab code available at http://www.i3s.unice.fr/˜comon/].

[17] P. Comon and B. Mourrain. Decomposition of quantics in sums of powers of linear forms. *Signal Process.*, 53(2):93–107, Sept. 1996.

[18] P. Comon and M. Rajih. Blind identification of under-determined mixtures based on the characteristic function. In *Proc. ICASSP'05*, volume 4, pages 1005–1008, Philadelphia, Mar. 2005.

[19] L. De Lathauwer. *Signal Processing Based on Multilinear Algebra*. PhD thesis, K.U.Leuven, E.E.Dept., Belgium, 1997.

[20] L. De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.*, 28(3):322–336, Dec. 2006.

[21] L. De Lathauwer and J. Castaing. Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. *IEEE Trans. Signal Processing*, 56(3):1096–1105, Mar. 2008.

[22] L. De Lathauwer, J. Castaing, and J.-F. Cardoso. Fourth-order cumulant based blind identification of underdetermined mixtures. *IEEE Trans. Signal Processing*, 55(6):2965–2973, June 2007.

[23] L. De Lathauwer, P. Comon, and B. De Moor. ICA algorithms for 3 sources and 2 sensors. In *Proc. Sixth Sig. Proc. Workshop on Higher Order Statistics*, pages 116–120, Caesarea, Israel, June 1999.

[24] L. De Lathauwer, B. De Moor, and J. Vandewalle. An algebraic ICA algorithm for 3 sources and 2 sensors. In *Proc. Xth European Signal Processing Conference (EUSIPCO 2000)*, Tampere, Finland, Sept. 2000.

[25] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-$(R_1, R_2, \ldots, R_n)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, Apr. 2000.

[26] L. De Lathauwer, B. De Moor, and J. Vandewalle. Independent component analysis and (simultaneous) third-order tensor diagonalization. *IEEE Trans. Signal Processing*, 49(10):2262–2271, Oct. 2001. [Matlab code available at ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/delathauwer/software/].

[27] L. De Lathauwer, B. De Moor, and J. Vandewalle. Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM J. Matrix Anal. Appl.*, 26:295–327, 2004.

[28] L. De Lathauwer, B. De Moor, and J. Vandewalle. A prewhitening-induced bound on the identification error in independent component analysis. *IEEE Trans. Circuits and Systems I*, 52(3):546–554, Mar. 2005.

[29] W. Gardner. *Introduction to Random Processes*. McGraw-Hill, 1990.

[30] G. Golub and C. Van Loan. *Matrix Computations*. 3rd ed., Johns Hopkins University Press, Baltimore, Maryland, 1996.

[31] R. Harshman. Foundations of the PARAFAC procedure: model and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[32] I. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, N.Y., 2nd edition, 2002.

[33] M. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Griffin, London, 1977.

[34] E. Kofidis and P. Regalia. On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM J. Matrix Anal. Appl.*, 23:863–884, 2002.

[35] P. Kroonenberg. *Applied Multiway Data Analysis*. Wiley, 2008.

[36] P. McCullagh. *Tensor Methods in Statistics*. Chapman and Hall, 1987.

[37] E. Moreau. A generalization of joint-diagonalization criteria for source separation. *IEEE Trans. on Signal Processing*, 49(3):530–541, Mar. 2001.

[38] C. Nikias and J. Mendel. Signal processing with higher-order spectra. *IEEE Signal Proc. Mag.*, 3:10–37, 1993.

[39] C. Nikias and A. Petropulu. *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[40] A. Smilde, R. Bro, and P. Geladi. *Multi-way Analysis. Applications in the Chemical Sciences*. John Wiley and Sons, Chichester, U.K., 2004.

[41] A. Souloumiac and J.-F. Cardoso. Performances en séparation de sources. In *Proc. GRETSI*, pages 321–324, Juan les Pins, France, 1993.

[42] B. Van Veen and K. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, 1988.

[43] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, and Signal Processing*, 33:387–392, 1986.

[44] A. Yeredor. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Trans. Signal Processing*, 50(7):1545–1553, July 2002.

[45] T. Zhang and G. Golub. Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.*, 23:534–550, 2001.