**KATHOLIEKE UNIVERSITEIT LEUVEN**
FACULTEIT DER TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee)

# SIGNAL PROCESSING BASED ON MULTILINEAR ALGEBRA

Promotoren:
Prof. dr. ir. J. Vandewalle
Prof. dr. ir. B. De Moor

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

**Lieven DE LATHAUWER**

September 1997

**KATHOLIEKE UNIVERSITEIT LEUVEN**
FACULTEIT DER TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee)

SIGNAL PROCESSING BASED ON MULTILINEAR ALGEBRA

Jury:
Prof. Y. Willems, vice-decaan, voorzitter
Prof. J. Vandewalle, promotor
Prof. B. De Moor, promotor
Prof. P. Comon (EURECOM, France)
Prof. J. McWhirter (DERA, U.K.)
Prof. S. Van Huffel
Prof. A. Bultheel

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

**Lieven DE LATHAUWER**

U.D.C. 681.3*I54

September 1997

© Katholieke Universiteit Leuven – Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

D/1997/7515/20

ISBN 90-5682-085-0

# Voorwoord

Bij het beëindigen van dit doctoraat is het gepast een aantal mensen te bedanken. Vooreerst gaat mijn dank uit naar de beide promotoren van dit onderzoek, Prof. Bart De Moor en Prof. Joos Vandewalle, die SISTA hebben uitgebouwd tot een uitmuntende onderzoeksgroep en een verrijkend en stimulerend milieu. Ik ben hen erkentelijk dat ze mij de kans hebben gegeven op verkenning te gaan op een misschien niet voor de hand liggend onbekend terrein. De interesse van Joos voor mijn onderzoek heeft een belangrijk aspect gevormd van het klimaat waarin ik heb gewerkt. Bart is de kampioen van de geestdriftige motiveringen en ik heb het door hem met veel flair propageren van tensoren in numerieke algebra-kringen weten te appreciëren.

Het leescomité, Prof. Sabine Van Huffel en Prof. Adhemar Bultheel, wil ik speciaal bedanken voor het op zeer korte tijd, in de examenperiode, zorgvuldig nalezen van een eerste versie van dit proefschrift.

I feel honoured that Prof. Pierre Comon and Prof. John McWhirter have accepted to be on the jury of this thesis.

Je voudrais remercier Pierre pour les moments agréables passés ensemble sur et autour de la scène SOE. Je garde des souvenirs excellents de ma visite à Sophia-Antipolis et ses environs en juillet '95.

I would like to express my gratitude to John, for inviting me and Bart to give a keynote talk at the 4th IMA Conference on Mathematics in Signal Processing, and for the vivid interest that he has always shown in my work. Also thanks to dr. Ian Proudler, who corrected some language errors.

Ik dank Prof. Yves Willems als voorzitter van de jury.

I would like to thank Prof. Andrew Cichocki, Prof. Asoke Nandi and Prof. Christian Jutten for inviting me to special conference sessions on Blind Source Separation. I also thank Prof. Michel Kinnaert for the invitation to submit a paper to *Journal A*. I thank Prof. Jean-François Cardoso, Prof. Gene Golub, Prof. Peter Ladefoged and Prof. Robert Ross for the interesting discussions,

feed-back and helpful suggestions. In het bijzonder zou ik dr. Alle-Jan van der Veen willen bedanken, voor het aanwijzen van referentie [266], en Jeroen Dehaene, voor de bereidwillige hulp i.v.m. principes van differentiaalmeetkunde.

Mijn bureaugenoten Peter, Jeroen, Yves, ex-bureaugenoten Filiep, Bart en de andere collega's (Willem!) wil ik bedanken voor de aangename sfeer. Een bijzonder woordje van dank aan Bart DS, bij wie ik nooit vergeefs ben gaan aankloppen bij computerproblemen, en Bart M, voor de administratieve coördinatie van SISTA. Uiteraard wil ik ook de secretaresses Ida, Rebecca, Ingrid, Rita, Lut, Patricia, Ann en de systeemploeg niet vergeten.

Ik dank ook het I.W.O.N.L. en het I.W.T. voor de financiële steun.

Het feit dat hier nu dit doctoraatswerk ligt, heeft ongetwijfeld heel wat te maken met de rol die mijn moeder en vader hebben gespeeld. Toch zou ik hen vooral willen bedanken om er te zíjn. Ik denk hierbij ook aan meme en pepe. Ik wil ook mijn vrienden bedanken, en een pluimpje geven aan mijn zus voor de assistentie bij de voorbereiding van de presentatie.

Gedurende mijn doctoraat heb ik me dag in, dag uit gedragen geweten door de liefde van Patricia. Mijn lieve schat: bedankt!

# Abstract

Numerous important results in signal processing and system theory have been obtained by means of first- and second-order tools: vector- and matrix algebra, from an algebraic point of view; or mean and variance, in a statistical framework. Beyond these boundaries, the emerging discipline of multilinear algebra allows one to deal with forms that are not necessarily linear or quadratic; whilst Higher-Order Statistics (HOS) is one mechanism for exploiting the characteristics of non-Gaussian signals.

The Singular Value Decomposition (SVD) of matrices and its second-order generalizations play a central role for the analysis and the solution of problems in mathematical engineering. However, generalizing the SVD proves to be difficult, due to its remarkable number of powerful properties; focusing on different aspects may lead to different tensor decompositions. We begin by examining the claim that the higher-order singular vectors correspond to the directions of extremal oriented energy, associated with the row vectors, the column vectors, etc. (HOSVD). Further, we investigate the least-squares approximation of a given tensor by a tensor with pre-specified column rank, row rank, ... and the least-squares approximation by a sum of possibly non-orthogonal rank-1 terms.

Independent Component Analysis (ICA) is a fundamental concept in HOS, aimed at the blind separation of a given dataset into a mixture of statistically independent components. An ICA can be performed by resorting to multilinear SVD-extensions. We establish links with the HOSVD, the best rank-1 approximation problem and the best rank-$R$ approximation.

# Korte Inhoud

Talrijke belangrijke resultaten in signaalverwerking en systeemtheorie zijn bekomen vanuit een eerste- of tweede-orde benadering: vector- en matrixalgebra, vanuit een algebraïsch standpunt, of verwachte waarde en variantie, in een statistisch kader. De opkomende discipline van multilineaire algebra laat aanvullend toe te werken met vormen die niet noodzakelijk lineair of kwadratisch zijn; aan de andere kant kunnen met Hogere-Orde Statistiek (HOS) de karakteristieken van niet-Gaussiaanse signalen geëploiteerd worden.

De Singuliere Waardenontbinding (SWO) van matrices en haar tweede-orde veralgemeningen spelen een centrale rol in de analyse en oplossing van wiskundige ingenieursproblemen. Het feit dat de SWO een merkwaardig aantal krachtige eigenschappen bundelt, blijft niet zonder meer behouden in een multilineair equivalent; de veralgemening van verschillende deelaspecten geeft aanleiding tot verschillende tensordecomposities. We nemen eerst als uitgangspunt dat de hogere-orde singuliere vectoren overeenkomen met de richtingen van extremale georiënteerde energie, geassocieerd met de rijvectoren, de kolomvectoren, enz. (HOSWO). Verder onderzoeken we de kleinste-kwadratenbenadering van een gegeven tensor door een tensor met vooraf gespecifieerde kolomrang, rijrang, ... en de kleinste-kwadratenbenadering door een som van mogelijks niet-orthogonale rang-1 termen.

Onafhankelijke Componenten Analyse (OCA) is een fundamenteel concept in HOS; het doel is de blinde scheiding van meerkanaals meetgegevens in een combinatie van statistisch onafhankelijke componenten. Een OCA kan uitgevoerd worden door gebruik te maken van multilineaire SWO-veralgemeningen. We ontwikkelen verbanden met de HOSWO, het beste rang-1 benaderingsprobleem en de de beste rang-$R$ benadering.



# Notation

In this section we list the symbols and acronyms that occur frequently in this thesis. The numbers in the last column refer to the page on which the symbol or concept in question is defined.

## List of symbols

| | | |
|---|---|---|
| $\|a\|$ | absolute value of $a \in \mathbb{R}$; modulus of $a \in \mathbb{C}$ | |
| $a^*$ | complex conjugate of $a$ | |
| $a \sim b$ | $a$ and $b$ belong to the same class of equivalence | |
| $[a, b]$ | closed interval | |
| $(a, b)$ | open interval | |
| $\mathbf{A}^T$ | transpose of $\mathbf{A}$ | |
| $\mathbf{A}^H$ | Hermitean transpose of $\mathbf{A}$ | |
| $\mathbf{A}^{-1}$ | inverse of $\mathbf{A}$ | |
| $\mathbf{A}^\dagger$ | Moore-Penrose pseudo-inverse of $\mathbf{A}$ | |
| $\langle \mathcal{A}, \mathcal{B} \rangle$ | scalar product of $\mathcal{A}$ and $\mathcal{B}$ | 16 |
| $\langle \mathcal{A} \rangle_{p,q}$ | contraction of $\mathcal{A}$ over the $p$th and $q$th index | 14 |
| $\langle \mathcal{A}, \mathcal{B} \rangle_{p,q}$ | inner product of $\mathcal{A}$ and $\mathcal{B}$ over the $p$th and $q$th index respectively | 14 |
| $\|A\|$ ( $\|\mathbf{A}\|$, $\|\mathcal{A}\|$) | Frobenius-norm of $A$ ($\mathbf{A}$ , $\mathcal{A}$) | 17 |
| $\|\mathbf{A}\|_{LF}$ ($\|\mathbf{A}\|_{UF}$) | Frobenius-norm of the lower-triangular (upper-triangular) part of $\mathbf{A}$ | 220 |
| $\nabla_X f$ | gradient of $f$ to $X$ | |
| $\|\mathbf{A}\|_{LFs}$ ($\|\mathbf{A}\|_{UFs}$) | Frobenius-norm of the strictly lower-triangular (upper-triangular) part of $\mathbf{A}$ | 220 |
| $\mathbf{A} \otimes \mathbf{B}$ | Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ | 219 |
| $\mathbf{A} \odot \mathbf{B}$ | Kathri-Rao product of $\mathbf{A}$ and $\mathbf{B}$ | 219 |
| $\mathcal{A} \circ \mathcal{B}$ | outer product of $\mathcal{A}$ and $\mathcal{B}$ | 13 |
| $\mathcal{A} \times_n \mathbf{A}$ | $n$-mode product of $\mathcal{A}$ and $\mathbf{A}$ | 14 |

| | | |
|---|---|---|
| $\mathbf{A}_{i_1:i_2,j_1:j_2}$ | submatrix of $\mathbf{A}$ with row index varying from $i_1$ to $i_2$ and column index varying from $j_1$ to $j_2$ | |
| $\mathcal{A}_{i=\alpha}$ | subtensor of $\mathcal{A}$, obtained by fixing the index $i$ to $\alpha$ | |
| $\mathbf{A}_{(n)}$ | standard matrix unfolding of $\mathcal{A}$ | 18 |
| $\mathcal{A}(X)$ | associated polynomial of $\mathcal{A}$ | 23 |
| $\mathbb{C}$ | set of the complex numbers | |
| $\mathbb{C}^I$ | set of complex-valued $I$-dimensional vectors | |
| $\mathbb{C}^{I\times J}$ | set of the $(I\times J)$-matrices with complex entries | 219 |
| $\mathbb{C}^{I_1\times I_2\times\ldots\times I_N}$ | set of complex-valued $(I_1\times I_2\times\ldots I_N)$-tensors | 10 |
| $c_N^X$ | $N$th order cumulant of the random variable $X$ | 31 |
| $\mathtt{Cum}\{X_1, X_2, \ldots, X_N\}$ | joint $N$th order cumulant of the stochastic variables $X_1, X_2, \ldots, X_N$ | 32 |
| $\mathcal{C}_N^X$ | $N$th order cumulant tensor of the stochastic vector $X$ | 31 |
| $\ldots \stackrel{\text{def}}{=} \ldots$ | $\ldots$ equals by definition $\ldots$ | |
| $\mathtt{dim}(\mathtt{V})$ | dimension of the vector space $\mathtt{V}$ | |
| $\mathtt{det}(\mathbf{A})$ | determinant of $\mathbf{A}$ | 219 |
| $\mathtt{diag}(a, b, \ldots)$ | diagonal matrix with diagonal entries $a, b, \ldots$ | 219 |
| $\mathtt{diag}(\mathbf{A})$ ($\mathtt{diag}(\mathcal{A})$) | diagonal part of matrix $\mathbf{A}$ (tensor $\mathcal{A}$) | 219 |
| $\mathtt{E}\{X\}$ | expected value of $X$ | |
| $\mathbf{G}$ | $(2\times 2)$ Givens rotation matrix | 223 |
| $\mathbf{G}_{ij}$ | Givens rotation matrix, affecting rows $i$ and $j$ | 223 |
| $\mathtt{Im}(x)$ | imaginary part of $x$ | |
| $\mathbf{I}$ | unit matrix | |
| $\mathcal{I}$ | unit tensor | 25 |
| $\mathtt{I}(X)$ | average mutual information of $X$ | 130 |
| $\mathbf{J}$ | $(2\times 2)$ Jacobi rotation matrix | 223 |
| $\mathbf{J}_{ij}$ | Jacobi rotation matrix, affecting rows $i$ and $j$ | 223 |
| $\mathtt{min}(a, b, \ldots)$ | minimum of $a, b, \ldots$ | |
| $\mathtt{max}(a, b, \ldots)$ | maximum of $a, b, \ldots$ | |
| $m_N^X$ | $N$th order moment of the random variable $X$ | 30 |
| $\mathtt{Mom}\{X_1, X_2, \ldots, X_N\}$ | joint $N$th order moment of the stochastic variables $X_1, X_2, \ldots, X_N$ | 30 |
| $\mathcal{M}_N^X$ | $N$th order moment tensor of the stochastic vector $X$ | 31 |
| $m_X$ | mean of $X$ | 30 |
| $\mathrm{O}(I)$ | manifold of orthogonal $(I\times I)$-matrices | 220 |
| $O(x)$ | any real function $f(x)$ such that $\limsup_{x\to 0} \lvert f(x)\rvert/x$ is finite | |
| $\mathtt{OE}(X, \mathbf{A})$ | oriented energy of $\mathbf{A}$ in the direction of unit-norm vector $X$ | 222 |



| | | |
|---|---|---|
| $\mathtt{OE}_\mathtt{n}(X, \mathcal{A})$ | $n$-mode oriented energy of $\mathcal{A}$ in the direction of unit-norm vector $X$ | 52 |
| $\mathtt{P}(i_1, i_2, \ldots, i_N)$ | permutation of $i_1, i_2, \ldots, i_N$ | 17 |
| $\mathtt{Prob}\{x\}$ | probability of realization of $x$ | |
| $p_X(x)$ | probability density function of $X$ | |
| $\mathbb{R}$ | set of the real numbers | |
| $\mathbb{R}_0$ | set of the real numbers except for 0: $\mathbb{R}_0 = \mathbb{R}\backslash\{0\}$ | |
| $\mathbb{R}^+$ | set of the nonnegative real numbers | |
| $\mathbb{R}^I$ | set of real-valued $I$-dimensional vectors | |
| $\mathbb{R}^{I\times J}$ | set of the $(I\times J)$-matrices with real entries | 219 |
| $\mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ | set of real-valued $(I_1\times I_2\times\ldots I_N)$-tensors | 10 |
| $\mathtt{R}(\mathbf{A})$ | range of $\mathbf{A}$ | 219 |
| $\mathtt{Re}(x)$ | real part of $x$ | |
| $R = \mathtt{rank}(\mathcal{A})$ | rank of $\mathcal{A}$ | 19 |
| $R_n = \mathtt{rank}_\mathtt{n}(\mathcal{A})$ | $n$-rank of $\mathcal{A}$ | 19 |
| $s_N^X(\omega_1, \omega_2, \ldots, \omega_{N-1})$ | $N$th order cumulant spectrum of $X(t)$ | 32 |
| $\mathrm{St}(J, I)$ | Stiefel manifold of $(I\times J)$-matrices | 220 |
| $\mathtt{sign}(a)$ | signum function | 65 |
| $\mathtt{span}\{X, Y, \ldots, Z\}$ | vector space spanned by $X, Y, \ldots, Z$ | |
| $\mathtt{skew}(\mathbf{A})$ | skew-symmetric part of matrix $\mathbf{A}$ | 219 |
| $\mathtt{trace}(\mathbf{A})$ | trace of $\mathbf{A}$ | 219 |
| $\mathrm{U}(I)$ | manifold of unitary $(I\times I)$-matrices | 220 |
| $\mathtt{upp}(\mathbf{A})$ | upper-diagonal part of matrix $\mathbf{A}$ | 219 |
| $\mathtt{var}\{X\}$ | variance of $X$ | 32 |
| $\gamma_X$ | skewness of $X$ | 32 |
| $\delta_{ij}$ | Kronecker-delta | 25 |
| $\delta(x)$ | Dirac-impulse | |
| $\frac{\delta f}{\delta x}$ | partial derivative of $f$ with respect to $x$ | |
| $\kappa_X$ | kurtosis of $X$ | 32 |
| $\sigma_i$ | $i$th singular value | 221 |
| $\sigma_i^{(n)}$ | $i$th $n$-mode singular value | 48 |
| $\sigma_X$ | standard deviation of $X$ | 32 |
| $\Pi_{i=i_1}^{i_2} a_i$ | product of the values $a_{i_1}$ to $a_{i_2}$ | |
| $\Pi_i a_i$ | product of the values $a_i$, for varying $i$ | |
| $\Sigma_{i=i_1}^{i_2} a_i$ | sum of the values $a_{i_1}$ to $a_{i_2}$ | |
| $\Sigma_i a_i$ | sum over the values $a_i$, for varying $i$ | |
| $\mathbf{\Phi}_X(\omega)$ | first characteristic function of $X$ | 29 |
| $\mathbf{\Psi}_X(\omega)$ | second characteristic function of $X$ | 30 |

**Remark:** The notation used in the thesis allows one to distinguish between scalars, vectors, matrices and higher-order tensors. Lower-case characters represent scalar values ($a, b, \ldots$). Non-boldface capitals ($A, B, \ldots$) are used for vectors. Boldface capitals ($\mathbf{A}$, $\mathbf{B}$, ... ) represent matrices. Calligraphic cap-

itals ($\mathcal{A}$, $\mathcal{B}$, ... ) represent tensors. This notation is consistently used for lower-order parts of a given structure. E.g. the entry with row index $i$ and column index $j$ in a matrix $\mathbf{A}$, $(\mathbf{A})_{ij}$, is symbolized by $a_{ij}$ (also $(A)_i = a_i$ and $(\mathcal{A})_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N}$); similarly, the $i$th column vector of a matrix $\mathbf{A}$ is called $A_i$. There are two exceptions to this rule. First, indices are in principle chosen among the symbols $i, j, k, p, q, r, n$; with this respect, $I, J, K, P, Q, R, N$ indicate the index upper-bounds, unless stated otherwise. Secondly, in statistical signal processing it is common to use capital and lower-case characters to make the distinction between a stochastic variable $X$ and a sample value $x$ taken by this variable. The scalar or vectorial nature of a stochastic variable will be clear from the context.

## Acronyms

| | | |
|---|---|---|
| ALS | Alternating Least Squares | 68 |
| BSPM | Body Surface Potential Mapping | 201 |
| BSS | Blind Source Separation | 123 |
| BSSS | Blind Source Subspace Separation | 205 |
| CANDECOMP | Canonical Decomposition | 92 |
| CL | Comon-Lacoume | 154 |
| DOA | Direction of Arrival | 199 |
| ECG | Electrocardiogram | 201 |
| EVD | Eigenvalue Decomposition | 220 |
| FECG | Fetal Electrocardiogram | 202 |
| FHR | Fetal Heart Rate | 202 |
| FHRV | Fetal Heart Rate Variability | 202 |
| HOEVD | Higher-Order Eigenvalue Decomposition | 57 |
| HOS | Higher-Order Statistics | 29 |
| HOSVD | Higher-Order Singular Value Decomposition | 47 |
| ICA | Independent Component Analysis | 122 |
| ISR | Interference to Signal Ratio | 130 |
| JADE | Joint Approximate Diagonalization of Eigenmatrices | 182 |
| LCMV | Linearly Constrained Minimum Variance | 131 |
| MD | Maximal Diagonality | 154 |
| MECG | Maternal Electrocardiogram | 203 |
| MVDR | Minimum Variance Distortionless Response | 130 |
| PARAFAC | Parallel Factors | 92 |
| PCA | Principal Component Analysis | 125 |
| PSK | Phase Shift Keying | 39 |
| QAM | Quadrature Amplitude Modulation | 39 |
| QRD | QR Decomposition | 220 |
| RMSE | Root Mean Square Error | 131 |
| SINR | Signal to Interference-plus-Noise Ratio | 130 |



| | | |
|---|---|---|
| SIR | Signal to Interference Ratio | 130 |
| SNR | Signal to Noise Ratio | 130 |
| STOTD | Simultaneous Third-Order Tensor Diagonalization | 174 |
| SVD | Singular Value Decomposition | 221 |

# Contents

**Voorwoord** **i**

**Abstract** **iii**

**Abstract in Dutch** **iv**

**Notation** **v**

**Contents** **xi**

**Summary in Dutch** **xix**

**1 Introduction** **1**

1.1 General overview . . . . . 3

1.2 Chapter by chapter overview . . . . . 5

**2 Introduction to Multilinear Algebra** **9**

2.1 Higher-order tensors . . . . . 10

2.2 Basic tensor operations . . . . . 13

2.3 Multiplication of a higher-order tensor by a matrix . . . . . 14

2.4 Scalar product, orthogonality, norm . . . . . 16

2.5 Matrix representation of a higher-order tensor . . . . . 17

2.6 Rank properties of a higher-order tensor . . . . . 19
2.7 Super-symmetric tensors . . . . . 22
2.8 Tensors as linear transformations . . . . . 23
2.8.1 Linear transformations . . . . . 24
2.8.2 Unit tensor . . . . . 25
2.8.3 Eigenvalue Decomposition . . . . . 26
2.8.4 Singular Value Decomposition . . . . . 27
2.9 Conclusion . . . . . 27

**3 Introduction to Higher-Order Statistics 29**
3.1 Characteristic functions . . . . . 29
3.2 Moments, cumulants and spectra . . . . . 30
3.2.1 Moments . . . . . 30
3.2.2 Cumulants . . . . . 31
3.2.3 Spectra . . . . . 32
3.2.4 Relations between moments and cumulants . . . . . 33
3.2.5 Properties of moments and cumulants . . . . . 34
3.3 Shape of the probability density function . . . . . 38
3.4 Examples . . . . . 39
3.5 Estimation of HOS . . . . . 40
3.5.1 Estimation of moments . . . . . 40
3.5.2 Estimation of cumulants . . . . . 42
3.6 Conclusion . . . . . 44

**4 The Higher-Order Singular Value Decomposition 45**
4.1 Introduction . . . . . 45
4.2 The Higher-Order Singular Value Decomposition . . . . . 46



4.3 Derivation of the HOSVD-theorem . . . . . 50
4.4 Properties . . . . . 51
4.5 Computation . . . . . 54
4.6 Higher-Order Eigenvalue Decomposition . . . . . 55
4.7 First-order perturbation analysis . . . . . 58
4.8 Conclusion . . . . . 61

**5 Best Rank-1 and Rank-$(R_1, R_2, \ldots, R_N)$ Approximation 63**
5.1 Power method and orthogonal iteration . . . . . 64
5.2 Higher-order power iteration . . . . . 65
5.2.1 Best rank-1 approximation . . . . . 66
5.2.2 A power algorithm . . . . . 68
5.2.3 Starting value . . . . . 72
5.2.4 Gradient expressions . . . . . 76
5.3 Higher-order orthogonal iteration . . . . . 77
5.3.1 Best rank-$(R_1, R_2, \ldots, R_N)$ approximation . . . . . 78
5.3.2 An orthogonal iteration . . . . . 80
5.3.3 Gradient expressions . . . . . 84
5.4 Higher-Order Hessenberg Decomposition . . . . . 85
5.5 Conclusion . . . . . 89

**6 The Canonical Decomposition 91**
6.1 Introduction . . . . . 92
6.2 CANDECOMP and simultaneous EVD . . . . . 95
6.3 CANDECOMP and simultaneous generalized Schur decomposition 99
6.4 Algorithms for the simultaneous generalized Schur decomposition 101
6.4.1 Extended QZ-iteration . . . . . 101

6.4.2 One-sided Givens iterations . . . 102
6.4.3 Two-sided Jacobi iterations . . . 105
6.4.4 Gradient-based optimization . . . 109
6.5 A matrix representation . . . 110
6.6 Least-squares model fitting . . . 112
6.6.1 ALS descent . . . 112
6.6.2 Gradient . . . 113
6.7 Simulations . . . 113
6.8 Conclusions . . . 118

**7 Independent Component Analysis and (HO)SVD 121**

7.1 Independent Component Analysis . . . 122
7.2 A two-stage procedure . . . 123
7.2.1 Step 1: prewhitening . . . 123
7.2.2 Step 2: determination of **Q** from the higher-order statistics of $Y$ . . . 125
7.3 Identifiability . . . 126
7.4 PCA versus ICA . . . 128
7.5 Contrast functions . . . 129
7.6 Performance of ICA . . . 130
7.6.1 Measures of performance . . . 130
7.6.2 Invariance and uniform performance . . . 131
7.6.3 Effect of prewhitening . . . 132
7.7 ICA by means of HOSVD . . . 135
7.7.1 Principle . . . 135
7.7.2 Properties . . . 135
7.7.3 Simulations . . . 138



7.8 Literature overview . . . 140
7.9 ICA by mere PCA? . . . 144
7.10 Conclusion . . . 148

**8 ICA and Higher-Order Best Rank-1 Approximation 153**

8.1 The Maximal Diagonality approach to ICA . . . 154
8.1.1 Basic idea . . . 154
8.1.2 Algorithm . . . 155
8.2 Fourth-order Jacobi-rotations and best rank-1 approximation . 160
8.2.1 Determination of a Jacobi-rotation . . . 161
8.2.2 Comparison . . . 164
8.2.3 Simulations . . . 165
8.3 Conclusion . . . 168

**9 ICA and (Simultaneous) Third-Order Tensor Diagonalization 171**

9.1 The MD technique for third-order cumulants . . . 172
9.2 ICA by Simultaneous Third-Order Tensor Diagonalization . . . 174
9.2.1 Basic idea . . . 174
9.2.2 Computation . . . 176
9.2.3 Properties . . . 178
9.2.4 Simulations . . . 180
9.3 The JADE-algorithm . . . 182
9.4 An algebraic framework for ICA . . . 185
9.5 Conclusion . . . 185

**10 ICA and CANDECOMP 187**

10.1 Higher-order-only ICA . . . 187
10.1.1 Concept . . . 187

10.1.2 Identifiability . . . . . 188
10.1.3 Simultaneous generalized Schur decomposition vs JADE 189
10.1.4 The equivalent of PCA in higher-order-only ICA . . . . 190
10.2 Simulations . . . . . 190
10.3 Conclusions . . . . . 193

**11 Applications** **195**
11.1 Factor analysis of tongue shapes . . . . . 195
11.1.1 Modelling the vocal tract during vowel production . . . 195
11.1.2 Analysis . . . . . 196
11.1.3 Discussion . . . . . 197
11.2 Blind beamforming in wireless communications . . . . . 199
11.3 FECG extraction by BSSS . . . . . 201
11.3.1 Introduction . . . . . 201
11.3.2 Extraction of the FECG from abdominal recordings . . 203
11.3.3 Extraction of the FECG with BSSS . . . . . 204
11.3.4 Example . . . . . 206
11.4 Conclusion . . . . . 209

**12 Conclusion** **211**
12.1 Contributions . . . . . 211
12.2 Suggestions for further research . . . . . 215

**A Matrix Definitions** **219**

**B The Multivariate Gaussian Distribution** **225**

**C Derivations** **227**



**Bibliography** **235**

# Signaalverwerking met Multilineaire Algebra

## Nederlandse samenvatting

### Hoofdstuk 1: Inleiding

Multilineaire algebra is de algebra van hogere-orde tensoren. In deze thesis kunnen hogere-orde tensoren intuïtief worden beschouwd als het meerdimensionele equivalent van matrices (tweede orde) of vectoren (eerste orde), t.t.z. als "blokken" van getallen, in drie of meer dimensies. De getalwaarden die een $N$e orde tensor bevat, zijn gedefinieerd met betrekking tot basissen, gekozen in $N$ referentie-vectorruimten. Coördinatentransformaties die een interessante voorstelling van de tensor induceren, geven aanleiding tot de definitie van verschillende types multilineaire decomposities; men kan zich gelijkaardige vragen stellen als voor matrices - de antwoorden zijn nu eens verwant, dan weer merkwaardig verschillend.

De huidige interesse van de signaalverwerkingsgemeenschap in multilineaire algebra hangt samen met het stijgende belang van het concept *Hogere-Orde Statistiek* (HOS), en de snelle ontwikkelingen in dit onderzoeksdomein. De basisgrootheden van HOS, hogere-orde momenten en cumulanten, zijn hogere-orde tensoren. Voor een vrij recent overzicht van workshops, speciale uitgaven van tijdschriften, boeken, enz. wordt de lezer verwezen naar [252].

Talrijke belangrijke resultaten in signaalverwerking en systeemtheorie werden bekomen a.h.v. eerste- en tweede-orde statistische beschrijvingen, d.w.z. in termen van verwachte waarde en variantie. Deze benadering houdt echter ook zekere beperkingen in: hoewel eerste- en tweede-orde statistiek adequaat zijn voor het karakteriseren van Gaussiaanse signalen, volstaan ze in het algemeen niet voor een volledige beschrijving van willekeurige verdelingen. Aan de andere

kant bestaat een interessante eigenschap van hogere-orde cumulanten erin, dat ze ongevoelig zijn voor additieve Gaussiaanse signaalcomponenten; dit houdt een voordeel in van cumulant-gebaseerde methodologieën in omgevingen onderhevig aan Gaussiaanse ruis. Bovendien, zoals later duidelijk zal worden, kan de structuur van HOS "rijker" genoemd worden dan die van hun tweede-orde equivalenten, hetgeen toelaat voorheen onopgeloste problemen aan te pakken.

Een belangrijke klasse van problemen houdt verband met de identificatie van lineaire systemen zonder het observeren van de ingangen. Dit soort probleem wordt "blinde identificatie" genoemd, of "blinde egalisatie", als het hoofddoel bestaat uit de reconstructie van de ingangssequentie. De oplossing kan niet worden gebaseerd op tweede-orde statistiek, vermits het vermogensspectrum op de eenheidscircel per definitie alleen reële waarden aanneemt, en bijgevolg geen fase-informatie draagt. Een verwant probleem is "blinde bronscheiding". Hier gaat het om de situatie waarin een onbekende lineaire combinatie van componentsignalen, uitgezonden door een aantal bronnen, na propagatie wordt opgevangen door een antennerij; het doel is de reconstructie van de individuele bronsequenties uit de meetgegevens. In deze context hebben orthogonale combinaties van de bronnen geen effect op de geobserveerde covariantiematrix. Niettemin kunnen beide problemen worden opgelost m.b.v. HOS. Andere benaderingen bestaan erin gebruik te maken van signaaleigenschappen zoals constante modulus [266], eindig alfabet [253], spectrale zelf-coherentie [4] of cyclostationariteit [187, 258, 199, 244].

Meer in het algemeen worden HOS en tensortechnieken gebruikelijk gesitueerd in het gebied voorbij de klassieke signaalverwerking, dat wordt benoemd als niet-Gaussiaans, niet-lineair en/of niet-stationair. Toepassingsgebieden zijn beeldverwerking, patroonherkenning, textuuranalyse, spraakverwerking, telecommunicatie, biomedische problemen, seismische prospectie, radar, sonar, astrofysica, neurale netwerken, computergebaseerde regeltechniek, enz., zoals een vluchtige blik op de literatuur duidelijk maakt.

Naast HOS is *veeltermtheorie* een tweede onderzoeksgebied dat wel vaart bij nieuwe ontwikkelingen in multilineaire algebra. Op dezelfde manier als vectoren en matrices een formalisme zijn voor lineaire resp. kwadratische/bilineaire vormen, komen hogere-orde tensoren overeen met homogene veeltermen of, meer algemeen, multilineaire vormen. Onder veeltermvorm zijn tensoren overigens ooit intensief bestudeerd geweest; de ontbinding van kwadratische, kubische, vierde-orde, ... vormen in een som van eenvoudige termen was een belangrijk onderwerp voor wiskundigen in de negentiende eeuw [67]. Het doel was hoofdzakelijk decomposities te vinden die op een exacte manier opgaan; voor tensoren lijken kleinste-kwadratenbenaderingen, en numerieke technieken in het algemeen, minder bestudeerd.

Ten derde is de ontwikkeling van tensormethoden uiteraard belangrijk voor de verwerking en analyse van *meerwegs-datatabellen*. Momenteel lijkt de reflex om



na te gaan of een meetexperiment best twee-wegs dan wel meer-wegs gegevens oplevert, en hoe een eventueel meerwegs-experiment kan worden georganiseerd, het meest ontwikkeld in Psychometrie en Chemometrie. In de voorbije decennia werden belangrijke datamodellen als het Tucker model [259], Meerwegs Factor Analyse [163] en de Canonieke Decompositie [49, 138] voorgesteld; ze zullen verder worden behandeld in deze thesis.

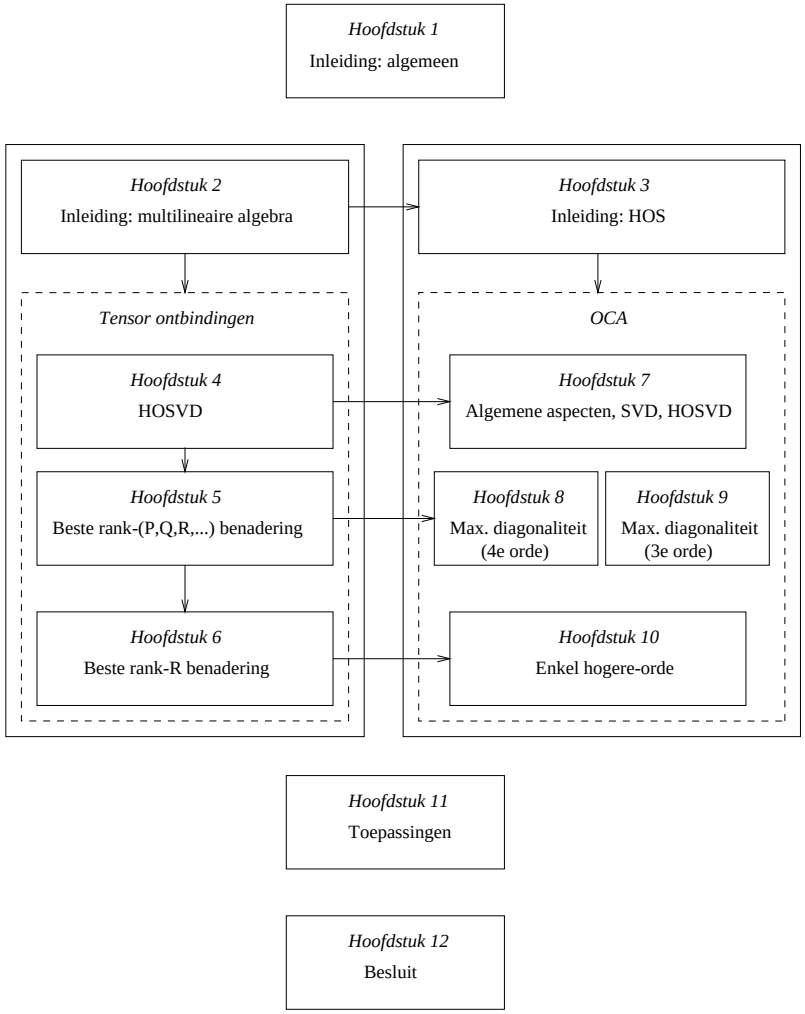
We vermelden ook dat tensoren een prominente rol spelen in de fysica [117, 145, 223, 280]: mechanica (bv. kinematica en dynamica van vloeistoffen), sterkteleer, elasticiteitsleer, relativiteitstheorie, ... Meestal ligt de nadruk echter op het "tensor" aspect, en minder op de "hogere-orde" specificatie; er is hoofdzakelijk belangstelling voor de manier waarop fysische parameters worden beïnvloed door coördinatentransformaties, en voor de ontwikkeling van basisonafhankelijke inzichten.

Er blijkt maar weinig interactie te zijn tussen de verschillende onderzoeksdomeinen waarvoor de ontwikkeling van multilineaire algebraïsche technieken relevant kan zijn. We hopen dat deze thesis kan helpen een brug te slaan.

De structuur van de thesis wordt getoond in Fig. 0.1. De linkerkant behandelt multilineaire algebraïsche onderwerpen; de rechterkant gaat meer specifiek over HOS. De verticale onderverdeling weerspiegelt de structuur van inleiding, theoretische ontwikkelingen, toepassingen en besluit.

Het is welbekend dat de *Singuliere Waardenontbinding* (SWO) van matrices en haar tweede-orde veralgemeningen een centrale rol spelen in de analyse en oplossing van wiskundige ingenieursproblemen; de literatuur over verwante probleemconcepten (lineaire/gestructureerde/totale kleinste-kwadraten, Principale Componenten Analyse (PCA), ... ), SWO-gebaseerde signaalverwerkingsmethoden (parameterschatting, identificatie, regeltechniek, ... ), numerieke algoritmen (snelle algoritmen, parallelle rekentechnieken, grote-schaal problemen, ... ), enz., is zeer uitgebreid [106, 262, 195]. Dit succes in het achterhoofd, is het hoofdthema van het eerder algebraïsche deel van deze thesis de veralgemening voor hogere-orde tensoren van de SWO en het daarbij aansluitende onderwerp van de beste rang-$R$ benadering. Hoewel de matrix decompositie een merkwaardig aantal krachtige eigenschappen bundelt (orthonormaliteit van de singuliere vectoren, diagonaliteit van de matrix van singuliere waarden, in essentie uniciteit, ... ), blijkt dit echter niet langer het geval op het multilineaire niveau. Ruwweg gesteld, kan de veralgemening van verschillende deelaspecten aanleiding geven tot verschillende tensordecomposities. Dit hangt samen met verschillen tussen matrices en tensoren m.b.t. het begrip rang (Hfdst. 4, 5, 6).

Het HOS-deel van deze thesis behandelt hoofdzakelijk het fundamentele probleem van *Onafhankelijke Componenten Analyse* (OCA), eerder "Blinde Bronscheiding" genoemd. Wiskundig gesproken, is de bedoeling een lineaire combi-




Figuur 0.1: Structuur van de thesis.

natie van statistisch onafhankelijke signalen op te splitsen in zijn componenten, zonder voorkennis van het mengmechanisme, noch van de signaalwaarden aangenomen door de onafhankelijke componenten. OCA is de natuurlijke verfijning van PCA, waar alleen wordt opgelegd dat de componenten ongecorreleerd zijn



(tweede-orde), hetgeen een rotationele vrijheidsgraad in de oplossingsverzameling laat. In deze thesis wordt het ICA-probleem aangepakt met multilineaire algebraïsche technieken; voor een literatuuroverzicht verwijzen we naar Sect. 7.8. Zoals PCA en SWO in verband staan, maakt OCA gebruik van tensoriële SWO-veralgemeningen. Het feit dat verscheidene types veralgemening mogelijk zijn, leidt tot verschillende oplossingsstrategieën voor OCA. Tussen de algebraïsche hoofdstukken links in Fig. 0.1, en de OCA-hoofdstukken rechts op het zelfde niveau, is er telkens een verband (Hfdst. 7, 8, 9, 10).

We bekijken de verschillende hoofdstukken nu wat meer in detail.

## Hoofdstuk 2: Inleiding Multilineaire Algebra

Dit hoofdstuk is het eerste van twee inleidende hoofdstukken. Het bevat het basismateriaal aangaande multilineaire algebra. Hfdst. 3 bestaat uit een inleiding tot HOS.

Zoals reeds vermeld, kunnen hogere-orde tensoren in deze thesis intuïtief worden beschouwd als "matrices met meerdere indices", die op een specifieke manier veranderen bij coördinatentransformaties (voor een formele definitie verwijzen we naar Def. 2.1.1). De vectorruimte van de reële $(I_1 \times I_2 \times \ldots \times I_N)$-tensoren wordt aangeduid als $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$; de complexe tegenhanger is $\mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$. Tensormanipulaties worden in de literatuur vaak weergegeven a.h.v. de index notatie en de Einstein sommatie conventie [155]. Dit formalisme is zeer algemeen en instructief, maar heeft het nadeel dat het slechts zelden gebruikt wordt in de context van numerieke algebra en signaalverwerking. Om de thesis meer toegankelijk te maken, maken we gebruik van een matrix-achtige notatie.

Er zijn grote verschillen tussen matrices en tensoren wat het begrip "rang" betreft. We maken gebruik van volgende definities:

**Definitie 0.0.1 (Rang-1 tensor)** *De rang van een $N$e-orde tensor $\mathcal{A}$ is gelijk aan 1, wanneer die tensor bestaat uit het uitwendig product van $N$ vectoren $U^{(1)}$, $U^{(2)}$, ..., $U^{(N)}$:*

$$\mathcal{A} = U^{(1)} \circ U^{(2)} \circ \ldots \circ U^{(N)}.$$

*d.w.z.* $a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$ *voor alle indexwaarden.*

**Definitie 0.0.2 (Rang)** *De rang van een willekeurige $N$e-orde tensor $\mathcal{A}$, voorgesteld door $R = \mathtt{rang}(\mathcal{A})$, is het kleinste aantal rang-1 tensoren die in een lineaire combinatie $\mathcal{A}$ opleveren.*

**Definitie 0.0.3 ($n$-rang)** *De $n$-mode vectoren van een $N$e-orde tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ zijn de $I_n$-dimensionele vectoren, bekomen uit $\mathcal{A}$ door de index $i_n$ te laten variëren, terwijl de andere indices constant blijven. De $n$-rang van $\mathcal{A}$, voorgesteld door $R_n = \mathtt{rank_n}(\mathcal{A})$, is de dimensie van de vectorruimte gegenereerd door de $n$-mode vectoren.*

De laatste definitie veralgemeent de begrippen "kolomvector (rijvector)" en "kolomrang (rijrang)" van matrices. In tegenstelling tot matrices, zijn de verschillende $n$-rangen van een hogere-orde tensor niet noodzakelijk gelijk; en als ze al gelijk zijn, dan kunnen ze nog steeds verschillen van de rang van de tensor. Uit de definities volgt dat $R_n \leqslant R$.

Scalair product, orthogonaliteit en Frobenius-norm kunnen worden gedefinieerd voor hogere-orde tensoren, op een manier analoog met matrices (Def. 2.4.1-2.4.4).

We geven nu een definitie van de vermenigvuldiging van een hogere-orde tensor met een matrix. Dit inwendig product laat o.a. toe het effect van basistransformaties op een gegeven tensor uit te drukken.

**Definitie 0.0.4** *Het $n$-mode product van een tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ en een matrix $\mathbf{U} \in \mathbb{C}^{J_n \times I_n}$, voorgesteld door $\mathcal{A} \times_n \mathbf{U}$, is een $(I_1 \times I_2 \times \ldots J_n \ldots \times I_N)$-tensor, gegeven door*

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \ldots j_n \ldots i_N} = \sum_{i_n} a_{i_1 i_2 \ldots i_n \ldots i_N} u_{j_n i_n},$$

*voor alle indexwaarden.*

De toepassing van tensortechnieken in de context van bv. HOS, vereist inzicht in symmetrie-eigenschappen. Reële tensoren worden "super-symmetrisch" genoemd als ze invariant zijn voor alle mogelijke permutaties van de indices. Voor complexe tensoren kan het begrip "super-symmetrie" worden gedefinieerd door met de tensor een homogene veelterm in meerdere veranderlijken te associëren, en te eisen dat dat de tensor voldoet aan gelijkaardige symmetrie-eigenschappen als de veelterm, m.b.t. permutaties van de veranderlijken.

Net zoals er een isomorfisme bestaat tussen matrix algebra en de algebra van lineaire vectorafbeeldingen, kunnen hogere-orde tensoren worden beschouwd als een formele voorstelling van lineaire afbeeldingen tussen een matrix- en een vectorruimte, een matrix- en een matrixruimte, een matrix- en een hogere-orde tensorruimte, enz. Veronderstellen we bijvoorbeeld een basis in de ruimte van $N_1$e-orde en $N_2$e-orde tensoren, dan kan een lineaire transformatie van $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_{N_1}}$ naar $\mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_{N_2}}$ worden voorgesteld door $\mathcal{C} \in \mathbb{C}^{I_1 \times \ldots I_{N_2} \times J_1 \times \ldots \times J_{N_1}}$, als volgt:



$$b_{i_1 i_2 \ldots i_{N_2}} = \sum_{j_1 j_2 \ldots j_{N_1}} c_{i_1 i_2 \ldots i_{N_2} j_1 j_2 \ldots j_{N_1}} a_{j_1 j_2 \ldots j_{N_1}}, \tag{0.1}$$

voor alle indexwaarden. Een zelfde tensor kan verschillende lineaire afbeeldingen voorstellen, afhankelijk van welke indices voor de sommatie worden geselecteerd. In analogie met matrices, kunnen bv. een Eigenwaardenontbinding (EWO) en een SWO van de lineaire afbeelding worden beschouwd. Dit eenvoudige principe vormt het vertrekpunt voor verschillende latere afleidingen.

## Hoofdstuk 3: Inleiding Hogere-Orde Statistiek

Dit hoofdstuk beschrijft het basismateriaal aangaande HOS. Eerst geven we een definitie van de basisgrootheden, nl. hogere-orde momenten en cumulanten.

**Definitie 0.0.5 (Moment)** *Het $N$e-orde moment $m_N^X$ van een reële stochastische variabele $X$ is gedefinieerd als*

$$m_N^X \stackrel{\text{def}}{=} \mathtt{E}\{X^N\}. \tag{0.2}$$

Het $N$e-orde moment van een reële stochastische vector $X$ is een $N$e-orde tensor, gedefinieerd als

$$\mathcal{M}_N^X \stackrel{\text{def}}{=} \mathtt{E}\{X \circ X \circ \ldots \circ X\}, \tag{0.3}$$

waarbij de verwachte waarde componentsgewijs wordt berekend. Voor complexe stochastische vectoren veronderstellen we bij conventie de volgende complexe symmetrie:

$$\begin{aligned} M_1^Z &= \mathtt{E}\{Z\}, \\ \mathbf{M}_2^Z &= \mathtt{E}\{Z \circ Z^*\}, \\ \mathcal{M}_3^Z &= \mathtt{E}\{Z \circ Z^* \circ Z\}, \\ \mathcal{M}_4^Z &= \mathtt{E}\{Z \circ Z^* \circ Z^* \circ Z\}. \end{aligned}$$

Voor reële toevalsgrootheden met gemiddelde 0, worden de cumulanten tot orde 4 gegeven door:

$$\mathtt{Cum}(X_1) = \mathtt{E}\{X_1\}, \tag{0.4}$$

$$\mathtt{Cum}(X_1, X_2) = \mathtt{E}\{X_1 X_2\}, \tag{0.5}$$

$$\mathtt{Cum}(X_1, X_2, X_3) = \mathtt{E}\{X_1 X_2 X_3\}, \tag{0.6}$$

$$\begin{aligned}\texttt{Cum}(X_1, X_2, X_3, X_4) &= \texttt{E}\{X_1X_2X_3X_4\} - \texttt{E}\{X_1X_2\}\texttt{E}\{X_3X_4\} \\ &\quad -\texttt{E}\{X_1X_3\}\texttt{E}\{X_2X_4\} \\ &\quad -\texttt{E}\{X_1X_4\}\texttt{E}\{X_2X_3\}. \end{aligned} \tag{0.7}$$

Voor reële stochastische variabelen met gemiddelde verschillend van 0, wordt $X_i$ in deze formules, behalve de eerste, vervangen door $X_i - \texttt{E}\{X_i\}$. Voor complexe toevalsgrootheden is het gebruikelijk in beide leden van Verg. (0.6) $X_2^*$ te beschouwen; in Verg. (0.7) worden zowel $X_2$ als $X_3$ complex toegevoegd. Cumulanten van stochastische vectoren worden analoog gedefinieerd als bij momenten.

Hogere-orde momenten en cumulanten van een stochastische vector zijn supersymmetrische hogere-orde tensoren. Zij transformeren op een multilineaire manier wanneer de vector wordt vermenigvuldigd met een matrix. Voor een reële variabele met een even verdeling, en voor een complexe variabele met een verdeling die punt-symmetrisch is rond de oorsprong, zijn de momenten en cumulanten van oneven orde gelijk aan 0. Ondanks hun minder voor de hand liggende definitie, zijn cumulanten vaak interessanter dan momenten:

- Tweede- en hogere-orde cumulanten zijn ongevoelig voor het gemiddelde van een toevalsgrootheid. Cumulanttensoren veranderen op een multilineaire manier bij willekeurige affiene transformaties, die een translatie kunnen omvatten.

- Kruiscumulanten van onafhankelijke variabelen zijn nul. Cumulanttensoren van een stochastische vector met statistisch onafhankelijke componenten zijn diagonaal.

- Voor onafhankelijke toevalsgrootheden zijn de cumulanten van een som gelijk aan de som van de cumulanten.

- Hogere-orde cumulanten van een Gaussiaanse toevalsgrootheid zijn nul. Technieken enkel gebaseerd op hoger-orde cumulanten, zijn in principe blind voor additieve Gaussiaanse ruis. Hogere-orde cumulanten geven een indicatie voor de mate waarin een toevalsgrootheid als Gaussiaans kan beschouwd worden.

In de praktijk zijn algoritmen gebaseerd op cumulantschattingen, bekomen door uitmiddeling over meetgegevens. Sect. 3.5 bevat een bondig overzicht van formules die de kwaliteit van een schatting beschrijven in functie van het aantal meetwaarden en de signaalkarakteristieken.



# Hoofdstuk 4: De Hogere-Orde Singuliere Waardenontbinding

In dit hoofdstuk onderzoeken we hoe de SWO van matrices kan veralgemeend worden voor hogere-orde tensoren, onder de voorwaarde dat in het multilineaire equivalent de $n$-mode vectoren een gelijkaardige rol spelen als de kolom- en rijvectoren in de matrix SWO. De centrale stelling is de volgende:

**Stelling 0.0.6 ($N$e-Orde Singuliere Waardenontbinding)** *Elke complexe $(I_1 \times I_2 \times \ldots \times I_N)$-tensor $\mathcal{A}$ kan geschreven worden als het product*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}, \tag{0.8}$$

*waarin:*

- $\mathbf{U}^{(n)} = \left[U_1^{(n)} U_2^{(n)} \ldots U_{I_n}^{(n)}\right]$ *een unitaire $(I_n \times I_n)$-matrix is,*

- $\mathcal{S}$ *een complexe $(I_1 \times I_2 \times \ldots \times I_N)$-tensor is, waarvan de deeltensoren $\mathcal{S}_{i_n=\alpha}$, bekomen door de $n$-de index gelijk te stellen aan de constante $\alpha$, de volgende eigenschappen hebben:*

  - *geheel-orthogonaliteit: twee deeltensoren $\mathcal{S}_{i_n=\alpha}$ en $\mathcal{S}_{i_n=\beta}$ zijn orthogonaal voor alle mogelijke waarden van $n$, $\alpha$ en $\beta$, met $\alpha \neq \beta$:*

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \textit{voor} \quad \alpha \neq \beta, \tag{0.9}$$

  - *ordening:*

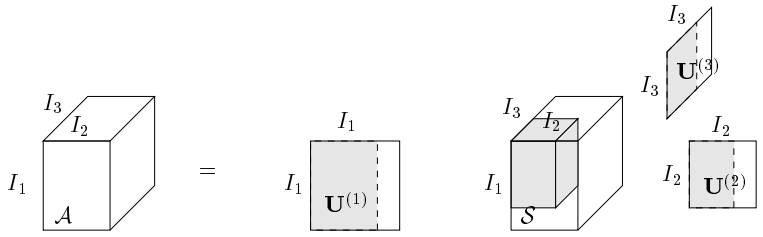
$$\|\mathcal{S}_{i_n=1}\| \geqslant \|\mathcal{S}_{i_n=2}\| \geqslant \ldots \geqslant \|\mathcal{S}_{i_n=I_n}\| \geqslant 0 \tag{0.10}$$

  *voor alle mogelijke waarden van $n$.*

*De Frobenius-norm $\|\mathcal{S}_{i_n=i}\|$, voorgesteld door $\sigma_i^{(n)}$, is de $i$-de $n$-mode singuliere waarde van $\mathcal{A}$ en de vector $U_i^{(n)}$ is de $i$-de $n$-mode singuliere vector. De decompositie wordt gevisualiseerd voor derde-orde tensoren in Fig. 0.2.*

In verband met het begrip "geheel-orthogonaliteit" merken we op dat pseudo-diagonaliteit van de kerntensor niet de geschikte manier is om de pseudo-diagonaliteit van de matrix van singuliere waarden te veralgemenen, vermits dit een decompositie zou opleveren die slechts in speciale gevallen mogelijk is.

In de psychometrische literatuur is de HOSWO van reële tensoren bekend als het Tucker model [259, 260], oorspronkelijk ontwikkeld voor de analyse van




Figuur 0.2: Voorstelling van de HOSWO voor een derde-orde tensor.

driewegs-gegevenstabellen. In dit hoofdstuk stellen we een vierkantswortelalgoritme voor als berekeningswijze en tonen we aan dat de decompositie een overtuigende veralgemening is van het SWO-concept. De eigenschappen van de HOSWO worden grondig geanalyseerd en geïnterpreteerd in een vergelijking met hun matrixequivalenten.

$\mathbf{U}^{(n)}$ kan worden berekend als de linker singuliere matrix van een matrix waarvan de kolommen de $n$-mode vectoren zijn (de volgorde speelt geen rol). De HOSWO van een $N$e-orde tensor leidt dus tot de berekening van $N$ verschillende matrix SWO's van matrices met dimensies $(I_n \times I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N)$ $(1 \leqslant n \leqslant N)$. Vervolgens kan de kerntensor $\mathcal{S}$ berekend worden door de matrices van singuliere vectoren naar het linkerlid te brengen in Verg. (0.8).

De HOSWO vertoont een gelijkaardige uniciteitseigenschap, verband met matrix EWO, interpretatie in termen van richtingen en waarden van extremale georiënteerde energie, eerste-orde perturbatiekarakteristiek, ... als de SWO.

Als een hogere-orde tensor door een indexpermutatie wordt afgebeeld op zichzelf of zijn complex toegevoegde, dan wordt deze symmetrie weerspiegeld door de HOSWO-componenten. De meest adequate manier om dit verband te analyseren is de decompositie van de permutatie in een sequentie van permutatiecycli. Als elk indexpaar verbonden is door een cyclus - een eigenschap die paarsgewijze symmetrie werd genoemd - dan zijn de verschillende matrices van hogere-orde singuliere vectoren ofwel gelijk ofwel complex toegevoegd. In analogie met de EWO van een reële symmetrische of complexe Hermitiaanse matrix kan dan de Hogere-Orde EWO (HOEWO) van een paarsgewijs symmetrische hogere-orde tensor worden gedefinieerd.



## Hoofdstuk 5: Hogere-Orde Beste Rang-1 en Rang-$(R_1, R_2, \ldots, R_N)$ Benadering

In dit hoofdstuk bespreken we een multilineaire veralgemening van het beste rang-$R$ benaderingsprobleem bij matrices. De vraag die we willen beantwoorden, is hoe een tensor in een optimale kleinste-kwadratenzin kan benaderd worden door een tensor met vooraf gespecifieerde $n$-mode rangen. In de literatuur van Psychometrie is de kleinste-kwadratenbenadering van een driewegsgegevenstabel door een gegevenstabel met gereduceerde kerntensor gekend als "drie-mode factor analyse" [161, 162, 163, 254]. In Hfdst. 5 wordt een vierkantswortelversie van het algoritme gegeven (Alg. 5.3.2) voor tensoren van willekeurige orde; de techniek wordt grondig geanalyseerd, er worden verbanden met matrixtechnieken gelegd, varianten voorgesteld en verbeteringen aangebracht.

Er zijn enkele merkwaardige verbanden, maar ook een aantal frappante verschilpunten tussen de beste rang-$R$ benadering van een matrix en de beste rang-$(R_1, R_2, \ldots, R_N)$ benadering van een $N$e-orde tensor:

- In tegenstelling tot de situatie bij matrices, kan de kostfunctie bij tensoren meerdere lokale optima hebben.

- Bij matrices kan de beste rang-$R$ benadering bekomen worden door de kleinste singuliere waarden gelijk te stellen aan nul, waarbij de grootste $R$ behouden blijven. Afknotting van de HOSWO kan wel een goede tensorbenadering opleveren, maar deze is over het algemeen suboptimaal. Onze simulaties suggereren echter dat er nog een - zwakke - link is: we hebben geobserveerd dat voor goed geconditioneerde problemen het HOSWO-resultaat over het algemeen in de vallei van de kostfunctie ligt waartoe ook het globale optimum behoort. Er is echter geen absolute garantie: er kunnen tegenvoorbeelden worden gegenereerd.

- Een manier om de dominante $R$-dimensionele deelruimten van de kolom- en rijruimte van een matrix te bepalen, is de iteratie van de machten $(R = 1)$ of de orthogonale iteratie $(R > 1)$. Hogere-orde veralgemeningen van deze technieken nemen de vorm aan van Alternerende Kleinste-Kwadratenalgoritmen (AKK algoritmen), waarmee de benadering kan verbeterd worden.

- Een iteratiestap van de hogere-orde methode van de machten bestaat in essentie uit een multilineaire afbeelding, i.p.v. een louter lineaire transformatie. Hogere-orde orthogonale iteraties kunnen bijkomend de schatting van een dominante deelruimte impliceren. Hierbij worden de deelruimten van de $n$-mode singuliere vectoren van de benadering mode per mode iteratief aangepast; de efficiëntie kan worden verhoogd door twee modes

tegelijk aan te passen, hetgeen kan gebeuren door voor een matrix het beste rang-$R$ benaderingsprobleem op te lossen.

- De Hessenberg Decompositie kan worden veralgemeend voor hogere-orde tensoren en gebruikt als voorbewerking, om de efficiëntie van hogere-orde iteraties van de machten en hogere-orde orthogonale iteraties te verhogen.

Voor super-symmetrische $(2 \times 2 \times \ldots \times 2)$-tensoren kan de berekening van de beste rang-1 benadering, van de super-symmetrische stationaire punten van de hogere-orde methode van de machten, en van de willekeurige stationaire punten van dit algoritme, worden geherformuleerd als de bepaling van de nulpunten van een veelterm.

We hebben ook expliciete uitdrukkingen opgesteld voor de gradiënt van de kostfunctie over de Stiefel manifold, waartoe de matrices van $n$-mode singuliere vectoren (of orthogonale transformaties ervan) van de benadering behoren (Sect. 5.2.4 en Sect. 5.3.3).

## Hoofdstuk 6: De Canonieke Decompositie

De HOSWO in Hfdst. 4 is niet erg instructief i.v.m. de rang van hogere-orde tensoren. Dit is een nadeel van unitaire (orthogonale) tensordecomposities in het algemeen. In dit hoofdstuk bestuderen we de volgende tensorontbinding [49, 138, 7, 19]:

**Definitie 0.0.7 (CANDECOMP)**
*Een Canonieke Decompositie of Parallelle-Factoren Decompositie van een tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ is een ontbinding van $\mathcal{A}$ als een lineaire combinatie van een minimaal aantal rang-1 termen:*

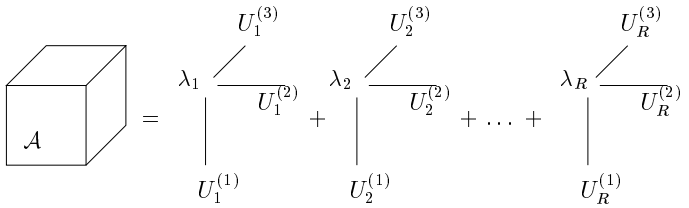
$$\mathcal{A} = \sum_{r}^{R} \lambda_r \, U_r^{(1)} \circ U_r^{(2)} \circ \ldots \circ U_r^{(N)}. \tag{0.11}$$

*Deze decompositie wordt gevisualiseerd voor derde-orde tensoren in Fig. 0.3.*

Deze decompositie is een multilineaire veralgemening van de diagonalisatie van een matrix door een equivalentie- of congruentietransformatie, echter met nogal afwijkende eigenschappen.

Uit de definitie is duidelijk dat de decompositie niet beïnvloed wordt door een permutatie van de rang-1 termen, noch door een scalering van de vectoren $U_r^{(n)}$ die gepaard gaat met de omgekeerde scalering van de coëfficiënten $\lambda_r$.






Figuur 0.3: Visuele voorstelling van de CANDECOMP voor een derde-orde tensor.

Los van deze triviale onbepaaldheden is de CANDECOMP eenduidig bepaald onder voorwaarden van lineaire onafhankelijkheid; dit staat in contrast met de situatie bij matrices, waar de ontbinding in een som van rang-1 termen doorgaans uniek wordt gemaakt door sterkere voorwaarden op te leggen, zoals orthogonaliteit.

De ontbinding van een gegevenstabel in een som van rang-1 termen, wordt soms ook het *Factor-Analyse* probleem genoemd. De bedoeling is de verschillende rang-1 termen te associëren met de verschillende "fysische mechanismen" die aan de waarden in de gegevenstabel ten grondslag liggen. Gezien het verschil in uniciteitseigenschappen tussen de matrix- en de tensordecompositie, kan de CANDECOMP van hogere-orde tensoren worden beschouwd als een belangrijke techniek in signaalverwerking.

In dit hoofdstuk bestuderen we het speciale maar belangrijke geval van een $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ met rang $R \leqslant \mathtt{min}\{I_1, I_2\}$ en 3-rang $R_3 \geqslant 2$. We nemen aan dat:

- de verzameling vectoren $\{U_r^{(1)}\}_{(1 \leqslant r \leqslant R)}$ lineair onafhankelijk is,
- de verzameling vectoren $\{U_r^{(2)}\}_{(1 \leqslant r \leqslant R)}$ lineair onafhankelijk is,
- en dat de verzameling $\{U_r^{(3)}\}_{(1 \leqslant r \leqslant R)}$ geen collineaire vectoren bevat.

Deze voorwaarden maken de CANDECOMP essentieel uniek. De uiteenzetting is beperkt tot reële derde-orde tensoren voor de eenvoud van notatie; de veralgemening naar hogere-orde tensoren en complexe waarden is voor de hand liggend.

Momenteel worden de CANDECOMP-factoren meestal berekend met een AKK afdalingsalgoritme, mogelijks geïnitialiseerd met een schatting bekomen via een matrix EWO [182, 227, 22, 225, 226].

De CANDECOMP kan echter ook berekend worden d.m.v. een simultane diagonalisatie, door een equivalentie- of congruentietransformatie, van een reeks matrices. Dit is numeriek betrouwbaarder dan de berekening van een enkelvoudige EWO, in het bijzonder voor matige signaal-ruisverhoudingen; het probleem kan trouwens geherformuleerd worden als een simultane EWO.

In de literatuur komt diagonalisatie door simultane congruentietransformatie ook voor in de afleiding van een Analytisch Constant Modulus Algoritme [266]; het werd vertaald in een simultane veralgemeende Schur-decompositie (de bepaling van twee orthogonale matrices $\mathbf{Q}$ en $\mathbf{Z}$ zodanig dat een reeks matrices $\{\mathbf{R}_k = \mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z}\}$, voor gegeven $\{\mathbf{V}_k\}$ ( $1 \leqslant k \leqslant K$), zo bovendiagonaals mogelijk is (in kleinste-kwadratenzin)), en vervolgens opgelost met een veralgemeend $QZ$-iteratieschema. In dit hoofdstuk hebben we ons toegelegd op éénzijdige Givens-iteraties, dubbelzijdige Jacobi-iteraties en gradiënt-gebaseerde optimisatie. Het blijkt dat de bepaling van een elementaire Givens-rotatie overeenkomt met de berekening van de dominante linker singuliere vector van een matrix met twee rijen. De bepaling van een elementaire Jacobi-rotatie leidt tot een stelsel van twee bikwadratische vergelijkingen in twee onbekenden, dat in essentie kan worden geformuleerd als het berekenen van de nulpunten van een veelterm van graad 8. Wat gradiënt-gebaseerde optimisatie betreft, hebben we een expliciete uitdrukking afgeleid voor de gradiënt van de kwadratische kostfunctie van de simultane veralgemeende Schur-decompositie, over de productmanifold van twee manifolds van orthogonale matrices (Subsect. 6.4.4).

Gradiënt-gebaseerde afdaling vormt ook een alternatief voor de AKK iteraties in een eventueel aansluitende minimisatie van de Frobenius-norm van het verschil tussen de oorspronkelijke tensor en zijn schatting. In dit verband hebben we een expliciete uitdrukking afgeleid voor de gradiënt van de kostfunctie over de productmanifold van manifolds van matrices waarvan de kolommen genormaliseerd zijn tot eenheidslengte (Sect. 6.6). Simulaties geven aan dat de bijkomende verbetering van de benadering die in deze stap wordt gerealiseerd, kleiner wordt naarmate het probleem beter geconditioneerd is.

We vermelden ook nog dat reguliere matrices waarvan de kolommen op arbitraire manier kunnen gescaleerd worden (bv. de factor-matrices van de CANDECOMP, de transfermatrix in OCA, ... ), op een eenduidige manier kunnen worden geparametriseerd als een koppel orthogonale matrices; deze voorstelling is interessant vanuit numeriek oogpunt.



# Hoofdstuk 7: Onafhankelijke Componenten Analyse en (HO)SWO

Beschouw het volgende statistische basismodel:

$$Y = \mathbf{M}X + N, \tag{0.12}$$

waarin $Y \in \mathbb{R}^I(\mathbb{C}^I)$ de *observatievector* of *uitgangsvector* genoemd wordt, $X \in \mathbb{R}^J(\mathbb{C}^J)$ de *bronvector* is, en $N \in \mathbb{R}^I(\mathbb{C}^I)$ additieve *ruis* voorstelt. $\mathbf{M} \in \mathbb{R}^{I\times J}(\mathbb{C}^{I\times J})$ is de *mengmatrix* of *transfermatrix*.

Het OCA-concept kan als volgt geformuleerd worden:

*Het doel van Onafhankelijke Componenten Analyse (OCA), of Blinde Bronscheiding (BBS), bestaat uit de schatting van de transfermatrix* $\mathbf{M}$ *en/of de overeenkomende realisaties van de bronvector* $X$*, enkel gebruikmakend van realisaties van de uitgangsvector* $Y$*, op basis van de volgende veronderstellingen:*

- *De kolommen van* $\mathbf{M}$ *zijn lineair onafhankelijk.*
- *De componenten van* $X$ *zijn onderling statistisch onafhankelijk, en statistisch onafhankelijk van de ruiscomponenten.*

De tweede veronderstelling is het sleutelingrediënt voor OCA. Het is een zeer sterke hypothese, maar toch vrij vanzelfsprekend in talrijke toepassingen. Vanuit een algebraïsch standpunt impliceert ze dat:

$$\mathbf{C}_Y = \mathbf{C}_X \times_1 \mathbf{M} \times_2 \mathbf{M}^* + \mathbf{C}_N, \tag{0.13}$$

$$\mathcal{C}_4^Y = \mathcal{C}_4^X \times_1 \mathbf{M} \times_2 \mathbf{M}^* \times_3 \mathbf{M}^* \times_4 \mathbf{M} + \mathcal{C}_4^N, \tag{0.14}$$

waarin de covariantiematrix $\mathbf{C}_X$ en de vierde-orde cumulant $\mathcal{C}_4^X$ diagonaal zijn (analoog voor andere tensorordes), en waarin $\mathcal{C}_4^N$ wegvalt als de ruis Gaussiaans is.

De norm van de kolommen van de mengmatrix kan onmogelijk bepaald worden, vermits een herschaling van deze vectoren kan gecompenseerd worden door de omgekeerde scalering van de bronsignaalwaarden. Ook de volgorde van de bronsignalen, die geen fysische betekenis heeft, kan niet geïdentificeerd worden. Voor bronvectoren met ten hoogste één Gaussiaanse component vormen deze onbepaaldheden de enige manier waarop een OCA-oplossing niet eenduidig bepaald is [64, 257].

De hierboven aangegeven veronderstellingen laten niet toe een onderscheid te maken tussen de signaalterm en de ruisterm in Verg. (0.12). Bijgevolg worden de bronsignalen geschat als $\hat{X}$, door een eenvoudige matrixvermenigvuldiging:

$$\hat{X} = \mathbf{W}^H Y. \tag{0.15}$$

$\mathbf{W}^H$ kan bijvoorbeeld de vorm aannemen van $\hat{\mathbf{M}}^\dagger$, waarin $\hat{\mathbf{M}}$ een schatting van de transfermatrix is. Meer algemeen kunnen verschillende technieken voor bundelvorming worden aangewend [274].

Verg. (0.13) geeft aanleiding tot een klassieke *Principale Componenten Analyse* (PCA), waardoor de bronnen en de mengmatrix tot op een unitaire (orthogonale) transformatie kunnen geschat worden. Om dit te illustreren, veronderstellen we even dat de bronnen een eenheidsvariantie hebben. Dan geldt (we laten hier de ruisterm weg voor de duidelijkheid):

$$\mathbf{C}_Y = \mathbf{M}\mathbf{M}^H. \tag{0.16}$$

Substitutie van de SWO van de mengmatrix, voorgesteld door $\mathbf{M} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}$, toont aan dat via de EWO van de geobserveerde covariantie de kolomruimte van $\mathbf{M}$ kan geschat worden, terwijl de factor $\mathbf{V}$ onbepaald blijft:

$$\mathbf{C}_Y = \mathbf{U}\mathbf{S}^2\mathbf{U}^H = (\mathbf{U}\mathbf{S})(\mathbf{U}\mathbf{S})^H. \tag{0.17}$$

Substitutie van de PCA-deelresultaten in Verg. (0.14) leidt (op de ruisterm na) tot een unitaire (orthogonale) transformatie van een hogere-orde cumulanttensor in een diagonaaltensor:

$$\mathcal{C}_4^Z = \mathcal{C}_4^X \times_1 \mathbf{V} \times_2 \mathbf{V}^* \times_3 \mathbf{V}^* \times_4 \mathbf{V}, \tag{0.18}$$

waarin $Z \stackrel{\text{def}}{=} \mathbf{S}^\dagger \cdot \mathbf{U}^H \cdot Y$. De onbekende unitaire (orthogonale) factor kan bijvoorbeeld geschat worden m.b.v. een HOEWO, vermits diagonaliteit een speciaal geval is van geheel-orthogonaliteit. In vergelijking met andere OCA-methoden is de rekenkost in deze benadering zeer laag. In volgende hoofdstukken zullen we zien dat de prijs in termen van nauwkeurigheid moet betaald worden, in vergelijking met meer rekenintensieve methoden. Een tweede nadeel is dat de vereiste identificeerbaarheidsvoorwaarden lichtjes zwaarder zijn dan degene die in het OCA-probleem op zich besloten liggen.

We merken op dat in het algemene twee-trapsschema voor OCA dat hierboven werd geschetst, de deelresultaten van de PCA-stap een rechtstreeks effect kunnen ondervinden van additieve Gaussiaanse ruis. De fout die in deze tweede-orde stap wordt geïntroduceerd, kan niet worden goedgemaakt in de hogere-orde stap, en brengt dus een bovengrens op de performantie van OCA-algoritmen met zich mee (Stelling 7.6.2 en 7.6.4).

Hoewel PCA op zich niet toelaat de mengmatrix en de bronsignalen op een blinde manier te identificeren, heeft ze in sommige gevallen toch een behoorlijke bronscheiding als resultaat. Een voorbeeld is de scheiding van bronnen met sterk verschillend vermogen (we nemen aan dat de kolommen van de transfermatrix normen van dezelfde grootte-orde hebben).



## Hoofdstuk 8: OCA en Hogere-Orde Beste Rang-1 Benadering

In Hfdst. 4 hebben we vermeld dat over het algemeen hogere-orde tensoren niet kunnen worden gediagonaliseerd d.m.v. unitaire (orthogonale) transformaties; daarom werd in de HOSWO de voorwaarde van geheel-orthogonaliteit vooropgesteld. Als een alternatief, zou de definitie van een tensoriële SWO kunnen gebaseerd worden op de optimale diagonalisatie, in kleinste-kwadratenzin, van de hogere-orde tensor. Dit criterium wordt aangeduid als *Maximale Diagonaliteit* (MD), of het *Comon-Lacoume* (CL) criterium.

In relatie met Verg. (0.18) definieert dit idee een nieuw OCA-algoritme. Deze benadering is conceptueel zeer belangrijk; ze kan worden geïnterpreteerd in termen van maximale waarschijnlijkheidsschatters en minimisatie van wederzijdse informatie in de geschatte bronsignalen [124, 125, 64]. Op basis van het MD-criterium valt het OCA-probleem ook in zijn volledige algemeenheid op te lossen. Het diagonaliseren van de hogere-orde cumulant kan gebeuren met een Jacobi-type iteratie [64]. Voor reële vierde-orde tensoren (het derde-orde geval wordt in volgend hoofdstuk behandeld) kan een elementaire rotatie bepaald worden door het berekenen van de nulpunten van een vierde-graadsveelterm; bij complexe tensoren is het probleem moeilijker en geeft het aanleiding tot een stelsel van twee veeltermvergelijkingen in twee onbekenden.

We kunnen echter aantonen dat een elementaire rotatie ook kan berekend worden via de beste rang-1 benadering van een super-symmetrische reële ($2 \times 2 \times 2 \times 2$)-tensor (in het reële geval) of een super-symmetrische reële ($3 \times 3 \times 3 \times 3$)-tensor (in het complexe geval). Deze benadering heeft drie voordelen: (1) ze kan de vorm aannemen van een sequentie van eenvoudige operaties (sommen en producten), (2) i.p.v. al de punten te berekenen waar de afgeleide van de kostfunctie nul is, mikt de methode rechtstreeks op de bepaling van het globale optimum, en (3) gedurende de berekening van de rotatie kan het verhogen van de waarde van de kostfunctie op een eenvoudige manier gevolgd worden. Een nadeel is dat de methode iteratief is; in deze zin is de techniek vooral interessant in het complexe geval.

## Hoofdstuk 9: OCA en (Simultane) Derde-Orde Tensor Diagonalisatie

Waar we in het vorige hoofdstuk MD-OCA gebaseerd op de vierde-orde cumulanttensor hebben behandeld, wordt in dit hoofdstuk het derde-orde geval onderzocht. De berekening van een elementaire Jacobi-rotatie is hier opnieuw equivalent met een beste rang-1 benaderingsprobleem, echter van een reële

symmetrische $(2 \times 2)$-matrix (in het reële geval) of van een reële symmetrische $(3 \times 3)$-matrix (in het complexe geval).

Deze, in vergelijking met het vierde-orde geval, gunstige rekenkost maakt het aantrekkelijk om ook voor vierde-orde cumulanten "iets derde-orde-achtig" te doen. Ook de optimale simultane diagonalisatie van een verzameling derde-orde tensoren komt neer op de beste rang-1 benadering van een $(2 \times 2)$- of $(3 \times 3)$-matrix; deze matrix bestaat uit een som van bijdragen van de individuele tensoren. Een OCA kan dan worden uitgevoerd d.m.v. de simultane diagonalisatie van derde-orde tensoren, die het bereik omspannen van de tensor $\mathcal{C}_4^Z$ in Verg. (0.18), geïnterpreteerd als een lineaire afbeelding van $\mathbb{C}^{J \times J \times J}$ naar $\mathbb{C}^J$. Deze methode van "Simultane Derde-Orde Tensor Diagonalisatie" (SDOTD) is verwant in filosofie met het JADE-algoritme [41]. Feitelijk vult SDOTD het gat op in een algebraïsch raamwerk voor PCA-gebaseerde OCA a.h.v. de vierde-orde cumulant. In de HOEWO-techniek, de JADE-techniek, de SDOTD-techniek en de MD-methode wordt de vierde-orde cumulant beschouwd als een verzameling van vectoren, matrices, derde-orde tensoren of een enkele vierde-orde tensor, die op één of andere manier worden gediagonaliseerd.

## Hoofdstuk 10: OCA en CANDECOMP

Verg. (0.14) is, op een ruisterm na, een CANDECOMP van de cumulant van $Y$. Deze decompositie kan dus worden aangewend voor een nieuwe OCA-techniek, die het voordeel heeft op PCA-gebaseerde OCA, dat hij asymptotisch ongevoelig is voor additieve Gaussiaanse perturbaties van de gegevens. Wanneer gebruik gemaakt wordt van cumulantschattingen, bekomen op basis van eindige datasequenties, dan is de nauwkeurigheid van de louter-hogere-orde versus de PCA-gebaseerde methodes onderwerp van een trade-off, veroorzaakt door het feit dat de schatting van HOS over het algemeen langere datareeksen vereist dan de schatting van tweede-orde statistieken. In een scenario met meer sensoren dan bronnen kan de dimensionaliteit van het probleem, zonder gebruik te maken van tweede-orde statistiek, herleid worden tot de grootte van het aantal bronsignalen, door het verschil tussen $n$-rang en rang van hogere-orde tensoren te exploiteren: voorafgaandelijk aan de berekening van de CANDECOMP, wordt eerst de beste rang-$(J, J, J, J)$ benadering van de cumulanttensor bepaald.



## Hoofdstuk 11: Toepassingen

In dit hoofdstuk illustreren we enkele typische aspecten van het CANDECOMP- en OCA-concept aan de hand van voorbeelden.

De eerste toepassing is te situeren in Linguïstiek. Misschien een beetje verrassend, blijkt de CANDECOMP een zeer geschikt en nauwkeurig model te zijn voor de beschrijving van de posities van de menselijke tong bij spraak [141]. In dit gebied gebeurt nogal wat onderzoek vanwege de grote interesse in neurale controle van de motoriek bij het genereren van spraak. Hogere-orde gegevenstabellen zijn het resultaat van het opmeten van de verplaatsing van het tongoppervlak op verschillende referentieposities, bij het uitspreken van verschillende klinkers, door een aantal proefpersonen. De toepassing is een goed voorbeeld van de hoge interpreteerbaarheid die de CANDECOMP in de praktijk vaak kenmerkt. De factoren kunnen geïnterpreteerd worden als (1) de genormaliseerde basispatronen waaruit een verplaatsing van het tongoppervlak bestaat, (2) de mate waarin elk basispatroon vereist is voor de tongverplaatsing bij verschillende klinkers (deze mode laat toe articulatie-eigenschappen van de verschillende klinkers op een sprekeronafhankelijke manier te vergelijken), en (3) de manier waarop individuele sprekers de basispatronen combineren (gemodelleerd als uniform over de verschillende klinkers). Om de analyse uit te voeren, kan men gebruik maken van de technieken behandeld in Hfdst. 6.

De meest aangehaalde toepassing van OCA is blinde bundelvorming en spatiale multiplexing in telecommunicatie, die capaciteitsverhoging en energiebesparing in draadloze communicatie tot doel hebben. We vermelden hier enkel dat, zelfs als een model van de manifold van de antennerij beschikbaar is, blinde methoden toch vaak de voorkeur verdienen, wegens hun inherente ongevoeligheid voor calibratiefouten en de meer geschikte manier van omgaan met multipad-propagatie.

Een belangrijk biomedisch signaalverwerkingsprobleem is de extractie van het foetaal elektrocardiogram uit meerkanaals-potentiaalmetingen op de huid van de moeder; we hebben OCA voorgesteld als een nieuwe manier om dit probleem aan te pakken. In vergelijking tot een eerder voorgestelde SWO-gebaseerde methode [26], vereist de hogere-orde OCA-stap bijkomend de schatting en (partiële) diagonalisatie van de cumulanttensor van de meetgegevens. OCA is hier een zeer ambitieuze methode: ze mikt op de volledige reconstructie van de verschillende statistisch onafhankelijke bio-elektrische bronsignalen én van de karakteristieken van hun propagatie naar de elektroden - beide aspecten leveren belangrijke medische informatie.

# Hoofdstuk 12: Besluit

Deze thesis levert een bijdrage tot de veralgemening van de SWO van matrices naar hogere-orde tensoren. Drie verschillende invalshoeken werden besproken: (1) de HOSWO, die het verband tussen de SWO-componenten en de karakteristieken van kolom- en rijvectoren veralgemeent (bv. wat georiënteerde energie betreft), (2) de beste rang-$(R_1, R_2, \ldots, R_N)$ benadering van een gegeven $N$e-orde tensor door een tensor waarvan de $n$-mode vectoren hoogstens een $R_n$-dimensionele ruimte omspannen ($1 \leqslant n \leqslant N$), en (3) de CANDECOMP, waar een gegeven tensor wordt ontbonden in een minimale som van mogelijks niet-orthogonale rang-1 termen.

Multilineaire algebraïsche technieken vinden een belangrijk toepassingsgebied in HOS, waar de basisgrootheden super-symmetrische hogere-orde tensoren zijn. In deze thesis hebben we ons hoofdzakelijk toegelegd op het fundamentele concept van OCA. Er werden nieuwe verbanden ontwikkeld met de HOEWO, de beste rang-1 benadering, de simultane diagonalisatie van hogere-orde tensoren en de CANDECOMP.

Numerieke multilineaire algebra is een zeer jonge discipline; in de komende jaren mogen verschillende nieuwe resultaten, uitbreidingen en verbeteringen worden verwacht. We raken tot slot enkele open problemen aan en geven richtingen aan voor verder onderzoek.

De uniciteitsvoorwaarden voor de CANDECOMP waarvan in Hfdst. 6 wordt uitgegaan, kunnen worden afgezwakt tot meer algemene patronen van lineaire onafhankelijkheid [165, 37, 67]. Een belangrijke eigenschap is dat voor een $(I_1 \times I_2 \times \ldots \times I_N)$-tensor, met $N > 2$, vaak meer dan $\texttt{max}\{I_1, I_2, \ldots, I_N\}$ rang-1 termen kunnen geïdentificeerd worden. Belangrijk werk is de ontwikkeling van numerieke algoritmen die deze gevallen aankunnen. Er moet ook verder onderzoek gebeuren naar de identificatie van de mengmatrix in OCA, wanneer er meer bronsignalen dan sensoren zijn.

Een belangrijke opmerking betreft het convergentiegedrag van de algoritmen die werden besproken. Momenteel zijn alleen in een zeer beperkt aantal gevallen expliciete uitdrukkingen voor de convergentiesnelheid beschikbaar. In de thesis en in de literatuur wordt ook vaak globale convergentie verondersteld, terwijl formele bewijzen ontbreken. In grote mate is dit te wijten aan het feit dat tensordecomposities een ingewikkelder structuur vertonen dan hun matrixequivalenten. Een voorbeeldje: de convergentie van de methode van de machten kan bij matrices op eenvoudige wijze worden geanalyseerd in termen van de EWO-componenten, terwijl bij hogere-orde tensoren het residu na beste rang-1 benadering veel moeilijker te parametriseren valt, en a priori nog steeds volle rang en volle $n$-rangen kan hebben. Vaak worden algoritmen ook afgeleid onder de veronderstelling dat de ruis klein is, hetgeen de aflei-



ding aanzienlijk kan vereenvoudigen. De globale performantie van algoritmen wordt doorgaans geanalyseerd aan de hand van "bewijzen door MATLAB", maar meer algemene theoretische afleidingen zouden het inzicht in tensormechanismen kunnen verhogen. We vermelden ook de nood aan verdere analyse van de aard van "moerassen" in CANDECOMP-algoritmen [220, 166], waar de convergentie plots zeer traag kan gaan, mogelijks gevolgd door de ontwikkeling van regularisatiemethoden om de performantie te verhogen.

De technieken die in deze thesis worden ontwikkeld, zijn hoofdzakelijk multilineaire veralgemeningen van de methode van de machten en Givens/Jacobi-iteratieschema's. Bijkomende resultaten en snelle algoritmen kunnen worden verwacht van de veralgemening van gesofisticeerder matrixtechnieken voor EWO/SWO-berekeningen, zoals bv. de $QR$-iteratie [130]. Een tweede belangrijke discipline die de numerieke multilineaire algebra kan ten goede komen, is het onderzoek naar de oplossing van stelsels van veeltermvergelijkingen.

De bespreking van BBS bleef in deze thesis beperkt tot de identificatie van scalaire mengsels van onderling onafhankelijke signalen. Meerkanaals blinde deconvolutie kan in vele gevallen worden opgesplitst in een scheidings- en een egalisatiestap [267], zodat BBS ook hier relevant is. Tensordecomposities kunnen mogelijks ook worden beschouwd als bouwblokken voor de (blinde) identificatie van Volterra-filters; hun matrixequivalenten spelen overigens een belangrijke rol in de identificatie van het lineaire toestandsruimtemodel [271, 272].

Met betrekking tot gradiënt-gebaseerde optimisatiestrategieën voor de berekening van tensordecomposities, merken we op dat de afleiding van tijdsdiscrete implementaties van tijdscontinue algoritmen, waarvoor de uitdrukkingen van de gradiënten in eerste instantie werden opgesteld, niet a priori voor de hand liggend hoeft te zijn; voor een bespreking en voorbeelden verwijzen we naar [77].

Multilineaire algebra en HOS hebben een groot toepassingsgebied. Blinde chemische analyse door een CANDECOMP van meerwegs-gegevenstabellen in Chemometrie [225, 226], en de verfijning van PCA tot OCA in vibro-akoestische operationele responsanalyse [211], zijn twee voorbeelden van ingenieursproblemen waar een tensorbenadering nuttig kan zijn.

# Chapter 1

# Introduction

Multilinear algebra is the algebra of higher-order tensors. In this thesis, higher-order tensors can intuitively be imagined as the multidimensional equivalent of matrices (second order) or vectors (first order), i.e. as "blocks" of numbers, in three or more dimensions. The entries of an $N$th order tensor are defined with respect to the bases chosen in $N$ reference vector spaces. By looking for coordinate transformations that induce an interesting representation of the tensor, one can define several types of multilinear decompositions; similar questions as for matrices can be raised - the answers are sometimes related, and sometimes remarkably different.

The current interest of the signal processing community in multilinear algebra, goes together with the growing importance of the concept of *Higher-Order Statistics* (HOS), and the fast developments in this field of research. The basic HOS quantities, higher-order moments and cumulants, are higher-order tensors. For a fairly recent overview of workshops, special journal issues, tutorials, books, edited books and bibliographies on HOS, the reader is referred to [252].

Numerous important results in signal processing and systems theory have been obtained with first- and second-order statistical descriptions, i.e. in terms of mean and variance. However, this approach also implies restrictions: although first- and second-order statistics are well-suited for the characterization of Gaussian signals, they are in general not sufficient for a complete description of arbitrary distributions. On the other hand, a nice feature of higher-order cumulants is that they are insensitive to additive Gaussian signal components; this constitutes an advantage of cumulant-based methodologies in Gaussian noise environments. In addition, as will become clear later on, the structure of HOS should be qualified as "richer" than their second-order counterparts, leading to the solution of previously unsolved problems.

A major class of problems is related to the identification of linear systems without observation of the inputs. This type of problem is called "blind identification", or "blind equalization", if reconstruction of the input sequence is the primary goal. The solution cannot be based on second-order statistics, as the power spectrum assumes by definition only real values on the unit circle, and hence does not carry information about the phase. A related problem is "blind source separation". Here one deals with the situation where an unknown linear combination of component signals, emitted by a number of sources, is captured after propagation by a sensor array; the aim is to reconstruct the individual source sequences from the data. In this context the observed covariance matrix remains unaffected by orthogonal combinations of the sources. Nevertheless both problems can be solved resorting to HOS. Other approaches are to resort to signal properties as constant modulus [266], finite alphabet [253], spectral self-coherence [4] or cyclostationarity [187, 258, 199, 244].

More in general, HOS and tensor-techniques are commonly situated in the area beyond classical signal processing, labeled as non-Gaussian, non-linear and/or non-stationary. Application areas include image processing, pattern recognition, texture analysis, speech processing, telecommunications, biomedical issues, seismic prospection, radar, sonar, astrophysics, neural networks, automatic control, vibro-acoustic analysis, etc., as a brief inspection of the literature reveals.

Besides HOS, a second research area that benefits from advances in multilinear algebra, is the *theory of polynomials*. In the same way as vectors and matrices are a formalism for linear and quadratic/bilinear forms respectively, higher-order tensors correspond to homogeneous polynomials or general multilinear forms. As a matter of fact, under polynomial form, tensors were once the subject of intensive study, even by researchers as famous as e.g. Gauss or Cayley; the decomposition of quadratic, cubic, quadric, ... forms in sums of simple terms, formed a major mathematical topic in the nineteenth century [67]. The aim was mainly to find decompositions that hold exactly; for tensors, least-squares approximations, and numerical techniques in general, seem to be less well-studied.

Thirdly, the development of tensor techniques is of course important for the processing and analysis of *multiway datasets*. At present, the reflex to evaluate whether a data acquisition campaign can best be organized to yield two-way or multi-way data records, seems to be most developed in Psychometrics and Chemometrics. In the past decades, important multiway data models as the Tucker model [259], Multiway Factor Analysis [163] and the Canonical Decomposition [49, 138] have been proposed; they will be further discussed in this thesis.

We also mention that tensors play a prominent role in physics [117, 145, 223, 280], like mechanics (e.g. fluid kinematics and dynamics), strength of materials,



theory of elasticity, theory of relativity, ... However, in most cases the stress is on the "tensor" aspect, and less on the "higher-order" specification; one is mainly concerned with the way in which physical quantities are influenced by coordinate transformations, and with the development of insights that hold in a basis-independent way.

It appears that there is little interaction between the different research areas to which the development of multilinear algebraic techniques can be relevant. We hope that this thesis can help building the bridge between those worlds.

## 1.1 General overview

The structure of this thesis is shown in Fig. 1.1. The left part deals with multilinear algebraic issues; the right part is more specifically related to HOS. The top-bottom subdivision reflects the structure of introduction, theoretical developments, applications, and conclusion.

It goes without saying that the *Singular Value Decomposition* (SVD) of matrices and its second-order generalizations play a central role for the analysis and the solution of problems in mathematical engineering; the literature on SVD-related problem concepts (linear/structured/total least squares, Principal Component Analysis (PCA), ... ), SVD-based signal processing methodologies (parameter estimation, identification, automatic control, ... ), numerical algorithms (fast algorithms, parallel computing, large-scale problems, ... ), etc., is abundant [106, 262, 195]. In view of this success, the main theme of the rather algebraic part of this thesis, is the generalization of the SVD and the close topic of best rank-$R$ approximation to higher-order tensors. However, where as the matrix decomposition concentrates a remarkable number of powerful properties (mutual orthonormality of the singular vectors, diagonality of the matrix of singular values, essential uniqueness, ... ), it turns out that this is no longer the case at the multilinear level. Loosely speaking, focusing on different matrix properties may lead to different tensor decompositions. This is much related to differences between matrices and tensors, as far as the concept of rank is concerned (Chapters 4, 5, 6).

The HOS-related part of this thesis is primarily dealing with the fundamental problem of *Independent Component Analysis* (ICA), earlier referred to as Blind Source Separation. Mathematically speaking, the goal is to split a linear mixture of statistically independent signals into its components, however without a priori knowledge of the mixing mechanism, neither of the values assumed by the independent components. ICA is the natural fine-tuning of PCA, in which it is only imposed that the components are uncorrelated (second-order), leaving a rotational degree of freedom in the set of possible solutions. In this thesis the ICA-problem is tackled with multilinear algebraic techniques; for a literature

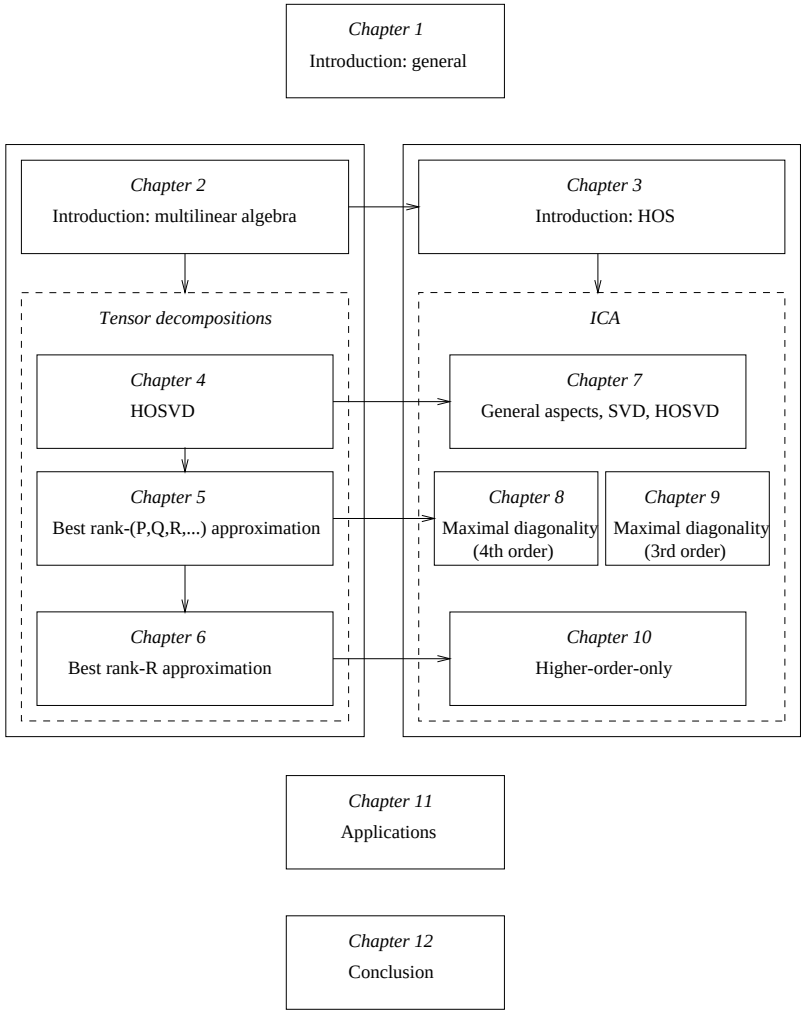



Figure 1.1: Organization of the thesis.

overview, we refer to Sect. 7.8. Like PCA and SVD are related, ICA makes use of tensorial SVD-extensions. As various generalizations are possible, this may lead to different solution strategies for ICA. The algebraic chapters on the left of Fig. 1.1, and the ICA-chapters at the same level on the right, are connected



(Chapters 7, 8, 9, 10).

## 1.2 Chapter by chapter overview

Chapters 2 and 3 contain the basic material with respect to multilinear algebra and HOS respectively.

The topics covered in Chapter 2 are: a formal definition of higher-order tensor, the introduction of basic tensor operations, tensor scalar product and Frobenius-norm, the meaning of rank in multilinear algebra, symmetry of higher-order tensors, and tensors as a matrix-like formalism to describe linear transformations between e.g. matrix spaces. The latter issue, discussed in Sect. 2.8, forms the starting point for the theoretical developments in Chapters 6, 9 and 10.

In Chapter 3 the basic quantities of HOS (moments, cumulants and spectra) are defined and their properties discussed. Sect. 3.5 treats the estimation of moments and cumulants from a given set of samples.

Chapters 4, 5, 6 deal with the generalization of the SVD of matrices, and the properties of this decomposition, in multilinear algebra. In Chapters 4 and 5 unitary (orthogonal) transformations are considered, where Chapter 6 is concerned with more general regular transformations.

In Chapter 4 the initial idea is to define, for a given $N$th order tensor, $N$ matrices of higher-order singular vectors that are related to the tensor's columns, rows, ... in exactly the same way as with matrices, i.e. the higher-order singular vectors should correspond to the directions of extremal (higher-order) oriented energy. By transformation to the bases of higher-order singular vectors, the property of all-orthogonality is induced, which generalizes the criterion of diagonality for a matrix of singular values. Diagonality is a special case of all-orthogonality, but on the other hand, it can be shown that it is impossible to diagonalize arbitrary higher-order tensors by coordinate transformations; this would e.g. conflict with a number of rank theorems. The resulting tensor decomposition, called Higher-Order Singular Value Decomposition (HOSVD) in this thesis, is highly analogous to the SVD of matrices. We analyze the properties of the decomposition and compare them to their matrix counterparts. Symmetry properties of the given tensor are reflected by the HOSVD components; this leads to a generalization of the symmetric matrix Eigenvalue Decomposition (EVD).

Chapter 5 deals with a multilinear generalization of the best rank-$R$ approximation problem, in the sense that the columns, rows, ... of the approximation should span a vector space of pre-specified dimension. For matrices, the solu-

tion is conceptually obtained by truncation of the SVD; for tensors, one can only state that truncation of the HOSVD may yield a good, but not necessarily optimal, approximation. The quality of this initial estimate can be enhanced by higher-order generalizations of the power method or orthogonal iteration method, or by gradient-based optimization over the Stiefel manifold; we present modifications, improvements and extensions of the current techniques.

A drawback of unitary (orthogonal) tensor decompositions is that they are not necessarily rank-revealing. In Chapter 6 we will discuss the approximation of a given higher-order tensor by a linear combination of $R$ possibly non-orthogonal rank-1 terms. The focus is on the development of Givens/Jacobi-type algorithms. If $R$ is small enough, the amount of work can be decreased by first reducing the dimensionality of the tensor to be decomposed, with the methods of Chapter 5.

Chapters 7, 8, 9, 10 deal with the generalization of PCA to ICA from a multilinear algebraic perspective. In Chapters 7, 8, 9 both second- and higher-order statistics are used to solve the problem; this involves unitary (orthogonal) tensor decompositions. Contrarily, higher-order-only algorithms lead to non-orthogonal multilinear decompositions (Chapter 10).

Chapter 7 treats the general aspects of ICA: problem definition, link with PCA, identifiability, the concept of contrast functions, literature overview. It is explained that ICA-algorithms can be devised as two-stage procedures, first consisting of a classical PCA-step, followed by a higher-order step in which the remaining rotational degree of freedom (an unknown unitary (orthogonal) matrix), is fixed. In Sect. 7.6 special attention is paid to performance issues, especially as far as the effect of the error, induced in the PCA-step, on the overall performance is concerned. Next, the chapter introduces a first new ICA-approach, by showing that the higher-order stage can take the form of a HOEVD (Chapter 4). Finally, we prove that the PCA-stage by itself can give a satisfactory quality of separation for uncorrelated sources with strongly different power.

Instead of imposing all-orthogonality, an other way to cope with the fact that arbitrary higher-order tensors cannot be diagonalized by unitary (orthogonal) transformations, consists of making the tensor as diagonal as possible (in Frobenius-sense). The equivalent approach to ICA is studied in Chapter 8, for the usual case in which the higher-order stage involves a fourth-order cumulant. It is shown that in a Jacobi-type algorithm, the computation of an elementary transformation can be formulated as a tensorial best rank-1 approximation problem, which establishes a link with Chapter 5; the relation even holds in a trivial manner for the computation by Jacobi-iteration of the symmetric matrix EVD.

Chapter 9 is the third-order counterpart of Chapter 8. We also discuss simultaneous optimal diagonalization of a set of third-order tensors as an approach to fourth-order-based ICA.



Chapter 10 is a short chapter, in which it is explained that ICA can also be carried out in a Gaussian-free way without resorting to second-order statistics, using the techniques of Chapter 6.

Chapter 11 contains a number of motivating application examples. First, the importance of the data model of Chapter 6 is illustrated by means of an application in Linguistics. A priori reduction of the dimensionality of the dataset involves the techniques of Chapters 7 and 8. Next, two applications of ICA are discussed. We cannot pass over the classical illustration, i.e. blind source separation in telecommunications; in Sect. 11.2, this issue is briefly touched on. Further, we propose ICA as a new approach to a biomedical problem, namely the separation of the fetal and the maternal electrocardiogram in antepartum abdominal potential recordings.

Chapter 12 is a summarizing conclusion, and contains directions for further research.

At the end of the text we have added a number of appendices. App. A contains some background material on matrix decompositions, used throughout the thesis. From the explanation above, it is clear that Gaussianity versus non-Gaussianity plays a crucial role in higher-order signal processing. For the ease of reference, it is explained in App. B how Gaussian distributions can be defined for multivariate stochastic processes. Finally, we have chosen to collect a number of rather technical derivations in App. C.

# Chapter 2

# Introduction to Multilinear Algebra

*This chapter is the first of two introductory chapters. It contains the basic material on multilinear algebra. Chapter 3 will consist of an introduction to Higher-Order Statistics.*

*We will start the exposition by giving a proper definition of higher-order tensors, which are the basic quantities in multilinear algebra. Depending on the scientific domain in which they are encountered, higher-order tensors may be approached in a different way. We will briefly situate and motivate our approach against the background of tensor literature (Sect. 2.1).*

*Our next concern is the development of tools to work with higher-order tensors. We will introduce:*

- *some basic tensor operations, like the inner and outer product (Sect. 2.2). The multiplication of a higher-order tensor by a matrix is a specific inner product, that will be used so frequently (e.g. to express basis transformations), that we will devote a separate section to it (Sect. 2.3).*
- *some geometry-related concepts, like the scalar product, mutual orthogonality and Frobenius-norm of higher-order tensors (Sect. 2.4).*
- *a generalization of row rank, column rank and rank to higher-order tensors (Sect. 2.6).*

*For our derivations we will need some additional material. In Sect. 2.5 we establish a convention to represent higher-order tensors in terms of matrices; such*

*a format is required to express tensor techniques in terms of matrix tools and software. Application of tensor techniques in e.g. the context of Higher-Order Statistics requires insight in symmetry properties. Compared to matrices, the issue of symmetry is much more involved in the higher-order case; we address the topic of super-symmetry in Sect. 2.7. Finally, in Sect. 2.8 we analyse how higher-order tensors can represent linear mappings between "tensor spaces". This section forms the starting-point of several further developments.*

## 2.1 Higher-order tensors

We will resort to the same definition of higher-order tensors as the one proposed in e.g. [180, 52, 233, 53, 177]:

**Definition 2.1.1 ($N$th order tensor)** *Let $\mathtt{V}_1, \mathtt{V}_2, \ldots, \mathtt{V}_N$ be $N$ Euclidean vector spaces with finite dimensions $I_1, I_2, \ldots, I_N$. Consider $N$ vectors $U_1 \in \mathtt{V}_1$, $U_2 \in \mathtt{V}_2, \ldots, U_N \in \mathtt{V}_N$. We denote by $U_1 \circ U_2 \circ \ldots \circ U_N$ the multilinear mapping on $\mathtt{V}_1 \times \mathtt{V}_2 \times \ldots \times \mathtt{V}_N$, defined by:*

$$(U_1 \circ U_2 \circ \ldots \circ U_N)(X_1, X_2, \ldots, X_N) = \langle U_1, X_1 \rangle_{\mathtt{V}_1} \langle U_2, X_2 \rangle_{\mathtt{V}_2} \ldots \langle U_N, X_N \rangle_{\mathtt{V}_N}, \tag{2.1}$$

*in which $\langle U_n, X_n \rangle_{\mathtt{V}_n}$ denotes the scalar product in $\mathtt{V}_n$, and $X_n$ is an arbitrary vector in $\mathtt{V}_n$ $(1 \leqslant n \leqslant N)$.*

*The space generated by all the elements $U_1 \circ U_2 \circ \ldots \circ U_N$ is called the tensor product space of $\mathtt{V}_1, \mathtt{V}_2, \ldots, \mathtt{V}_N$. An element of the tensor product space is called an $N$th order tensor (over $\mathtt{V}_1 \times \mathtt{V}_2 \times \ldots \times \mathtt{V}_N$).*

For $\mathtt{V}_n = \mathbb{R}^{I_n}$ $(1 \leqslant n \leqslant N)$, the tensor product space is called the space of real-valued $(I_1 \times I_2 \times \ldots \times I_N)$-tensors, denoted by $\mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$. Its complex equivalent is represented by $\mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$.

Depending on the scientific domain in which they are encountered, and the goal for which they are used, higher-order tensors may be approached in different ways. Important is the field of *Tensor Analysis*, where mathematical objects are studied that behave in a particular way under coordinate transformations. The transformation mechanism is called covariant or contravariant, and raises the following definition of tensors. Assume $P$ vector fields $\mathtt{V}_p$, of dimension $I_p$, in which coordinate systems $(x^{(p)}_{i_p})$ have been defined $(1 \leqslant p \leqslant P; 1 \leqslant i_p \leqslant I_p)$. Equivalently assume $Q$ vector fields $\mathtt{W}_q$, of dimension $J_q$, in which coordinate systems $(y^{(q)}_{j_q})$ have been defined $(1 \leqslant q \leqslant Q; 1 \leqslant j_q \leqslant J_q)$. Consider an ordered array $\mathcal{T}$ of $\prod_{pq} I_p J_q$ scalar functions, defined over $\mathtt{V}_1 \times \ldots \times \mathtt{V}_P \times \mathtt{W}_1 \times \ldots \times \mathtt{W}_Q$.



**Definition 2.1.2** *Consider coordinate transformations from $(x^{(p)}_{i_p})$ to $(\tilde{x}^{(p)}_{i_p})$ $(1 \leqslant p \leqslant P)$ and from $(y^{(q)}_{j_q})$ to $(\tilde{y}^{(q)}_{j_q})$ $(1 \leqslant q \leqslant Q)$. Denote the components of $\mathcal{T}$ by $t^{j_1 j_2 \ldots j_Q}_{i_1 i_2 \ldots i_P}$; in the new coordinate systems $\mathcal{T}$ is represented as $\tilde{\mathcal{T}}$, with components $\tilde{t}^{j_1 j_2 \ldots j_Q}_{i_1 i_2 \ldots i_P}$. $\mathcal{T}$ is called a mixed tensor of order $P+Q$, contravariant of order $P$ and covariant of order $Q$, if its components obey the following law of transformation:*

$$\tilde{t}^{j_1 j_2 \ldots j_Q}_{i_1 i_2 \ldots i_P} = \sum_{s_1 \ldots s_Q ; r_1 \ldots r_P} t^{s_1 s_2 \ldots s_Q}_{r_1 r_2 \ldots r_P} \frac{\partial y^{(1)}_{s_1}}{\partial \tilde{y}^{(1)}_{j_1}} \frac{\partial y^{(2)}_{s_2}}{\partial \tilde{y}^{(2)}_{j_2}} \cdots \frac{\partial y^{(Q)}_{s_Q}}{\partial \tilde{y}^{(Q)}_{j_Q}} \frac{\partial \tilde{x}^{(1)}_{i_1}}{\partial x^{(1)}_{r_1}} \frac{\partial \tilde{x}^{(2)}_{i_2}}{\partial x^{(2)}_{r_2}} \cdots \frac{\partial \tilde{x}^{(P)}_{i_P}}{\partial x^{(P)}_{r_P}}, \tag{2.2}$$

*for all $i_1, i_2, \ldots, i_P, j_1, j_2, \ldots, j_Q$.*

According to Def. 2.1.2, whether $\mathcal{T}$ is a tensor or not, depends on the class of coordinate transformations under consideration. In this thesis multilinear techniques are studied and consequently only linear coordinate transformations are considered. Moreover the entries of a tensor are assumed to be constants in $\mathbb{R}$ or $\mathbb{C}$.

**Example 2.1.3** Consider the vector space $\mathtt{V} = \mathbb{C}^I$ and the dual space $\mathtt{W}$ of linear forms over $\mathtt{V}$. Any vector $V$ is a contravariant tensor of order 1 and any linear form is a covariant tensor of order 1. A matrix $\mathbf{A}$, corresponding to a transformation of $\mathtt{V}$ is a tensor of order 2, once covariant (column index) and once contravariant (row index). The link between a vector space and its dual space of linear forms on one hand, and contravariance resp. covariance on the other hand, extends to orders higher than 2. □

The principle established in Ex. 2.1.3 makes it also possible to associate co- and contravariance properties to Def. 2.1.1. If $\mathtt{V}_n$ in Def. 2.1.1 is assumed to be the dual space of an other vector space, say $\mathtt{W}_n$, then the higher-order tensor is transformed in a covariant way with respect to basis transformations in $\mathtt{W}_n$. Otherwise, if $\mathtt{V}_n$ is considered as the "original" vector space, the higher-order tensor is transformed in a contravariant way with respect to basis transformations in $\mathtt{V}_n$.

The algebra of higher-order tensors requires the manipulation of quantities of which the components are sometimes characterized by a large number of indices. In literature often the so-called *index notation* is used, which was introduced by Einstein in the development of the general theory of relativity.

Briefly, the index notation consists of the following. Coordinates of a contravariant vector $V$ of an $I$-dimensional vector space $\mathtt{V}$, with respect to a given basis $\mathtt{B}_\mathtt{V}$, are represented by the symbol $v_i$, i.e. by means of subscripts. Superscripts are used for the coordinates of covariant vectors. Generally, if a tensor

$\mathcal{T}$ is $P$ times contravariant and $Q$ times covariant, it is symbolized by $t^{j_1 j_2 \ldots j_Q}_{i_1 i_2 \ldots i_P}$, as in Def. 2.1.2. The most important aspect of the index notation is the *Einstein summation convention* [155]. This holds an implicit summation over any index that appears twice, in a contravariant/covariant combination, related to a vector space and its dual space. It is not allowed to use an index more than twice in a given expression.

By Def. 2.1.2 (also by Def. 2.1.1) the tensor-concept was introduced in a basis-independent way. As a consequence a tensor equality does not depend on the specific basis chosen. In addition, an Einstein summation has a basis-independent meaning as well. This high level of abstraction is a major asset of working in a tensor framework, and gives it a touch of beauty.

**Example 2.1.4** Consider the linear mapping of a vector $V \in \mathbb{C}^I$ to the vector $W \in \mathbb{C}^I$, represented by multiplication with a matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$:

$$w_j = a^i_j v_i, \tag{2.3}$$

in which $j$ is a free index. This is the equivalent in index-notation of the common equation $W = \mathbf{A}V$. If after coordinate transformation $V$, $W$ and $\mathbf{A}$ are represented by $\tilde{v}_i$, $\tilde{w}_j$ and $\tilde{a}^i_j$, then holds:

$$\tilde{w}_j = \tilde{a}^i_j \tilde{v}_i, \tag{2.4}$$

which takes exactly the same form as Eq. (2.3). □

**Example 2.1.5** A nice illustration of the power of the Einstein summation convention is the invariance of the trace of a square matrix under coordinate transformations of the underlying vector space:

$$\texttt{trace}(\mathbf{A}) = a^i_i. \tag{2.5}$$

As this equation is an equality between two 0th-order tensors, it is basis-independent. E.g. in the basis of $\mathbf{A}$'s eigenvectors, the trace shows to be equal to the sum of eigenvalues. □

On the other hand, in signal processing and (numerical) linear algebra the co- and contravariant characteristics of vectors, linear forms and matrices are usually dealt with in an implicit way. To a large extent this is due to a different viewpoint: roughly, in mathematical engineering the emphasis is on the computation of a reliable solution, given a certain basis, rather than on the conceptual analysis of basis-independent properties.

In this thesis new tracks in signal processing and (numerical) linear algebra, involving higher-order tensors, are explored. It is our experience that, in our context, the explicit distinction between covariance and contravariance rather makes the exposition more severe, than that it induces additional insight. Therefore we will consider higher-order tensors intuitively as ordered higher-dimensional tables of numerical values, given with respect to a number of bases, where the covariant or contravariant character of a given index will be clear from the context. Indices will take the form of subscripts. To facilitate the access from conventional algebra and signal processing, and to highlight analogies between matrix and higher-order tensor concepts, we will introduce a matrix-like notation.



Tensors pop up in diverse scientific domains, inspired by scientific-geometrical considerations as well as inherently linked to physical laws. In physics numerous examples of tensor mechanisms can be found in mechanics, fluidum kinematics and dynamics, theory of elasticity, theory of strength of materials, theory of relativity, etc. . . . [223, 280]. Famous scientists like Gauss, Kronecker, Cayley, Weyl and Hilbert have indirectly contributed to the development of tensor algebra through research on homogeneous polynomials: their results can be casted in a tensor framework via the isomorphism between homogeneous polynomials and super-symmetric tensors (see Sect. 2.7), in the same way as quadratic forms and symmetric matrices are related [67]. Tensors are also widely used in differential geometry, a.o. by the work of Gauss and Riemann. Differential geometry is concerned with describing the local behaviour of curves and surfaces in space; the tensor concept leads e.g. to notions of curvature and torsion that are independent of the choice of coordinate system [248, 279]. For a detailed exposition on the use of tensor methods in statistics we refer to [188].

## 2.2 Basic tensor operations

The real- or complex-valued tensors of a given dimensionality form a linear vector space, which implies the definition of a sum and a multiplication with a scalar. Naturally, we will consider the component-wise sum and scalar multiplication. In addition, the outer and inner product of two tensors, and the contraction of a given tensor will be defined.

**Definition 2.2.1 (Outer product)** *The outer product $\mathcal{A} \circ \mathcal{B}$ of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_P}$ and a tensor $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_Q}$, is defined by*

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

*for all values of the indices.*

**Example 2.2.2** An $N$th-order tensor $\mathcal{A}$, equal to the outer product of $N$ vectors $U^{(1)}$, $U^{(2)}$, ..., $U^{(N)}$, is given by the element-wise equation

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

□

**Definition 2.2.3 (Inner product)** *The inner product $\langle \mathcal{A}, \mathcal{B} \rangle_{p,q}$ of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_P}$ and a tensor $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_Q}$, over a common index $i_p = j_q = u$, is defined by*

$$(\langle \mathcal{A}, \mathcal{B} \rangle_{p,q})_{i_1 i_2 \ldots i_{p-1} i_{p+1} \ldots i_P j_1 j_2 \ldots j_{q-1} j_{q+1} \ldots j_Q} \stackrel{\text{def}}{=} \sum_u a_{i_1 i_2 \ldots u \ldots i_P} b_{j_1 j_2 \ldots u \ldots j_Q}.$$

More generally, it is possible to define inner products over several indices.

**Definition 2.2.4 (Contraction)** *The contraction $\langle \mathcal{A} \rangle_{p,q}$ of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_P}$, over the indices $i_p$ and $i_q$, is defined by*

$$(\langle \mathcal{A} \rangle_{p,q})_{i_1 i_2 \ldots i_{p-1} i_{p+1} \ldots i_{q-1} i_{q+1} \ldots i_P} \stackrel{\text{def}}{=} \sum_u a_{i_1 i_2 \ldots i_{p-1}\, u\, i_{p+1} \ldots i_{q-1}\, u\, i_{q+1} \ldots i_P}.$$

More generally, it is possible to define contractions over several indices.

**Remark 2.2.5** An implicit condition on the sum, in order to be well-defined, is that the different tensor terms have the same covariant/contravariant structure, related to the same underlying vector spaces. As far as the inner product and the contraction are concerned, the summation should involve a covariant/contravariant index pair, related to the same vector space. A tensor obtained by basic tensor operations inherits its covariant/contravariant structure from its components. ◇

## 2.3 Multiplication of a higher-order tensor by a matrix

An inner product that will be frequently used throughout this thesis, is the multiplication of a higher-order tensor by a matrix. It will e.g. allow us to express the effect of basis transformations on a given tensor.



**Definition 2.3.1** *The $n$-mode product of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $\mathbf{U} \in \mathbb{C}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is an $(I_1 \times I_2 \times \ldots J_n \ldots \times I_N)$-tensor given by*

$$\mathcal{A} \times_n \mathbf{U} \stackrel{\text{def}}{=} \langle \mathcal{A}, \mathbf{U} \rangle_{n,2},$$

*i.e.*

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \ldots j_n \ldots i_N} = \sum_{i_n} a_{i_1 i_2 \ldots i_n \ldots i_N} u_{j_n i_n}$$

*for all index values.*

In the literature, the $n$-mode product of a tensor and a matrix, as an inner product, often takes the form of an Einstein summation convention, with a basis-independent interpretation. For our purpose however, the use of the $\times_n$-symbol [260] will more clearly demonstrate the analogy between matrix and tensor decompositions. It is our experience that a conceptual insight in tensorial decompositions is easier induced by using the $\times_n$-notation than by the use of element-wise summations.

**Corollary 2.3.2** *Given the tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and the matrices $\mathbf{F} \in \mathbb{C}^{J_n \times I_n}$, $\mathbf{G} \in \mathbb{C}^{J_m \times I_m}$, one has:*

$$(\mathcal{A} \times_n \mathbf{F}) \times_m \mathbf{G} = (\mathcal{A} \times_m \mathbf{G}) \times_n \mathbf{F} = \mathcal{A} \times_n \mathbf{F} \times_m \mathbf{G}.$$

**Corollary 2.3.3** *Given the tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and the matrices $\mathbf{F} \in \mathbb{C}^{J_n \times I_n}$, $\mathbf{G} \in \mathbb{C}^{K_n \times J_n}$, one has:*

$$(\mathcal{A} \times_n \mathbf{F}) \times_n \mathbf{G} = \mathcal{A} \times_n (\mathbf{G} \cdot \mathbf{F}).$$

**Example 2.3.4** For the matrices $\mathbf{F} \in \mathbb{R}^{I_1 \times I_2}$, $\mathbf{U} \in \mathbb{R}^{J_1 \times I_1}$, $\mathbf{V} \in \mathbb{R}^{J_2 \times I_2}$ the matrix product $\mathbf{U} \cdot \mathbf{F} \cdot \mathbf{V}^T$ can be written as $\mathbf{F} \times_1 \mathbf{U} \times_2 \mathbf{V}$. The $\times_n$-notation overcomes the problem of finding "generalized transposes" in the multilinear case and allows to formulate e.g. the SVD in a completely symmetric way. In the example the columns of $\mathbf{U}$ are associated to the "1-mode space" of $\mathbf{F}$; in exactly the same way the columns of $\mathbf{V}$ are associated to the "2-mode space". For complex matrices the product $\mathbf{U} \cdot \mathbf{F} \cdot \mathbf{V}^H$ is consequently represented by $\mathbf{F} \times_1 \mathbf{U} \times_2 \mathbf{V}^*$. □

Fig. 2.1 visualizes the equation $\mathcal{A} = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$ for third-order tensors $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times J_3}$ and $\mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. Unlike the conventional way to visualize second-order matrix products, $\mathbf{U}^{(2)}$ has not been transposed, for reasons of symmetry. Multiplication with $\mathbf{U}^{(1)}$ involves linear combinations of

the "horizontal matrices" (index $i_1$ fixed) in $\mathcal{B}$. Stated otherwise, multiplication of $\mathcal{B}$ with $\mathbf{U}^{(1)}$ means that every column of $\mathcal{B}$ (indices $i_2$ and $i_3$ fixed) has to be multiplied from the left with $\mathbf{U}^{(1)}$. Multiplication with $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ can be expressed in a similar way. Visualization schemes like Fig. 2.1 have proven to be very useful to gain insight in tensor techniques.

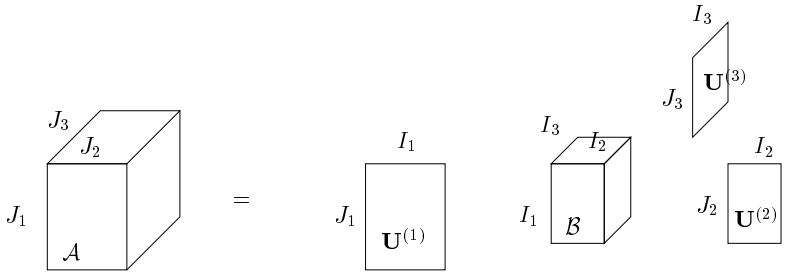



Figure 2.1: Visualization of the multiplication of a third-order tensor $\mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ with matrices $\mathbf{U}^{(1)} \in \mathbb{C}^{J_1 \times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{C}^{J_2 \times I_2}$ and $\mathbf{U}^{(3)} \in \mathbb{C}^{J_3 \times I_3}$.

## 2.4 Scalar product, orthogonality, norm

To be able to express geometrical properties in a vector space of higher-order tensors, it is necessary to generalize the well-known definitions of scalar product, orthogonality and Frobenius-norm to tensors of arbitrary order. We assume orthogonal bases in $\mathbb{C}^{I_1}$, $\mathbb{C}^{I_2}$, ..., $\mathbb{C}^{I_N}$.

**Definition 2.4.1 (Scalar product)** *The scalar product $\langle \mathcal{A}, \mathcal{B} \rangle$ of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ is defined as*

$$\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{\text{def}}{=} \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} b^*_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N}.$$

The tensor scalar product is an extension of the classical scalar product of vectors $X$ and $Y$:

$$\langle X, Y \rangle = Y^H X.$$

Another example is the scalar product of matrices. If we consider the matrices $\mathbf{F}, \mathbf{G} \in \mathbb{C}^{I_1 \times I_2}$, then their scalar product is given by:

$$\langle \mathbf{F}, \mathbf{G} \rangle = \sum_{i_1} \sum_{i_2} g^*_{i_1 i_2} f_{i_1 i_2} = \mathtt{trace}(\mathbf{G}^H \mathbf{F}).$$



**Remark 2.4.2** Without going into detail, we remark that the scalar product could as well have been defined as an inner product of two tensors, over all the indices. The two tensors are $\mathcal{A}$, and the so-called *dual tensor* of $\mathcal{B}$ (or vice-versa), where the dual tensor is a generalization of the complex conjugated transpose in the vector case [155]. ◇

Due to the generalization of the scalar product it becomes possible to treat tensors in a geometrical way. We have e.g.:

**Definition 2.4.3 (Orthogonality)** *Tensors of which the scalar product equals* 0*, are mutually orthogonal.*

**Definition 2.4.4 (Frobenius-norm)** *The Frobenius-norm of a tensor $\mathcal{A}$ is given by*

$$\|\mathcal{A}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}.$$

The Frobenius-norm can be interpreted as a measure for the "size" of the tensor. The square of this norm can be seen as the "energy" in the tensor.

## 2.5 Matrix representation of a higher-order tensor

Some results will be derived by formulating a tensorial problem in terms of matrix/vector-subproblems. Hence it will be useful to represent higher-order tensors in a matrix format. (An other reason why we have resorted to matrix notation, was the fact that standard software packages for numerical algebra and signal processing, like MATLAB [186], have not been supporting arrays with more than 2 indices for a long time (despite the fact that languages like C, Fortran, ... do not prohibit the use of multi-indexed arrays); only recently, MATLAB 5.0 allows multiway data structures.) Therefore we define "matrix unfoldings" of a given tensor $\mathcal{A}$, with respect to a certain permutation $\mathtt{P}(1, 2, \ldots, N)$ of $(1, 2, \ldots, N)$ (the permuted values are briefly indicated as $(\mathtt{P}(1), \mathtt{P}(2), \ldots, \mathtt{P}(N))$):

**Definition 2.5.1 (Matrix unfolding)**
*Assume an $N$th-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$. For a particular choice of $\overline{N} \in \{1, 2, \ldots, N\}$, the matrix unfolding $\mathbf{A}_{(I_{\mathtt{P}(1)} I_{\mathtt{P}(2)} \ldots I_{\mathtt{P}(\overline{N})} \times I_{\mathtt{P}(\overline{N}+1)} I_{\mathtt{P}(\overline{N}+2)} \ldots I_{\mathtt{P}(N)})} \in \mathbb{C}^{(I_{\mathtt{P}(1)} I_{\mathtt{P}(2)} \ldots I_{\mathtt{P}(\overline{N})} \times I_{\mathtt{P}(\overline{N}+1)} I_{\mathtt{P}(\overline{N}+2)} \ldots I_{\mathtt{P}(N)})}$, linked with a permutation $\mathtt{P}(1, 2, \ldots, N)$ of $(1, 2, \ldots, N)$, contains the element $a_{i_1 i_2 \ldots i_N}$ at the position with row index*

$(i_{\mathtt{P}(1)}-1)I_{\mathtt{P}(2)}I_{\mathtt{P}(3)}\ldots I_{\mathtt{P}(\overline{N})}+(i_{\mathtt{P}(2)}-1)I_{\mathtt{P}(3)}I_{\mathtt{P}(4)}\ldots I_{\mathtt{P}(\overline{N})}+\ldots+i_{\mathtt{P}(\overline{N})}$ *and column index* $(i_{\mathtt{P}(\overline{N}+1)}-1)I_{\mathtt{P}(\overline{N}+2)}I_{\mathtt{P}(\overline{N}+3)}\ldots I_{\mathtt{P}(N)}+(i_{\mathtt{P}(\overline{N}+2)}-1)I_{\mathtt{P}(\overline{N}+3)}I_{\mathtt{P}(\overline{N}+4)}\ldots I_{\mathtt{P}(N)}+\ldots+i_{\mathtt{P}(N)}$.

A standard matrix unfolding, corresponding to $\overline{N}=1$ and permuted indices $(n,1,2,\ldots,n-1,n+1,\ldots,N)$, will briefly be represented by $\mathbf{A}_{(n)}$.

The unfolding operation involved with the construction of $\mathbf{A}_{(1)}$, $\mathbf{A}_{(2)}$ and $\mathbf{A}_{(3)}$ is visualized in Fig. 2.2, for the third-order case.

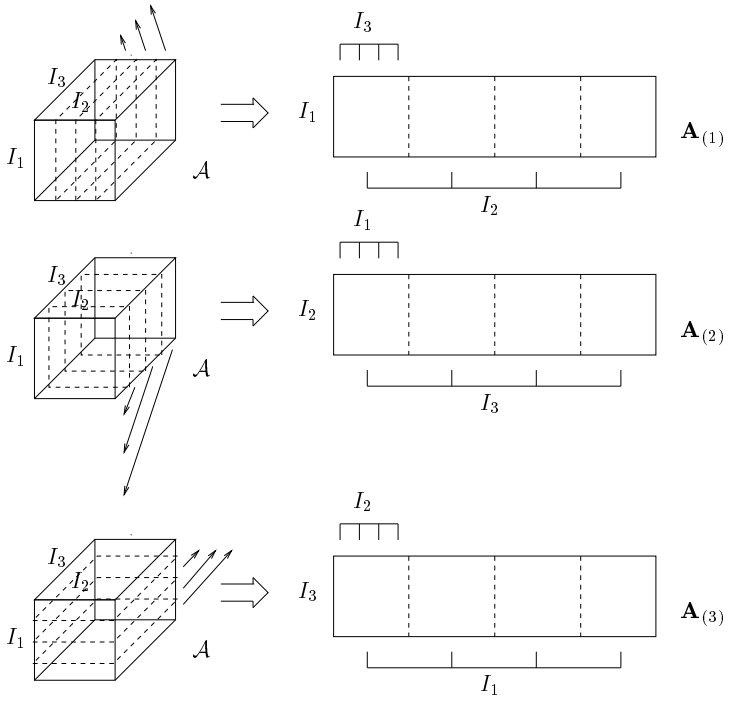



Figure 2.2: Unfolding of the $(I_1\times I_2\times I_3)$-tensor $\mathcal{A}$ to the $(I_1\times I_2I_3)$-matrix $\mathbf{A}_{(1)}$, the $(I_2\times I_3I_1)$-matrix $\mathbf{A}_{(2)}$ and the $(I_3\times I_1I_2)$-matrix $\mathbf{A}_{(3)}$ $(I_1=I_2=I_3=4)$.

**Example 2.5.2** Define a tensor $\mathcal{A}\in\mathbb{R}^{3\times2\times3}$ by $a_{111}=a_{112}=a_{113}=a_{211}=a_{123}=-a_{212}=-a_{313}=1$, $a_{311}=a_{122}=a_{321}=a_{322}=a_{323}=-a_{213}=-a_{222}=2$, $a_{312}=0$, $a_{121}=a_{221}=4$, $a_{223}=-5$. The matrix unfolding $\mathbf{A}_{(1)}$ is



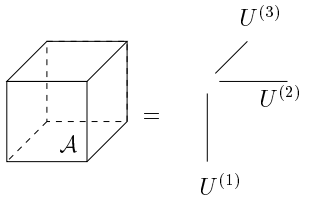



Figure 2.3: Visualization of a third-order rank-1 tensor $\mathcal{A}$, consisting of the outer product of three vectors $U^{(1)}$, $U^{(2)}$ and $U^{(3)}$.

given by

$$\mathbf{A}_{(1)}=\left(\begin{array}{ccc|ccc}1&1&1&4&2&1\\1&-1&-2&4&-2&-5\\2&0&-1&2&2&2\end{array}\right).$$

□

## 2.6 Rank properties of a higher-order tensor

There are major differences between matrices and higher-order tensors when rank properties are concerned. As will be explained later on, these differences directly affect the way in which tensorial decompositions differ from matrix decompositions. We will use the following definitions:

**Definition 2.6.1 (Rank-1 tensor)** *An $N$th-order tensor $\mathcal{A}$ has rank 1 when it equals the outer product of $N$ vectors $U^{(1)}$, $U^{(2)}$, ..., $U^{(N)}$:*

$$\mathcal{A}=U^{(1)}\circ U^{(2)}\circ\ldots\circ U^{(N)}.$$

A third-order rank-1 tensor $\mathcal{A}=U^{(1)}\circ U^{(2)}\circ U^{(3)}$ will be visualized as in Fig. 2.3. As shown in Ex. 2.2.2, the component of $\mathcal{A}$ at the position corresponding to the indices $i_1$, $i_2$ and $i_3$, equals the product of the corresponding entries in $U^{(1)}$, $U^{(2)}$ and $U^{(3)}$.

**Definition 2.6.2 (Rank)** *The rank of an arbitrary $N$th-order tensor $\mathcal{A}$, represented by $R=\mathtt{rank}(\mathcal{A})$, is the minimal number of rank-1 tensors that yield $\mathcal{A}$ in a linear combination.*

**Definition 2.6.3 ($n$-rank)** *The $n$-mode vectors of an $N$th-order tensor $\mathcal{A}\in\mathbb{C}^{I_1\times I_2\times\ldots\times I_N}$ are the $I_n$-dimensional vectors obtained from $\mathcal{A}$ by varying the*

*index $i_n$ and keeping the other indices fixed. The n-rank of $\mathcal{A}$, represented by $R_n = \mathtt{rank_n}(\mathcal{A})$, is the dimension of the vector space generated by the n-mode vectors.*

This definition generalizes the term "column (row) vector" and the concept "column (row) rank" of matrices. In contrast to the matrix case, the different $n$-ranks of a higher-order tensor are not necessarily the same; and when they are equal, they can still be different from the rank of the tensor. From the definitions it is clear that always $R_n \leqslant R$.

The $n$-ranks of a higher-order tensor can be determined by means of matrix techniques:

**Property 2.6.1** *The n-mode vectors of $\mathcal{A}$ are the column vectors of the matrix unfolding $\mathbf{A}_{(n)}$ and*

$$\mathtt{rank_n}(\mathcal{A}) = \mathtt{rank}(\mathbf{A}_{(n)}).$$

The determination of a tensor rank is a much harder problem since it treats the tensor as a global quantity, rather than as a collection of $n$-mode vectors. By way of illustration: the determination of the maximal rank value over the set of $(I_1 \times I_2 \times \ldots \times I_N)$-tensors is still an open problem in literature (it is not bounded by $\mathtt{min}\{I_1, I_2, \ldots, I_N\}$). [67] gives an overview of some partial results obtained for super-symmetric tensors in the context of invariant theory. The paper contains a.o. a tensor-independent rank upper-bound, an algorithm to compute maximal generic ranks and a complete discussion of the case of super-symmetric $(2 \times 2 \times \ldots \times 2)$-tensors. A remarkable difference with matrices, is that there is not an indisputable "generic rank" for higher-order tensors: e.g. for a uniform probability distribution over the unit sphere in $\mathbb{R}^{2\times 2\times 2}$ (normalization to unit norm does not affect the rank value), the probability that the rank of a random tensor equals 2, and the probability that it equals 3, are each different from zero. (Rank values with a non-vanishing probability w.r.t. a uniform distribution over the unit sphere in $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, will all be called "generic" in this thesis.) Another striking difference is that the rank value is more dependent on the algebraic field to which the entries belong; as such, the rank of a real-valued tensor in the field of complex numbers is not necessarily equal to the rank of the same tensor in the field of real numbers [167]. In [142] it is shown that the determination of the rank for third-order tensors with coefficients in a finite field is already an NP-complete problem; for third-order rational-valued tensors the problem is NP-hard. This contrasts with a fairly simple second-order equivalent: the rank of a general real- or complex-valued matrix can already be determined by e.g. QR-factorization (assuming infinite precision arithmetic).



Let us now analyze two examples in detail.

**Example 2.6.4** Consider the $(2 \times 2 \times 2)$-tensor $\mathcal{A}$ defined by

$$\begin{cases} a_{111} = a_{221} = a_{112} = a_{222} = 1 \\ a_{211} = a_{121} = a_{212} = a_{122} = 0. \end{cases}$$

It is easy to verify that $R_1 = R_2 = 2$ but $R_3 = 1$. The rank is not larger than 2, since

$$\mathcal{A} = E_1 \circ E_1 \circ (E_1 + E_2) + E_2 \circ E_2 \circ (E_1 + E_2),$$

in which

$$E_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This decomposition is minimal, since $r_1$ and $r_2$ give a rank lower-bound. □

**Example 2.6.5** Consider the $(2 \times 2 \times 2)$-tensor $\mathcal{A}$ defined by

$$\begin{cases} a_{211} = a_{121} = a_{112} = 1 \\ a_{111} = a_{222} = a_{122} = a_{212} = a_{221} = 0. \end{cases}$$

The 1-rank, 2-rank and 3-rank are equal to 2. On the other hand we will now show that the rank equals 3. First, the rank cannot be larger than 3, since

$$\mathcal{A} = E_2 \circ E_1 \circ E_1 + E_1 \circ E_2 \circ E_1 + E_1 \circ E_1 \circ E_2, \tag{2.6}$$

in which

$$E_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

It remains to be proved that Eq. (2.6) is a decomposition in a minimal linear combination of rank-1 tensors. Assume that the rank equals 2, and that a minimal decomposition of $\mathcal{A}$ is given by

$$\mathcal{A} = X_1 \circ Y_1 \circ Z_1 + X_2 \circ Y_2 \circ Z_2, \tag{2.7}$$

with $X_1, Y_1, Z_1, X_2, Y_2, Z_2 \in \mathbb{R}^2$. In terms of $\mathbf{X} = (X_1 X_2)$, $\mathbf{Y} = (Y_1 Y_2)$, $\mathbf{Z} = (Z_1 Z_2)$, and $\mathbf{D_1} = \mathtt{diag}(z_{11}, z_{12})$, $\mathbf{D_2} = \mathtt{diag}(z_{21}, z_{22})$, Eq. (2.7) can be re-expressed as the following set of matrix equations:

$$\mathbf{X} \cdot \mathbf{D_1} \cdot \mathbf{Y}^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.8}$$

$$\mathbf{X} \cdot \mathbf{D_2} \cdot \mathbf{Y}^T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.9}$$

Eqs. (2.8,2.9) imply that $\mathtt{rank}(\mathbf{D_1}) = 2$ and $\mathtt{rank}(\mathbf{D_2}) = 1$: $\mathbf{X}$ and $\mathbf{Y}$ have full rank, since $r_1 = r_2 = 2$. Without losing generality we assume that $z_{22} = 0$. Now consider the EVD

$$(\mathbf{X}\mathbf{D_2}\mathbf{Y}^T) \cdot (\mathbf{X}\mathbf{D_1}\mathbf{Y}^T)^{-1} = \mathbf{X} \cdot (\mathbf{D_2}\mathbf{D_1})^{-1} \cdot \mathbf{X}^{-1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This equation implies that $X_1$ is proportional to $(\ 1 \quad 0\ )^T$, while the first row of $\mathbf{X}^{-1}$ is proportional to $(\ 0 \quad 1\ )$. This is only possible when $\mathbf{X}$ is a rank-1 matrix, which conflicts with $R_1 = 2$. □

**Remark 2.6.6** For an $N$th order tensor $\mathcal{A}$, consider the $(N-1)$th order subtensors $\mathcal{A}_{i_n=\alpha}$, obtained by fixing the $n$th index to $\alpha$. Clearly, the quantity $\tilde{r}_n$, defined as the minimal cardinality of a single set of $(N-1)$th order rank-1 tensors that can be linearly combined to any of the tensors $\mathcal{A}_{i_n=\alpha}$, is a generalization of the concepts of row and column rank of matrices, alternative to the definition of $n$-rank. For this alternative definition, it is not so hard to prove that $\tilde{r}_1 = \tilde{r}_2 = \ldots = \tilde{r}_N = r$ [167]. However it is of less importance to this thesis; it will not be used in further developments. ◇

## 2.7 Super-symmetric tensors

Higher-order tensors that are invariant under symmetry operators, play an important role in this thesis, especially in the context of Higher-Order Statistics. In this section we introduce the concept of super-symmetry, which is a stringent tensorial generalization of real and Hermitean matrix symmetry. Other generalizations are possible, e.g. in Section 4.6 we will define the weaker concept of pair-wise symmetry.

Assume an $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I \times I \times \cdots \times I}$. Matrix symmetry is generalized in a natural way by claiming that any component $a_{i_1 i_2 \ldots i_N}$ remains unaffected by arbitrary index permutations. In the complex case the generalization of complex symmetry is somewhat less straightforward. In [37] the symmetry properties of a fourth-order cumulant tensor of a complex-valued stochastic vector are analyzed, and the label "super-symmetry" is put forward to indicate that the cumulant shows all the possible basis-independent symmetries. Basis-independent symmetries are related with permutations of covariant indices, or with permutations of contravariant indices, or with the mutual interchanging of the whole set of covariant indices with the whole set of contravariant indices [31].

To distinguish between real symmetry and complex Hermitean symmetry in the case of e.g. higher-order statistics, we can also relate the concept of super-symmetry to the tensorial representation of homogeneous polynomials. Assume



a vector of unknowns $X = (x_1 x_2 \ldots x_I)$ in $\mathbb{R}^I$ ($\mathbb{C}^I$). We associate a homogeneous polynomial to a given higher-order tensor in the following way:

**Definition 2.7.1 (Associated polynomial)** *The associated polynomial $\mathcal{A}(X)$ of an $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I \times I \times \ldots \times I}$ is obtained as*

$$\mathcal{A}(X) \stackrel{\text{def}}{=} \mathcal{A} \times_1 X \times_2 X \ldots \times_N X = \sum_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N} x_{i_1} x_{i_2} \ldots x_{i_N}.$$

*An associated polynomial $\mathcal{A}(X)$ of an $N$th-order tensor $\mathcal{A} \in \mathbb{C}^{I \times I \times \ldots \times I}$ is obtained as*

$$\mathcal{A}(X) \stackrel{\text{def}}{=} \mathcal{A} \times_1 \tilde{X}^{(1)} \times_2 \tilde{X}^{(2)} \ldots \times_N \tilde{X}^{(N)} = \sum_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N} \tilde{x}^{(1)}_{i_1} \tilde{x}^{(2)}_{i_2} \ldots \tilde{x}^{(N)}_{i_N},$$

*in which $\tilde{X}^{(n)} = (\tilde{x}^{(n)}_1 \tilde{x}^{(n)}_2 \ldots \tilde{x}^{(n)}_I)$ is either equal to $X$ or to $X^*$ $(1 \leqslant n \leqslant N)$.*

**Definition 2.7.2 (Super-symmetry)** *Consider an $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I \times I \times \ldots \times I}$ ($\mathbb{C}^{I \times I \times \ldots \times I}$) and an associated polynomial $\mathcal{A}(X)$, as in Def. 2.7.1. $\mathcal{A}$ is called a super-symmetric tensorial representation of $\mathcal{A}(X)$ if*

- *for any permutation $\mathtt{P}$ such that $\tilde{x}^{(1)}_{\mathtt{P}(i_1)} \tilde{x}^{(2)}_{\mathtt{P}(i_2)} \ldots \tilde{x}^{(N)}_{\mathtt{P}(i_N)} = \tilde{x}^{(1)}_{i_1} \tilde{x}^{(2)}_{i_2} \ldots \tilde{x}^{(N)}_{i_N}$, holds that $a_{\mathtt{P}(i_1 i_2 \ldots i_N)} = a_{i_1 i_2 \ldots i_N}$, and*
- *for any permutation $\mathtt{P}$ such that $\tilde{x}^{(1)}_{\mathtt{P}(i_1)} \tilde{x}^{(2)}_{\mathtt{P}(i_2)} \ldots \tilde{x}^{(N)}_{\mathtt{P}(i_N)} = (\tilde{x}^{(1)}_{i_1} \tilde{x}^{(2)}_{i_2} \ldots \tilde{x}^{(N)}_{i_N})^*$, holds that $a_{\mathtt{P}(i_1 i_2 \ldots i_N)} = a^*_{i_1 i_2 \ldots i_N}$.*

The link with the covariance/contravariance structure of $\mathcal{A}$ is that we only consider index permutations $\mathtt{P}$ with a basis-independent meaning in Def. 2.7.2. The definition of super-symmetry is independent of the particular basis in which the unknown $X$ is expressed. On the other hand, whether a complex-valued tensor is super-symmetric or not, depends on the way in which the associated polynomial has been defined - this definition depends on the context (see Sect. 3.2.1). Finally, we remark that a homogeneous polynomial, associated to a super-symmetric tensor, takes only real values.

## 2.8 Tensors as linear transformations

In this section we will explain how higher-order tensors can be used to represent "linear mappings between tensor spaces" (e.g. matrix/matrix space, vector/matrix space, matrix/higher-order tensor space, etc.). This generalizes the matrix formalism for linear mappings between vector spaces.

First we show how "higher-order linear mappings" can be formalized in tensor notation. We show that the composition of such transformations leads to the inner product of the corresponding tensors, which gives a wider view on the classical definition of the matrix product. Wondering which higher-order objects have an image in their own "direction", leads naturally to an EVD. Finally we will also describe a generalized SVD. For some background material on the corresponding matrix decompositions, we refer to App. A.

The direct tensor approach to higher-order linear mappings offers, besides the mathematical convenience, the advantage that symmetries (in e.g. higher-order cumulants) and position-related information (in e.g. images) remain present in the formalism. The results of this section might seem pretty simple (actually they are), but nevertheless they are conceptually very powerful (see e.g. [31], where it is shown that the tensor formalism, applied to a fourth-order cumulant, reveals a whole arsenal of new insights - invisible when resorting to standard matrix algebra - in the context of Blind Source Separation). The contents of this section form the starting-point of several techniques to be developed in the subsequent chapters.

### 2.8.1 Linear transformations

In the same way as there exists an isomorphic link between matrix algebra and the algebra of linear vector mappings, higher-order tensors can be regarded as formal representations of linear mappings between a matrix and a vector space, a matrix and a matrix space, a matrix and a higher-order tensor space, etc. E.g. assuming a basis in the space of $N_1$th-order and $N_2$th-order tensors, a linear transformation of $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_{N_1}}$ to $\mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_{N_2}}$ can be represented by $\mathcal{C} \in \mathbb{C}^{I_1 \times \ldots I_{N_2} \times J_1 \times \ldots \times J_{N_1}}$, according to the equation

$$\begin{aligned} \mathcal{B} &= \langle \mathcal{C}, \mathcal{A} \rangle_{N_2+1,\ldots,N_2+N_1;1,\ldots,N_1} \\ &\Updownarrow \\ b_{i_1 i_2 \ldots i_{N_2}} &= \sum_{j_1 j_2 \cdots j_{N_1}} c_{i_1 i_2 \ldots i_{N_2} j_1 j_2 \ldots j_{N_1}} a_{j_1 j_2 \ldots j_{N_1}}, \end{aligned} \tag{2.10}$$

in which the inner product subscript indicates a particular choice of $N_1$ summation indices. The same tensor can represent several linear mappings, depending on which indices have been selected for the summation. A mapping can be implemented as a matrix multiplication, with the appropriate matrix unfolding of $\mathcal{C}$.

**Remark 2.8.1** The transformation mechanism being an inner product, the summation is assumed to involve covariant/contravariant index pairs. ◇



The *range* $\mathtt{R}$ of the linear mapping can be defined as the linear vector space of all tensors in $\mathbb{C}^{I_1 \times I_2 \times \ldots \times I_{N_2}}$ that are the image, through Eq. (2.10), of a tensor $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_{N_1}}$. The *rank* of the linear mapping is defined as the dimension of the range. (This quantity is not necessarily equal to the $n$-rank or the rank of the tensor, as defined by Def. 2.6.3 and Def. 2.6.3 respectively; in the further developments we will clearly indicate when a given tensor is considered as the formalism for a linear mapping, if confusion is possible). The *kernel* $\mathtt{K}$ of the linear mapping is the linear vector space of all tensors $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_{N_1}}$ that are mapped to the 0-tensor. The dimensions of kernel and range are related by the following dimension theorem:

$$\mathtt{dim}(\mathtt{K}) + \mathtt{dim}(\mathtt{R}) = \mathtt{dim}(\mathbb{C}^{J_1 \times J_2 \times \ldots \times J_{N_1}}) = J_1 J_2 \ldots J_{N_1} \tag{2.11}$$

(in which the dimension of the source space $\mathbb{C}^{J_1 \times J_2 \times \ldots \times J_{N_1}}$ is computed over the field of complex numbers).

**Theorem 2.8.2** *The composition of linear mappings is represented by the inner product of the corresponding tensors over their common indices.*

**Proof:** Assume the tensors $\mathcal{X} \in \mathbb{C}^{I_1 \times \ldots I_{N_1}}$, $\mathcal{Y} \in \mathbb{C}^{J_1 \times \ldots J_{N_2}}$, $\mathcal{Z} \in \mathbb{C}^{K_1 \times \ldots K_{N_3}}$ and compose the mappings

$$\mathcal{Y} = \langle \mathcal{A}, \mathcal{X} \rangle_{N_2+1,\ldots,N_2+N_1;1,\ldots,N_1} \text{ and } \mathcal{Z} = \langle \mathcal{B}, \mathcal{Y} \rangle_{N_3+1,\ldots,N_3+N_2;1,\ldots,N_2},$$

with $\mathcal{A} \in \mathbb{C}^{J_1 \times \ldots J_{N_2} \times I_1 \times \ldots I_{N_1}}$ and $\mathcal{B} \in \mathbb{C}^{K_1 \times \ldots K_{N_3} \times J_1 \times \ldots J_{N_2}}$. Now consider the overall mapping from $\mathcal{X}$ to $\mathcal{Z}$, represented by $\mathcal{C} \in \mathbb{C}^{K_1 \times \ldots K_{N_3} \times I_1 \times \ldots I_{N_1}}$:

$$\mathcal{Z} = \langle \mathcal{C}, \mathcal{X} \rangle_{N_3+1,\ldots,N_3+N_1;1,\ldots,N_1}.$$

Since the summation of the inner product is associative, we have:

$$\mathcal{C} = \langle \mathcal{B}, \mathcal{A} \rangle_{N_3+1,\ldots,N_3+N_2;1,\ldots,N_2}.$$

□

This theorem generalizes the classical result for matrices, where a linear mapping composition $f_C = f_B \circ f_A$ is represented by the matrix product $\mathbf{C} = \mathbf{B} \cdot \mathbf{A}$.

### 2.8.2 Unit tensor

An important type of tensor is the unit tensor $\mathcal{I}$, used to represent the identity mapping. It can be constructed from the outer product of classical identity matrices $\mathbf{I}$. Since its components, given by $\delta_{i_1 j_1} \delta_{i_2 j_2} \ldots \delta_{i_N j_N}$, in which $\delta_{ij}$ is the Kronecker delta (i.e. $\delta_{ij} = 1$ iff $i = j$, and 0 otherwise), do not depend on

a specific choice of basis, the unit tensor is called *isotropic* [188] (we assume that the linear mapping is obtained by summation over e.g. all the $j$-indices).

We remark that the components of an $N$th order unit tensor are *not* given by $\delta_{i_1 i_2 \ldots i_N}$, in which we assume that the delta-symbol is 1 when all the indices are equal, and 0 otherwise. The latter tensor is not even isotropic, such that it is only well-defined if a particular basis has been specified as well. This simple fact is important for the discussion in Chapters 7 and 10, where it will allow to separate statistically independent sources with exactly the same probability distribution, in the context of Independent Component Analysis.

**Example 2.8.3** We illustrate the basis-dependence for a tensor with components $\delta_{i_1 i_2 j_1 j_2}$, called $\mathcal{D}$. Assume it is transformed according to $\tilde{\mathcal{D}} = \mathcal{D} \times_1 \mathbf{X^{-1}} \times_2 \mathbf{X^{-1}} \times_3 \mathbf{X} \times_4 \mathbf{X}$. Component-wise the transformation can be expressed as:

$$(\tilde{\mathcal{D}})_{p_1 p_2 q_1 q_2} = \sum_{i_1 i_2 j_1 j_2} \delta_{i_1 i_2 j_1 j_2} (\mathbf{X^{-1}})_{p_1 i_1} (\mathbf{X^{-1}})_{p_2 i_2} (\mathbf{X})_{q_1 j_1} (\mathbf{X})_{q_2 j_2}.$$

The components $\tilde{d}_{ppqq}$ e.g., with $p \neq q$, are given by

$$\tilde{d}_{ppqq} = \sum_i (\mathbf{X^{-1}})^2_{pi} (\mathbf{X})^2_{qi},$$

which is generally different from 0. □

### 2.8.3 Eigenvalue Decomposition

**Theorem 2.8.4** *Assume a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \ldots I_N \times I_1 \times \ldots \times I_N}$, associated with a linear mapping via the inner product over the last $N$ indices. If*

$$a_{i_1 \ldots i_N j_1 \ldots j_N} = a^*_{j_1 \ldots j_N i_1 \ldots i_N}$$

*for all index values, then $\mathcal{A}$ has $P = \prod_{n=1}^{N} I_n$ real eigenvalues $\lambda_p$ and $P$ mutually orthonormal eigentensors $\mathcal{E}_p \in \mathbb{C}^{I_1 \times \ldots \times I_N}$. The Eigenvalue Decomposition associated with the linear mapping is given by*

$$\mathcal{A} = \sum_p^P \lambda_p \, \mathcal{E}_p \circ \mathcal{E}_p^*. \tag{2.12}$$

**Proof:** The tensor decomposition can be found by unfolding $\mathcal{A}$ to the level of matrices, followed by computing the EVD of the matrix that is obtained, and finally re-introducing the tensor notation for the pseudo-matrices in the decomposition. □



The eigentensors $\mathcal{E}_p$ form an orthonormal basis in $\mathbb{C}^{I_1 \times \ldots \times I_N}$. They show the property that they are simply scaled by linear transformation under $\mathcal{A}$. They have an image in their own "direction", as it were. The amount of amplification is given by the corresponding eigenvalue $\lambda_p$.

### 2.8.4 Singular Value Decomposition

**Theorem 2.8.5** *Assume a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \ldots I_{N_2} \times J_1 \times \ldots \times J_{N_1}}$, associated with a linear mapping via the inner product over the last $N_1$ indices. If $P = \mathtt{min}(\prod_{n=1}^{N_2} I_n, \prod_{n=1}^{N_1} J_n)$, the Singular Value Decomposition associated with this linear mapping is given by*

$$\mathcal{A} = \sum_p^P \sigma_p \, \mathcal{U}_p \circ \mathcal{V}_p^*, \tag{2.13}$$

*in which $\sigma_p \in \mathbb{R}$, $\mathcal{U}_p \in \mathbb{C}^{I_1 \times \ldots \times I_{N_2}}$ and $\mathcal{V}_p \in \mathbb{C}^{J_1 \times \ldots \times J_{N_1}}$ satisfy $\sigma_p \geqslant 0$, $\langle \mathcal{U}_p, \mathcal{U}_q \rangle = \delta_{pq}$ and $\langle \mathcal{V}_p, \mathcal{V}_q \rangle = \delta_{pq}$.*

**Proof:** Cf. Theor. 2.8.4. □

The SVD gives important insight in the mechanism of higher-order linear mappings. The right singular tensors $\mathcal{V}_p$ form an orthonormal basis for the tensors in the source space $\mathbb{C}^{J_1 \times \ldots \times J_{N_1}}$. They are mapped to the corresponding left singular tensors $\mathcal{U}_p$, scaled by the singular values $\sigma_p$. The left singular tensors form an orthonormal basis for the image space $\mathbb{C}^{I_1 \times \ldots \times I_{N_2}}$.

## 2.9 Conclusion

This section contained the basic material of multilinear algebra that will be used in this thesis.

We paid special attention to the definition of a higher-order tensor, and we situated our approach among those described in literature. Tensors will be addressed from an engineering viewpoint, but from time to time the mathematical aesthetics will be indicated. Roughly speaking, higher-order tensors will be considered as multi-indexed arrays of numerical values, that change in a peculiar way under coordinate transformations. The transformation mechanism, labeled as covariant or contravariant, will be clear from the context.

We introduced some basic tensor operations (like the inner and outer product) and some geometric concepts (like the scalar product and the Frobenius-norm).

We stressed the fact that for higher-order tensors row rank, column rank, ... are not necessarily equal, which lead to the separate definition of $n$-rank and rank of a tensor. We defined the concept of super-symmetry; to distinguish between real symmetry and complex Hermitean symmetry, we linked symmetry properties to the tensorial representation of homogeneous polynomials, as an alternative to the approach described in literature. We explained how tensors can represent linear mappings, and consequently derived a.o. an EVD and SVD.

The most versatile and informative way to deal with higher-order tensors, and tensorial operations, is the index notation and the Einstein summation convention. This formalism however, has the disadvantage that it is only rarely used in the context of numerical algebra and signal processing. To make the thesis more accessible, we introduced a matrix-like notation for higher-order concepts. This idea, and our approach to inner and outer products, has been adopted in e.g. [134].

# Chapter 3

# Introduction to Higher-Order Statistics

*This chapter contains the basic material on Higher-Order Statistics (HOS).*

*In Sect. 3.1 we define the first and second characteristic function of a random variable. Sect. 3.2 shows how the characteristic functions "generate" higher-order moments and cumulants, which are the basic quantities of HOS. For stochastic processes, Fourier transformation of higher-order cumulants leads to the definition of higher-order spectra. Further in Sect. 3.2, the links between moments and cumulants are explicited and the properties of each are analyzed. Sect. 3.3 explains how the shape of a probability density function is influenced by the cumulants up to order 4. Sect. 3.4 contains the moments and cumulants, up to order 4, for some probability densities that will be used for illustrations in later chapters. Finally, Sect. 3.5 treats the estimation of moments and cumulants from a given set of samples and therefore contains the link to real-world applications.*

## 3.1 Characteristic functions

**Definition 3.1.1 (First characteristic function)** *The first characteristic (or generating) function $\Phi_X(\omega)$ of a real stochastic variable $X$ is defined as the expected value of $e^{j\omega X}$, with $\omega \in \mathbb{R}$:*

$$\Phi_X(\omega) \stackrel{\text{def}}{=} \mathtt{E}\{e^{j\omega X}\}. \tag{3.1}$$

This equation implies that the first characteristic function is the complex con-

jugate of the Fourier transform of the probability density function $p_X(x)$ of $X$:

$$\Phi_X(\omega) = \int_{-\infty}^{+\infty} p_X(x) e^{j\omega x} dx. \tag{3.2}$$

The generalization of the first characteristic function for a real stochastic vector X is given by

$$\Phi_X(\omega) = \mathtt{E}\{e^{j\omega^T X}\},$$

in which $\omega$ is a real vector.

**Definition 3.1.2 (Second characteristic function)** *The second characteristic (or generating) function $\Psi_X(\omega)$ of a real stochastic variable (vector) $X$ is defined as the Neperian logarithm of the first characteristic function:*

$$\Psi_X(\omega) \stackrel{\text{def}}{=} \ln \Phi_X(\omega), \tag{3.3}$$

Both characteristic functions contain the same information.

## 3.2 Moments, cumulants and spectra

### 3.2.1 Moments

**Definition 3.2.1 (Moment)** *The $N$th order moment $m_N^X$ of a real stochastic variable $X$ is defined as the expected value of the $N$th power of $X$:*

$$m_N^X \stackrel{\text{def}}{=} \mathtt{E}\{X^N\}. \tag{3.4}$$

The first-order moment is the *mean* $m_X$ of the stochastic variable. The second-order moment is known as *correlation*.

The $N$th order moment can be obtained from the coefficient of $\omega^N$ in the Taylor series expansion of the first generating function $\Phi_X(\omega)$:

$$m_N^X = (\frac{1}{j^N}) \frac{d^N \Phi_X(\omega)}{d\omega^N}\bigg|_{\omega=0}. \tag{3.5}$$

In the multivariable case the joint moment of a set of real random variables $X_1, X_2, \ldots, X_N$ is defined as the expected value of their product:

$$\mathtt{Mom}(X_1, X_2, \ldots, X_N) \stackrel{\text{def}}{=} \mathtt{E}\{X_1 X_2 \ldots X_N\}. \tag{3.6}$$



This moment can be derived from the Taylor series expansion of the corresponding first generating function. For different random variables the second-order moment is known as *cross-correlation*.

The $N$th order moment of a real stochastic vector $X$ is an $N$th order tensor, defined by

$$\mathcal{M}_N^X \stackrel{\text{def}}{=} \mathtt{E}\{X \circ X \circ \ldots \circ X\}, \tag{3.7}$$

in which the expectation over the $N$th order outer product is performed in a component-wise manner.

For a stochastic process $X(t)$, stationary up to $N$th order, $N$th order moments are defined according to:

$$m_N^X(\tau_1, \tau_2, \ldots, \tau_{N-1}) \stackrel{\text{def}}{=} \mathtt{Mom}(X(t), X(t+\tau_1), \ldots, X(t+\tau_{N-1})). \tag{3.8}$$

Moments of a complex random variable $Z$ can be defined resorting to Eq. (3.6), where $X_n$ either equals $Z$ or $Z^*$. Several definitions are possible, depending on the number of complex conjugations. Unless stated otherwise, we will assume the following standard definitions for orders 1 up to 4:

$$\begin{aligned} m_1^Z &= \mathtt{E}\{Z\}, \\ m_2^Z &= \mathtt{E}\{|Z|^2\}, \\ m_3^Z &= \mathtt{E}\{|Z|^2 Z\}, \\ m_4^Z &= \mathtt{E}\{|Z|^4\}. \end{aligned}$$

For complex stochastic vectors we will adopt the following consistent convention:

$$\begin{aligned} M_1^Z &= \mathtt{E}\{Z\}, \\ \mathbf{M}_2^Z &= \mathtt{E}\{Z \circ Z^*\}, \\ \mathcal{M}_3^Z &= \mathtt{E}\{Z \circ Z^* \circ Z\}, \\ \mathcal{M}_4^Z &= \mathtt{E}\{Z \circ Z^* \circ Z^* \circ Z\}. \end{aligned}$$

### 3.2.2 Cumulants

**Definition 3.2.2 (Cumulant)** *The $N$th order cumulant $c_N^X$ of a real stochastic variable $X$ is defined resorting to the Taylor series expansion of the second generating function $\Psi_X(\omega)$:*

$$c_N^X \stackrel{\text{def}}{=} (\frac{1}{j^N}) \frac{d^N \Psi_X(\omega)}{d\omega^N}\bigg|_{\omega=0}. \tag{3.9}$$

In the multivariable case the joint cumulant $\texttt{Cum}(X_1, X_2, \ldots, X_N)$ can be derived from the multivariate second generating function. Explicit expressions in terms of moments will be given in Section 3.2.4. Cumulants of complex random variables can be defined using similar conventions as for moments. Cumulants of random vectors and stochastic processes can also be defined in an analogous way.

The first-order cumulant of a random variable is again the *mean* $m_X$ of this variable. The second-order cumulant equals the *variance* $\texttt{Var}\{X\} \stackrel{\text{def}}{=} \texttt{E}\{|X - m_X|^2\}$ (the square root is the standard deviation $\sigma_X$). In the multivariate case the second-order cumulant corresponds to the *covariance* matrix $\texttt{E}\{(X - M_X)(X - M_X)^H\}$. In the case of different random variables the term *cross-covariance* is used (the cross-covariance of $X$ and $Y$ is defined by $\texttt{E}\{(X - m_X)(Y - m_Y)^*\}$ or $\texttt{E}\{(X - M_X)(Y - M_Y)^H\}$, for the univariate and multivariate case respectively.)

**Definition 3.2.3 (Skewness and kurtosis)** *The skewness $\gamma_X$ and kurtosis $\kappa_X$ of a stochastic variable $X$ are defined as the third- resp. fourth-order cumulant of $X$, normalized by its variance $\sigma_X^2$ in the following way:*

$$\gamma_X = c_3^X / \sigma_X^{3/2}, \tag{3.10}$$

$$\kappa_X = c_4^X / \sigma_X^4. \tag{3.11}$$

In some publications the normalization is omitted, and "skewness" and "kurtosis" just used as synonyms for third- and fourth-order cumulants.

Cumulants were first defined by the Danish scientist Thiele, around 1889 (for an English exposition, see [255]). In 1929 they were rediscovered by Fisher [13], who also noticed their superiority over moments (see Section 3.2.5) and investigated the computation by sample averaging.

### 3.2.3 Spectra

**Definition 3.2.4 (Spectrum)** *If the $N$th order cumulants of a discrete-time stochastic process $X(t)$, stationary up to $N$th order, satisfy the condition*

$$\sum_{\tau_1=-\infty}^{+\infty} \cdots \sum_{\tau_{N-1}=-\infty}^{+\infty} |c_N^X(\tau_1, \tau_2, \ldots, \tau_{N-1})| < \infty, \tag{3.12}$$

*then the $(N-1)$-dimensional Fourier transform of the cumulant sequence,*

$$s_N^X(\omega_1, \omega_2, \ldots, \omega_{N-1}) =$$



$$\sum_{\tau_1=-\infty}^{+\infty} \cdots \sum_{\tau_{N-1}=-\infty}^{+\infty} c_N^X(\tau_1, \tau_2, \ldots, \tau_{N-1}) \exp(-j(\omega_1\tau_1 + \ldots + \omega_{N-1}\tau_{N-1})), \tag{3.13}$$

*exists, is continuous, and defined as the $N$th order (cumulant) spectrum of $X(t)$.*

The second-, third- and fourth-order spectra are called *power spectrum*, resp. *bispectrum* and *trispectrum*.

**Remark 3.2.5** One can also define *moment spectra* by Fourier transformation [206]. They are often interesting in deterministic signal processing, where the expectation operator is replaced by time-averaging, but will not be used in this thesis. ◇

### 3.2.4 Relations between moments and cumulants

Since the first and the second characteristic function contain the same information, moments and cumulants give two different representations of the same reality, but mathematical derivations might be simpler or harder depending on the choice one makes.

Of course there are links between moments and cumulants. E.g., for zero-mean real random variables the cumulants up to order 4 are given by:

$$\texttt{Cum}(X_1) = \texttt{E}\{X_1\}, \tag{3.14}$$

$$\texttt{Cum}(X_1, X_2) = \texttt{E}\{X_1 X_2\}, \tag{3.15}$$

$$\texttt{Cum}(X_1, X_2, X_3) = \texttt{E}\{X_1 X_2 X_3\}, \tag{3.16}$$

$$\begin{aligned}\texttt{Cum}(X_1, X_2, X_3, X_4) = {} & \texttt{E}\{X_1 X_2 X_3 X_4\} - \texttt{E}\{X_1 X_2\}\texttt{E}\{X_3 X_4\} \\ & -\texttt{E}\{X_1 X_3\}\texttt{E}\{X_2 X_4\} \\ & -\texttt{E}\{X_1 X_4\}\texttt{E}\{X_2 X_3\}.\end{aligned} \tag{3.17}$$

In the case of real stochastic variables of which the mean is different from 0, $X_i$ has to be replaced in these formulas, except the first one, by $X_i - \texttt{E}\{X_i\}$. For complex random variables (see Sections 3.2.1 and 3.2.2) it is more convenient to consider $X_2^*$ in both sides of Eq. (3.16); in Eq. (3.17) $X_2$ and $X_3$ are complex conjugated. Note that in our notation "$\texttt{Cum}$" (and "$\texttt{Mom}$") are operators that act on real and complex random variables in the same way.

The general relation between joint moments and cumulants of order $N$, for a

set of random variables $X_1$, $X_2$, ... , $X_N$, is given by

$$\mathtt{Cum}(X_1,\dots,X_N) = \sum (-1)^N (N-1)!\, \mathtt{E}\{\prod_{i\in \mathtt{A}_1} X_i\}\, \mathtt{E}\{\prod_{i\in \mathtt{A}_2} X_i\} \dots \mathtt{E}\{\prod_{i\in \mathtt{A}_N} X_i\}, \tag{3.18}$$

where the summation involves all possible partitions $\{\mathtt{A}_1, \mathtt{A}_2, \dots, \mathtt{A}_N\}$ of the integers from 1 to $N$ [181].

### 3.2.5 Properties of moments and cumulants

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, that are not shared with higher-order moments, so that in practice cumulants are more frequently used. We enumerate some of the most interesting properties. Their proof is based on the definition of moments and cumulants.

Assume $I$ stochastic variables $X_1$, $X_2$, ... , $X_I$, then we have:

**Property 3.2.1 (Scaling)** *If $X_1$, $X_2$, ... , $X_I$ are multiplied with constants $a_1$, $a_2$, ... , $a_I$, then we have:*

$$\mathtt{Mom}(a_1X_1, a_2X_2, \dots, a_IX_I) = (\prod_{i=1}^{I} a_i)\, \mathtt{Mom}(X_1, X_2, \dots, X_I), \tag{3.19}$$

$$\mathtt{Cum}(a_1X_1, a_2X_2, \dots, a_IX_I) = (\prod_{i=1}^{I} a_i)\, \mathtt{Cum}(X_1, X_2, \dots, X_I). \tag{3.20}$$

**Property 3.2.2 (Sum)** *Moments or cumulants of a sum are the sum of the moments or cumulants:*

$$\mathtt{Mom}(X_1 + Y_1, X_2, \dots, X_N) =$$

$$\mathtt{Mom}(X_1, X_2, \dots, X_N) + \mathtt{Mom}(Y_1, X_2, \dots, X_N), \tag{3.21}$$

$$\mathtt{Cum}(X_1 + Y_1, X_2, \dots, X_N) =$$

$$\mathtt{Cum}(X_1, X_2, \dots, X_N) + \mathtt{Cum}(Y_1, X_2, \dots, X_N), \tag{3.22}$$

*in which $Y_1$ is a real stochastic variable.*



**Property 3.2.3 (Multilinearity)** *If a real stochastic vector $X$, with components $X_1$, $X_2$, ... , $X_I$, is transformed into a stochastic vector $\tilde{X}$ by a matrix multiplication $\tilde{X} = \mathbf{A} \cdot X$, with $\mathbf{A} \in \mathbb{R}^{J\times I}$, then we have:*

$$\mathcal{M}_N^{\tilde{X}} = \mathcal{M}_N^X \times_1 \mathbf{A} \times_2 \mathbf{A} \dots \times_N \mathbf{A}, \tag{3.23}$$

$$\mathcal{C}_N^{\tilde{X}} = \mathcal{C}_N^X \times_1 \mathbf{A} \times_2 \mathbf{A} \dots \times_N \mathbf{A}, \tag{3.24}$$

*for arbitrary order $N$. For complex stochastic vectors $\mathbf{A}$ should be complex conjugated, depending on the definition of the moment or cumulant.*

The multilinearity property follows from the scaling and sum property. Eqs. (3.23) and (3.24) take the form of the transformation law (2.2), which defined tensors. This transformation law still holds when $X$ is transformed in a non-linear way. Actually higher-order moments and cumulants are more than higher-order tensors, since Eqs. (3.23) and (3.24) are not only valid for coordinate transformations, but hold also when $\mathbf{A}$ is an arbitrary matrix.

**Property 3.2.4 (Symmetry)** *Moments and cumulants are symmetric in their arguments, i.e.*

$$\mathtt{Mom}(X_1, X_2, \dots, X_I) = \mathtt{Mom}(X_{\mathtt{P}(1)}, X_{\mathtt{P}(2)}, \dots, X_{\mathtt{P}(I)}), \tag{3.25}$$

$$\mathtt{Cum}(X_1, X_2, \dots, X_I) = \mathtt{Cum}(X_{\mathtt{P}(1)}, X_{\mathtt{P}(2)}, \dots, X_{\mathtt{P}(I)}), \tag{3.26}$$

*in which $\mathtt{P}$ is an arbitrary permutation of $(1, 2, \dots, I)$.*

As a consequence, moments and cumulants of a stochastic vector are super-symmetric higher-order tensors. For a complex stochastic vector the precise form of the super-symmetry depends on the exact definition of the moment or cumulant.

**Property 3.2.5 (Even distribution)** *If a real random variable $Y$ has an even probability density function $p_Y(y)$, i.e. $p_Y(y)$ is symmetric around the origin, then the odd moments and cumulants of $Y$ vanish. The same holds for a complex random variable $Z$, provided $p_Z(z)$ is point-symmetric around the origin.*

**Property 3.2.6 (Translation)** *For an arbitrary constant $a$ we have:*

$$\mathtt{Mom}(a + X_1, X_2, \dots, X_I) = a\,\mathtt{Mom}(X_2, \dots, X_I) + \mathtt{Mom}(X_1, \dots, X_I), \tag{3.27}$$

$$\mathtt{Cum}(a + X_1) = a + \mathtt{Cum}(X_1), \tag{3.28}$$

$$\mathtt{Cum}(a + X_1, X_2, \dots, X_I) = \mathtt{Cum}(X_1, X_2, \dots, X_I). \tag{3.29}$$

Cumulants of order higher than 1 are blind to the mean of a random variable, which does not hold for moments. As a consequence, the effect of a transformation $\tilde{X} = X_o + \mathbf{A}X$, in which $X_o$ is a vector of constants, still takes the simple form of Eq. (3.24) for second- and higher-order cumulants. The transformation of an $N$th order moment also involves the moments of orders lower than $N$.

**Property 3.2.7 (Partitioning of independent variables)** *If a proper subset of the $I$ stochastic variables $X_1$, $X_2$, ... , $X_I$ is independent of the other variables, then we have:*

$$\texttt{Cum}(X_1, X_2, \dots, X_I) = 0. \tag{3.30}$$

This property does not hold in general for moments. A consequence of the property is, that any higher-order cumulant of a stochastic vector with mutually independent components is a diagonal tensor. This very strong algebraic condition is in fact the basis of all the techniques for Independent Component Analysis that will be developed in this thesis. Another implication is that the cumulants of an i.i.d. process are delta-functions:

$$c_N^W(\tau_1, \tau_2, \dots, \tau_{N-1}) = \bar{c}_N^W \, \delta(\tau_1)\delta(\tau_2)\dots\delta(\tau_{N-1}),$$

for an i.i.d. process $W(t)$, with $\bar{c}_N^W$ the $N$th order cumulant of $W(t)$ at a single time instance $t_0$.

**Property 3.2.8 (Sum of independent variables)** *If the random variables $X_1$, $X_2$, ... , $X_I$ are mutually independent from the variables $Y_1$, $Y_2$, ... , $Y_I$, then we have:*

$$\begin{aligned}&\texttt{Cum}(X_1 + Y_1, X_2 + Y_2, \dots, X_k + Y_k) = \\ &\texttt{Cum}(X_1, X_2, \dots, X_k) + \texttt{Cum}(Y_1, Y_2, \dots, Y_k).\end{aligned} \tag{3.31}$$

This property is an immediate consequence of the previous and the sum property. It does not hold for moments; as a matter of fact, it explains the term "cumulant".

**Property 3.2.9 (Non-Gaussianity)** *If $Y$ is a Gaussian variable with the same mean and variance as a given stochastic variable $X$, then holds for $N \geqslant 3$:*

$$c_N^X = m_N^X - m_N^Y. \tag{3.32}$$

Higher-order cumulants of a Gaussian variable are 0. In combination with Prop. 3.2.2, we observe that higher-order cumulants have the interesting property to be blind for additive Gaussian noise. This is the basis of several HOS-based techniques for noise reduction, and will be the key idea behind the development in Chapter 10. In fact, Eq. (3.32) also implies that higher-order



cumulants give an indication to which extent a random variable can be considered as Gaussian.

**Remark 3.2.6** For a formal definition of multivariate Gaussian distributions we refer to Appendix B. ◇

**Remark 3.2.7** The relationship between cumulants and Gaussianity is further detailed by the Gram-Charlier series expansion and the Edgeworth series expansion of a probability density function $p_X(x)$, around its Gaussian approximation $\hat{p}_X(x)$ with the same mean and variance [188].

For a standardized distribution $p_X(x)$ of a real random variable $X$, with mean $m_X = 0$ and variance $\sigma_X^2 = 1$, the *Gram-Charlier series expansion* is given by

$$\begin{aligned}p_X(x) &= \hat{p}_X(x)\{1 + \frac{1}{3!}c_3^X h_3(x) + \frac{1}{4!}c_4^X h_4(x) \\ &\quad + \frac{1}{5!}c_5^X h_5(x) + \frac{1}{6!}[c_6^X + 10(c_3^X)^2]h_6(x) + \dots\},\end{aligned} \tag{3.33}$$

in which $h_i(x)$ represents the $i$th Hermite polynomial. The Hermite polynomials form an orthogonal basis w.r.t. the zero-mean unit-variance Gaussian distribution as weight function, in the space of continuous real functions. They are given by [188]:

$$\begin{aligned}h_1(x) &= x, \\ h_2(x) &= x^2 - 1, \\ h_3(x) &= x^3 - 3x, \\ h_4(x) &= x^4 - 6x^2 + 3, \\ h_5(x) &= x^5 - 10x^3 + 15x, \\ h_6(x) &= x^6 - 15x^4 + 45x^2 - 15, \\ &\vdots \\ h_{k+1}(x) &= xh_k(x) - \frac{dh_k(x)}{dx}.\end{aligned}$$

Now assume that $X$ is obtained as the sum of $P$ independent random variables with finite cumulants. The *central limit theorem* of Cramér [277] states that the $N$th order cumulant of $X$ decays as $P^{\frac{2-N}{2}}$, and hence - as is well-known - an infinite sum of mutually independent variables has a Gaussian distribution. Re-ordering the terms of the Gram-Charlier series expansion, in order of importance, yields the *Edgeworth series expansion* of $p_X(x)$ around $\hat{p}_X(x)$. For a standardized distribution we have:

$$\begin{aligned} p_X(x) &= \hat{p}_X(x)[1 \\ &\quad + \frac{1}{3!}c_3^X h_3(x) \\ &\quad + \frac{1}{4!}c_4^X h_4(x) + \frac{10}{6!}(c_3^X)^2 h_6(x) \\ &\quad + \frac{1}{5!}c_5^X h_5(x) + \frac{35}{7!}c_3^X c_4^X h_7(x) + \frac{280}{9!}(c_3^X)^3 h_9(x) \\ &\quad + o(P^{-3/2})]. \end{aligned} \tag{3.34}$$

◇

As a conclusion, it is usually preferable to work with cumulants instead of moments, because:

- Second- and higher-order cumulants are insensitive to the mean of a random variable. Cumulant tensors change in a multilinear way under arbitrary affine transformations that can involve a translation.
- Cross-cumulants of independent variables are zero. Cumulant tensors of a random vector with statistically independent components are diagonal.
- For independent random variables, the cumulants of a sum are the sum of the cumulants.
- Higher-order cumulants of a Gaussian random variable are zero. Higher-order-only techniques, based on cumulants, are conceptually blind for Gaussian noise. Higher-order cumulants indicate to which extent a random variable can be considered as Gaussian.

## 3.3 Shape of the probability density function

The cumulants of a real random variable $X$ can be obtained from the Taylor series expansion of its second characteristic function . Since $\Psi_X(\omega)$ is an alternative representation of the probability density function $p_X(x)$, the cumulant values affect the shape of $p_X(x)$:

- The mean of $X$, $m_X$, is the $x$-coordinate of the centre of gravity of a uniform laminar body of unit mass, bounded by the $x$-axis and $p_X(x)$.



- The variance of $X$, $\sigma_X^2$, is the moment of inertia of this laminar body around the axis $x = m_X$. It gives a quadratic measure of the "distance" of $p_X(x)$ around $x = m_X$.
- The skewness of $X$, $\gamma_X$, indicates asymmetry of $p_X(x)$ around $x = m_X$: Prop. 3.2.5 implies that for a symmetric distribution all the cumulants of odd order vanish. The reverse is also true, under the condition that $\Phi_X(\omega)$ is analytic at the origin, since Eq. (3.5) implies that $\Phi_X(\omega)$, and hence $p_X(x)$, is completely determined by its moments (and hence by its cumulants) if $\Phi_X(\omega)$ is analytic at the origin. If the condition of analyticity is not fulfilled, different probability density functions might have the same moments for all orders (w.r.t. Eq. (3.5)) (e.g. [119], Section VII.3), unsymmetric distributions might have vanishing cumulants for all orders (e.g. [157], Exercise 3.26), and higher-order statistics might not even exist (e.g. Student's distribution [188]).
- The kurtosis of $X$, $\kappa_X$, is often interpreted in terms of *super-Gaussianity* and *sub-Gaussianity* of the probability density function $p_X(x)$, depending on the relative position of its tails and the tails of its best Gaussian approximation $\hat{p}_X(x)$ (other definitions can be found in literature; for a detailed discussion we refer to [66]). More specifically, if the even part of $p_X(x)$ cuts $\hat{p}_X(x)$ two times in the interval $(-\infty, +\infty)$, then a positive resp. negative kurtosis implies super- resp. sub-Gaussianity.

## 3.4 Examples

In this section we list some precomputed moment and cumulant values, for a number of statistical distributions that will be used in simulations troughout this thesis.

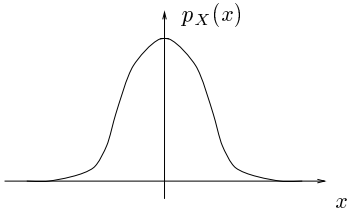
Table 3.1 contains the moments and cumulants, up to order 4, of the real Gaussian, uniform, and exponential distributions.

Table 3.2 addresses some typical types of modulation, in the context of telecommunications: it contains the moments and cumulants of QAM4 (Quadrature Amplitude Modulation), QAM16, and continuous PSK (Phase Shift Keying) signals. For a complex number $a_R + ja_I$ in the signal space diagrams of Table 3.2, the corresponding modulated signal waveform is given by

$$s(t) = \mathtt{Re}[(a_R + ja_I)\, u(t)\, e^{j2\pi f_c t}],$$

in which $u(t)$ is the signal pulse, designed on the basis of the transmission channel characteristics, and $f_c$ the carrier frequency. For more details we refer to e.g. [217].

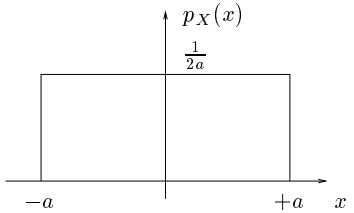
**Gaussian distribution**




$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})$

| $n$ | $m_n^X$ | $c_n^X$ |
|---|---|---|
| 1 | 0 | 0 |
| 2 | $\sigma^2$ | $\sigma^2$ |
| 3 | 0 | 0 |
| 4 | $3\ \sigma^4$ | 0 |

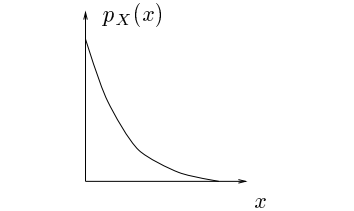
**Uniform distribution**




$p_X(x) = \frac{1}{2a} \qquad (x \in [-a, +a])$

| $n$ | $m_n^X$ | $c_n^X$ |
|---|---|---|
| 1 | 0 | 0 |
| 2 | $a^2/3$ | $a^2/3$ |
| 3 | 0 | 0 |
| 4 | $3a^4/5$ | $-2a^4/15$ |

**Exponential distribution**




$p_X(x) = a \exp(-ax) \qquad (x \geqslant 0)$

| $n$ | $m_n^X$ | $c_n^X$ |
|---|---|---|
| 1 | $1/a$ | $1/a$ |
| 2 | $2/a^2$ | $1/a^2$ |
| 3 | $6/a^3$ | $2/a^3$ |
| 4 | $24/a^4$ | $6/a^4$ |

Table 3.1: The moments and cumulants, up to order 4, of a Gaussian, a uniform, and an exponential distribution.

## 3.5 Estimation of HOS

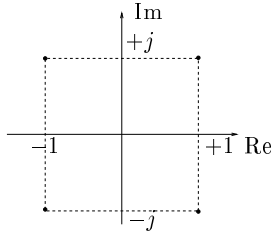
### 3.5.1 Estimation of moments

Given $T$ independent samples $x_t$ $(1 \leqslant t \leqslant T)$ of a real stochastic variable $X$, it is natural to estimate the moment $m_N^X$ by the average $\hat{m}_N^X$:

$$\hat{m}_N^X \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=1}^{T} x_t^N. \tag{3.35}$$



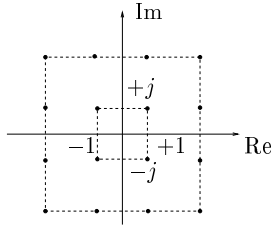
**QAM4**




$\texttt{Prob}(x + jy) = \frac{1}{4}$
$x \in \{\pm 1\}$ and $y \in \{\pm 1\}$

| $n$ | $m_n^X$ | $c_n^X$ |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 2 | 2 |
| 3 | 0 | 0 |
| 4 | 4 | -8 |

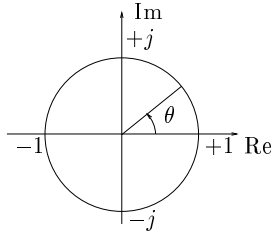
**QAM16**




$\texttt{Prob}(x + jy) = \frac{1}{16}$
$x \in \{\pm 1, \pm 3\}$ and $y \in \{\pm 1, \pm 3\}$

| $n$ | $m_n^X$ | $c_n^X$ |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 10 | 10 |
| 3 | 0 | 0 |
| 4 | 132 | $-168$ |

**Continuous PSK**




$p_\Theta(\theta) = \frac{1}{2\pi} \qquad (\theta \in (-\pi, +\pi])$

| $n$ | $m_n^X$ | $c_n^X$ |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 1 | 1 |
| 3 | 0 | 0 |
| 4 | 1 | $-2$ |

Table 3.2: The signal space diagrams of QAM4, QAM16 and continuous PSK signals, and the corresponding moments and cumulants up to order 4.

**Theorem 3.5.1** *The sample estimate $\hat{m}_N^X$ has the following properties:*

- *As $T \to \infty$, $\hat{m}_N^X$ converges to $m_N^X$ with probability 1.*

- *$\hat{m}_N^X$ is unbiased:* $\mathtt{E}\{\hat{m}_N^X\} = m_N^X$.
- *The variance of $\hat{m}_N^X$ is given by:*

$$\mathtt{Var}\{\hat{m}_N^X\} = \frac{1}{T}\,[m_{2N}^X - (m_N^X)^2]. \tag{3.36}$$

- *$\hat{m}_N^X$ is asymptotically normal.*

**Proof:** See [235], Chapter 2. □

The inverse of the variance can be considered as a quadratic measure of the absolute precision of the estimate; it is proportional to the number of samples. The relative precision can be characterized in terms of the *efficiency* $w_k$ [147], defined as (note that a *small* "efficiency", corresponds to an *accurate* estimation):

$$w_k \stackrel{\text{def}}{=} \frac{\mathtt{Var}\{\hat{m}_N^X\}}{(m_N^X)^2}. \tag{3.37}$$

The way in which the efficiency depends on the order $N$, for a given number of samples $T$, is revealed by the following theorem:

**Theorem 3.5.2** *For all orders $N \in \mathbb{N}$ we have:*

- $w_{2N+1} \geqslant w_{2N}$.
- $w_{2N+2} \geqslant w_{2N}$.
- $w_{2N+2} \geqslant w_{2N+1}$ *if* $X \geqslant 0$.

**Proof:** See [15]. □

This theorem indicates that it becomes generally harder to estimate higher-order moments as the order increases. However the fourth-order case is not necessarily harder than the third-order case, the sixth-order case is not necessarily harder than the fifth-order case, and so on.

### 3.5.2 Estimation of cumulants

Sample cumulant values can be obtained from the estimates of moments through the relations (3.14-3.17). However, these cumulant estimates are generally biased. E.g. the expected value of the variance estimator

$$\tilde{m}_2^X \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{m}_X)^2$$



is given by $\mathtt{E}\{\tilde{m}_X^2\} = \sigma_X^2 \cdot (T-1)/T$. Compensation leads to the definition of *$\kappa$-statistics* [66]. Up to order 4 they are given by

$$\hat{c}_2^X \stackrel{\text{def}}{=} \frac{T}{T-1}\,\tilde{m}_2^X, \tag{3.38}$$

$$\hat{c}_3^X \stackrel{\text{def}}{=} \frac{T^2}{(T-1)(T-2)}\,\tilde{m}_3^X, \tag{3.39}$$

$$\hat{c}_4^X \stackrel{\text{def}}{=} \frac{T^2}{(T-1)(T-2)(T-3)}\left[(T+1)\tilde{m}_4^X - 3(T-1)(\tilde{m}_2^X)^2\right], \tag{3.40}$$

in which $\tilde{m}_N^X$ is given by

$$\tilde{m}_N^X \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{m}_X)^N.$$

**Theorem 3.5.3** *The $\kappa$-statistics $\hat{c}_N^X$ have the following properties:*

- *As $T \to \infty$, $\hat{c}_N^X$ converges to $c_N^X$ with probability 1.*
- *$\hat{c}_N^X$ is unbiased:* $\mathtt{E}\{\hat{c}_N^X\} = c_N^X$.
- *The variances of $\hat{c}_N^X$, up to order 4, are given by*

$$\mathtt{Var}\{\hat{c}_2^X\} = \frac{c_4^X}{T} + \frac{2\,(c_2^X)^2}{T-1}, \tag{3.41}$$

$$\mathtt{Var}\{\hat{c}_3^X\} = \frac{c_6^X}{T} + \frac{9\,[c_4^X c_2^X + (c_3^X)^2]}{T-1} + \frac{6\,T\,(c_2^X)^3}{(T-1)(T-2)}, \tag{3.42}$$

$$\mathtt{Var}\{\hat{c}_4^X\} = \frac{c_8^X}{T} + \frac{16\,c_6^X c_2^X + 48\,c_5^X c_3^X + 34\,(c_4^X)^2}{T-1} + \frac{8\,T\,[9c_4^X(c_2^X)^2 + 18(c_3^X)^2 c_2^X]}{(T-1)(T-2)} + \frac{24T(T+1)(c_2^X)^4}{(T-1)(T-2)(T-3)}. \tag{3.43}$$

- *$\hat{c}_N^X$ is asymptotically normal.*

**Proof:** See [66], and the references therein. □

The expression for the variance of a higher-order sample cumulant allows to estimate the number of independent samples, required to obtain a cumulant estimate with a given absolute precision. In case the samples are perturbed by an additive Gaussian noise term, the variance of its sample cumulant should be taken into account as well; it takes the form of the last term in Eqs. (3.42-3.43).

**Remark 3.5.4** The sample skewness and kurtosis are obtained in analogy with Eqs. (3.10) and (3.11). For Gaussian random variables these estimators are unbiased. There exists no unbiased skewness or kurtosis estimator for arbitrary probability distributions. ◇

**Remark 3.5.5** The concepts of this section can be generalized for complex random variables, random vectors, stochastic processes, etc. For more details we refer to [66, 121, 206]. ◇

## 3.6 Conclusion

This section contained the basic material of HOS that will be used in this thesis.

We defined higher-order moments, cumulants and spectra and analyzed their links. We enumerated the most important properties of moments and cumulants, and concluded that cumulants are more interesting than moments. Higher-order cumulants have e.g. the advantage to be blind for perturbations of the random variable by additive Gaussian noise. This allows to devise higher-order-only cumulant-based techniques that enhance signal-to-noise ratios, like the methods derived in Chapter 10. In practice algorithms are based on cumulant estimates, obtained by sample averaging. Therefore we also briefly addressed the issue of estimating HOS from a set of samples.

Higher-order moments and cumulants of a multivariate random variable are super-symmetric higher-order tensors. They change in a multilinear way, when the random vector is multiplied with a matrix. Higher-order cumulants of a random vector, of which the components are mutually statistically independent, are diagonal. This suggests that EVD-like tensor techniques are useful tools for the analysis of higher-order cumulants. The development of such techniques is one of the main objectives of this thesis.

# Chapter 4

# The Higher-Order Singular Value Decomposition

*In this chapter we present a multilinear generalization of the Singular Value Decomposition (SVD), called the Higher-Order Singular Value Decomposition (HOSVD).*

*The text is organized as follows. Sect. 4.1 contains an introduction. The HOSVD-model is first presented in Sect. 4.2; a formal derivation is given in Sect. 4.3. In Sect. 4.4 we discuss some well-known SVD-properties and demonstrate that they have striking higher-order counterparts. Sect. 4.5 explains how the decomposition can be computed. In Sect. 4.6 we investigate how the HOSVD is influenced by symmetry properties of the higher-order tensor; in analogy with the Eigenvalue Decomposition (EVD) of symmetric (Hermitean) matrices we define a Higher-Order Eigenvalue Decomposition (HOEVD) for "pair-wise symmetric" higher-order tensors. Sect. 4.7 contains a first-order perturbation analysis of the HOSVD.*

## 4.1 Introduction

The SVD of matrices is one of the most fruitful developments in the world of linear algebra and signal processing and has a long history of a steadily growing success (for a detailed historical survey we recommend [250] as well as [146], Chapter 3)[1]. By now it has become well-known that the SVD concentrates a remarkable number of powerful matrix properties. We list a number of basic

[1] With thanks to Ning Hu, Washington University, St. Louis, U.S.A., for providing us these references.

properties, that will be generalized in Sect. 4.4, in App. A. It comes as no surprise that, since the development of reliable numerical algorithms for its calculation, SVD has evolved from a powerful concept to an indispensable tool in the field of mathematical engineering [268, 269, 106, 262, 195].

The evolution towards the integration of multivariate statistical methods and advanced numerical algorithms in modern signal processing is supported by major technological evolutions. Recent years showed a popularization of multichannel data acquisition systems and a spectacular increase in computational power (the fact that computation speed is doubled every $1\frac{1}{2}$ year is known as "Moore's Law"). In this context researchers started to develop interesting SVD-generalizations, e.g. the (tree of) Product and Quotient SVD's [101] (consisting of a set of decompositions that can be combined to obtain the SVD of a product or a quotient of matrices), the Riemannian SVD [104] (appearing in the Weighted and Structured Total Least Squares problem), the max-algebraic SVD [108] (related to the analysis of Discrete Event Systems), etc.

The tensorial generalization presented in this chapter first appeared in psychometrics as the Tucker model, originally developed to obtain a method for searching for relations in three-way data tables [259, 260]. In this chapter we will present a square-root version of the algorithm, valid for real- and complex-valued tensors of arbitrary order. We will show that the result is a convincing multilinear generalization of the SVD-concept itself - in this thesis it will be called the *Higher-Order Singular Value Decomposition* (HOSVD). The properties of the decomposition will thoroughly be analyzed and interpreted in terms of their matrix counterparts.

The HOSVD will play an important role in further chapters. In Chapter 7 we show that the decomposition provides a new conceptual approach to the problem of Blind Source Separation, also known as Independent Component Analysis. In Chapter 5 we will use the decomposition to compute an initial value for a tensorial equivalent of the Power Method, with which the best rank-1 approximation of a given tensor can be calculated. Chapter 6 shows that the HOSVD can be used for dimensionality reduction in higher-order Factor Analysis-type problems.

## 4.2 The Higher-Order Singular Value Decomposition

In this section a singular value decomposition model is proposed for $N$th-order tensors. Comparison to the structure of matrix SVD shows the complete analogy between both cases. For convenience we first repeat the model for matrices, using a similar notation:



**Theorem 4.2.1 (Matrix Singular Value Decomposition)**
*Every complex $(I_1 \times I_2)$-matrix $\mathbf{F}$ can be written as the product*

$$\mathbf{F} = \mathbf{U}^{(1)} \cdot \mathbf{S} \cdot \mathbf{V}^{(2)^H} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{V}^{(2)^*} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}, \tag{4.1}$$

*in which:*

- $\mathbf{U}^{(1)} = \left[U_1^{(1)} U_2^{(1)} \dots U_{I_1}^{(1)}\right]$ *is a unitary* $(I_1 \times I_1)$*-matrix,*
- $\mathbf{U}^{(2)} = \left[U_1^{(2)} U_2^{(2)} \dots U_{I_2}^{(2)}\right]$ $(= \mathbf{V}^{(2)^*})$ *is a unitary* $(I_2 \times I_2)$*-matrix,*
- $\mathbf{S}$ *is an* $(I_1 \times I_2)$*-matrix with the properties of:*
  - *pseudodiagonality:*

$$\mathbf{S} = \mathtt{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\mathtt{min}(I_1, I_2)}), \tag{4.2}$$

  - *ordering:*

$$\sigma_1 \geqslant \sigma_2 \geqslant \dots \geqslant \sigma_{\mathtt{min}(I_1, I_2)} \geqslant 0. \tag{4.3}$$

*The $\sigma_i$ are singular values of $\mathbf{F}$ and the vectors $U_i^{(1)}$ and $U_i^{(2)}$ are respectively an ith left and an ith right singular vector. The decomposition is visualized in Fig. 4.1.*

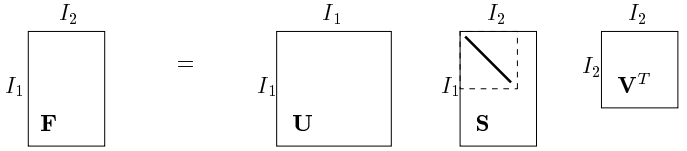



Figure 4.1: Visualization of the matrix SVD.

Now we state:

**Theorem 4.2.2 ($N$th-Order Singular Value Decomposition)**
*Every complex $(I_1 \times I_2 \times \dots \times I_N)$-tensor $\mathcal{A}$ can be written as the product*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)}, \tag{4.4}$$

*in which:*

- $\mathbf{U}^{(n)} = \left[U_1^{(n)} U_2^{(n)} \dots U_{I_n}^{(n)}\right]$ *is a unitary* $(I_n \times I_n)$*-matrix,*

- *$\mathcal{S}$ is a complex $(I_1 \times I_2 \times \ldots \times I_N)$-tensor of which the subtensors $\mathcal{S}_{i_n=\alpha}$, obtained by fixing the nth index to $\alpha$, have the properties of:*

  - *all-orthogonality: two subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all possible values of $n$, $\alpha$ and $\beta$ subject to $\alpha \neq \beta$:*

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when} \quad \alpha \neq \beta, \tag{4.5}$$

  - *ordering:*

$$\|\mathcal{S}_{i_n=1}\| \geqslant \|\mathcal{S}_{i_n=2}\| \geqslant \ldots \geqslant \|\mathcal{S}_{i_n=I_n}\| \geqslant 0 \tag{4.6}$$

  *for all possible values of $n$.*

*The Frobenius-norm $\|\mathcal{S}_{i_n=i}\|$, symbolized by $\sigma_i^{(n)}$, is the ith $n$-mode singular value of $\mathcal{A}$ and the vector $U_i^{(n)}$ is the ith $n$-mode singular vector. The decomposition is visualized for third-order tensors in Fig. 4.2.*

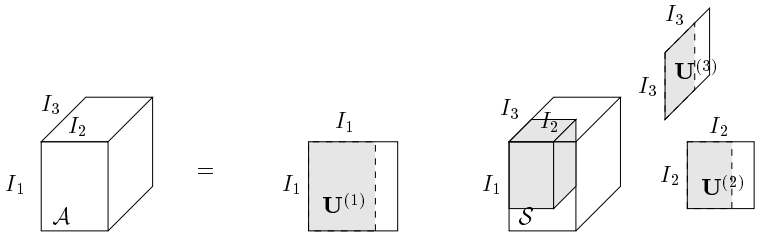



Figure 4.2: Visualization of the HOSVD for a third-order tensor.

The proof of the HOSVD-theorem is given in Section 4.3.

From a comparison of the matrix and tensor theorem it is clear that in the higher-order case the "core matrix" $\mathbf{S}$ is replaced by the "core tensor" $\mathcal{S}$. Instead of being pseudo-diagonal, which would mean that non-zero elements could only occur when the indices $i_1 = i_2 = \ldots = i_N$, $\mathcal{S}$ is in general a full tensor, but it obeys the weaker condition of all-orthogonality instead. The actual role of the singular values is taken over by the Frobenius-norms of the $(N-1)$th-order subtensors of the core tensor. By definition the $n$-mode singular values are positive and real, like in the matrix case. On the other hand the entries of $\mathcal{S}$ are not necessarily positive in general; they can even be complex, when $\mathcal{A}$ is a complex-valued tensor.

Reflections on rank easily reveal that pseudo-diagonality of the core tensor cannot be the appropriate generalization of pseudo-diagonality of the core matrix. Let us consider a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ of which the core tensor $\mathcal{S}$



is pseudo-diagonal, with $I$ non-zero elements on the diagonal. Since the singular vectors of any given mode are orthogonal, it follows that $\texttt{rank}_1(\mathcal{A}) = \texttt{rank}_2(\mathcal{A}) = \ldots \texttt{rank}_N(\mathcal{A}) = \texttt{rank}(\mathcal{A})(= I)$. As illustrated in Section 2.6, none of the equations holds in general. One comes to the same conclusion by noticing that pseudo-diagonality of the core tensor would imply that the decomposition would show but $I\,(\sum I_n + 1 - N(I+1)/2)$ degrees of freedom, while the original tensor contains $I_1 I_2 \ldots I_N$ independent entries. Only in the second-order case both quantities are equal, for $I = \texttt{min}(I_1, I_2)$. In the next section we will prove that relaxation of the pseudo-diagonality condition to all-orthogonality yields a decomposition which is always possible.

A mere matrix representation of the HOSVD can be obtained by unfolding $\mathcal{A}$ and $\mathcal{S}$ in model equation (4.4):

$$\mathbf{A}_{\left(I_{\mathtt{P}(1)} \times I_{\mathtt{P}(2)} I_{\mathtt{P}(3)} \cdots I_{\mathtt{P}(N)}\right)} =$$
$$\mathbf{U}^{(\mathtt{P}(1))} \cdot \mathbf{S}_{\left(I_{\mathtt{P}(1)} \times I_{\mathtt{P}(2)} I_{\mathtt{P}(3)} \cdots I_{\mathtt{P}(N)}\right)} \cdot \left[\mathbf{U}^{(\mathtt{P}(2))} \otimes \mathbf{U}^{(\mathtt{P}(3))} \otimes \ldots \otimes \mathbf{U}^{(\mathtt{P}(N))}\right]^T,$$

in which $\otimes$ denotes the Kronecker product (App. A).

The dyadic decomposition can be generalized by expressing the model equation by means of rank-1 tensors:

$$\mathcal{A} = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} s_{i_1 i_2 \ldots i_N}\, U_{i_1}^{(1)} \circ U_{i_2}^{(2)} \circ \ldots \circ U_{i_N}^{(N)}.$$

In connection with the discussion on pseudo-diagonality versus all-orthogonality, we remark that the summation generally involves $r_1 r_2 \ldots r_N$ terms, in which $r_n$ is the highest index for which $\|\mathcal{S}_{i_n=r_n}\| > 0$ in Eq. (4.6). In Fig. 4.3 this decomposition is visualized for third-order tensors.

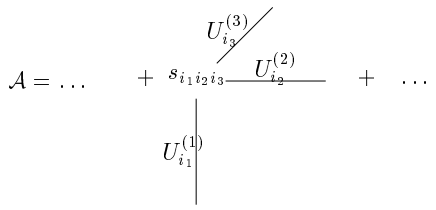



Figure 4.3: Visualization of the higher-order dyadic decomposition for a third-order tensor.

## 4.3 Derivation of the HOSVD-theorem

The derivation establishes the connection between the HOSVD of a tensor $\mathcal{A}$ and the matrix SVD of a matrix unfolding of $\mathcal{A}$. It is given in terms of real-valued tensors; the complex case is completely analogous but more cumbersome from a notational point of view.

Consider two $(I_1 \times I_2 \times \ldots \times I_N)$-tensors $\mathcal{A}$ and $\mathcal{S}$, related by

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}, \tag{4.7}$$

in which $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ are orthogonal matrices. Eq. (4.7) can be expressed in matrix format as

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \mathbf{S}_{(n)} \cdot \left[\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)} \ldots \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \ldots \otimes \mathbf{U}^{(N)}\right]^T. \tag{4.8}$$

Now consider the particular case where $\mathbf{U}^{(n)}$ is obtained from the SVD of $\mathbf{A}_{(n)}$ as

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \Sigma^{(n)} \cdot \mathbf{V}^{(n)^T}, \tag{4.9}$$

in which $\mathbf{V}^{(n)}$ is orthogonal and $\Sigma^{(n)} = \mathtt{diag}\{\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots, \sigma_{I_n}^{(n)}\}$, where

$$\sigma_1^{(n)} \geqslant \sigma_2^{(n)} \geqslant \ldots \geqslant \sigma_{I_n}^{(n)} \geqslant 0. \tag{4.10}$$

We call $r_n$ the highest index for which $\sigma_{r_n}^{(n)} > 0$. Taking into account that the Kronecker factor in Eq. (4.8) is orthogonal, comparison of (4.8) and (4.9) shows that

$$\mathbf{S}_{(n)} = \Sigma^{(n)} \cdot \mathbf{V}^{(n)^T} \cdot \left[\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)} \ldots \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \ldots \otimes \mathbf{U}^{(N)}\right]. \tag{4.11}$$

This equation implies, for arbitrary orthogonal matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, ... , $\mathbf{U}^{(n-1)}$, $\mathbf{U}^{(n+1)}$, ... , $\mathbf{U}^{(N)}$, that

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \qquad \text{when} \qquad \alpha \neq \beta, \tag{4.12}$$

and

$$\|\mathcal{S}_{i_n=1}\| = \sigma_1^{(n)} \geqslant \|\mathcal{S}_{i_n=2}\| = \sigma_2^{(n)} \geqslant \ldots \geqslant \|\mathcal{S}_{i_n=I_n}\| = \sigma_{I_n}^{(n)} \geqslant 0, \tag{4.13}$$

and, if $r_n < I_n$,

$$\|\mathcal{S}_{i_n=r_n+1}\| = \sigma_{r_n+1}^{(n)} = \ldots = \|\mathcal{S}_{i_n=I_n}\| = \sigma_{I_n}^{(n)} = 0. \tag{4.14}$$

By constructing $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(n-1)}, \mathbf{U}^{(n+1)}, \ldots, \mathbf{U}^{(N)}$ in a similar way as $\mathbf{U}^{(n)}$, $\mathcal{S}$ can be made satisfy all the conditions of the HOSVD-theorem. On the other hand, as can be seen from (4.8)-(4.14), all matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ and tensors $\mathcal{S}$ satisfying the HOSVD-theorem can be found by the SVD of $\mathbf{A}_{(1)}$, $\mathbf{A}_{(2)}$, ... , $\mathbf{A}_{(N)}$, where $\mathcal{S}$ follows from (4.7).



## 4.4 Properties

In this section we generalize a number of well-known SVD-properties to the HOSVD. The proofs are outlined at the end of the section. The corresponding matrix properties are briefly listed in App. A.

**Property 4.4.1 (Uniqueness)** *When the n-mode singular values are different, then the n-mode singular vectors are determined up to multiplication with a unit-modulus factor. When $U_\alpha^{(n)}$ is multiplied by $e^{j\theta}$, then $\mathcal{S}_{i_n=\alpha}$ has to be multiplied by the inverse factor $e^{-j\theta}$.*

*The n-mode singular vectors corresponding to the same n-mode singular value can be replaced by any unitary linear combination. The corresponding subtensors $\{\mathcal{S}_{i_n=\alpha}\}$ have to be combined in the inverse way. Formally, $\mathbf{U}^{(n)}$ can be replaced by $\mathbf{U}^{(n)} \cdot \mathbf{Q}$, in which $\mathbf{Q}$ is a block-diagonal matrix, consisting of unitary blocks, where the block-partitioning corresponds to the partitioning of $\mathbf{U}^{(n)}$ in sets of n-mode singular vectors with identical n-mode singular value. At the same time $\mathcal{S}$ has to be replaced by $\mathcal{S} \times_n \mathbf{Q}^H$.*

*For real-valued tensors uniqueness is up to the sign, resp. multiplication with an orthogonal matrix.*

This property implies that the HOSVD shows similar uniqueness properties as its matrix counterpart. Next, we claim that the HOSVD is not only a formal equivalent of the matrix SVD, but that it really establishes a multilinear SVD-framework, containing the matrix decomposition in the special case of second-order tensors.

**Property 4.4.2 (Generalization)** *The tensor SVD of a second-order tensor boils down (up to the underdetermination) to its matrix SVD.*

**Property 4.4.3 ($n$-rank)** *Let the HOSVD of $\mathcal{A}$ be given as in Theor. 4.2.2, and let $r_n$ be equal to the highest index for which $\|\mathcal{S}_{i_n=r_n}\| > 0$ in Eq. (4.6), then one has:*

$$R_n = \mathtt{rank_n}(\mathcal{A}) = r_n.$$

Just like the matrix SVD reveals the column (row) rank of a given matrix, HOSVD shows the $n$-rank values for a given tensor, even in a numerical sense, as pointed out by the perturbation analysis in Section 4.7.

**Property 4.4.4 (Structure)** *Let the HOSVD of $\mathcal{A}$ be given as in Theor. 4.2.2, then one has:*

*- the n-mode vector space* $\mathtt{R}(\mathbf{A}_{(n)}) = \mathtt{span}\{U_1^{(n)}, \ldots, U_{R_n}^{(n)}\}$,
*- the orthogonal complement of* $\mathtt{R}(\mathbf{A}_{(n)})$*, the left n-mode null space* $\mathtt{N}(\mathbf{A}_{(n)}^H) = \mathtt{span}\{U_{R_n+1}^{(n)}, \ldots, U_{I_n}^{(n)}\}$.

**Property 4.4.5 (Norm)** *Let the HOSVD of $\mathcal{A}$ be given as in Theor. 4.2.2, then holds:*

$$\begin{aligned} \|\mathcal{A}\|^2 &= \sum_{i=1}^{R_1} \left(\sigma_i^{(1)}\right)^2 = \ldots = \sum_{i=1}^{R_N} \left(\sigma_i^{(N)}\right)^2 \\ &= \|\mathcal{S}\|^2. \end{aligned}$$

Multiplication of a higher-order tensor by an orthogonal matrix leaves its "energy" unchanged. Moreover, multiplication with the matrices of singular vectors gives a detailed geometrical picture of the energy distribution over the $n$-mode vector space. With this respect we first define the $n$-mode oriented energy of a higher-order tensor:

**Definition 4.4.1 ($n$-mode oriented energy)** *The n-mode oriented energy of an Nth-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ in the direction of a unit-norm vector $X$, represented by* $\mathtt{OE_n}(X, \mathcal{A})$*, is the oriented energy of the set of n-mode vectors, i.e.*

$$\mathtt{OE_n}(X, \mathcal{A}) = \|X^H \mathbf{A}_{(n)}\|^2.$$

We now state:

**Property 4.4.6 (Oriented energy)** *The directions of extremal n-mode oriented energy correspond to the n-mode singular vectors, with extremal energy value equal to the corresponding squared n-mode singular value.*

In the same way as the SVD of a matrix $\mathbf{F}$ is related to the EVD of the Hermitean (real symmetric) positive (semi)definite matrices $\mathbf{F}\mathbf{F}^H$ and $\mathbf{F}^H\mathbf{F}$, the HOSVD of a higher-order tensor $\mathcal{A}$ is related to the EVD of Hermitean (real symmetric) positive (semi)definite matrices, constructed from $\mathcal{A}$. The construction is clarified in Fig. 4.4: just like the entries of $\mathbf{F}\mathbf{F}^H$ consist of the mutual scalar products of the rows of $\mathbf{F}$, the matrix $\mathbf{H}^{(1)}$ in the property below is computed from the scalar products of the "horizontal matrices" of $\mathcal{A}$, in the third-order case.



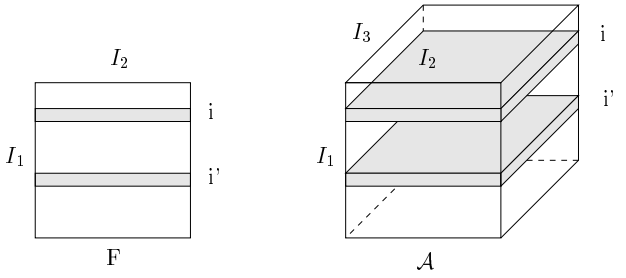



Figure 4.4: Construction of $\mathbf{H}^{(1)}$ for (a) a matrix and (b) a third-order tensor.

**Property 4.4.7 (Link between HOSVD and matrix EVD)** *Let the HOSVD of $\mathcal{A}$ be given as in Theor. 4.2.2. Define $\mathbf{H}^{(n)} \stackrel{\text{def}}{=} \mathbf{A}_{(n)} \cdot \mathbf{A}_{(n)}^H$, i.e. $\mathbf{H}^{(n)}$ contains on position $(i, i')$ the scalar product $\langle \mathcal{A}_{i_n=i}, \mathcal{A}_{i_n=i'} \rangle$. If the EVD of $\mathbf{H}^{(n)}$ is given by*

$$\mathbf{H}^{(n)} = \mathbf{U}^{(n)} \mathbf{D}^{(n)} \mathbf{U}^{(n)^H}, \tag{4.15}$$

*then $\mathbf{U}^{(n)}$ contains the n-mode singular vectors of $\mathcal{A}$. Moreover the scalar product $\langle \mathcal{S}_{i_n=i}, \mathcal{S}_{i_n=i'} \rangle$ is the $(i, i')$th element of $\mathbf{D}^{(n)}$. Different subtensors of $\mathcal{S}$ are mutually orthogonal.*

**Property 4.4.8 (Link between HOSVD and matrix SVD)** *Let the HOSVD of $\mathcal{A}$ be given as in Theor. 4.2.2. Then*

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \Sigma^{(n)} \cdot \mathbf{V}^{(n)^H}$$

*is an SVD of $\mathbf{A}_{(n)}$, where the diagonal matrix $\Sigma^{(n)} \in \mathbb{R}^{I_n \times I_n}$ and the unitary matrix $\mathbf{V}^{(n)} \in \mathbb{C}^{I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N \times I_n}$ are defined according to*

$$\Sigma^{(n)} \stackrel{\text{def}}{=} \mathtt{diag}\{\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots, \sigma_{I_n}^{(n)}\},$$

$$\mathbf{V}^{(n)^H} \stackrel{\text{def}}{=} \tilde{\mathbf{S}}_{(n)} \cdot \left[\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)} \ldots \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \ldots \otimes \mathbf{U}^{(N)}\right]^T,$$

*in which $\tilde{\mathbf{S}}_{(n)}$ is a normalized version of $\mathbf{S}_{(n)}$, with the rows scaled to unit-length:*

$$\mathbf{S}_{(n)} = \Sigma^{(n)} \cdot \tilde{\mathbf{S}}_{(n)}.$$

This property establishes the link between the HOSVD and the SVD of linear mappings, proposed in Subsect. 2.8.4: $\mathbf{U}^{(n)}$, $\Sigma^{(n)}$ and $\mathbf{V}^{(n)}$ correspond to the components of the linear mapping from $\mathbb{C}^{I_1 \times I_2 \times \ldots I_{n-1} \times I_{n+1} \ldots \times I_N}$ to $\mathbb{C}^{I_n}$, defined by summation over the indices $i_1, i_2, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N$ (in Eq. (2.13) we have that $\mathcal{U}_p = U_p^{(n)}$, $\sigma_p = \sigma_p^{(n)}$ and $(\mathcal{V}_p)_{i_1,i_2,\ldots,i_{n-1},i_{n+1},\ldots,i_N}$ equal to the entry of $\mathbf{V}^{(n)}$ with column index $p$ and row index given by $(i_1-1)I_1I_2\ldots I_{n-1}I_{n+1}\ldots I_N + (i_2-1)I_2\ldots I_{n-1}I_{n+1}\ldots I_N + \ldots + (i_{n-1}-1)I_{n-1}I_{n+1}\ldots I_N + (i_{n+1}-1)I_{n+1}\ldots I_N + \ldots + i_N$).

**Proofs:** The proofs are established by a further analysis of the derivation in Section 4.3; the key idea behind this derivation is explicited in Prop. 4.4.8.
Prop. 4.4.1: is proved by investigating the uniqueness conditions.
Prop. 4.4.2: follows by applying the same procedure to second-order tensors. Note that the 2-mode singular vectors are *defined* as the complex conjugates of the right matrix singular vectors. The only case leading to a non-trivial difference between the matrix and the tensor decomposition is the one with multiple coinciding singular values, where the underdetermination of the HOSVD does not impose full diagonality on the matrix of singular values.
Prop. 4.4.3: follows from the combination of Prop. 4.4.8 and Prop. 2.6.1. In the same way the deduction of Prop. 4.4.4 and Prop. 4.4.6 is trivial.
Prop. 4.4.5: we have $\|\mathcal{A}\|^2 = \|\mathbf{A}_{(n)}\|^2$ which, in turn, equals $\sum_{i=1}^{r_n} \|\mathcal{S}_{i_n=i}\|^2 = \|\mathcal{S}\|^2$.
Prop. 4.4.7: follows from the link between matrix SVD and HOSVD, combined with the relation between matrix SVD and EVD. □

## 4.5 Computation

Prop. 4.4.8 shows that $\mathbf{U}^{(n)}$ can directly be computed as the left singular matrix of any matrix of which the columns are given by the $n$-mode vectors. Hence HOSVD of an $N$th-order tensor leads to the computation of $N$ different matrix SVD's of matrices with size $(I_n \times I_1I_2\ldots I_{n-1}I_{n+1}\ldots I_N)(1 \leqslant n \leqslant N)$.

Afterwards the core tensor $\mathcal{S}$ can be computed by bringing the matrices of singular vectors to the left side of Eq. (4.4):

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}. \tag{4.16}$$

This can be computed in matrix format as e.g.:

$$\mathbf{S}_{\left(I_{\mathtt{P}(1)} \times I_{\mathtt{P}(2)} I_{\mathtt{P}(3)} \ldots I_{\mathtt{P}(N)}\right)} = \mathbf{U}^{(\mathtt{P}(1))^T} \cdot \mathbf{A}_{\left(I_{\mathtt{P}(1)} \times I_{\mathtt{P}(2)} I_{\mathtt{P}(3)} \ldots I_{\mathtt{P}(N)}\right)} \cdot \left[\mathbf{U}^{(\mathtt{P}(2))} \otimes \mathbf{U}^{(\mathtt{P}(3))} \otimes \ldots \otimes \mathbf{U}^{(\mathtt{P}(N))}\right], \tag{4.17}$$

in which $\mathtt{P}$ represents a permutation.



**Example 4.5.1** Consider the $(3 \times 2 \times 3)$-tensor $\mathcal{A}$ of Ex. 2.5.2. The 1-mode singular vectors are the columns of the left singular matrix of $\mathbf{A}_{(1)}$; in the same way $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ can be obtained:

$$\mathbf{U}^{(1)} = \begin{pmatrix} 0.1893 & -0.7572 & 0.6252 \\ 0.9816 & 0.1622 & -0.1008 \\ 0.0251 & -0.6327 & -0.7740 \end{pmatrix},$$

$$\mathbf{U}^{(2)} = \begin{pmatrix} 0.3256 & 0.9455 \\ 0.9455 & -0.3256 \end{pmatrix},$$

$$\mathbf{U}^{(3)} = \begin{pmatrix} -0.6564 & 0.7055 & -0.2673 \\ 0.2546 & 0.5406 & 0.8018 \\ 0.7102 & 0.4582 & -0.5345 \end{pmatrix}.$$

The core tensor of the HOSVD then follows from application of Eq. (4.17); unfolded it is equal to:

$$\mathbf{S}_{(1)} = \left(\begin{array}{ccc|ccc} -7.1846 & 0.2235 & 0.0000 & 0.0615 & -0.4906 & 0.0000 \\ -0.1703 & -5.9024 & 0.0000 & 0.7654 & -0.0991 & 0.0000 \\ 0.0035 & 0.2675 & 0.0000 & 2.1071 & 0.3350 & 0.0000 \end{array}\right).$$

The core tensor is all-orthogonal: the rows of $\mathbf{S}_{(1)}$ are mutually orthogonal, but also the matrices formed by the first and the last three columns, as well as the three matrices formed by columns 1 and 4, 2 and 5, and 3 and 6. This boils down to orthogonality of respectively the "horizontal", "frontal" and "vertical" matrices of $\mathcal{A}$, defined in analogy with Fig. 4.4 (right).

The core tensor is ordered: its matrices are put in order of non-increasing Frobenius-norm. The Frobenius-norms give the singular values of $\mathcal{A}$:

mode 1: 7.2051 , 5.9551 and 2.1502
mode 2: 9.3063 and 2.3221
mode 3: 7.5284, 5.9433 and 0

The sums of the squared $n$-mode singular values are all equal to 92, which is the squared Frobenius-norm of $\mathcal{A}$. The third 3-mode singular value equals 0, since the 3-mode singular vectors (the rows of the left and right submatrix of $\mathbf{A}_{(1)}$) only span a two-dimensional space. □

## 4.6 Higher-Order Eigenvalue Decomposition

An interesting property of the HOSVD is that it reflects symmetry properties of the tensor that is decomposed:

**Theorem 4.6.1** *Let the HOSVD of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ be given as in Theor. 4.2.2. Consider a permutation $\mathtt{P}$ of its indices and the decomposition of this permutation in a sequence of permutation cycles $\mathtt{C}_1$, $\mathtt{C}_2$, ... We assume, without losing generality (possibly by redefining the ordering of the indices of $\mathcal{A}$), that $\mathtt{P}(i_1, i_2, \ldots, i_N) = (\mathtt{C}_1(i_1, i_2, \ldots, i_{n_1}), \mathtt{C}_2(i_{n_1+1}, i_{n_1+2}, \ldots, i_{n_2}), \ldots) = (i_2, i_3, \ldots, i_{n_1}, i_1; i_{n_1+2}, \ldots, i_{n_2}, i_{n_1+1}; \ldots)$.*

*If $a_{i_1 i_2 \ldots i_N} = a_{\mathtt{P}(i_1 i_2 \ldots i_N)}$ then one has:*

$$\mathbf{U}^{(1)} = \mathbf{U}^{(2)} = \ldots = \mathbf{U}^{(n_1)};\ \mathbf{U}^{(n_1+1)} = \mathbf{U}^{(n_1+2)} = \ldots = \mathbf{U}^{(n_2)};\ \ldots$$

*If $a_{i_1 i_2 \ldots i_N} = a^*_{\mathtt{P}(i_1 i_2 \ldots i_N)}$ then one has:*

$$\mathbf{U}^{(1)} = \mathbf{U}^{(2)^*} = \mathbf{U}^{(3)} = \ldots\ ;\ \mathbf{U}^{(n_1+1)} = \mathbf{U}^{(n_1+2)^*} = \mathbf{U}^{(n_1+3)} = \ldots;\ \ldots$$

*For permutation cycles with an odd number of elements the matrix of singular vectors is real.*

*The core tensor $\mathcal{S}$ exhibits the same permutation symmetry as $\mathcal{A}$.*

**Proof:** Construct the higher-order tensors $\mathcal{A}'$, $\mathcal{A}''$, ... by repeatedly permuting the indices by $\mathtt{P}$:

$$a'_{i_1, i_2, \ldots, i_N} = a_{\mathtt{P}(i_1, i_2, \ldots, i_N)}; \qquad a''_{i_1, i_2, \ldots, i_N} = a_{\mathtt{P}(\mathtt{P}(i_1, i_2, \ldots, i_N))}; \qquad \ldots$$

For real symmetry we have that $\mathcal{A} = \mathcal{A}' = \mathcal{A}'' = \ldots$; for Hermitean symmetry we have that $\mathcal{A} = (\mathcal{A}')^* = \mathcal{A}'' = \ldots$ The same equalities hold for the matrix unfoldings $\mathbf{A}_{(n_1)}$, $\mathbf{A}'_{(n_1)}$, $\mathbf{A}''_{(n_1)}$, ..., as well as for $\mathbf{A}_{(n_2)}$, $\mathbf{A}'_{(n_2)}$, $\mathbf{A}''_{(n_2)}$, ..., etc. Hence the same symmetry is shown by the left singular matrices of these matrix unfoldings, which correspond to $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, ..., resp. $\mathbf{U}^{(n_1+1)}$, $\mathbf{U}^{(n_1+2)}$, ..., etc.
If $\mathtt{P}$ corresponds to Hermitean symmetry, and e.g. $\mathtt{C}_1$ is a cycle with an odd number of elements, then it can be shown that $\mathbf{U}^{(1)} = \mathbf{U}^{(1)^*}$ by repeating the construction above $n_1$ times.
For an analysis of the symmetry of $\mathcal{S}$ we write down an element-wise form of Eq. (4.16) (we take the example of Hermitean symmetry):

$$s_{j_1 j_2 \ldots j_N} = \sum_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(n_1)^*}_{i_1 j_1} u^{(n_1)}_{i_2 j_2} u^{(n_1)^*}_{i_3 j_3} \ldots u^{(n_2)^*}_{i_{n_1+1} j_{n_1+1}} u^{(n_2)}_{i_{n_1+2} j_{n_1+2}} u^{(n_2)^*}_{i_{n_1+3} j_{n_1+3}} \ldots \tag{4.18}$$

Permutation of the indices by $\mathtt{P}$ yields the following element:

$$s_{\mathtt{P}(j_1 j_2 \ldots j_N)} = \sum_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(n_1)^*}_{i_1 j_2} u^{(n_1)}_{i_2 j_3} u^{(n_1)^*}_{i_3 j_4} \ldots$$



$$u^{(n_2)^*}_{i_{n_1+1} j_{n_1+2}} u^{(n_2)}_{i_{n_1+2} j_{n_1+3}} u^{(n_2)^*}_{i_{n_1+3} j_{n_1+4}} \ldots \tag{4.19}$$

Invoking the Hermitean symmetry of $\mathcal{A}$, and comparing to Eq. (4.18) it follows that $s_{\mathtt{P}(j_1 j_2 \ldots j_N)} = s^*_{j_1 j_2 \ldots j_N}$. □

Many applications show highly symmetric higher-order tensors. As an example, higher-order moments and cumulants of a random vector are super-symmetric, in the sense of Def. 2.7.2. The same holds e.g. for the symmetric tensorial representation of homogeneous polynomials [67]. According to Theor. 4.6.1 this super-symmetry is reflected in the HOSVD. However, it appears that the $i$-mode singular vectors, for different $i$, are already fully related under a weaker condition, that we defined as *pair-wise symmetry*:

**Definition 4.6.2** *A higher-order tensor $\mathcal{A} \in \mathbb{C}^{I \times I \times \ldots \times I}$ is called pair-wise symmetric when, for every pair of indices $(i_{n_1}, i_{n_2})$ there exists a permutation $\mathtt{P}_{n_1 n_2}$ such that $i_{n_1} = \mathtt{P}_{n_1 n_2}(i_{n_2})$ and either $a_{\mathtt{P}_{n_1 n_2}(i_1 i_2 \ldots i_N)} = a_{i_1 i_2 \ldots i_N}$ or $a_{\mathtt{P}_{n_1 n_2}(i_1 i_2 \ldots i_N)} = a^*_{i_1 i_2 \ldots i_N}$.*

Like super-symmetric higher-order tensors, pair-wise symmetric higher-order tensors are a tensorial equivalent of symmetric and Hermitean matrices. It is well-known that at the matrix level symmetry is reflected by the form of the EVD, which is also equal to the SVD, generically up to the ordering and the sign of the singular values and vectors. This leads in a very natural way to the definition of a *Higher-Order Eigenvalue Decomposition* (HOEVD) for pair-wise symmetric higher-order tensors:

**Theorem 4.6.3 ($N$th-Order Pair-Wise Symmetric EVD)** *Every pair-wise symmetric $(I \times I \times \ldots \times I)$-tensor $\mathcal{A}$ can be written as the product*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}, \tag{4.20}$$

*in which:*

- *$\mathcal{S}$ is an all-orthogonal $(I \times I \times \ldots \times I)$-tensor with the same pair-wise symmetry as $\mathcal{A}$,*
- *$\mathbf{U}^{(1)} = \mathbf{U} = [U_1 U_2 \ldots U_I]$ is a unitary $(I \times I)$-matrix, equal to $\mathbf{U}^{(n)}$ or $\mathbf{U}^{(n)^*}$ $(1 \leqslant n \leqslant N)$, depending on the symmetry of $\mathcal{A}$ as stated in Theor. 4.6.1.*

*The Frobenius-norms of the subtensors of $\mathcal{S}$, obtained by fixing one index to $i$, are the $N$th-order eigenvalues and are symbolized by $\lambda_i$. The vectors $U_i$ are the $N$th-order eigenvectors.*

The computation of the HOEVD of a pair-wise symmetric $N$th-order ($I \times I \times \ldots \times I$)-tensor, reduces to the SVD of an ($I \times I^{N-1}$)-matrix unfolding.

**Example 4.6.4** Consider a complex third-order tensor $\mathcal{A}$ that satisfies $a_{\mathtt{P}(i_1 i_2 i_3)} = a_{i_2 i_3 i_1} = a^*_{i_1 i_2 i_3}$. According to Theor. 4.6.1, and taking into account that $\mathtt{P}$ is a cycle with three elements, the HOSVD of $\mathcal{A}$ is given by

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U} \times_2 \mathbf{U} \times_3 \mathbf{U},$$

where $\mathbf{U}$ is real and $\mathcal{S}$ satisfies the same symmetries as $\mathcal{A}$. In fact, it is pretty obvious in this example that $\mathbf{U}$ is real: $a_{i_1 i_2 i_3} = a^*_{i_2 i_3 i_1} = a_{i_3 i_1 i_2} = a^*_{i_1 i_2 i_3}$ implies that $\mathcal{A}$ itself, and hence the whole decomposition, is real. □

**Example 4.6.5** Consider the fourth-order cumulant tensor $\mathcal{C}$ of a complex random vector $X$, defined in accordance with the conventions of Sect. 3.2.1 and 3.2.2:

$$c_{i_1 i_2 i_3 i_4} = \mathtt{Cum}\{x_{i_1}, x^*_{i_2}, x^*_{i_3}, x_{i_4}\}. \tag{4.21}$$

The permutation symmetries of $\mathcal{C}$ can be generated by the permutation pair $\mathtt{P}_1$ and $\mathtt{P}_2$, satisfying

$$c_{\mathtt{P}_1(i_1 i_2 i_3 i_4)} = c_{i_4 i_2 i_3 i_1} = c_{i_1 i_2 i_3 i_4}, \tag{4.22}$$

$$c_{\mathtt{P}_2(i_1 i_2 i_3 i_4)} = c_{i_2 i_1 i_4 i_3} = c^*_{i_1 i_2 i_3 i_4}. \tag{4.23}$$

Eq. (4.23) implies that $\mathbf{U}^{(1)} = \mathbf{U}^{(2)^*}$ and that $\mathbf{U}^{(3)} = \mathbf{U}^{(4)^*}$; on the other hand Eq. (4.22) implies that $\mathbf{U}^{(1)} = \mathbf{U}^{(4)}$. Hence the HOSVD of $\mathcal{C}$ reveals the structure of Eq. (4.21):

$$\mathcal{C} = \mathcal{S} \times_1 \mathbf{U} \times_2 \mathbf{U}^* \times_3 \mathbf{U}^* \times_4 \mathbf{U},$$

with $\mathcal{S}$ satisfying Eqs. (4.22) and (4.23) as well. □

## 4.7 First-order perturbation analysis

In this section the first-order derivatives of the $n$-mode singular values and vectors with respect to the elements of the original tensor are computed. The expressions can be used to analyze the numerical sensitivity and the statistical performance of HOSVD-based signal processing algorithms (e.g. the ICA-algorithm of Sect. 7.7), especially for high signal-to-noise ratios: if the output of these algorithms can be considered as a function of the $n$-mode singular values and vectors, then the derivatives of the output could be computed via the chain rule.



Due to the relationship between the matrix and tensor SVD, a perturbation analysis of the HOSVD can be based on existing results that have been developed in numerical linear algebra. Let $\mathcal{A}$ be a real or complex higher-order tensor, the elements of which are infinitely often continuously differentiable functions of a real parameter $p$. Like for matrices, a singular value and vector are infinitely often continuously differentiable functions of $p$ in the neighbourhood of a nominal parameter $p_0$ where that singular value is isolated and non-zero. In case of a zero singular value the singular value function need not be differentiable, due to the definition of a singular value as a norm, being a non-negative number. A similar effect appears when several singular value functions cross each other, due to the ordering constraint on the HOSVD. However such problems van be avoided by dropping the sign and ordering convention. With this goal we define, in analogy to the "Unordered-Unsigned Singular Value Decomposition" (USVD) of [99] an *Unordered-Unsigned Higher-Order Singular Value Decomposition* (UHOSVD):

**Theorem 4.7.1 (Unordered-Unsigned HOSVD)** *If the elements of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ are infinitely often continuously differentiable functions of a real parameter $p$, then there exist real functions $f_i^{(n)} : \mathbb{R} \to \mathbb{R} \quad (1 \leqslant i \leqslant I_n)$ that are infinitely often continuously differentiable, such that, for all $p \in \mathbb{R}$:*

$$\{\sigma_i^{(n)}(p)\,(1 \leqslant i \leqslant I_n)\} = \{|f_i^{(n)}|\,(1 \leqslant i \leqslant I_n)\}.$$

*The functions $f_i^{(n)}$ are called unordered, unsigned $n$-mode singular value functions.*

For the analytic continuation of a set of $n$-mode singular vectors at a nominal parameter value, where they correspond to a multiple $n$-mode singular value, the term *preferred $n$-mode singular vectors* will be used. These vectors are given by the following theorem:

**Theorem 4.7.2 (Preferred $n$-mode singular vectors)** *Let the HOSVD of $\mathcal{A}$ be given as in Theor. 4.2.2, the ordering of the $n$-mode singular vectors being of no importance. Let $\sigma^{(n)}$ be an $n$-mode singular value of multiplicity $m$ and let an SVD of $\mathbf{A}_{(n)}$ be given by:*

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \Sigma^{(n)} \cdot \mathbf{V}^{(n)^H} = \left[\mathbf{U}_1^{(n)} \mathbf{U}_2^{(n)}\right] \cdot \begin{bmatrix} \sigma^{(n)} \mathbf{I}_m & 0 \\ 0 & \Sigma_2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_1^{(n)^H} \\ \mathbf{V}_2^{(n)^H} \end{bmatrix}, \tag{4.24}$$

*in which $\mathbf{U}^{(n)}$, $\Sigma^{(n)}$ and $\mathbf{V}^{(n)}$ are defined in accordance to Prop. 4.4.8. In the partitioning $\mathbf{U}_1^{(n)}$ and $\mathbf{V}_1^{(n)}$ correspond to the singular value $\sigma^{(n)}$.*

*Consider a first-order perturbation of* $\mathcal{A}$ *as* $\mathcal{A} + \epsilon\mathcal{B}$, $\epsilon \in \mathbb{R}$, *and define the matrices* $\mathbf{B}_{j_1 j_2}$, $j_1, j_2 = 1, 2$ *as*

$$\mathbf{B}_{j_1 j_2} = \mathbf{U}_{j_1}^{(n)^H} \mathbf{B}_{(n)} \mathbf{V}_{j_2}^{(n)}.$$

*There are two cases,* $\sigma^{(n)} \neq 0$ *and* $\sigma^{(n)} = 0$*:*

Case 1: $\sigma^{(n)} \neq 0$:

*Consider the eigenvalue problem*

$$\frac{\mathbf{B}_{11} + \mathbf{B}_{11}^H}{2} = \mathbf{X} \cdot \Lambda \cdot \mathbf{X}^H.$$

*Then the* $m$ *preferred* $n$*-mode singular vectors are given by the columns of* $\mathbf{U}_1^{(n)} \cdot \mathbf{X}$. *(The* $m$ *preferred right singular vectors in (4.24) are given by the columns of* $\mathbf{V}_1^{(n)} \cdot \mathbf{X}$*.)*

Case 2: $\sigma^{(n)} = 0$:

*Consider an SVD of* $\mathbf{B}_{11}$

$$\mathbf{B}_{11} = \mathbf{X} \cdot \Lambda \cdot \mathbf{Y}^H.$$

*Then the* $m$ *preferred* $n$*-mode singular vectors are given by the columns of* $\mathbf{U}_1^{(n)} \cdot \mathbf{X}$. *(The* $m$ *preferred right singular vectors in (4.24) are given by the columns of* $\mathbf{V}_1^{(n)} \cdot \mathbf{Y}$*.)*

In analogy with [154, 99] we now state the following perturbation theorem:

**Theorem 4.7.3 (First-order perturbation of the HOSVD)** *Consider a higher-order tensor* $\mathcal{A}(\epsilon)$*, the elements of which are infinitely often continuously differentiable functions of a real parameter* $\epsilon$*. Let the first-order approximation of the Taylor series expansion of* $\mathcal{A}(\epsilon)$ *around* $\epsilon = 0$ *be given by* $\mathcal{A} + \epsilon\mathcal{B}$*. Represent the HOSVD of* $\mathcal{A}$ *as in Theor. 4.2.2. Let an SVD of* $\mathbf{A}_{(n)}$ *be given by*

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \cdot \Sigma^{(n)} \cdot \mathbf{V}^{(n)^H},$$

*in which* $\mathbf{U}^{(n)}$, $\Sigma^{(n)}$ *and* $\mathbf{V}^{(n)}$ *are defined in accordance to Prop. 4.4.8. In case of multiple* $n$*-mode singular values,* $\mathbf{U}^{(n)}$ *and* $\mathbf{V}^{(n)}$ *contain preferred basis vectors. Define the matrix* $\mathbf{B}$ *as*

$$\mathbf{B} = \mathbf{U}^{(n)^H} \mathbf{B}_{(n)} \mathbf{V}^{(n)}$$

*Then one has:*



- *The slopes of the* $n$*-mode unordered-unsigned singular value functions are given by the diagonal elements of* $\mathbf{B}$*.*
- *The first-order approximation of the Taylor series expansion of the* $n$*-mode singular vector functions is given by* $\mathbf{U}^{(n)} + \epsilon\mathbf{U}^{(n)}\Omega$*, in which* $\Omega$ *is a skew-Hermitean matrix defined by:*

$$\begin{aligned}
\omega_{ij} &= 0 \\
&\quad \textit{if } i = j, \\
\omega_{ij} &= (\sigma_i^{(n)} b_{ij}^* + \sigma_j^{(n)} b_{ji}) / (\sigma_i^{(n)^2} - \sigma_j^{(n)^2}) \\
&\quad \textit{if } \sigma_i^{(n)} \neq \sigma_j^{(n)}, \\
\omega_{ij} &= \lim_{\epsilon \to 0} \left[ (\sigma_i^{(n)}(\epsilon) b_{ij}^* + \sigma_j^{(n)}(\epsilon) b_{ji}) / (\sigma_i^{(n)^2}(\epsilon) - \sigma_j^{(n)^2}(\epsilon)) \right] \\
&\quad \textit{if } \sigma_i^{(n)}(\epsilon) = \sigma_j^{(n)}(\epsilon) \textit{ for the isolated value } \epsilon = 0, \\
\omega_{ij} &= -\omega_{ji} = \textit{arbitrary} \\
&\quad \textit{if } \sigma_i^{(n)}(\epsilon) = \sigma_j^{(n)}(\epsilon) \textit{ in a neighbourhood of } \epsilon = 0 \ (i \neq j).
\end{aligned}$$

## 4.8 Conclusion

In this chapter we have investigated the problem of generalizing the SVD of matrices to higher-order tensors, under the condition that the multilinear equivalent should address the so-called $n$-mode vectors in a similar way as the SVD does for column and row vectors. It was shown that this implies that the role of the matrix of singular values has to be assumed by a tensor with the property of all-orthogonality. The resulting decomposition, which was called the Higher-Order SVD (HOSVD), is always possible for real or complex $N$th-order tensors, and when applied to matrices, it reduces - up to some trivial indeterminacies - to the matrix SVD.

In psychometric literature the HOSVD of real-valued tensors is known as the Tucker model, originally developed for the analysis of three-way data tables. In this chapter, we have presented a square-root version of the algorithm, valid for real- and complex-valued tensors of arbitrary order, and shown that the result is a convincing generalization of the SVD-concept itself. The properties of the decomposition have thorougly been analyzed and interpreted in terms of their matrix counterparts.

From the fact that $n$-mode vectors and column (row) vectors play the same role in HOSVD resp. SVD, it followed that the matrices of singular vectors can be computed from the sets of $n$-mode vectors in the same way as in the second-order case. In other words, the HOSVD of an $N$th-order tensor boils

down to $N$ matrix SVD's. As a consequence, several matrix properties have a very clear higher-order counterpart. Uniqueness, link with EVD, directions and values of extremal oriented energy, etc. have been investigated and explicit expressions for a first-order perturbation analysis have been presented.

The link between tensor symmetry and HOSVD was also investigated: if a higher-order tensor is transformed into itself or into its complex conjugate by a permutation of the indices, then this symmetry is reflected by the HOSVD-components. It turned out that the most adequate way to analyze this relation is the decomposition of the permutation in a sequence of permutation cycles. When every pair of indices is connected by some cycle, a property which was called pair-wise symmetry, then the different matrices of higher-order singular vectors are either equal or complex conjugated. In analogy to the EVD of a real symmetric or complex Hermitean matrix the Higher-Order EVD (HOEVD) of a pair-wise symmetric higher-order tensor was defined.

# Chapter 5

# Best Rank-1 and Rank-$(R_1, R_2, \ldots, R_N)$ Approximation

*In this chapter we discuss a multilinear generalization of the best rank-R approximation problem for matrices. The question we aim to answer, is how a given tensor can be approximated, in an optimal least-squares sense, by a tensor with pre-specified n-mode ranks.*

*For matrices, the solution is conceptually obtained by truncation of the SVD. In Sect. 5.1 we briefly remind the power iteration technique for the computation of the dominant singular triplet, and the orthogonal iteration technique for the computation of the dominant R-dimensional subspaces of row and column space. We also remind the Hessenberg Decomposition, which can be used in a preprocessing step, to reduce the computational load of power and orthogonal iterations.*

*As far as higher-order tensors are concerned, Sect. 5.2 deals with the best rank-1 approximation problem; in Sect. 5.3 the more general case is addressed in which the pre-specified n-mode ranks are arbitrary. Higher-order generalizations of the power method and the orthogonal iteration method are discussed; they implement an Alternating Least Squares (ALS) optimization strategy. On the other hand, explicit gradient expressions are provided, for use in gradient-based optimization methods. Truncation of the HOSVD provides an interesting starting value for the optimization routine.*

*In Sect. 5.4 we derive a tensorial generalization of the Hessenberg Decomposition, to decrease the computational cost of higher-order power and orthogonal*

*iterations.*

*The concepts of this chapter are formulated in terms of real-valued tensors, for clarity. They can be generalized for complex tensors.*

## 5.1 Power method and orthogonal iteration

Before presenting a higher-order power method and orthogonal iteration, we briefly remind the definition of their matrix counterparts [131].

Assume a matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$, with EVD given by $\mathbf{A} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{E}^{-1}$, in which $\mathbf{D} = \mathtt{diag}(\lambda_1, \lambda_2, \ldots, \lambda_I)$ and $\mathbf{E} = [E_1 E_2 \ldots E_I]$. Assume that $|\lambda_1| > |\lambda_2| \geqslant \ldots \geqslant |\lambda_I|$. Given a unit-norm vector $U^{(0)} = \sum_i \alpha_i E_i$ with $\alpha_1 \neq 0$, the *power method* consists of the following iteration:

$$\begin{aligned} \tilde{U}^{(k+1)} &= \mathbf{A} U^{(k)}; \\ \lambda^{(k+1)} &= \|\tilde{U}^{(k+1)}\|; \\ U^{(k+1)} &= \tilde{U}^{(k+1)} / \lambda^{(k+1)}. \end{aligned}$$

In each step, the components of $U^{(k)}$ in the direction of the various eigenvectors, are scaled in a way proportional to the corresponding eigenvalues. Eventually the power method converges to the dominant eigenvector and eigenvalue.

Now assume that $|\lambda_1| \geqslant \ldots |\lambda_p| > |\lambda_{p+1}| \ldots \geqslant |\lambda_I|$ and consider an $(I \times p)$-matrix $\mathbf{U}^{(0)}$ with orthonormal columns. The power method can be generalized in a straightforward way to obtain the $p$-dimensional dominant eigenspace:

$$\begin{aligned} \tilde{\mathbf{U}}^{(k+1)} &= \mathbf{A} \mathbf{U}^{(k)}; \\ \mathbf{U}^{(k+1)} \mathbf{R}^{(k+1)} &= \tilde{\mathbf{U}}^{(k+1)}; \end{aligned}$$

in which the second step is a QR-factorization. This algorithm is called an *orthogonal iteration.*

A power iteration can also be used to estimate a dominant singular triplet in the SVD of a matrix. The algorithm is of the type:

$$\begin{aligned} \tilde{V}^{(k+1)} &= \mathbf{A} U^{(k)}; \\ \lambda^{(k+1)} &= \|\tilde{V}^{(k+1)}\|; \\ V^{(k+1)} &= \tilde{V}^{(k+1)} / \lambda^{(k+1)}; \\ \tilde{U}^{(k+1)} &= \mathbf{A}^H V^{(k)}; \\ \mu^{(k+1)} &= \|\tilde{U}^{(k+1)}\|; \\ U^{(k+1)} &= \tilde{U}^{(k+1)} / \mu^{(k+1)}. \end{aligned}$$



$U^{(k)}$ and $V^{(k)}$ converge to the dominant right and left singular vector respectively; $\lambda^{(k)}$ and $\mu^{(k)}$ both converge to the dominant singular value. Evidently, the orthogonal iteration can be modified in a similar way.

Now reconsider the computation of the dominant eigenvector or eigenspace of a square matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$. One way to decrease the computational cost of power and orthogonal iterations, is by first computing the *Hessenberg Decomposition* of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{H} \cdot \mathbf{Q}^H, \tag{5.1}$$

in which $\mathbf{Q} \in \mathbb{C}^{I \times I}$ is unitary, and $\mathbf{H} \in \mathbb{C}^{I \times I}$ is *upper Hessenberg* (i.e. its elements below the subdiagonal are zero). This decomposition can be computed in a finite number of steps and, in addition, power and orthogonal iterations for an upper Hessenberg matrix can be implemented in a relatively efficient way. The eigenvalues of $\mathbf{A}$ and $\mathbf{H}$ are the same; the eigenmatrix of $\mathbf{H}$ is given by $\mathbf{Q}^H \mathbf{E}$, in which $\mathbf{E}$ is the eigenmatrix of $\mathbf{A}$.

The upper Hessenberg form can e.g. be obtained by Householder reduction. This means that $\mathbf{Q}$ is found as a product of Householder matrices $\mathbf{P}_1 \cdot \ldots \cdot \mathbf{P}_{I-2}$: left multiplication with $\mathbf{P}_i$ zeroes the $i$th column below the subdiagonal, while leaving the first $i$ elements unchanged. A *Householder matrix* is a unitary matrix $\mathbf{P} \in \mathbb{C}^{I \times I}$ of the form

$$\mathbf{P} = \mathbf{I} - 2WW^H/(W^H W), \tag{5.2}$$

in which $W \in \mathbb{C}^I$ is non-zero. Given a non-zero vector $X \in \mathbb{C}^I$, $\mathbf{P}X$ can be made a multiple of $Z_1 = (1, 0, \ldots 0)^T$, by choosing

$$W = X + \mathtt{sign}(\mathtt{Re}(x_1)) \|X\| Z_1. \tag{5.3}$$

(The signum function is defined by $\mathtt{sign}(a) = 1$ if $a \geqslant 0$, $\mathtt{sign}(a) = -1$ if $a < 0$.) For the storage of $W$, it can be normalized such that $w_1 = 1$. Multiplication of a vector with the Householder matrix $\mathbf{P}$ can be done efficiently by first computing the inner product of the vector and $W$, rather than first determining the full form of $\mathbf{P}$.

A matrix $\mathbf{A}$ can be reduced to a tridiagonal form $\mathbf{T} = \mathbf{Q}_1^H \cdot \mathbf{A} \cdot \mathbf{Q}_2$, by computing the unitary matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ in a similar way as $\mathbf{Q}$ above. This is interesting in the computation of a dominant singular subspace in the SVD of $\mathbf{A}$. An Hermitean matrix can be tridiagonalized with $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}$, because of the symmetry.

## 5.2 Higher-order power iteration

In this section, we investigate how a given tensor can be approximated, in an optimal least-squares sense, by a tensor of rank 1. In Subsect. 5.2.1, the

problem is formalized in two different ways and analyzed with the technique of Lagrange multipliers. In Subsect. 5.2.2, we show that an approximation can gradually be enhanced with a higher-order generalization of the power method. For super-symmetric $(2 \times 2 \times \ldots \times 2)$-tensors, the determination of the best rank-1 approximation, and of the other stationary points of the higher-order power algorithm, is reformulated as a polynomial rooting problem. Contrarily to the matrix case, truncation of the HOSVD does not necessarily yield the best rank-1 approximation, but it makes sense to initialize an optimization routine with the truncate; this forms the topic of Subsect. 5.2.3. In Subsect. 5.2.4 we give an explicit expression for the gradient of the least-squares cost function, which can be used, as an alternative for the higher-order power method, in gradient-based descent algorithms.

In [231] it has independently been observed that the third-order equivalent of the power iteration converges to a stationary solution; this result was consequently used to obtain an expansion of third-order tensors in mutually orthogonal rank-1 terms. However, the paper did not mention e.g. the monotonous convergence properties of the algorithm, which can be interpreted as an Alternating Least Squares algorithm, and neither did it take into account the presence of spurious stationary points.

The results in this section form an important special case of the results to be discussed in Sect. 5.3. They will be used to devise a new computation scheme for ICA in Chapter 8.

### 5.2.1 Best rank-1 approximation

The problem we want to solve, can mathematically be described as follows:

*Given a real Nth-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, find a scalar $\lambda$ and unit-norm vectors $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$ such that the rank-1 tensor $\hat{\mathcal{A}} \stackrel{\text{def}}{=} \lambda\, U^{(1)} \circ U^{(2)} \circ \ldots \circ U^{(N)}$ minimizes the least-squares cost function*

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 \tag{5.4}$$

*over the manifold of rank-1 tensors.*

This constrained optimization problem can be analyzed using the technique of Lagrange multipliers. Therefore, we consider the following combination of $f$ with the constraint terms:

$$\tilde{f} \stackrel{\text{def}}{=} \sum_{i_1 i_2 \ldots i_N} (a_{i_1 i_2 \ldots i_N} - \lambda u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N})^2 + \sum_n \lambda^{(n)} (\sum_{i_n} (u^{(n)}_{i_n})^2 - 1), \tag{5.5}$$

in which $\lambda^{(n)}$ $(1 \leqslant n \leqslant N)$ are the Lagrange multipliers. Setting the derivative



with respect to $u^{(n)}_{i_n}$ equal to zero, yields:

$$\lambda \sum_{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(1)}_{i_1} \ldots u^{(n-1)}_{i_{n_1}} u^{(n+1)}_{i_{n+1}} \ldots u^{(N)}_{i_N} =$$

$$\lambda^{(n)} u^{(n)}_{i_n} + \lambda^2 u^{(n)}_{i_n} \sum_{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N} (u^{(1)}_{i_1})^2 \ldots (u^{(n-1)}_{i_{n_1}})^2 (u^{(n+1)}_{i_{n+1}})^2 \ldots (u^{(N)}_{i_N})^2. \tag{5.6}$$

Derivation with respect to $\lambda^{(n)}$ and $\lambda$ respectively, yields:

$$\sum_{i_n} (u^{(n)}_{i_n})^2 = 1 \tag{5.7}$$

$$\sum_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N} = \lambda \sum_{i_1 i_2 \ldots i_N} (u^{(1)}_{i_1})^2 (u^{(2)}_{i_2})^2 \ldots (u^{(N)}_{i_N})^2. \tag{5.8}$$

Combination of Eq. (5.7) with Eq. (5.8) and with the right-hand side of Eq. (5.6), yields:

$$\sum_{i_1 i_2 \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N} = \lambda, \tag{5.9}$$

$$\lambda \sum_{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(1)}_{i_1} \ldots u^{(n-1)}_{i_{n_1}} u^{(n+1)}_{i_{n+1}} \ldots u^{(N)}_{i_N} = (\lambda^2 + \lambda^{(n)}) u^{(n)}_{i_n}. \tag{5.10}$$

Combination of Eq. (5.7) with the right-hand side of Eq. (5.10), and comparison to Eq. (5.9), yields:

$$\sum_{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N} a_{i_1 i_2 \ldots i_N} u^{(1)}_{i_1} \ldots u^{(n-1)}_{i_{n_1}} u^{(n+1)}_{i_{n+1}} \ldots u^{(N)}_{i_N} = \lambda u^{(N)}_{i_N}. \tag{5.11}$$

Summarizing, the Lagrange equations correspond to $(1 \leqslant n \leqslant N)$:

$$\mathcal{A} \times_1 {U^{(1)}}^T \ldots \times_{n-1} {U^{(n-1)}}^T \times_{n+1} {U^{(n+1)}}^T \ldots \times_N {U^{(N)}}^T = \lambda\, U^{(n)}, \tag{5.12}$$

$$\mathcal{A} \times_1 {U^{(1)}}^T \times_2 {U^{(2)}}^T \ldots \times_N {U^{(N)}}^T = \lambda, \tag{5.13}$$

$$\|U^{(n)}\| = 1. \tag{5.14}$$

**Remark 5.2.1** For the best approximation of a super-symmetric tensor $\mathcal{A}$ by a super-symmetric rank-1 tensor $\hat{\mathcal{A}} = \lambda\, U \circ U \circ \ldots \circ U$, the derivation is analogous. The equivalent of Eq. (5.12) is:

$$\mathcal{A} \times_1 U^T \ldots \times_{n-1} U^T \times_{n+1} U^T \ldots \times_N U^T = \lambda\, U. \tag{5.15}$$

$\diamond$

The best rank-1 approximation problem can also be formulated as follows:

**Theorem 5.2.2** *Assume a real Nth-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, then the minimization of the cost function of Eq. (5.4) is equivalent to the maximization, over the unit-norm vectors $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$, of the function*

$$g(U^{(1)}, U^{(2)}, \ldots, U^{(N)}) = |\mathcal{A} \times_1 U^{(1)^T} \times_2 U^{(2)^T} \ldots \times_N U^{(N)^T}|^2. \qquad (5.16)$$

*If the scalar $\lambda$, corresponding to the Frobenius-norm of $\hat{\mathcal{A}}$ in Eq. (5.4), is chosen in accordance with Eq. (5.13), then the functions of Eq. (5.4) and Eq. (5.16) are related by:*

$$f = \|\mathcal{A}\|^2 - g. \qquad (5.17)$$

**Proof:** We have:

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 = \|\mathcal{A}\|^2 - 2\langle \mathcal{A}, \hat{\mathcal{A}} \rangle + \|\hat{\mathcal{A}}\|^2.$$

According to the definition of $\lambda$, the value taken by $\langle \mathcal{A}, \hat{\mathcal{A}} \rangle$ equals $\lambda^2$. Since $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$ have unit-norm, $\|\hat{\mathcal{A}}\|^2 = \lambda^2$ as well. Combination with the definition of $g$ proves the theorem. □

**Remark 5.2.3** The optimal choice of $\lambda$, given by Eq. (5.13), implies mutual orthogonality of $\hat{\mathcal{A}}$ and the error term $\mathcal{A} - \hat{\mathcal{A}}$: $\langle \hat{\mathcal{A}}, \mathcal{A} - \hat{\mathcal{A}} \rangle = 0$, as indicated in the proof of the theorem. ◇

**Remark 5.2.4** Theor. 5.2.2 is the higher-order generalization of the fact that the computation of the best rank-1 approximation of a matrix $\mathbf{A}$ is equivalent to the determination of unit-vectors $U$ and $V$ such that $|U^T \mathbf{A} V|^2$ is maximal. ◇

### 5.2.2 A power algorithm

For the actual computation of the best rank-1 approximation of $\mathcal{A}$, the Lagrange equations and their derivation can be interpreted in a constructive way. Imagine that the vectors $U^{(1)}, \ldots, U^{(n-1)}, U^{(n+1)}, \ldots, U^{(N)}$ are fixed and that $f$ in Eq. (5.4) is merely a quadratic function in the unknown unconstrained vector $\lambda U^{(n)}$. Eq. (5.12) then simply shows how the optimal $\lambda U$ can be computed, for fixed $U^{(1)}, \ldots, U^{(n-1)}, U^{(n+1)}, \ldots, U^{(N)}$.

The full set of Lagrange equations ($1 \leqslant n \leqslant N$) can be linked to an Alternating Least Squares (ALS) algorithm for the (local) minimization of $f(\hat{\mathcal{A}})$: in each step the estimate of the scalar $\lambda$ and the estimate of one of the vectors



**Algorithm 5.2.2**
*Higher-Order Power Method*

1. Initial values:
   $U_0^{(n)}$ is the dominant left singular vector of $\mathbf{A}_{(n)}$ ($2 \leqslant n \leqslant N$); or repeat the algorithm for several initial values.

2. Iterate until convergence:

   - $\tilde{U}_{k+1}^{(1)} = \mathcal{A} \times_2 U_k^{(2)^T} \times_3 U_k^{(3)^T} \ldots \times_N U_k^{(N)^T}$;
     $\lambda_{k+1}^{(1)} = \|\tilde{U}_{k+1}^{(1)}\|$;
     $U_{k+1}^{(1)} = \tilde{U}_{k+1}^{(1)} / \lambda_{k+1}^{(1)}$;
   - $\tilde{U}_{k+1}^{(2)} = \mathcal{A} \times_1 U_{k+1}^{(1)^T} \times_3 U_k^{(3)^T} \ldots \times_N U_k^{(N)^T}$;
     $\lambda_{k+1}^{(2)} = \|\tilde{U}_{k+1}^{(2)}\|$;
     $U_{k+1}^{(2)} = \tilde{U}_{k+1}^{(2)} / \lambda_{k+1}^{(2)}$;
   - ...
   - $\tilde{U}_{k+1}^{(N)} = \mathcal{A} \times_1 U_{k+1}^{(1)^T} \times_2 U_{k+1}^{(2)^T} \ldots \times_{N-1} U_{k+1}^{(N-1)^T}$;
     $\lambda_{k+1}^{(N)} = \|\tilde{U}_{k+1}^{(N)}\|$;
     $U_{k+1}^{(N)} = \tilde{U}_{k+1}^{(N)} / \lambda_{k+1}^{(N)}$.

$U^{(1)}, U^{(2)}, \ldots, U^{(N)}$ are optimized, while the other vector estimates are kept constant. The result is shown in step 2 of Alg. 5.2.2. We remark that the expression

$$\tilde{U}_{k+1}^{(n)} = \mathcal{A} \times_1 U_{k+1}^{(1)^T} \ldots \times_{n-1} U_{k+1}^{(n-1)^T} \times_{n+1} U_k^{(n+1)^T} \ldots \times_N U_k^{(N)^T}$$

can be written in matrix format as follows:

$$\tilde{U}_{k+1}^{(n)} = \mathbf{A}_{(n)} \cdot (U_{k+1}^{(1)} \otimes \ldots \otimes U_{k+1}^{(n-1)} \otimes U_k^{(n+1)} \otimes \ldots \otimes U_k^{(N)}).$$

Clearly this technique is a higher-order extension of the power method for matrices, described in Sect. 5.1. The termination criterion could be formulated in terms of the accuracy of components (e.g. $|U_{k+1}^{(n)^T} U_k^{(n)}| < \epsilon$, for $1 \leqslant n \leqslant N$), or in terms of the quality of the fit (e.g. $\lambda_{k+1}^{(N)} - \lambda_k^{(N)} < \epsilon$).

**Remark 5.2.5** If $\mathcal{A}$ is super-symmetric, Eq. (5.15) can as well be interpreted as an update of the unknown $n$-mode vector $\lambda U^{(n)}$, while the other vectors $U^{(1)}, \ldots, U^{(n-1)}, U^{(n+1)}, \ldots, U^{(N)}$ are fixed to $U$. However, such an update breaks the symmetry of the estimate, unless $\lambda$ and $U$ already corresponded to

a super-symmetric solution of the Lagrange equation. A symmetric version of the algorithm, based on the iteration

$$\tilde{U}_{k+1} = \mathcal{A} \times_1 U_k{}^T \times_2 U_k{}^T \ldots \times_{N-1} U_k{}^T, \tag{5.18}$$

for which $U_k^{(1)} = U_k^{(2)} = \ldots = U_k^{(N)} = U_k$, is unreliable since it does not necessarily decrease the cost function $f(\hat{\mathcal{A}})$ in a monotonous way.

The fact that a power iteration produces unsymmetric intermediate results is not observed for symmetric matrices: at the second order level, each iteration step only involves the knowledge of one intermediate vector, such that it is simply impossible to compare the different $n$-mode vector estimates at a given iteration step. On the other hand, minimization of the cost function of Eq. (5.4), while keeping the symmetry, can e.g. take the form of a gradient descent algorithm (for the expression of the gradient, we refer to Subsect. 5.2.4). ◇

**Remark 5.2.6** As a matter of fact, it is perfectly possible to update the estimates of two of the vectors $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$ at the same time. E.g. for fixed values of $U^{(1)}, \ldots, U^{(n-1)}, U^{(n+2)}, \ldots, U^{(N)}$, Theor. 5.2.2 implies that the optimal $U^{(n)}$ and $U^{(n+1)}$ can be found as the dominant left and right singular vectors of the matrix $\mathcal{A} \times_1 U^{(1)^T} \ldots \times_{n-1} U^{(n-1)^T} \times_{n+2} U^{(n+2)^T} \ldots \times_N U^{(N)^T}$, as $g = |U^{(n)^T} \cdot (\mathcal{A} \times_1 U^{(1)^T} \ldots \times_{n-1} U^{(n-1)^T} \times_{n+2} U^{(n+2)^T} \ldots \times_N U^{(N)^T}) \cdot U^{(n+1)}|^2$; the corresponding value of $g$ is the squared dominant singular value. Computation of the dominant singular triplet by a matrix power algorithm would correspond to an iteration between substeps $n$ and $n+1$ of step 2 in Alg. 5.2.2; for more efficient methods we refer to [71]. Note that, in the terminology of Alg. 5.2.2, $U_k^{(n)}$ and $U_k^{(n+1)}$ can be considered as first approximations of $U_{k+1}^{(n)}$ and $U_{k+1}^{(n+1)}$. ◇

**Example 5.2.7** Fig. 5.1 visualizes step 2 of Alg. 5.2.2 for a $(2 \times 2 \times 2)$-example. For different choices of $\theta_0$, determining initial vectors $U_0^{(2)} = U_0^{(3)} = (\ \cos\theta_0 \quad \sin\theta_0\ )^T$, we plotted, after each iteration, the angle $\theta_k^{(3)}$ in $U_k^{(3)} = (\ \cos\theta_k^{(3)} \quad \sin\theta_k^{(3)}\ )^T$ versus the angle $\theta_k^{(2)}$ in $U_k^{(2)} = (\ \cos\theta_k^{(2)} \quad \sin\theta_k^{(2)}\ )^T$. The angles were normalized to the interval $(-\pi/2, +\pi/2]$ (the sign of the vectors does not matter, as can be seen from the definition of the function $g$; as far as $f$ is concerned, the sign can be incorporated in the scalar $\lambda$). The tensor $\mathcal{A}$ that we consider, is super-symmetric and defined by:

$$\begin{cases} a_{111} = 1.5578, & a_{222} = 1.1226, \\ a_{112} = -2.4443, & a_{221} = -1.0982. \end{cases}$$

We observe that Alg. 5.2.2 leads to unsymmetric intermediate results, not located on the main diagonal of the figure.



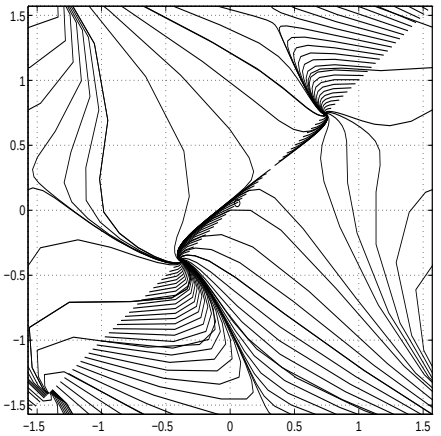



Figure 5.1: Visualization of Alg. 1 for a typical example of a super-symmetric tensor in $\mathbb{R}^{2\times2\times2}$. Abscis: the angle $\theta_k^{(2)}$ in $U_k^{(2)} = (\ \cos\theta_k^{(2)} \quad \sin\theta_k^{(2)}\ )^T$ (in radials). Ordinate: the angle $\theta_k^{(3)}$ in $U_k^{(3)} = (\ \cos\theta_k^{(3)} \quad \sin\theta_k^{(3)}\ )^T$ (in radials). Both angles are normalized to the interval $(-\pi/2, +\pi/2]$. The mark shows the initial guess obtained by HOSVD. The global optimum is $(-0.3860, -0.3860)$.

We also remark that there are two stable $((-0.3860, -0.3860), (0.7413, 0.7413))$ and two unstable $((-1.4052, -1.4052), (0.3347, 0.3347))$ symmetric stationary points of the higher-order power algorithm. The first three of them correspond to the three solutions of the Lagrange equation

$$\mathcal{A} \times_2 U^T \times_3 U^T = \lambda U \tag{5.19}$$

on the unit circle $\|U\| = 1$. The solution (0.3347,0.3347) is more special: even after convergence the three substeps of each iteration produce unsymmetric results, but these effects compensate such that the overall power iteration is symmetric. The global optimum is $(-0.3860, -0.3860)$. □

**Remark 5.2.8** In connection with Ex. 5.2.7, we first prove that there can be at most three distinct symmetric solutions to the Lagrange equations for a super-symmetric $(2 \times 2 \times 2)$-tensor.
For convenience we use the notation $U = (c\ s)^T$. Eq. (5.19) corresponds to the following set of equations:

$$\begin{aligned} a_{111}c^2 + 2a_{112}cs + a_{122}s^2 &= \lambda c, \\ a_{211}c^2 + 2a_{212}cs + a_{222}s^2 &= \lambda s. \end{aligned}$$

Elimination of $\lambda$, and claiming that $\|U\|^2 = 1$, yields:

$$\begin{aligned} a_{211}c^3 + (2a_{212} - a_{111})c^2 s + (a_{222} - 2a_{112})cs^2 - a_{122}s^3 &= 0, \\ c^2 + s^2 &= 1. \end{aligned}$$

On the unit circle a solution is intirely defined by $t = s/c$, each time corresponding to the vectors $(c\ s)^T$ and $(-c\ -s)^T$, which are not essentially different. The solutions can be found by rooting the polynomial

$$a_{211} + (2a_{212} - a_{111})t + (a_{222} - 2a_{112})t^2 - a_{122}t^3 = 0,$$

which has at most three distinct real roots. The same expression shows where the derivative of $\tilde{g} = \mathcal{A} \times_1 U^T \times_2 U^T \times_3 U^T$ w.r.t. $\alpha$ is equal to zero. For an $N$th order super-symmetric $(2 \times 2 \times \ldots \times 2)$-tensor, there can be at most $N$ distinct symmetric solutions to the Lagrange equations.
Now let us consider solutions to the Lagrange equations of the type

$$\mathcal{A} \times_2 U^T \times_3 U^T = \lambda U^{(1)}, \tag{5.20}$$

$$\mathcal{A} \times_2 U^{(1)^T} \times_3 U^T = \lambda U. \tag{5.21}$$

for a super-symmetric $(2 \times 2 \times 2)$-tensor $\mathcal{A}$. This is the case where a higher-order power iteration yields a stationary solution $U^{(2)} = U^{(3)} = U$, while producing an intermediate vector $U^{(1)}$ that is possibly different from $U$, e.g. see the solution (0.3347,0.3347) in Ex. 5.2.7. Eqs. (5.20) and (5.21) can be combined as

$$\mathcal{B} \times_2 U^T \times_3 U^T \times_4 U^T = \lambda U, \tag{5.22}$$

in which $\mathcal{B} \stackrel{\text{def}}{=} \langle \mathcal{A}, \mathcal{A} \rangle_{1,1}$. Although $\mathcal{B}$ itself is not super-symmetric, the solutions to Eq. (5.22) can be determined as above; at most four distinct real solutions are possible. For an $N$th order super-symmetric $(2 \times 2 \times \ldots \times 2)$-tensor, there can be at most $2N - 2$ distinct solutions to the Lagrange equations

$$\mathcal{A} \times_2 U^T \times_3 U^T \ldots \times_N U^T = \lambda U^{(1)}, \tag{5.23}$$

$$\mathcal{A} \times_2 U^{(1)^T} \times_3 U^T \ldots \times_N U^T = \lambda U. \tag{5.24}$$

◇

### 5.2.3 Starting value

Ex. 5.2.7 illustrates a major difference between the second-order and the higher-order problem of best rank-1 approximation: for tensors, the Lagrange equations can have several stable solutions. A descent algorithm, like the higher-order power method, will only reveal the global optimum if the starting point is in the attraction region of the global optimum.



Considering the fact that the best rank-1 approximation of a matrix is obtained by truncating its SVD after the first singular value, and taking into account the ordering constraint in the HOSVD, it is clear that truncation of the HOSVD after the first $n$-mode singular value $(1 \leqslant n \leqslant N)$ may yield a good rank-1 approximation for tensors. An other important difference between matrices and higher-order tensors is that this approximation is not necessarily the globally optimal one (e.g. in Fig. 5.1 the estimate obtained by truncation of the HOSVD is indicated with a small circle). However, in our simulations we have observed that the HOSVD-estimate usually belongs to the attraction region of the best rank-1 approximation. Hence we propose to use the HOSVD-estimate as the starting value for the power iteration derived above; we assume that this iteration will not cause "jumps" to other attraction regions. However, there is not an absolute guarantee: we have been able to generate special cases in which monotonous descent techniques eventually lead to a local optimum with a close to optimal fit, but nevertheless different from the global optimum. With this respect, repeating the optimization procedure for several initial values could be envisaged.

The complete algorithm, for a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, is presented in Alg. 5.2.2.

**Example 5.2.9** Consider a tensor $\mathcal{B} \in \mathbb{R}^{2\times 2\times 2\times 2}$, of which the non-zero entries are given by:

$$\begin{cases} b_{1111} = 25.1, & b_{1212} = 25.6, \\ b_{2121} = 24.8, & b_{2222} = 23. \end{cases}$$

Due to the symmetries $b_{ijkl} = b_{kjil}$ and $b_{ijkl} = b_{ilkj}$, the best rank-1 approximation of $\mathcal{B}$ has the form $\lambda U^{(1)} \circ U^{(2)} \circ U^{(1)} \circ U^{(2)}$; we assume that $U^{(1)} = (\cos\alpha\ \sin\alpha)^T$ and $U^{(2)} = (\cos\beta\ \sin\beta)^T$. The function $\tilde{g}_1 \stackrel{\text{def}}{=} \mathcal{B} \times_1 U^{(1)^T} \times_2 U^{(2)^T} \times_3 U^{(1)^T} \times_4 U^{(2)^T}$ corresponds to a weighted arithmetic mean of the non-zero enties of $\mathcal{B}$:

$$\begin{aligned} \tilde{g}_1(\alpha, \beta) &= 25.1 \cos^2\alpha \cos^2\beta + 25.6 \cos^2\alpha \sin^2\beta \\ &\quad + 24.8 \sin^2\alpha \cos^2\beta + 23 \sin^2\alpha \sin^2\beta. \end{aligned}$$

The global maximum of this function is 25.6, obtained for $\alpha = 0$ and $\beta = \pi/2$; there are no essentially different local maxima. According to Theor. 5.2.2, setting $\lambda = 25.6$, $U^{(1)} = (1\ 0)^T$ and $U^{(2)} = (0\ 1)^T$, gives the best rank-1 approximation of $\mathcal{B}$. On the other hand, $\mathcal{B}$ is all-orthogonal and ordered; truncation of the HOSVD yields the vectors $U^{(1)} = (1\ 0)^T$ and $U^{(2)} = (1\ 0)^T$, which belong to the correct attraction region in a trivial way.
Now consider a tensor $\mathcal{A} \in \mathbb{R}^{2\times 2\times 2\times 2}$, equal to $\mathcal{B}$, except for the entries

$$a_{1121} = 0.3, \quad a_{2111} = 0.3.$$

$\mathcal{A}$ has similar symmetries as $\mathcal{B}$. The function $\tilde{g}_2 \stackrel{\text{def}}{=} \mathcal{A} \times_1 U^{(1)^T} \times_2 U^{(2)^T} \times_3 U^{(1)^T} \times_4 U^{(2)^T}$, given by

$$\tilde{g}_2(\alpha, \beta) = \tilde{g}_1(\alpha, \beta) + 0.6 \cos^2 \beta \cos \alpha \sin \alpha, \tag{5.25}$$

is visualized in Fig. 5.2 with a mesh and a contour plot. The extra term in Eq. (5.25) is small enough such that $\tilde{g}_1$ and $\tilde{g}_2$ are similar - in particular the global maximum is the same - but it induces a local maximum (namely $(0.5536, 0)$) on the axis $\beta = 0$. On the contour plot, the critical points of $\tilde{g}_2$, i.e. where the gradient is zero, are indicated with "+"-marks; the global maximum is indicated with a small circle. $\mathcal{A}$ is still nearly all-orthogonal, with the same ordering as $\mathcal{B}$. Truncation of the HOSVD leads to vectors that are still close to $(1\ 0)^T$; this corresponds to the "x"-mark on the contour plot. Because of the symmetry of $\tilde{g}_2$ around the axis $\beta = 0$, a gradient ascent starting from the first HOSVD-components will not converge to the global maximum but to the local optimum $(0.5536, 0)$. The same holds for the higher-order power algorithm: after each iteration the intermediate result was plotted as a dot in the contour plot of Fig. 5.2. Altering the ordering of the substeps in Alg. 5.2.2 yields comparable results. (Starting from the HOSVD-guess, the global maximum could be found however by alternating, like indicated in Remark 5.2.6, between simultaneous updates of the mode-1 and mode-3 vector on one hand, and on the other hand the mode-2 and mode-4 vector, beginning with the former.) It is clear that this example is not an isolated case: the conditions that ensure that the HOSVD-guess is in the wrong attraction region, are still satisfied for sufficiently small perturbations of $\mathcal{A}$. □

**Example 5.2.10** In accordance with Remark 5.2.8, the super-symmetric solutions to the Lagrange equations for a super-symmetric tensor $\mathcal{A} \in \mathbb{R}^{2\times2\times2\times2}$, are given by

$$\begin{aligned} a_{1222}t^4 + (3a_{1122} - a_{2222})t^3 + 3(a_{1112} - a_{1222})t^2 + \\ (a_{1111} - 3a_{1122})t - a_{1112} = 0, \end{aligned} \tag{5.26}$$

in which $t = \sin\alpha / \cos\alpha$, with $U = (\cos\alpha\ \ \sin\alpha)^T$. The extra solutions to Eqs. (5.23 - 5.24) are given by

$$(a_{1122} + a_{2222})t^2 + 2(a_{1112} + a_{1222})t + a_{1111} + a_{1122} = 0. \tag{5.27}$$

Let the roots of Eq. (5.26) be given by $t_1$, $t_2$, $t_3$, $t_4$, and the roots of Eq. (5.27) by $t'_1$, $t'_2$. The relationships between the coefficients in Eq. (5.26) imply that

$$t_4(t_1 + t_2 + t_3 + 3t_1t_2t_3) + t_1t_2 + t_1t_3 + t_2t_3 + 3 = 0.$$

As a limited experiment to investigate the position of the HOEVD-truncate for real super-symmetric $(2 \times 2 \times 2 \times 2)$-tensors, we consider the cases in which Eq. (5.26) has 4 distinct real roots, with $t_1$, $t_2$, $t_3$ ($|t_1| > |t_2| > |t_3|$) integers



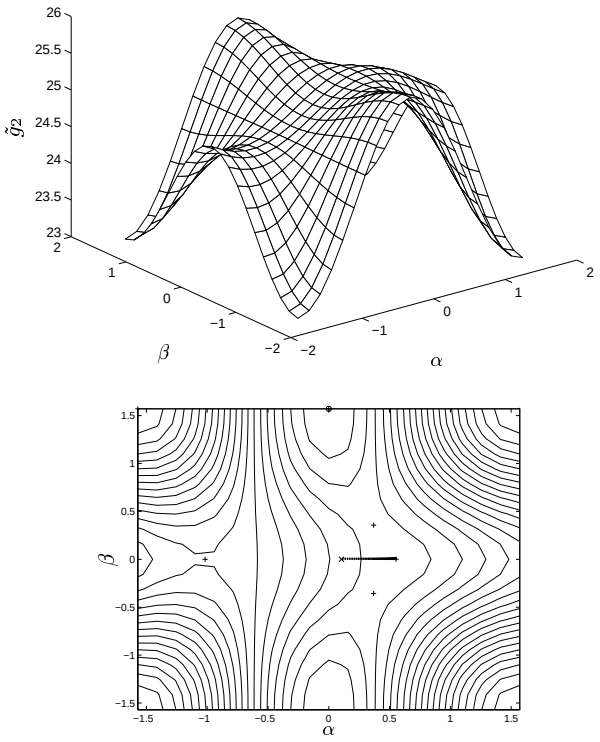



Figure 5.2: Visualization of the quality of the best rank-1 approximation of the partially symmetric tensor $\mathcal{A}$ in Ex. 5.2.9, for different values of $\alpha$ and $\beta$. The "x"-mark shows the initial guess obtained by HOSVD. The global optimum is indicated with an "o"-mark. The "+"-marks correspond to the critical points of the function $\tilde{g}_2$, defined in the example. The intermediate results after each iteration cycle of Alg. 5.2.2, initialized with the first HOSVD-components, are plotted as a sequence of dots.

between -10 and 10, and with $|t'_1 + t'_2|$ equal to an integer between 1 and 20; we do not consider the cases where $t_1 + t_2 + t_3 + 3t_1t_2t_3 = 0$ or $t'_1 + t'_2 = 0$, as they are most often incompatible with Eqs. (5.26-5.27); in addition, we assume that $a_{1222} = 1$. In more than 99.9 % of these cases, the HOEVD-truncate is in the same valley of the function $g$ in Theor. 5.2.2 as the global optimum ($g$ is considered as a function in $\alpha$, for $U^{(1)} = \ldots = U^{(4)} = U$). In the remaining

cases, $g$ shows two peaks with a comparable amplitude (we call the amplitudes $g_1$ and $g_2$, with $g_1 > g_2$); the value $(g_1 - g_2)/g_1 = 0.08$ on the average. The tensor defined by

$$\begin{cases} a_{1111} = -3.1936, & a_{1112} = 2.1053, \quad a_{1122} = 1.4267, \\ a_{1222} = 1.0000, & a_{2222} = -0.9831. \end{cases}$$

is a typical example. In Fig. 5.3 we plot $g$ as a function of $\alpha$. The critical points where the derivative of $g$ vanishes, are indicated with a "+"-mark. The HOEVD-truncate corresponds to $\alpha = 0.3033$, immediately right to the root $\alpha = 0.2784$, and has been indicated with an "x"-mark. The global maximum is reached for $\alpha = -0.2573$ ("o"-mark). □

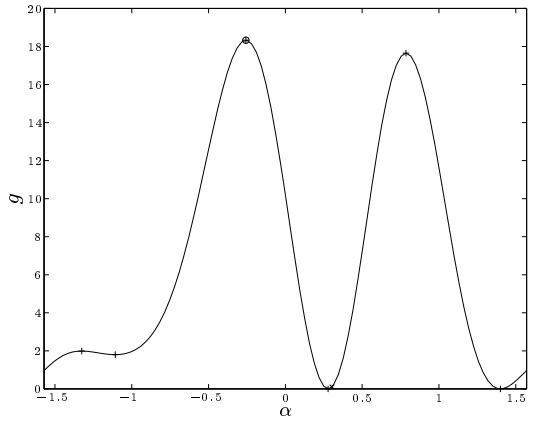



Figure 5.3: Visualization of the quality of the best rank-1 approximation of the super-symmetric tensor $\mathcal{A}$ in Ex. 5.2.10, for different values of $\alpha$. The "x"-mark shows the initial guess obtained by HOEVD. The global optimum is indicated with an "o"-mark. The "+"-marks correspond to the critical points of the function $g$.

The complete algorithm, for a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, is presented in Alg. 5.2.2.

### 5.2.4 Gradient expressions

Instead of minimizing the cost function $f$ in the valley of the global optimum by an ALS technique such as the Higher-Order Power Method, it is also possible to use other descent algorithms. With this respect, we give an explicit expression for the gradient of $f$.



Consider the derivative of the quadratic term in Eq. (5.5), corresponding to the value of $f$. This derivative can be written as follows:

$$\begin{aligned} \nabla_{U^{(n)}} f &= -2\lambda \left(\mathcal{A} \times_1 {U^{(1)}}^T \ldots \times_{n-1} {U^{(n-1)}}^T \right. \\ &\quad \left. \times_{n+1} {U^{(n+1)}}^T \ldots \times_N {U^{(N)}}^T - \lambda U^{(n)}\right), \end{aligned} \tag{5.28}$$

in which we assumed that the vectors $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$ have unit norm. If we claim that $\lambda$ takes its optimal value, given by Eq. (5.13), then the derivative is orthogonal to $U^{(n)}$. Hence it corresponds to the gradient of $f$, with respect to $U^{(n)}$, over the unit sphere. The gradient of $g$ is given by the same expression, with an inverted sign.

For a super-symmetric tensor $\mathcal{A}$ the derivative of $f$ has $N$ terms of the form of Eq. (5.28), since $U^{(1)}, U^{(2)}, \ldots, U^{(N)} = U$. Because of the symmetry, all these terms are equal. Hence the gradient takes the following form:

$$\nabla_U f = -2N\lambda \left(\mathcal{A} \times_2 U^T \times_3 U^T \ldots \times_N U^T - \lambda U\right), \tag{5.29}$$

in which $\lambda$ is given by the optimal value $\lambda = \mathcal{A} \times_1 U^T \times_2 U^T \ldots \times_N U^T$.

## 5.3 Higher-order orthogonal iteration

In this section we generalize the best rank-1 approximation problem of the previous section, in the sense that the approximation should now have pre-specified mode-1, mode-2, ... ranks. Such an approximate tensor will briefly be called a rank-$(R_1, R_2, \ldots, R_N)$ tensor. The derivation of the computational procedure follows a similar scheme as in the previous section. In Subsect. 5.3.1, we give two related formal definitions of the approximation problem. Subsect. 5.3.2 is devoted to the actual computation of the solution; the resulting algorithm can be interpreted as a higher-order generalization of the orthogonal iteration method. In Subsect. 5.3.3, we give an explicit expression for the gradient of the least-squares cost function.

It was Tucker who noticed that truncation of his model does not necessarily provide an estimate with an optimal least-squares fit [260]. Later on, Kroonenberg, de Leeuw and ten Berghe [161, 162, 163, 254] devised an ALS method to improve the fit, known as three-mode factor analysis in Psychometrics (the generalization to orders higher than three is briefly indicated in [163]). In this section, we will present a square-root version of the algorithm, valid for tensors of arbitrary order, and interpret the result as a multilinear generalization of the technique of orthogonal iterations; the derivation in Subsect. 5.3.1 and 5.3.2 follows roughly the same lines as in the work by Kroonenberg. We propose to base

the computation scheme on the calculation of dominant subspaces, rather than individual singular vectors, resulting in a higher efficiency; on the other hand, we explain that the efficiency can further be increased by updating two modes at the same time. We pay special attention to the case of super-symmetric higher-order tensors. An important remark concerns the initialization of the algorithm. Kroonenberg, de Leeuw and ten Berghe initialized their algorithm with a truncated Tucker model, but is was indicated that only local optimization was guaranteed. In Subsect. 5.2.3 we have demonstrated that initializing optimization routines with a truncated HOSVD does not always lead to the global optimum, but it is our experience that defective cases are rarely met. Finally, in Subsect. 5.3.3 we give explicit expressions of the gradient of the cost function over the Stiefel manifold.

### 5.3.1 Best rank-$(R_1, R_2, \ldots, R_N)$ approximation

In this section we generalize the best rank-1 approximation problem of the previous section to the best approximation by a tensor with pre-specified mode-1, mode-2, ... ranks. Formally, the problem we want to solve, can be formulated as follows:

*Given a real $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, find a tensor $\hat{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, with $\texttt{rank}_1(\hat{\mathcal{A}}) = R_1$, $\texttt{rank}_2(\hat{\mathcal{A}}) = R_2$, ..., $\texttt{rank}_N(\hat{\mathcal{A}}) = R_N$, that minimizes the least-squares cost function*

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 \tag{5.30}$$

The $n$-rank conditions imply that $\hat{\mathcal{A}}$ can be decomposed as

$$\hat{\mathcal{A}} = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}, \tag{5.31}$$

in which $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times R_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{I_2 \times R_2}$, ..., $\mathbf{U}^{(N)} \in \mathbb{R}^{I_N \times R_N}$ each have orthonormal columns and $\mathcal{B} \in \mathbb{R}^{R_1 \times R_2 \times \ldots \times R_N}$.

Actually it is sufficient to determine the matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, ..., $\mathbf{U}^{(N)}$ for the optimization of $f$. For any estimate of these matrices, the optimal tensor $\mathcal{B}$ is given by the following theorem:

**Theorem 5.3.1** *Assume a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ and consider a tensor $\hat{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ as in Eq. (5.31). For given matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, ..., $\mathbf{U}^{(N)}$, the tensor $\mathcal{B}$ that optimizes $f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2$, is given by*

$$\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}. \tag{5.32}$$



**Proof:** The optimization of

$$f = \|\mathcal{A} - \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}\|^2$$

for the unknown elements of $\mathcal{B}$, is merely a classical linear least-squares problem. It corresponds to the least-squares estimation of the solution of the following set of linear equations, which is possibly overdetermined:

$$\mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)} = \mathcal{A}.$$

Taking into account that $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, ..., $\mathbf{U}^{(N)}$ have orthonormal columns, these matrices can be brought to the other side of the equation by ($n$-mode) multiplication with the transposed matrices. □

**Remark 5.3.2** Theor. 5.3.1 is the multidimensional equivalent of the optimal choice of $\lambda$ in Eq. (5.13). ◇

In analogy with Theor. 5.2.2 we state that the best rank-$(R_1, R_2, \ldots, R_N)$ approximation problem can also be formulated as follows:

**Theorem 5.3.3** *Assume a real $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, then the minimization of the cost function of Eq. (5.30) is equivalent to the maximization, over the matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ with orthonormal columns, of the function*

$$g(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}) = \|\mathcal{A} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}\|^2 \tag{5.33}$$

*If the tensor $\mathcal{B}$, of which the Frobenius-norm equals the Frobenius-norm of $\hat{\mathcal{A}}$ in Eq. (5.31), is chosen in accordance with Eq. (5.32), then the functions of Eq. (5.30) and Eq. (5.33) are related by:*

$$f = \|\mathcal{A}\|^2 - g \tag{5.34}$$

(The Stiefel manifold of real $(I \times R)$-matrices with mutually orthonormal columns will further be symbolized as $\mathrm{St}(R, I)$; see also App. A.)

**Proof:** We look for the relations between the definition of $g$ and the terms in the following expression for $f$:

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 = \|\mathcal{A}\|^2 - 2\langle \mathcal{A}, \hat{\mathcal{A}} \rangle + \|\hat{\mathcal{A}}\|^2.$$

First, the definition of the inner product allows to write:

$$\langle \mathcal{A}, \hat{\mathcal{A}} \rangle \quad = \quad \langle \mathcal{A}, \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)} \rangle$$

$$\begin{aligned} &= \langle \mathcal{A} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \ldots \times_N \mathbf{U}^{(N)^T}, \mathcal{B} \rangle \\ &= \|\mathcal{B}\|^2. \end{aligned}$$

Next, since $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ have orthonormal columns, they do not affect the Frobenius-norm:

$$\|\hat{\mathcal{A}}\|^2 = \|\mathcal{B}\|^2.$$

Substitution of the preceding expressions in the equation for $f$, yields:

$$f(\hat{\mathcal{A}}) = \|\mathcal{A}\|^2 - \|\mathcal{B}\|^2.$$

Combination with the definition of $g$ proves the theorem. □

**Remark 5.3.4** Basically, the best rank-$(R_1, R_2, \ldots, R_N)$ approximation problem consists of the determination of a reduced-$n$-rank tensor $\hat{\mathcal{A}}$ that explains as much of the "energy" of a given tensor $\mathcal{A}$ as possible under the given $n$-rank constraints. Theor. 5.3.3 implies that this problem is equivalent to the explicit maximization of the energy of the approximation. ◇

**Remark 5.3.5** The optimal choice of $\mathcal{B}$, given by Eq. (5.32), implies mutual orthogonality of $\hat{\mathcal{A}}$ and the error term $\mathcal{A} - \hat{\mathcal{A}}$: $\langle \hat{\mathcal{A}}, \mathcal{A} - \hat{\mathcal{A}} \rangle = 0$, as indicated in the proof of the theorem. ◇

### 5.3.2 An orthogonal iteration

For the actual computation of the best rank-$(R_1, R_2, \ldots, R_N)$ approximation of $\mathcal{A}$, let us interpret the function $g$ in the following way. Imagine that the matrices $\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(n-1)}, \mathbf{U}^{(n+1)}, \ldots, \mathbf{U}^{(N)}$ are fixed and that $g$ in Eq. (5.33) is merely a quadratic expression in the components of the unknown matrix $\mathbf{U}^{(n)}$ with orthonormal columns. We have:

$$g = \|\mathbf{U}^{(n)^T} \cdot \tilde{\mathbf{U}}^{(n)}\|^2, \tag{5.35}$$

in which

$$\tilde{\mathbf{U}}^{(n)} \stackrel{\text{def}}{=} \mathcal{A} \times_1 \mathbf{U}^{(1)^T} \ldots \times_{n-1} \mathbf{U}^{(n-1)^T} \times_{n+1} \mathbf{U}^{(n+1)^T} \ldots \times_N \mathbf{U}^{(N)^T}. \tag{5.36}$$

Hence the columns of $\mathbf{U}^{(n)}$ can be found as an orthonormal basis for the dominant subspace of the column space of $\tilde{\mathbf{U}}^{(n)}$.

Repeating this procedure for different mode numbers leads to an ALS algorithm for the (local) maximization of $f(\hat{\mathcal{A}})$: in each step the estimate of one of the matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ is optimized, while the other matrix estimates



**Algorithm 5.3.2**
*Higher-Order Orthogonal Iteration*

1. Initial values: $\mathbf{U}_0^{(n)} \in \mathrm{St}(R_n, I_n)$ of which the columns form a basis for the dominant $R_n$-dimensional left singular subspace of $\mathbf{A}_{(n)}$ $(2 \leqslant n \leqslant N)$; or repeat the algorithm for several initial values.

2. Iterate until convergence:

   - $\tilde{\mathbf{U}}_{k+1}^{(1)} = \mathcal{A} \times_2 \mathbf{U}_k^{(2)^T} \times_3 \mathbf{U}_k^{(3)^T} \ldots \times_N \mathbf{U}_k^{(N)^T}$;
     Maximize over $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times R_1}$ with $\mathbf{U}^{(1)^T}\mathbf{U}^{(1)} = \mathbf{I}$:
     $h(\mathbf{U}^{(1)}) = \|\mathbf{U}^{(1)^T} \cdot \tilde{\mathbf{U}}_{k+1}^{(1)}\|$;
     $\mathtt{max}(h(\mathbf{U}^{(1)})) = h(\mathbf{U}_{\max}^{(1)}) = \lambda_{k+1}^{(1)}$;
     $\mathbf{U}_{k+1}^{(1)} = \mathbf{U}_{\max}^{(1)}$;
   - $\tilde{\mathbf{U}}_{k+1}^{(2)} = \mathcal{A} \times_1 \mathbf{U}_{k+1}^{(1)^T} \times_3 \mathbf{U}_k^{(3)^T} \ldots \times_N \mathbf{U}_k^{(N)^T}$;
     Maximize over $\mathbf{U}^{(2)} \in \mathbb{R}^{I_2 \times R_2}$ with $\mathbf{U}^{(2)^T}\mathbf{U}^{(2)} = \mathbf{I}$:
     $h(\mathbf{U}^{(2)}) = \|\mathbf{U}^{(2)^T} \cdot \tilde{\mathbf{U}}_{k+1}^{(2)}\|$;
     $\mathtt{max}(h(\mathbf{U}^{(2)})) = h(\mathbf{U}_{\max}^{(2)}) = \lambda_{k+1}^{(2)}$;
     $\mathbf{U}_{k+1}^{(2)} = \mathbf{U}_{\max}^{(2)}$;
   - ...
   - $\tilde{\mathbf{U}}_{k+1}^{(N)} = \mathcal{A} \times_1 \mathbf{U}_{k+1}^{(1)^T} \times_2 \mathbf{U}_{k+1}^{(2)^T} \ldots \times_{N-1} \mathbf{U}_{k+1}^{(N-1)^T}$;
     Maximize over $\mathbf{U}^{(N)} \in \mathbb{R}^{I_N \times R_N}$ with $\mathbf{U}^{(N)^T}\mathbf{U}^{(N)} = \mathbf{I}$:
     $h(\mathbf{U}^{(N)}) = \|\mathbf{U}^{(N)^T} \cdot \tilde{\mathbf{U}}_{k+1}^{(N)}\|$;
     $\mathtt{max}(h(\mathbf{U}^{(N)})) = h(\mathbf{U}_{\max}^{(N)}) = \lambda_{k+1}^{(N)}$;
     $\mathbf{U}_{k+1}^{(N)} = \mathbf{U}_{\max}^{(N)}$.

are kept constant. The result is shown in step 2 of Alg. 5.3.2. Note that Eq. (5.36) can be written in matrix format as follows:

$$\tilde{\mathbf{U}}^{(n)} = \mathbf{A}_{(n)} \cdot (\mathbf{U}^{(1)} \otimes \ldots \otimes \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \otimes \ldots \otimes \mathbf{U}^{(N)}).$$

Clearly this technique is a higher-order extension of the orthogonal iteration for matrices, described in Sect. 5.1. A major difference, in some sense also with the higher-order power method, is that each iteration step not only involves the computation of a multilinear transformation, but also the computation of a dominant subspace. Consequently, like for matrices, a numerically reliable basis for this subspace is obtained by orthonormalization of an original set of basis vectors.

As far as super-symmetric higher-order tensors are concerned, higher-order

orthogonal iterations - like higher-order power iterations - yield unsymmetric intermediate results.

Like for higher-order power iterations, it makes sense to initialize the higher-order orthogonal iteration with the rank-$(R_1, R_2, \ldots, R_N)$-estimate obtained by truncation of the HOSVD of $\mathcal{A}$. We refer to Subsect. 5.2.3.

The complete higher-order orthogonal iteration algorithm, for a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, is presented in Alg. 5.3.2.

With respect to Alg. 5.3.2, we remark:

- It is not required that the columns of $\mathbf{U}_k^{(n)}$ ($k \geqslant 0$ and $1 \leqslant n \leqslant N$) coincide with the left singular vectors of the matrix from which they are derived. It is sufficient to compute an arbitrary orthonormal basis for the $R_n$-dimensional dominant subspace.
- The projection of $\mathbf{U}_k^{(n)}$ on the column space of $\tilde{\mathbf{U}}_{k+1}^{(n)}$ gives a first estimate of the dominant $R_n$-dimensional subspace, in the calculation of $\mathbf{U}_{k+1}^{(n)}$. For a discussion of fast subspace computation methods, we refer to [71].
- The stop criterion used in the algorithm for the computation of a dominant subspace, can be based on the value of $h(\mathbf{U}^{(n)})$, which corresponds with the square root of the function $g$ that has to be optimized (see Eq. (5.35)). In terms of the accuracy of the components, the termination criterion can take the form of e.g. $\|\mathbf{U}_{k+1}^{(n)^T} \cdot \mathbf{U}_k^{(n)}\|^2 > (1-\epsilon)R_n$ $(1 \leqslant n \leqslant N)$.

**Remark 5.3.6** In analogy with Remark 5.2.6, it is possible to update the estimates of e.g. $\mathbf{U}^{(n)}$ and $\mathbf{U}^{(n+1)}$ simultaneously, as the dominant left and right singular matrices of the matrix $\mathcal{A} \times_1 \mathbf{U}^{(1)^T} \ldots \times_{n-1} \mathbf{U}^{(n-1)^T} \times_{n+2} \mathbf{U}^{(n+2)^T} \ldots \times_N \mathbf{U}^{(N)^T}$. Note however that, if $R_n > R_{n+1}$, only $R_{n+1}$ columns of $\mathbf{U}^{(n)}$ can be estimated in this way (let us assume that these vectors are the columns of a matrix $\mathbf{V} \in \mathrm{St}(R_{n+1}, I_n)$). However, given an estimate $\mathbf{U}_k^{(n)} \in \mathrm{St}(R_n, I_n)$, a new estimate $\mathbf{U}_{k+1}^{(n)} \in \mathrm{St}(R_n, I_n)$ can be obtained by adding $R_n - R_{n+1}$ columns that are mutually orthonormal, as well as orthogonal to the columns of $\mathbf{V}$, such that the column spaces of $\mathbf{U}_k^{(n)}$ and $\mathbf{U}_{k+1}^{(n)}$ are as close as possible; as a criterion, we consider the sum of squared cosines of the principal angles between both subspaces [269]. The extra columns can be chosen in the column space of $\mathbf{U}_k^{(n)}$ itself; they can be found as the columns of $\mathbf{U}_k^{(n)} \cdot \mathbf{W}$, in which the columns of $\mathbf{W} \in \mathrm{St}(R_n - R_{n+1}, R_n)$ form an orthonormal basis for the kernel of $\mathbf{V}_1^T \cdot \mathbf{U}_k^{(n)}$. $\diamond$

**Example 5.3.7** Consider a tensor $\mathcal{A} \in \mathbb{R}^{3 \times 2 \times 2}$, defined by the following matrix unfolding:



$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} 0 & -1 & 1 & 4 \\ 2 & -2 & 3 & -5 \\ 4 & 3 & 5 & -6 \end{array}\right)$$

The HOSVD-components of $\mathcal{A}$ are given by:

$$\mathbf{U}^{(1)} = \begin{pmatrix} -0.2465 & -0.4993 & 0.8306 \\ 0.5217 & 0.6539 & 0.5479 \\ 0.8167 & -0.5684 & -0.0994 \end{pmatrix}, \quad \mathbf{U}^{(2)} = \begin{pmatrix} 0.1715 & 0.9852 \\ 0.9852 & -0.1715 \end{pmatrix},$$

$$\mathbf{U}^{(3)} = \begin{pmatrix} 0.5105 & 0.8599 \\ -0.8599 & 0.5105 \end{pmatrix},$$

$$\mathbf{S}_{(1)} = \left(\begin{array}{cc|cc} 10.0470 & 1.0849 & -0.9583 & 4.4300 \\ 1.1651 & -2.4648 & 1.4908 & -1.7164 \\ 0.3847 & 2.1766 & 2.2365 & -0.9218 \end{array}\right).$$

The higher-order singular values equal:
mode 1: 11.0753 , 3.5498 and 3.2768,
mode 2: 10.6975 and 5.6181,
mode 3: 10.5162, 5.9506 and 0.
Truncation of the HOSVD after the first term in each mode, gives a rank-1 approximation of $\mathcal{A}$ with Frobenius-norm 10.0470. This value is increased to $\lambda = 10.1693$ by a higher-order power iteration (Alg. 5.2.2). The evolution of the norm during the iteration is given by the solid line in Fig. 5.4. The best rank-1 approximation $\lambda\, V^{(1)} \circ V^{(2)} \circ V^{(3)}$ is further given by:

$$V^{(1)} = \begin{pmatrix} -0.2515 \\ 0.6035 \\ 0.7567 \end{pmatrix}, \qquad V^{(2)} = \begin{pmatrix} 0.1344 \\ 0.9909 \end{pmatrix}, \qquad V^{(3)} = \begin{pmatrix} 0.5765 \\ -0.8171 \end{pmatrix}.$$

The HOSVD of the best rank-$(2,2,1)$ approximation of $\mathcal{A}$, represented by $\mathcal{B} \times_1 \mathbf{W}^{(1)} \times_2 \mathbf{W}^{(2)} \times_3 W^{(3)}$, is given by:

$$\mathbf{W}^{(1)} = \begin{pmatrix} -0.2789 & -0.4141 \\ 0.5984 & -0.7806 \\ 0.7511 & 0.4681 \end{pmatrix}, \qquad \mathbf{B}_{(1)} = \left(\begin{array}{c|c} 10.1473 & 0.0000 \\ 0.0000 & 2.7607 \end{array}\right),$$

$$\mathbf{W}^{(2)} = \begin{pmatrix} 0.0982 & -0.9952 \\ 0.9952 & 0.0982 \end{pmatrix}, \qquad W^{(3)} = \begin{pmatrix} 0.5105 \\ -0.8599 \end{pmatrix}.$$

The Frobenius-norm of this approximation equals 10.5162. The evolution of the norm during the higher-order orthogonal iteration (Alg. 5.3.2) is given by the dashed line in Fig. 5.4.

$\square$

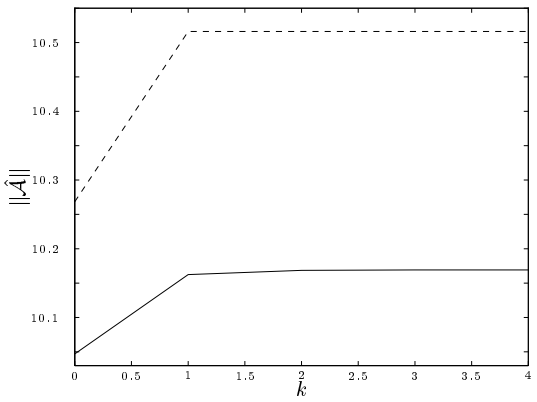



Figure 5.4: Evolution of the norm of the approximation $\hat{\mathcal{A}}$ of the tensor $\mathcal{A}$ in Ex. 5.3.7, during the higher-order power iteration or higher-order orthogonal iteration, as a function of the iteration step $k$. Solid: computation of the best rank-1 approximation. Dashed: computation of the best rank-$(2,2,1)$ approximation.

### 5.3.3 Gradient expressions

In section 5.2.4 we derived an expression for the gradient of the best rank-1 approximation cost function. In this section we will derive the equivalent formula for the best rank-$(R_1, R_2, \ldots, R_N)$ cost function $g$ (Eq. (5.35)). Both the unsymmetric and the super-symmetric case are addressed. The result can be used for continuous-time optimization of $g$ [77], or, after integration, in discrete-time implementations. With respect to higher-order orthogonal iterations, it should be noted that gradient-based descent is more sensitive to the specific shape of the cost function than ALS descent.

We start with the case in which $\mathcal{A}$ is unsymmetric:

**Theorem 5.3.8** *Consider the function $g$, defined by Theor. 5.3.3. The gradient of $g$ w.r.t. $\mathbf{U}^{(n)}$, over the Stiefel manifold $St(R_n, I_n)$, takes the following form:*

$$\nabla_{\mathbf{U}^{(n)}} g = 2\left(\mathbf{I} - \mathbf{U}^{(n)}{\mathbf{U}^{(n)}}^T\right) \cdot \tilde{\mathbf{U}}^{(n)} \cdot {\tilde{\mathbf{U}}^{(n)}}^T \cdot \mathbf{U}^{(n)}, \tag{5.37}$$

*in which*

$$\tilde{\mathbf{U}}^{(n)} = \mathcal{A} \times_1 {\mathbf{U}^{(1)}}^T \ldots \times_{n-1} {\mathbf{U}^{(n-1)}}^T \times_{n+1} {\mathbf{U}^{(n+1)}}^T \ldots \times_N {\mathbf{U}^{(N)}}^T. \tag{5.38}$$



**Proof:** From the interpretation of $g$ in Eq. (5.35), we have:

$$g = \texttt{trace}({\mathbf{U}^{(n)}}^T \cdot \tilde{\mathbf{C}}^{(n)} \cdot \mathbf{U}^{(n)}),$$

in which $\tilde{\mathbf{C}}^{(n)} \stackrel{\text{def}}{=} \tilde{\mathbf{U}}^{(n)} \cdot {\tilde{\mathbf{U}}^{(n)}}^T$. For this representation of $g$, the gradient over the Stiefel manifold is e.g. derived in [77], p. 106:

$$\nabla_{\mathbf{U}^{(n)}} g = 2\left(\mathbf{I} - \mathbf{U}^{(n)}{\mathbf{U}^{(n)}}^T\right) \cdot \tilde{\mathbf{C}}^{(n)} \cdot \mathbf{U}^{(n)}.$$

□

In the case where $\mathcal{A}$ is super-symmetric, the gradient has a similar form, since it can be considered as a superposition of terms like Eq. (5.37). We state:

**Theorem 5.3.9** *For a real super-symmetric $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I \times I \times \ldots I}$, consider the following function $g$ of $\mathbf{U} \in St(R, I)$:*

$$g(\mathbf{U}) = \|\mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \ldots \times_N \mathbf{U}^T\|^2. \tag{5.39}$$

*The gradient of $g$, over the Stiefel manifold, takes the following form:*

$$\nabla_{\mathbf{U}} g = 2N\left(\mathbf{I} - \mathbf{U}\mathbf{U}^T\right) \cdot \tilde{\mathbf{U}} \cdot \tilde{\mathbf{U}}^T \cdot \mathbf{U}, \tag{5.40}$$

*in which*

$$\tilde{\mathbf{U}} = \mathcal{A} \times_2 \mathbf{U}^T \times_3 \mathbf{U}^T \ldots \times_N \mathbf{U}^T. \tag{5.41}$$

**Proof:** Because of the super-symmetry, the gradient consists of $N$ equal terms, given in Eq. 5.37. App. C contains a proof in which the effect of infinitesimal changes of $\mathbf{U}$ on the value of $g$ is investigated, and expressed in terms of the gradient of $g$ w.r.t. $\mathbf{U}$. □

## 5.4 Higher-Order Hessenberg Decomposition

Like for matrices, the computational cost of the higher-order power and orthogonal iterations can be decreased by first zeroing a number of elements of the tensor under consideration. Therefore we introduce a higher-order equivalent of the Hessenberg Decomposition in this section, and we explain how it can be computed by Householder reduction. For super-symmetric tensors, the Householder reduction leads to a higher-order equivalent of tridiagonalization.

First we generalize the definition of upper Hessenberg matrices to higher-order tensors:

**Definition 5.4.1 (Higher-order upper Hessenberg form)**
*A tensor $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is in $n$-mode upper Hessenberg form iff its matrix unfolding $\mathbf{H}_{(n)}$ is upper Hessenberg. $\mathcal{H}$ is full upper Hessenberg iff all the matrix unfoldings $\mathbf{H}_{(n)}$ $(1 \leqslant n \leqslant N)$ are upper Hessenberg.*

The full upper Hessenberg form is visualized in Fig. 5.5, for a $(5 \times 5 \times 5)$-tensor. In this figure, the filled parts contain zero entries. For matrices, the full upper Hessenberg form is a tridiagonal matrix.

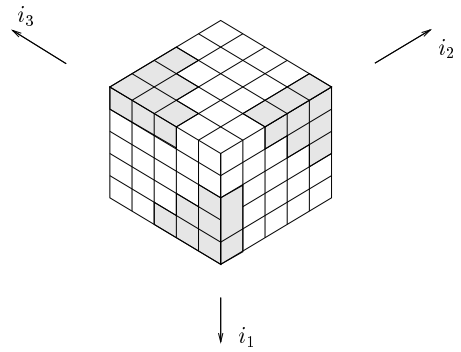



Figure 5.5: Visualization of the full upper Hessenberg form for a $(5 \times 5 \times 5)$-tensor. The filled tensor parts, which belong to the slices for which the index $i_1 = 1$, $i_2 = 1$ or $i_3 = 1$, consist of zeros.

Now we generalize the definition of the Hessenberg Decomposition of matrices:

**Definition 5.4.2 (Higher-Order Hessenberg Decomposition)**
*The $n$-mode Hessenberg Decomposition of an $N$th order tensor $\mathcal{A} \in \mathbb{R}^{I \times I \times \ldots \times I}$ is defined as*

$$\mathcal{A} = \mathcal{H} \times_1 \mathbf{U} \times_2 \mathbf{U} \ldots \times_N \mathbf{U}, \tag{5.42}$$

*in which:*

- $\mathbf{U} = [U_1 U_2 \ldots U_I]$ *is an orthogonal $(I \times I)$-matrix,*
- $\mathcal{H} \in \mathbb{R}^{I \times I \times \ldots \times I}$ *is an $n$-mode upper Hessenberg tensor.*

*The full Hessenberg Decomposition of a real $N$th order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is defined as*

$$\mathcal{A} = \mathcal{H} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}, \tag{5.43}$$

*in which:*



- $\mathbf{U}^{(n)} = \left[U_1^{(n)} U_2^{(n)} \ldots U_{I_n}^{(n)}\right]$ *is an orthogonal $(I_n \times I_n)$-matrix,*
- $\mathcal{H} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is a full upper Hessenberg tensor.*

The $n$-mode Hessenberg Decomposition can be computed in a finite number of steps by Householder reduction, where $\mathbf{U}$ is found as a product of Householder matrices $\mathbf{P}_1 \cdot \ldots \cdot \mathbf{P}_{I-2}$. First, $\mathbf{P}_1$ is determined such that $n$-mode multiplication with $\mathcal{A}$ zeroes the elements $a_{1\ldots131\ldots1}$ to $a_{1\ldots1I1\ldots1}$ (the index $i_n$ is varied), while leaving the element $a_{1\ldots111\ldots1}$ unchanged. The latter condition implies that the zeroed elements are not refilled by the subsequent multiplication $\mathcal{A}' = \mathcal{A} \times_1 \mathbf{P}_1 \times_2 \mathbf{P}_1 \ldots \times_N \mathbf{P}_1$. Next, $\mathbf{P}_2$ is obtained by zeroing the elements $a_{1\ldots241\ldots1}$ to $a'_{1\ldots2I1\ldots1}$ (again, the index $i_n$ is varied), while the elements $a'_{1\ldots211\ldots1}$ and $a'_{1\ldots221\ldots1}$ are left unchanged. The latter condition now implies that the zeros obtained with $\mathbf{P}_1$ are not refilled by the $n$-mode product $\mathcal{A}' \times_n \mathbf{P}_2$; on the other hand, it also implies that multiplication with $\mathbf{P}_2$ over the other modes is safe, in terms of the zeros obtained so far. The $n$-mode Hessenberg Decomposition is finally obtained by repeating the procedure $I - 2$ times.

In matrix terms, this computation corresponds to the determination of an orthogonal matrix $\mathbf{U}$, such that

$$\mathbf{H}_{(n)} = \mathbf{U} \cdot \mathbf{A}_{(n)} \cdot (\mathbf{U} \otimes \mathbf{U} \otimes \ldots \otimes \mathbf{U})^T$$

is upper Hessenberg. This is realized by zeroing the subsequent columns below the subdiagonal by left multiplication with Householder matrices; right multiplication with the corresponding Kronecker products does not affect the structure.

The full Hessenberg Decomposition is found by computing all the $\mathbf{U}^{(n)}$ as a product of Householder matrices $\mathbf{P}_1^{(n)} \cdot \ldots \cdot \mathbf{P}_{I-2}^{(n)}$, in the analogous way. The zero-pattern realized in preceding steps is not overwritten if the Householder matrices are computed in the cyclic order $\mathbf{P}_1^{(1)}$, $\mathbf{P}_1^{(2)}$, $\ldots$, $\mathbf{P}_1^{(N)}$, $\mathbf{P}_2^{(1)}$, $\mathbf{P}_2^{(2)}$, $\ldots$, $\mathbf{P}_2^{(N)}$, $\ldots$ (Note that, if e.g. $I_n > I_{n-1} + 2$, after zeroing the entries $a_{1\ldots1I_{n-1}(I_{n-1}+2)1\ldots1}$ to $a_{1\ldots1I_{n-1}I_n1\ldots1}$, the procedure is continued with zeroing the entries $a_{1\ldots11(I_{n-1}+3)21\ldots1}$ to $a_{1\ldots11I_n21\ldots1}$.)

Super-symmetry of $\mathcal{A}$ implies that the $n$-mode Hessenberg Decomposition equals the full Hessenberg Decomposition, where $\mathbf{U}^{(1)} = \mathbf{U}^{(2)} = \ldots = \mathbf{U}^{(N)} = \mathbf{U}$. The full invariance under index permutations even leads to the following property:

**Corollary 5.4.3** *Assume a super-symmetric $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I \times I \times \ldots \times I}$ and the higher-order Hessenberg Decomposition of Eq. (5.42). Now define $N(N-1)/2$ matrices $\mathbf{H}^{(n_1 n_2)}$ by the element-wise equation*

$$h^{(n_1 n_2)}_{i_{n_1} i_{n_2}} = h_{1\ldots i_{n_1}\ldots 1\ldots i_{n_2}\ldots 1} \qquad (n_1 > n_2).$$

*These matrices are tridiagonal.*

The zero-pattern of $\mathcal{H}$ is visualized in Fig. 5.6 for a third-order tensor. The filled parts in this figure contain zero entries.

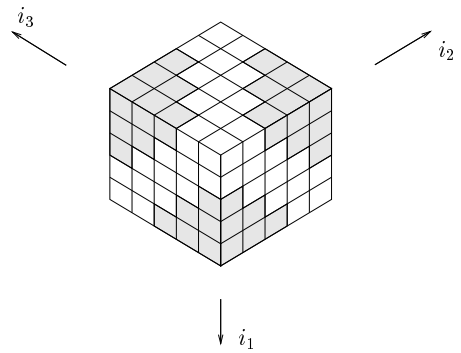



Figure 5.6: Visualization of the tridiagonal structure of a super-symmetric $(5 \times 5 \times 5)$-tensor, obtained by higher-order Hessenberg Decomposition. The filled tensor parts, which belong to the slices for which the index $i_1 = 1$, $i_2 = 1$ or $i_3 = 1$, consist of zeros.

The best rank-$(R_1, R_2, \ldots, R_N)$ approximation of a tensor $\mathcal{A}$ and its full upper Hessenberg form $\mathcal{H}$ are related by $n$-mode multiplication with $\mathbf{U}^{(n)}$ in Eq. (5.43) $(1 \leqslant n \leqslant N)$. This follows from repeated application of the following theorem:

**Theorem 5.4.4** *For a real $N$th-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$, consider the function $g$ defined by Eq. (5.33). Define the analogous function $\tilde{g}$ for the tensor $\tilde{\mathcal{A}} = \mathcal{A} \times_n \mathbf{Q}$, in which $\mathbf{Q}$ is orthogonal. Then we have:*

$$g(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}) = \tilde{g}(\mathbf{U}^{(1)}, \ldots, \mathbf{Q}\mathbf{U}^{(n)}, \ldots, \mathbf{U}^{(N)}). \tag{5.44}$$

**Proof:** The theorem is an immediate consequence of the fact that $\mathbf{Q}$ is orthogonal: the proof follows from substitution of

$$\mathcal{A} = (\mathcal{A} \times_n \mathbf{Q}) \times_n \mathbf{Q}^T$$

in the definition of $g$. □

Finally we remark that the *relative* speed-up of power or orthogonal iterations, caused by an a priori full Hessenberg Decomposition, becomes smaller for larger tensor order $N$: the matrices in which zeros are obtained form a relatively decreasing part of the tensor, as $N$ is increased. (In Figs. 5.5 and 5.6 these matrices are the matrices that are visible; in the second-order case the matrix itself is made tridiagonal. More in general, for an unsymmetric $N$th order $(I \times I \times \ldots \times I)$-tensor the number of entries that is zeroed, is given by $N(I-2)(I-1)/2$, on a total number of $I^N$ elements; if the tensor is super-symmetric, $N(N-1)(I-2)(I-1)/2$ zeros are obtained.) However, counting the degrees of freedom in the model of the (higher-order) Hessenberg Decomposition shows that this effect is inevitable.



## 5.5 Conclusion

There are some remarkable relations, but also some striking differences between the best rank-$R$ approximation of a matrix and the best rank-$(R_1, R_2, \ldots, R_N)$ approximation of an $N$th-order tensor:

- Unlike the matrix case, the least-squares cost function can show several local optima at the tensorial level.
- For matrices, the best rank-$R$ approximation is obtained by setting the smallest singular values equal to zero, while keeping the $R$ largest ones. Truncation of the HOSVD may yield a good tensor approximation, but this approximation is generically suboptimal. However, our simulations suggest that there is still a weak link: it is observed that for well-conditioned problems the HOSVD-estimate generally belongs to the valley of the least-squares cost function, corresponding to the global optimum.
- One way to obtain the dominant $R$-dimensional subspaces of row and column space of a matrix, is the power iteration $(R = 1)$ or orthogonal iteration $(R \neq 1)$. Higher-order generalizations of these techniques take the form of ALS algorithms, to enhance the fit.
- A higher-order power iteration step basically consists of a multi-linear mapping, instead of merely a linear transformation. In addition, higher-order orthogonal iterations can involve the estimation of a dominant subspace.

In psychometric literature, the least-squares approximation of a three-way dataset by a dataset with reduced core tensor, is known as three-mode factor analysis. In this chapter, we have further contributed to a linear/multilinear algebraic framework for the least-squares approximation of tensors of arbitrary order. In the square-root algorithm that we presented, an iteration step was based on a multilinear mapping, followed by the estimation of a dominant subspace; the technique was interpreted as a higher-order generalization of the method of orthogonal iterations. Aiming at a higher efficiency, we have

indicated how two modes can be updated at the same time. We have demonstrated that starting the iteration from a truncated HOSVD does not lead to the global optimum in all possible cases. We have paid special attention to the case of super-symmetric tensors, in view of applications in HOS. Explicit expressions for the gradient of the cost function over the Stiefel manifold have been provided. We have extensively discussed the best rank-1 approximation problem; this case will be important in Chapter 8. For super-symmetric $(2 \times 2 \times \ldots \times 2)$-tensors, the determination of the best rank-1 approximation, and the calculation of the super-symmetric stationary points and of the other stationary points of the higher-order power algorithm, have been reformulated as polynomial rooting problems. To improve the efficiency of higher-order power and orthogonal iterations, we proposed a preprocessing step, based on a higher-order generalization of the Hessenberg Decomposition.

# Chapter 6

# The Canonical Decomposition

*In Chapter 4 we have emphasized that the HOSVD is not necessarily rank-revealing. This is a drawback of unitary (orthogonal) tensor decompositions in general. In this chapter we will study the decomposition of a given tensor as a linear combination of a minimal number of, possibly non-orthogonal, rank-1 terms. This type of decomposition is often called "Canonical Decomposition" (CANDECOMP) or "Parallel Factors" model (PARAFAC). It is a multilinear generalization of diagonalizing a matrix by an equivalence or congruence transformation, however with thoroughly different properties, e.g. as far as uniqueness is concerned.*

*Sect. 6.1 is a brief introduction to the subject, with a formal definition of the CANDECOMP-concept and an overview of the main current computational techniques. In Sect. 6.2 we establish a computational link between the tensor decomposition and the simultaneous diagonalization of a set of matrices by equivalence or congruence; this problem might also be looked at as a simultaneous matrix EVD. The fact that the CANDECOMP usually involves non-orthogonal factor matrices is numerically disadvantageous. By reformulating the problem as a simultaneous generalized Schur decomposition, the unknowns are restricted to the manifold of orthogonal matrices in Sect. 6.3. The actual computation of the simultaneous generalized Schur decomposition is considered in Sect. 6.4. Complementarily to Sect. 6.3, Sect. 6.5 contains a note on the reconstruction of the CANDECOMP factors from the components of the simultaneous generalized Schur decomposition, if the symmetry is such that at least two modes can mutually be interchanged; this important case e.g. applies to the CANDECOMP-application of Chapter 10. Computation of the simultaneous generalized Schur decomposition is in general only equivalent to least-squares*

*fitting of the CANDECOMP-model if that model is exactly valid; Sect. 6.6 discusses least-squares matching techniques. The algorithms of this chapter are illustrated with some simulations in Sect. 6.7.*

## 6.1 Introduction

The "Canonical Decomposition" (CANDECOMP), or "Parallel Factors" model (PARAFAC), is defined as follows:

**Definition 6.1.1 (CANDECOMP)** *A Canonical Decomposition or Parallel Factors Decomposition of a tensor $\mathcal{A} \in \mathbb{C}^{(I_1 \times I_2 \times \ldots \times I_N)}$ is a decomposition of $\mathcal{A}$ as a linear combination of a minimal number of rank-1 terms:*

$$\mathcal{A} = \sum_{r}^{R} \lambda_r \, U_r^{(1)} \circ U_r^{(2)} \circ \ldots \circ U_r^{(N)}. \tag{6.1}$$

*The decomposition is visualized for third-order tensors in Fig. 6.1.*

Both terms originate from Psychometrics [49, 138]. Later on, the decomposition model was also applied in Chemometrics [7]. A good tutorial of the current state-of-the-art in Psychometrics and Chemometrics is [19].

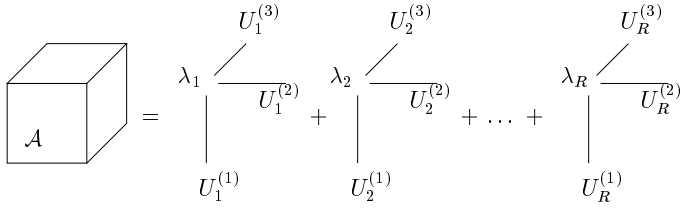



Figure 6.1: Visualization of the CANDECOMP for a third-order tensor.

This decomposition can be considered as the tensorial generalization of the diagonalization of matrices by equivalence transformation (unsymmetric case) or by congruence transformation (symmetric case). However, its properties are thoroughly different from their second-order counterparts.

First, recall from Sect. 2.6 that, even if non-orthogonal rank-1 terms are allowed, the minimal number of terms is not bounded by $\mathtt{min}\{I_1, I_2, \ldots, I_N\}$ in general; it is usually larger and depends also on the tensor order. For rank-related bibliographic pointers we refer to the same section. The fact that small perturbations of a given rank-$R$ tensor $\mathcal{A}$, with $R$ smaller than the possible



"generic ranks" (see also Sect. 2.6), may result in a tensor with rank higher than $R$, makes us look for an approximation of a tensor $\mathcal{A}$ by a sum of $R$ rank-1 components in this chapter, rather than solving Eq. (6.1) exactly.

Secondly, the uniqueness properties of the CANDECOMP are very different from (and also much more complicated than) their matrix equivalents. The theorems of [67] allow one to determine the dimensionality of the set of valid decompositions for generic tensors. Uniqueness in the third-order case is thoroughly studied in [165], from a combinatorial algebraic perspective. In [37] complex fourth-order cumulants are addressed. Here we will restrict ourselves to some remarks of a more general nature, that are of direct importance to this thesis. From the CANDECOMP-definition it is clear that the decomposition is insensitive to:

- a permutation of the rank-1 terms,
- a scaling of the vectors $U_r^{(n)}$, combined with the inverse scaling of the coefficients $\lambda_r$.

Apart from these trivial indeterminacies, the CANDECOMP is unique under conditions of linear independence, if the set of possible decompositions has dimensionality zero. Contrarily, the decomposition of a matrix $\mathbf{A}$ in a sum of $\mathtt{rank}(\mathbf{A})$ rank-1 terms is usually made unique by imposing stronger (e.g. orthogonality) constraints. In addition, for an essentially unique CANDECOMP the number of terms $R$ can exceed $\mathtt{min}\{I_1, I_2, \ldots, I_N\}$.

It is common practice to look for the CANDECOMP-components by straightforward minimization of the quadratic cost function

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 \tag{6.2}$$

over all rank-$R$ tensors $\hat{\mathcal{A}}$, that we will parametrize as

$$\hat{\mathcal{A}} = \sum_{r}^{R} \hat{\lambda}_r \, \hat{U}_r^{(1)} \circ \hat{U}_r^{(2)} \circ \ldots \circ \hat{U}_r^{(N)}. \tag{6.3}$$

It is e.g. possible to resort to an alternating least-squares algorithm, in which the vector estimates are updated mode per mode [51] (see Sect. 6.6). Optimization routines are observed to converge more easily if the rank is generic and the number of solutions is big. If the rank is not generic, the convergence often fails [67]. If the CANDECOMP-model is only approximately valid, the risk of finding a spurious local optimum can be diminished by repeating the optimization for a number of randomly chosen initial values. The decision on whether the global optimum has been found or not, usually relies on heuristics. The process of iterating over different starting values is very time-consuming. In

addition, if the directions of some of the $n$-mode vectors in the CANDECOMP-model ($1 \leqslant n \leqslant N$) are close, then it is quite unlikely that this configuration is found from a random start [182]. Some alternative initializations are discussed in [50]. The rank itself is usually determined by repeating the procedure for different values of $R$, and comparing the results. An alternative, also based on heuristics, is the evaluation of split-half experiments [138].

ALS iterations can be very slow. In addition, it is sometimes observed that the algorithm moves through a *swamp*: after initial convergence, the convergence speed drastically decreases and remains small for several iteration steps, after which it may suddenly increase again. The nature of swamps and how they can be avoided forms a topic of ongoing research [220, 166]. To cope with the slow convergence, a number of acceleration methods have been proposed [139, 140, 158]. One could e.g. make use of a prediction technique, in which estimates of previous iteration steps are extrapolated to forecast new estimates [19]. The fastest method at present is described in [213]. It consists of a Gauss-Newton method, in which all the CANDECOMP-factors are updated simultaneously; in addition, the inherent indeterminacy of the decomposition has been fixed by adding a quadratic regularization constraint on the component entries.

The deterministic equivalent of the direct optimization approach, consisting of setting the gradient of $f$ to zero and solving the resulting set of equations, is computationally hard as well: a set of $R(I_1+I_2+\ldots I_N)-R(N-1)$ polynomial equations of degree $2N-1$, in $R(I_1+I_2+\ldots I_N)-R(N-1)$ independent unknowns, has to be solved (to determine this dimensionality, imagine that the indeterminacy has been overcome by incorporating the factor $\lambda_r$ ($1 \leqslant r \leqslant R$) in one of the vectors of the $r$th outer product, and by fixing one non-zero entry in the other vectors).

An interesting alternative procedure, for the important special case where $R \leqslant \mathtt{min}\{I_1, I_2, \ldots, I_N\}$, has been proposed in [182]. Similar results have been proposed in [227, 22, 225, 226]. It was explained that, if Eq. (6.1) is exactly valid, the decomposition can be found by a simple matrix EVD. When $\mathcal{A}$ is only known with limited accuracy, a least-squares matching of both sides of Eq. (6.1) can now be initialized with the EVD-result. This technique forms the starting point for the developments in Sect. 6.3.

Some promising computation schemes, at this moment only formulated in terms of cumulant tensors, have been developed as means to solve the Higher-Order-Only ICA-problem (see Chapter 7). Together with other ICA-methods, these techniques are overviewed in Sect. 7.8.

The decomposition of a dataset as a sum of rank-1 terms is sometimes called the *Factor Analysis* problem. With the decomposition, one aims at relating the different rank-1 terms to the different "physical mechanisms" that have contributed to the dataset. We repeat that Factor Analysis of matrices is, as



such, essentially underdetermined. The extra conditions (maximal variance, orthonormality, ... ) that are usually imposed to guarantee uniqueness, are not seldom physically irrelevant. In a wide range of parameters, this is not the case for the higher-order decomposition; the weaker conditions of linear independence to ensure uniqueness, often have a physical meaning. This makes the CANDECOMP of higher-order tensors to an important signal processing tool. An example will be worked out in Chapter 11. In Chap. 10 we will apply the technique that will be developed in Sect. 6.3, in the context of Higher-Order-Only ICA.

In this chapter, we will study the special but important case of an $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ with rank $R \leqslant \mathtt{min}\{I_1, I_2\}$ and 3-rank $R_3 \geqslant 2$. (If $R_3 = 1$, then the different matrices obtained from $\mathcal{A}$ by fixing the index $i_3$ are proportional, and the CANDECOMP reduces to the diagonalization of one of these matrices by congruence or equivalence.) We assume that:

- the set $\{U_r^{(1)}\}_{(1 \leqslant r \leqslant R)}$ is linearly independent (i.e. no vector can be written as a linear combination of the other vectors),
- the set $\{U_r^{(2)}\}_{(1 \leqslant r \leqslant R)}$ is linearly independent,
- and that the set $\{U_r^{(3)}\}_{(1 \leqslant r \leqslant R)}$ does not contain collinear vectors (i.e. no vector is a scalar multiple of an other vector).

It can be verified that these conditions are sufficient to make the CANDECOMP essentially unique [182] (see also Sect. 6.2). The exposition is restricted to real-valued third-order tensors for notational convenience; the generalization to higher tensor orders and complex values is straightforward. An outline of the exposition is presented as Alg. 6.1.

## 6.2 CANDECOMP and simultaneous EVD

We start the derivation of a computation scheme with associating to $\mathcal{A}$ a linear transformation of the vector space $\mathbb{R}^{I_3}$ to the matrix space $\mathbb{R}^{I_1 \times I_2}$, in the following way:

$$\mathbf{V} = \mathcal{A}(W) \quad \Longleftrightarrow \quad v_{i_1 i_2} = \sum_{i_3} a_{i_1 i_2 i_3} w_{i_3}, \tag{6.4}$$

for all index values. Substitution of Eq. (6.4) in Eq. (6.1) shows that the image of $W$ can easily be expressed in terms of the CANDECOMP-components:

$$\mathbf{V} = \mathbf{U}^{(1)} \cdot \mathbf{D} \cdot \mathbf{U}^{(2)^T}, \tag{6.5}$$

**Algorithm 6.1**
*CANDECOMP*

(1. Perform an initial best rank-$(R,R,R_3)$ approximation of $\mathcal{A}$: maximize $g(\mathbf{X}^{(1)},\mathbf{X}^{(2)},\mathbf{X}^{(3)}) = \|\mathcal{A} \times_1 \mathbf{X}^{(1)^T} \times_2 \mathbf{X}^{(2)^T} \times_3 \mathbf{X}^{(3)^T}\|^2$ over $\mathbf{X}^{(1)} \in \mathrm{St}(R,I_1)$, $\mathbf{X}^{(2)} \in \mathrm{St}(R,I_2)$ and $\mathbf{X}^{(3)} \in \mathrm{St}(R,I_3)$; $\mathtt{max}(g(\mathbf{X}^{(1)},\mathbf{X}^{(2)},\mathbf{X}^{(3)})) = g(\mathbf{X}^{(1)}_{\max},\mathbf{X}^{(2)}_{\max},\mathbf{X}^{(3)}_{\max})$. $\mathcal{B} = \mathcal{A} \times_1 \mathbf{X}^{(1)^T}_{\max} \times_2 \mathbf{X}^{(2)^T}_{\max} \times_3 \mathbf{X}^{(3)^T}_{\max}$. Continue for $\mathcal{B}$ with steps 2, 3, 4, (5) below. $\hat{\mathcal{A}} = \hat{\mathcal{B}} \times_1 \mathbf{X}^{(1)}_{\max} \times_2 \mathbf{X}^{(2)}_{\max} \times_3 \mathbf{X}^{(3)}_{\max}$. (Perform step 5 for $\hat{\mathcal{A}}$.))

2. Associate to $\mathcal{A}$ a linear mapping from $\mathbb{R}^{I_3}$ to $\mathbb{R}^{I_1 \times I_2}$ (Eq. (6.4)). Determine $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ such that $\mathtt{span}\{\mathbf{V}_1,\mathbf{V}_2,\ldots,\mathbf{V}_K\} = \mathtt{R}\{\mathcal{A}\}$.

3. Compute orthogonal $\mathbf{Q},\mathbf{Z}$ and (approximately) upper triangular $\{\mathbf{R}_k\}_{(1 \leqslant k \leqslant K)}$ from the simultaneous generalized Schur Decomposition of Eqs. (6.17-6.19):

   - extended $QZ$-iteration (Subsect.6.4.1, [266]), or
   - maximize $g(\mathbf{Q},\mathbf{Z}) = \|\mathbf{Q}\cdot\mathbf{V}_1\cdot\mathbf{Z}\|^2_{UF} + \ldots + \|\mathbf{Q}\cdot\mathbf{V}_K\cdot\mathbf{Z}\|^2_{UF}$ by means of one-sided Givens iterations (Subsect.6.4.2) or two-sided Jacobi-iterations (Subsect.6.4.3).

4. Compute $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ from $\{\mathbf{R}_k\}_{(1 \leqslant k \leqslant K)}$ and $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ (and $\mathbf{Q}$ , $\mathbf{Z}$) (Sect. 6.3) or from $\mathbf{Q}$ , $\mathbf{Z}$ ($\mathbf{U}^{(1)} = \mathbf{U}^{(2)}$) (Sect.6.5).

(5. Minimize $f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2$:

   - ALS descent (Subsect.6.6.1), or
   - gradient descent (Subsect.6.6.2).)

in which we have used the following notations:

$$\mathbf{U}^{(n)} \stackrel{\mathrm{def}}{=} [U_1^{(n)} U_2^{(n)} \ldots U_{I_n}^{(n)}], \tag{6.6}$$

$$\mathbf{D} \stackrel{\mathrm{def}}{=} \mathtt{diag}\{(\lambda_1,\lambda_2,\ldots,\lambda_R)\} \cdot \mathtt{diag}\{\mathbf{U}^{(3)^T} W\}. \tag{6.7}$$

Any matrix in the range of the mapping can be diagonalized by equivalence with the matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$. (If $\mathcal{A}$ does not change under permutation of its first two indices, then any matrix in the range can be diagonalized by congruence with the matrix $\mathbf{U}^{(1)} = \mathbf{U}^{(2)}$.) If the range is spanned by the matrices $\mathbf{V}_1$, $\mathbf{V}_2$, ... , $\mathbf{V}_K$, then we should solve the following simultaneous decomposition:

$$\mathbf{V}_1 = \mathbf{U}^{(1)} \cdot \mathbf{D}_1 \cdot \mathbf{U}^{(2)^T}, \tag{6.8}$$



$$\mathbf{V}_2 = \mathbf{U}^{(1)} \cdot \mathbf{D}_2 \cdot \mathbf{U}^{(2)^T}, \tag{6.9}$$

$$\vdots$$

$$\mathbf{V}_K = \mathbf{U}^{(1)} \cdot \mathbf{D}_K \cdot \mathbf{U}^{(2)^T}, \tag{6.10}$$

in which $\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_K$ are diagonal. A possible choice of $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ consists of the "matrix slices" $\{\mathbf{A}_i\}_{(1 \leqslant i \leqslant I_3)}$, obtained by fixing the index $i_3$ to $i$ (see Fig. 6.2); the corresponding vectors $\{W_i\}_{(1 \leqslant i \leqslant I_3)}$ are the canonical unit vectors. An other possible choice consists of the $K$ dominant left singular matrices of the mapping in Eq. (6.4).

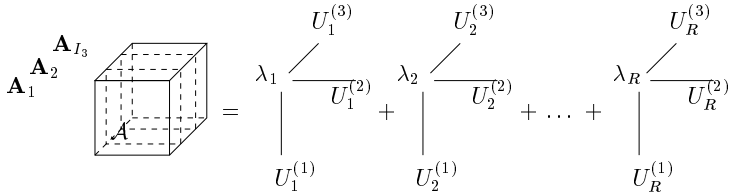



Figure 6.2: Definition of matrix slices for the computation of the CANDECOMP by simultaneous diagonalization.

If the CANDECOMP-model is exactly satisfied, then its terms can be computed from two of the equations in (6.8-6.10). Let us assume that e.g. the matrix $\mathbf{V}_1$ has full rank (this is the case for a generic choice of $W_1$, as we assume that $\mathbf{U}^{(3)}$ does not contain a zero column). Combination of the first two equations then leads to the following EVD:

$$(\mathbf{V}_2 \cdot \mathbf{V}_1^{\dagger}) \cdot \mathbf{U}^{(1)} = \mathbf{U}^{(1)} \cdot (\mathbf{D}_2 \cdot \mathbf{D}_1^{-1}) \tag{6.11}$$

If the diagonal elements of $\mathbf{D}_2 \cdot \mathbf{D}_1^{-1}$ are mutually different (this is also generically the case; remind that we assume for identifiability that $\mathbf{U}^{(3)}$ does not contain collinear columns), then the EVD reveals the columns of $\mathbf{U}^{(1)}$, up to an irrelevant scaling and/or permutation. Once $\mathbf{U}^{(1)}$ is known, $\mathbf{U}^{(2)}$ can be obtained, up to a scaling of its columns, by solving e.g. Eq. (6.8) as a linear set of equations in the unknown matrix $\mathbf{U}^{(2)} \cdot \mathbf{D}_1$. Finally, the matrix $\mathbf{U}^{(3)} \cdot \mathtt{diag}\{(\lambda_1,\lambda_2,\ldots,\lambda_R)\}$ is found by solving the CANDECOMP-model as a linear set of equations, for given matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$. (Remark that the assumptions that we have made for identifiability in Sect. 6.1, indeed allow to obtain the CANDECOMP in an essentially unique way.) If the CANDECOMP-model is only approximately satisfied, then the estimates can be used to initialize an additional optimization algorithm for the minimization of cost function (6.2). This EVD-approach is a variant of the techniques described in [182, 227, 22, 225, 226].

It is intuitively clear however, that it is preferable to exploit all the available information by taking into account all the equations in (6.8-6.10). This leads

to a *simultaneous EVD*:

$$\begin{aligned}(\mathbf{V}_2 \cdot \mathbf{V}_1^\dagger) \cdot \mathbf{U}^{(1)} &= \mathbf{U}^{(1)} \cdot (\mathbf{D}_2 \cdot \mathbf{D}_1^{-1}), && (6.12)\\ (\mathbf{V}_3 \cdot \mathbf{V}_1^\dagger) \cdot \mathbf{U}^{(1)} &= \mathbf{U}^{(1)} \cdot (\mathbf{D}_3 \cdot \mathbf{D}_1^{-1}), && (6.13)\\ &\vdots \\ (\mathbf{V}_K \cdot \mathbf{V}_1^\dagger) \cdot \mathbf{U}^{(1)} &= \mathbf{U}^{(1)} \cdot (\mathbf{D}_K \cdot \mathbf{D}_1^{-1}). && (6.14)\end{aligned}$$

One could solve these EVD's separately, and retain the solution that leads to the best CANDECOMP-estimate. However, it is safer from a numerical point of view to solve Eqs. (6.12-6.14) simultaneously, in some optimal sense, especially when the perturbation of the matrices $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ (with respect to their ideal values in an exact CANDECOMP) may have caused eigenvalues to cross each other. This is illustrated in the next example; a symmetric version of the example can be found in [21].

**Example 6.2.1** Consider the following matrix pair:

$$\mathbf{V}_1 = \begin{pmatrix} 1-\epsilon & 0 & 0 & 0 \\ 0 & 1+\epsilon & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \qquad \mathbf{V}_2 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1-\epsilon & 0 \\ 0 & 0 & 0 & 1+\epsilon \end{pmatrix},$$

in which $\epsilon \in \mathbb{R}$ is small. For $\epsilon = 0$, the two matrices have a common eigenmatrix:

$$\mathbf{E} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

If $\epsilon \neq 0$, $\mathbf{E}$ still nearly diagonalizes $\mathbf{V}_1$ and $\mathbf{V}_2$:

$$\mathbf{V}_1 \cdot \mathbf{E} = \mathbf{E} \cdot \mathtt{diag}([1\ 1\ 2\ 3]) + O(\epsilon), \qquad \mathbf{V}_2 \cdot \mathbf{E} = \mathbf{E} \cdot \mathtt{diag}([2\ 3\ 1\ 1]) + O(\epsilon).$$

On the other hand, for $\epsilon \neq 0$, the distinct eigenmatrices $\mathbf{E}_1$ and $\mathbf{E}_2$, of $\mathbf{V}_1$ and $\mathbf{V}_2$ respectively, are not suitable for diagonalization of the other matrix:

$$\mathbf{V}_1 \cdot \mathbf{E}_2 = \mathbf{E}_2 \cdot \mathtt{diag}([1\ 1\ 2\ 3]) + O(1), \qquad \mathbf{V}_2 \cdot \mathbf{E}_1 = \mathbf{E}_1 \cdot \mathtt{diag}([2\ 3\ 1\ 1]) + O(1).$$

□

Simultaneous matrix decompositions receive a growing interest in numerical matrix algebra. In [21] a Jacobi-technique is presented for the simultaneous diagonalization of commuting pairs of normal matrices by unitary (orthogonal)



transformations. In [41, 48] the least-squares diagonalization of a set of Hermitean (symmetric) matrices is addressed; this technique will be discussed in Sect. 9.3. The (real-valued) continuous-time QZ-problem is considered in [56]. Simultaneous congruence transformations of positive definite matrices forms the topic of [120]. An other approach, applicable to arbitrary matrices, is given in [266] and will be further investigated in the next sections.

**Remark 6.2.2** Under our identifiability assumptions, the CANDECOMP-model (6.1) implies that $R_1 = \mathtt{rank}_1(\mathcal{A}) = R = \mathtt{rank}_2(\mathcal{A}) = R_2$ and that $R_3 = \mathtt{rank}_3(\mathcal{A}) = \mathtt{rank}(\mathbf{U}^{(3)})$. If $R < \mathtt{max}(I_1, I_2)$, or $R_3 < I_3$, then an a priori *model reduction* of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ to a tensor $\mathcal{B} \in \mathbb{R}^{R \times R \times R_3}$ decreases the computational load of the actual determination of the CANDECOMP. If $\mathcal{A}$ and $\mathcal{B}$ are related by

$$\mathcal{A} = \mathcal{B} \times_1 \mathbf{X}^{(1)} \times_2 \mathbf{X}^{(2)} \times_3 \mathbf{X}^{(3)}, \qquad (6.15)$$

in which $\mathbf{X}^{(1)} \in \mathbb{R}^{I_1 \times R}$, $\mathbf{X}^{(2)} \in \mathbb{R}^{I_2 \times R}$ and $\mathbf{X}^{(3)} \in \mathbb{R}^{I_3 \times R_3}$ each have mutually orthonormal columns, then the optimal rank-$R$ approximation $\hat{\mathcal{B}}$ of $\mathcal{B}$ and the optimal rank-$R$ approximation $\hat{\mathcal{A}}$ of $\mathcal{A}$ are related in the same way:

$$\hat{\mathcal{A}} = \hat{\mathcal{B}} \times_1 \mathbf{X}^{(1)} \times_2 \mathbf{X}^{(2)} \times_3 \mathbf{X}^{(3)}, \qquad (6.16)$$

since $n$-mode multiplication with the column-wise orthonormal matrices $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, $\mathbf{X}^{(3)}$ does not change the cost function $f$ (Eq. (6.2)). If the CANDECOMP-model is exactly satisfied, then any orthonormal basis of the mode-1, mode-2 and mode-3 vectors of $\mathcal{A}$ gives a suitable $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, $\mathbf{X}^{(3)}$ respectively. In practice however, $R = R_1 = R_2$ and $R_3$ will be estimated as the number of significant mode-1 / mode-2 and mode-3 singular values of $\mathcal{A}$. This step becomes more difficult as the tensor order increases. By way of illustration, assume that $\mathcal{A}$ is given by Eq. (6.1), with $\mathbf{U}^{(1)} = \mathbf{U}^{(2)} = \ldots = \mathbf{U}^{(N)} = \mathbf{Q} \cdot \Omega$, in which $\mathbf{Q}$ is orthogonal and $\Omega = \mathtt{diag}\{\omega_1, \ldots, \omega_R\}$ diagonal. In this example, the $n$-mode singular values are given by $\lambda_r \omega_r^N$. An optimal rank-$(R, R, R_3)$ approximation of $\mathcal{A}$, before computing the optimal rank-$R$ approximation, can be achieved with the methods of the previous chapter. ◇

## 6.3 CANDECOMP and simultaneous generalized Schur decomposition

Without loss of generality we assume that $I_1 = I_2 = R$ and $I_3 = R_3 = K$ (see model reduction, in the previous section).

The fact that the unknown matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ are basically arbitrary regular matrices, makes them hard to deal with in a proper numerical way. In

this section, we will reformulate the problem in terms of orthogonal unknowns. Therefore, we can make an appeal to the technique established in [266], where the symmetric equivalent of Eqs. (6.8-6.10) was encountered in the derivation of an analytical Constant Modulus algorithm.

Introducing a $QR$-factorization $\mathbf{U}^{(1)} = \mathbf{Q}^T \mathbf{R}'$ and an $RQ$-decomposition $\mathbf{U}^{(2)^T} = \mathbf{R}'' \mathbf{Z}^T$ leads to a set of matrix equations that we will call a *simultaneous generalized Schur decomposition* (a set of two of the equations below is called "generalized Schur decomposition" [130]):

$$\begin{aligned} \mathbf{Q} \cdot \mathbf{V}_1 \cdot \mathbf{Z} &= \mathbf{R}_1 = \mathbf{R}' \cdot \mathbf{D}_1 \cdot \mathbf{R}'', & (6.17) \\ \mathbf{Q} \cdot \mathbf{V}_2 \cdot \mathbf{Z} &= \mathbf{R}_2 = \mathbf{R}' \cdot \mathbf{D}_2 \cdot \mathbf{R}'', & (6.18) \\ &\vdots & \\ \mathbf{Q} \cdot \mathbf{V}_K \cdot \mathbf{Z} &= \mathbf{R}_K = \mathbf{R}' \cdot \mathbf{D}_K \cdot \mathbf{R}'', & (6.19) \end{aligned}$$

in which $\mathbf{Q}, \mathbf{Z} \in \mathbb{R}^{R \times R}$ are orthogonal and $\mathbf{R}', \mathbf{R}'', \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathbb{R}^{R \times R}$ are upper triangular. The new problem consists of the determination of $\mathbf{Q}$ and $\mathbf{Z}$ such that $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K$ are "as upper triangular as possible" (in least-squares sense). The function $g$ to be maximized, is given by:

$$g(\mathbf{Q}, \mathbf{Z}) = \|\mathbf{Q} \cdot \mathbf{V}_1 \cdot \mathbf{Z}\|_{UF}^2 + \|\mathbf{Q} \cdot \mathbf{V}_2 \cdot \mathbf{Z}\|_{UF}^2 + \ldots + \|\mathbf{Q} \cdot \mathbf{V}_K \cdot \mathbf{Z}\|_{UF}^2. \quad (6.20)$$

The decomposition is visualized in Fig. 6.3.

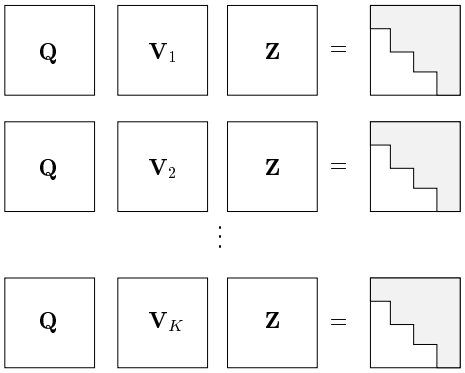



Figure 6.3: Visualization of a simultaneous generalized Schur decomposition.

Once $\mathbf{Q}$ and $\mathbf{Z}$ are known (the actual computation of the simultaneous generalized Schur decomposition will be discussed in the next section), the matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ can be estimated as follows. (Note that Sect. 6.5 contains an alternative reconstruction method, for the special case where $\mathbf{U}^{(1)} = \mathbf{U}^{(2)}$.)



First assume that $K \geqslant R$. From Eqs. (6.8-6.10) we have that:

$$\sum_{k=1}^{K} c_{rk} \mathbf{V}_k = \mu \, U_r^{(1)} U_r^{(2)^T}, \quad (6.21)$$

with $\mu \in \mathbb{R}_0$ and $c_{rk} \in \mathbb{R}$ $(1 \leqslant k \leqslant K)$, iff just the $r$th diagonal entry of $\sum_{k=1}^{K} c_{rk} \mathbf{D}_k$ is non-zero. All the linear combinations of $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K$ that give a rank-1 matrix, and hence the matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$, can be found by determining the coefficients $c_{rk}$ $(1 \leqslant k \leqslant K;\ 1 \leqslant r \leqslant R)$ by:

$$\mathbf{C} = \mathbf{D}^{\dagger}, \quad \mathbf{D} \stackrel{\text{def}}{=} \begin{pmatrix} (\mathbf{D}_1)_{11} & \ldots & (\mathbf{D}_1)_{RR} \\ (\mathbf{D}_2)_{11} & \ldots & (\mathbf{D}_2)_{RR} \\ \vdots & & \vdots \\ (\mathbf{D}_K)_{11} & \ldots & (\mathbf{D}_K)_{RR} \end{pmatrix}. \quad (6.22)$$

It is not necessary to compute $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_K$ from Eqs. (6.17-6.19): since $\mathbf{R}'$ and $\mathbf{R}''$ have full rank, linear combinations of $\{\mathbf{D}_k\}_{(1 \leqslant k \leqslant K)}$ are rank-1 iff the same linear combinations of $\{\mathbf{R}_k\}_{(1 \leqslant k \leqslant K)}$ have just one non-zero entry on the main diagonal. Hence, we determine $\mathbf{C}$ as:

$$\mathbf{C} = \mathbf{R}^{\dagger}, \quad \mathbf{R} \stackrel{\text{def}}{=} \begin{pmatrix} (\mathbf{R}_1)_{11} & \ldots & (\mathbf{R}_1)_{RR} \\ (\mathbf{R}_2)_{11} & \ldots & (\mathbf{R}_2)_{RR} \\ \vdots & & \vdots \\ (\mathbf{R}_K)_{11} & \ldots & (\mathbf{R}_K)_{RR} \end{pmatrix}. \quad (6.23)$$

If $K < R$, then an other approach has to be followed, but it is still possible to estimate $\mathbf{R}'$ and $\mathbf{R}''$ from Eqs. (6.17-6.19), and then combine them with $\mathbf{Q}$ and $\mathbf{Z}$. If we assume that the main diagonals of $\mathbf{R}'$ and $\mathbf{R}''$ only contain entries equal to 1, then $\mathbf{D}_k = \texttt{diag}(\mathbf{R}_k)$ $(1 \leqslant k \leqslant K)$. The strictly upper diagonal elements of $\mathbf{R}'$ and $\mathbf{R}''$ can be estimated by subsequently solving in a least-squares sense the equations related to the entries $(R-1, R)$, $(R-2, R-1)$, $(R-2, R)$, ..., $(1,2)$, $(1,3)$, ..., $(1, R)$ in (6.17-6.19), w.r.t. the unknowns $r'_{R-1,R}$ and $r''_{R-1,R}$, $r'_{R-2,R-1}$ and $r''_{R-2,R-1}$, $r'_{R-2,R}$ and $r''_{R-2,R}$, ..., $r'_{1,2}$ and $r''_{1,2}$, $r'_{1,3}$ and $r''_{1,3}$, ..., $r'_{1,R}$ and $r''_{1,R}$ respectively.

## 6.4 Algorithms for the simultaneous generalized Schur decomposition

### 6.4.1 Extended QZ-iteration

For the actual computation of the simultaneous generalized Schur decomposition, an extended QZ-iteration was proposed in [266]. For the optimization

of the function $g$ in (6.20), $\mathbf{Q}$ and $\mathbf{Z}$ are updated alternatingly. In each step, the optimal $\mathbf{Q}$ (given $\mathbf{Z}$, or vice-versa) is estimated as a product of matrices $\mathbf{H}_1\mathbf{H}_2\ldots\mathbf{H}_{R-1}$, that form the equivalent of Householder matrices for the computation of a simple QR-decomposition [130]. E.g. as far as $\mathbf{Q}$ is concerned, $\mathbf{H}_1$ maximally reduces (in least-squares sense) the below-diagonal norm of the first columns of the instantaneous estimates of $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K$. After multiplication with $\mathbf{H}_1$, $\mathbf{H}_2$ minimizes the below-diagonal norm of the second columns, without further affecting the first rows, and so on.

The resulting scheme is observed to find a good estimate of the global optimum in a limited number of steps, if the CANDECOMP-model is exactly satisfied. However, even moderate perturbations can cause the algorithm to end up in good estimates of the theoretical matrices $\mathbf{Q}$ and $\mathbf{Z}$, that do not globally minimize the below-diagonal norm. It is also possible that at some point in the iteration (e.g. initially, or after approximate convergence), the algorithm starts to increase the below-diagonal norm. The reason for this behaviour is that the way in which $\mathbf{Q}$ and $\mathbf{Z}$ are computed, does not imply monotonic convergence in terms of the function $g$ in Eq. (6.20): e.g. it is possible that the matrix $\mathbf{H}_1$ increases the Frobenius-norm of the part of columns 2 to $R-1$ below the diagonal. As an alternative, we investigate a Givens- and Jacobi-type approach in Subsect. 6.4.2 and 6.4.3, to ensure monotonic convergence; in Subsect. 6.4.4 we provide an explicit expression for the gradient of the cost function, with a view to gradient-based optimization.

### 6.4.2 One-sided Givens iterations

Let us first investigate an iterative optimization of the function $g(\mathbf{Q}, \mathbf{Z})$, by separately updating the estimates of $\mathbf{Q}$ and $\mathbf{Z}$ at a given iteration step by multiplication with a Givens rotation matrix. The purpose of each elementary rotation is to make the current estimates of $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K$ simultaneously as upper triangular as possible.

First, the estimates of $\mathbf{Q}$ and $\mathbf{Z}$ are initialized as any orthogonal matrix, e.g. as the identity matrix: $\mathbf{Q}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{I}$. The estimates of $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K$ are initialized accordingly: $\mathbf{R}_1^{(0)} = \mathbf{V}_1$, $\mathbf{R}_2^{(0)} = \mathbf{V}_2$, $\ldots$, $\mathbf{R}_K^{(0)} = \mathbf{V}_K$. (As an alternative, initial matrices $\mathbf{Q}^{(0)}$ and $\mathbf{Z}^{(0)}$ could be obtained from the generalized Schur decomposition corresponding to two of the equations in (6.17-6.19) [130].)

Assume that at iteration step $l+1$, the estimates $\mathbf{Q}^{(l)}$, $\mathbf{Z}^{(l)}$ and $\mathbf{R}_1^{(l)}, \ldots, \mathbf{R}_K^{(l)}$ are available. Let $\mathbf{G}_{ij} \in \mathbb{R}^{R\times R}$ represent an elementary Givens rotation matrix that affects rows $i$ and $j$, i.e. $\mathbf{G}_{ij}$ equals the identity matrix, except for the entries

$$(\mathbf{G}_{ij})_{ii} = (\mathbf{G}_{ij})_{jj} = \cos\alpha,$$



$$(\mathbf{G}_{ij})_{ji} = -(\mathbf{G}_{ij})_{ij} = \sin\alpha,$$

in which $\alpha$ is the rotation angle (assume e.g. that $j > i$). An update of $\mathbf{Q}^{(l)}$ takes the form of $\mathbf{Q}^{(l+1)} = \mathbf{G}_{ij} \cdot \mathbf{Q}^{(l)}$; at the same time $\mathbf{R}_1^{(l)}, \mathbf{R}_2^{(l)}, \ldots, \mathbf{R}_K^{(l)}$ are updated as $\mathbf{R}_1^{(l+1)} = \mathbf{G}_{ij} \cdot \mathbf{R}_1^{(l)}$, $\mathbf{R}_2^{(l+1)} = \mathbf{G}_{ij} \cdot \mathbf{R}_2^{(l)}$, $\ldots$, $\mathbf{R}_K^{(l+1)} = \mathbf{G}_{ij} \cdot \mathbf{R_K}^{(l)}$. Similarly, an update of $\mathbf{Z}^{(l)}$ should take the form of $\mathbf{Z}^{(l+1)} = \mathbf{Z}^{(l)} \cdot {\mathbf{G}'_{ij}}^T$, where the Givens rotation matrix $\mathbf{G}'_{ij}$ is defined in the same way as $\mathbf{G}_{ij}$, in terms of an angle $\beta$; here $\mathbf{R}_1^{(l)}, \mathbf{R}_2^{(l)}, \ldots, \mathbf{R}_K^{(l)}$ are updated as $\mathbf{R}_1^{(l+1)} = \mathbf{R}_1^{(l)} \cdot {\mathbf{G}'_{ij}}^T$, $\mathbf{R}_2^{(l+1)} = \mathbf{R}_2^{(l)} \cdot {\mathbf{G}'_{ij}}^T$, $\ldots$, $\mathbf{R}_K^{(l+1)} = \mathbf{R}_K^{(l)} \cdot {\mathbf{G}'_{ij}}^T$.

Let us focus on the updating of $\mathbf{Q}^{(l)}$ by multiplication with $\mathbf{G}_{ij}$. The Givens-rotation should be determined such that it minimizes the below-diagonal norm of all the $i$th and $j$th rows in $\mathbf{R}_1^{(l+1)}, \ldots, \mathbf{R}_K^{(l+1)}$. If we define

$$\mathbf{E}_{ij}^{(l)} = \begin{pmatrix} (\mathbf{R}_1^{(l)})_{i,\,i:j-1} & \ldots & (\mathbf{R}_K^{(l)})_{i,\,i:j-1} \\ (\mathbf{R}_1^{(l)})_{j,\,i:j-1} & \ldots & (\mathbf{R}_K^{(l)})_{j,\,i:j-1} \end{pmatrix} \tag{6.24}$$

using MATLAB-notation, then $\mathbf{G}_{ij}$ should be determined such that after multiplication the second row of $\mathbf{E}_{ij}^{(l+1)}$ has minimal Frobenius-norm (see Fig. 6.4; the parts of rows $i$ and $j$ that are not contained in $\mathbf{E}_{ij}^{(l)}$ do not influence the below diagonal norm). Hence the submatrix

$$\mathbf{G} \stackrel{\text{def}}{=} \begin{pmatrix} (\mathbf{G}_{ij})_{ii} & (\mathbf{G}_{ij})_{ij} \\ (\mathbf{G}_{ij})_{ji} & (\mathbf{G}_{ij})_{jj} \end{pmatrix}$$

can be obtained as the transpose of the left singular matrix of $\mathbf{E}_{ij}^{(l)}$: in this way the norm of the second row equals the smallest singular value of $\mathbf{E}_{ij}^{(l)}$, which is the best one can do. The estimate of $\mathbf{Z}$ can be updated analogously.

Now we visualize a typical example of the function that has to be optimized.

**Example 6.4.1** Consider the matrices $\mathbf{V}_1$, $\mathbf{V}_2$, and $\mathbf{V}_3$, constructed as in Eqs. (6.8-6.10), for given

$$\mathbf{U}^{(1)} = \mathbf{U}^{(2)} = \begin{pmatrix} -1.2 & -0.3 & -1.5 \\ 0.5 & -1.6 & -1.3 \\ -1.1 & -0.2 & -0.1 \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{D}_1 &= \texttt{diag}(-0.8, -1.1, 0.6), \\ \mathbf{D}_2 &= \texttt{diag}(0.3, 0.4, -1.5), \\ \mathbf{D}_3 &= \texttt{diag}(0.2, 0.4, -1.4). \end{aligned}$$

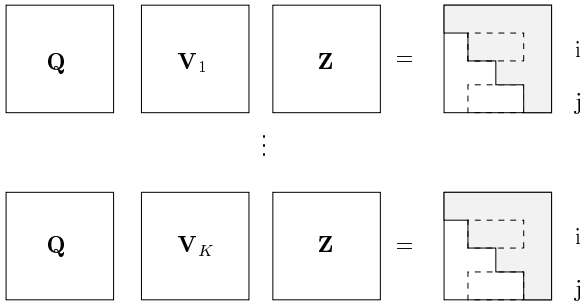



Figure 6.4: Under Givens rotation of rows $i$ and $j$, only the indicated parts affect the below-diagonal norm.

Fig. 6.5 shows the effect of the multiplication $\mathbf{V}_1' = \mathbf{G}_{12}\mathbf{V}_1{\mathbf{G}'_{12}}^T$, $\mathbf{V}_2' = \mathbf{G}_{12}\mathbf{V}_2{\mathbf{G}'_{12}}^T$, and $\mathbf{V}_3' = \mathbf{G}_{12}\mathbf{V}_3{\mathbf{G}'_{12}}^T$ on the joint below-diagonal Frobenius-norm. This norm is plotted on the right (with a minus sign, for clearness): $\overline{g}(\alpha,\beta) = g(\mathbf{G}_{12}(\alpha), \mathbf{G}'_{12}(\beta)^T) - \|\mathbf{V}_1\|^2 - \|\mathbf{V}_2\|^2 - \|\mathbf{V}_3\|^2$. In the left plot the contour lines of this function are drawn (the function values of the solid lines differ with 10 units; for the dashed lines the interval is 1.5 units). The "o"-marks indicate the global minima of the joint below-diagonal norm; the two distinct values correspond to the possible orderings of the two first columns of $\mathbf{A}$. The "+"-marks correspond to the local minima of $\overline{g}$. The two "x"-marks are saddle-points. Remark that $\overline{g}$ is rather flat in the neighbourhood of the minima. □

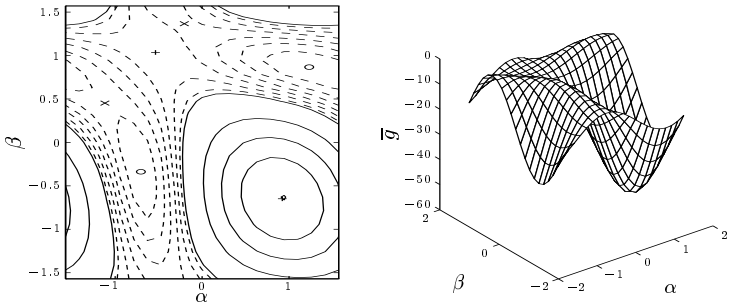



Figure 6.5: Effect of left ($\alpha$) and right ($\beta$) Givens-rotation of the set of matrices in Ex. 6.4.1 on their joint below-diagonal Frobenius-norm ($-\overline{g}$). The "o"-marks indicate the global Frobenius-norm minima.



As a cycling strategy, we propose to alternate between cyclic sweeps for $\mathbf{Q}$ and $\mathbf{Z}$. As an alternative, one could think of a QZ-iteration-type method, where one aims at finding a good estimate of the optimal $\mathbf{Q}$, given $\mathbf{Z}$ or vice versa, by alternating between sequences of sweeps for $\mathbf{Q}$ and $\mathbf{Z}$. Through simulations we have observed that the latter approach is less efficient on the average. The extra cost of "marginally" fine-tuning the fit in the individual $\mathbf{Q}$-space (or $\mathbf{Z}$-space) can in general not be motivated against the more significant "original" improvement of the overall fit that can be achieved by switching between both spaces, even if only two sweeps are carried out each time. This will be illustrated in Sect. 6.7 with a number of simulations.

A major drawback of the Givens method, with respect to the extended QZ-algorithm discussed at the beginning of this section, is that the algorithm sometimes gets stuck far away from the global optimum. This depends on the starting point, and on the number of $\mathbf{Q}$-sweeps and $\mathbf{Z}$-sweeps performed in each iteration step. See also Sect. 6.7.

### 6.4.3 Two-sided Jacobi iterations

Now let us consider a two-sided updating of the simultaneous generalized Schur decomposition, i.e. $\mathbf{Q}^{(l)}$ and $\mathbf{Z}^{(l)}$ are updated at the same time. The entries on the strictly lower diagonal part of $\mathbf{R}_k^{(l+1)} = \mathbf{G}_{ij} \cdot \mathbf{R}_k^{(l)} \cdot {\mathbf{G}'_{ij}}^T$ that are changed by Jacobi-rotation, are given by:

$$(\mathbf{R}_k^{(l+1)})_{ir} = \cos\alpha(\mathbf{R}_k^{(l)})_{ir} - \sin\alpha(\mathbf{R}_k^{(l)})_{jr}, \qquad 1 \leqslant r < i, \tag{6.25}$$

$$(\mathbf{R}_k^{(l+1)})_{jr} = \sin\alpha(\mathbf{R}_k^{(l)})_{ir} + \cos\alpha(\mathbf{R}_k^{(l)})_{jr}, \qquad 1 \leqslant r < j,\ r \neq i, \tag{6.26}$$

$$\begin{aligned}(\mathbf{R}_k^{(l+1)})_{ji} &= \sin\alpha(\cos\beta(\mathbf{R}_k^{(l)})_{ii} - \sin\beta(\mathbf{R}_k^{(l)})_{ij}) \\ &\quad + \cos\alpha(\cos\beta(\mathbf{R}_k^{(l)})_{ji} - \sin\beta(\mathbf{R}_k^{(l)})_{jj}),\end{aligned} \tag{6.27}$$

$$(\mathbf{R}_k^{(l+1)})_{ri} = \cos\beta(\mathbf{R}_k^{(l)})_{ri} - \sin\beta(\mathbf{R}_k^{(l)})_{rj}, \qquad i < r \leqslant R,\ r \neq j, \tag{6.28}$$

$$(\mathbf{R}_k^{(l+1)})_{rj} = \sin\beta(\mathbf{R}_k^{(l)})_{ri} + \cos\beta(\mathbf{R}_k^{(l)})_{rj}, \qquad j < r \leqslant R. \tag{6.29}$$

Hence, at iteration step $l$, the maximization of the function $g$ in Eq. (6.20) is equivalent to the minimization of

$$h(\alpha,\beta) = \sum_{k=1}^{K}[(\mathbf{R}_k^{(l+1)})_{ji}^2 + \sum_{r=i+1}^{j-1}((\mathbf{R}_k^{(l+1)})_{ri}^2 + (\mathbf{R}_k^{(l+1)})_{jr}^2)] \tag{6.30}$$

(the other entries do not affect the below diagonal norm; see Fig. 6.6).

Substitution of Eqs. (6.26-6.28) in Eq. (6.30) gives the following explicit ex-

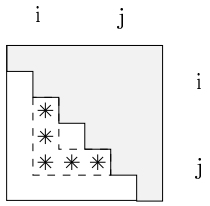



Figure 6.6: Under Jacobi rotation of rows/columns $i$ and $j$, the below-diagonal norm can only be affected by the indicated transformed elements.

pression for $h$:

$$h(\alpha,\beta) = \sum_{k=1}^{K}\sum_{n=1}^{5} h_{kn}(\alpha,\beta), \tag{6.31}$$

in which:

$$\begin{aligned}
h_{k1}(\alpha,\beta) &= \sin^2\alpha\,[\cos^2\beta\,(\mathbf{R}_k^{(l)})_{ii}^2 + \sin^2\beta\,(\mathbf{R}_k^{(l)})_{ij}^2 \\
&\quad -2\sin\beta\cos\beta\,(\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{ij}], && (6.32)\\
h_{k2}(\alpha,\beta) &= 2\sin\alpha\cos\alpha\,\{\cos^2\beta\,(\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{ji} + \sin^2\beta\,(\mathbf{R}_k^{(l)})_{ij}(\mathbf{R}_k^{(l)})_{jj} \\
&\quad -\sin\beta\cos\beta\,[(\mathbf{R}_k^{(l)})_{ij}(\mathbf{R}_k^{(l)})_{ji} + (\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{jj}]\}, && (6.33)\\
h_{k3}(\alpha,\beta) &= \cos^2\alpha\,[\cos^2\beta\,(\mathbf{R}_k^{(l)})_{ji}^2 + \sin^2\beta\,(\mathbf{R}_k^{(l)})_{jj}^2 \\
&\quad -2\sin\beta\cos\beta\,(\mathbf{R}_k^{(l)})_{ji}(\mathbf{R}_k^{(l)})_{jj}], && (6.34)\\
h_{k4}(\alpha,\beta) &= (\sin^2\alpha+\cos^2\alpha)\sum_{r=i+1}^{j-1}[\cos^2\beta\,(\mathbf{R}_k^{(l)})_{ri}^2 + \sin^2\beta\,(\mathbf{R}_k^{(l)})_{rj}^2 \\
&\quad -2\sin\beta\cos\beta\,(\mathbf{R}_k^{(l)})_{ri}(\mathbf{R}_k^{(l)})_{rj}], && (6.35)\\
h_{k5}(\alpha,\beta) &= (\sin^2\beta+\cos^2\beta)\sum_{r=i+1}^{j-1}[\cos^2\alpha\,(\mathbf{R}_k^{(l)})_{jr}^2 + \sin^2\alpha\,(\mathbf{R}_k^{(l)})_{ir}^2 \\
&\quad +2\sin\alpha\cos\alpha\,(\mathbf{R}_k^{(l)})_{ir}(\mathbf{R}_k^{(l)})_{jr}]. && (6.36)
\end{aligned}$$

Setting the partial derivatives of $h$, with respect to $\alpha$ and $\beta$, equal to zero, leads to a set of biquadratic equations in $\tan\alpha$ and $\tan\beta$:

$$\begin{aligned}
\frac{1}{\cos^2\alpha\cos^2\beta}\cdot\frac{\partial h}{\partial\alpha} &= b_1(\beta)\tan^2\alpha + b_2(\beta)\tan\alpha - b_1(\beta) = 0, && (6.37)\\
\frac{1}{\cos^2\alpha\cos^2\beta}\cdot\frac{\partial h}{\partial\beta} &= b_3(\beta)\tan^2\alpha + b_4(\beta)\tan\alpha + b_5(\beta) = 0, && (6.38)
\end{aligned}$$



in which $b_n(\beta) = \sum_{k=1}^{K} b_{kn}(\beta)$, with:

$$\begin{aligned}
b_{k1}(\beta) &= \tan^2\beta\,\{(\mathbf{R}_k^{(l)})_{ij}^2 - (\mathbf{R}_k^{(l)})_{jj}^2 + \sum_{r=i+1}^{j-1}[(\mathbf{R}_k^{(l)})_{ir}^2 - (\mathbf{R}_k^{(l)})_{jr}^2]\} \\
&\quad +2\tan\beta\,[(\mathbf{R}_k^{(l)})_{ji}(\mathbf{R}_k^{(l)})_{jj} - (\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{ij}] \\
&\quad +\{(\mathbf{R}_k^{(l)})_{ii}^2 - (\mathbf{R}_k^{(l)})_{ji}^2 + \sum_{r=i+1}^{j-1}[(\mathbf{R}_k^{(l)})_{ir}^2 - (\mathbf{R}_k^{(l)})_{jr}^2]\}, && (6.39)\\
b_{k2}(\beta) &= \tan^2\beta\,[(\mathbf{R}_k^{(l)})_{ij}(\mathbf{R}_k^{(l)})_{jj} + \sum_{r=i+1}^{j-1}(\mathbf{R}_k^{(l)})_{ir}(\mathbf{R}_k^{(l)})_{jr}] \\
&\quad -\tan\beta\,[(\mathbf{R}_k^{(l)})_{ij}(\mathbf{R}_k^{(l)})_{ji} + (\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{jj}] \\
&\quad +[(\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{ji} + \sum_{r=i+1}^{j-1}(\mathbf{R}_k^{(l)})_{ir}(\mathbf{R}_k^{(l)})_{jr}], && (6.40)\\
b_{k3}(\beta) &= (\tan^2\beta-1)\,[(\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{ij} + \sum_{r=i+1}^{j-1}(\mathbf{R}_k^{(l)})_{ri}(\mathbf{R}_k^{(l)})_{jr}] \\
&\quad +\tan\beta\,\{(\mathbf{R}_k^{(l)})_{ij}^2 - (\mathbf{R}_k^{(l)})_{ii}^2 + \sum_{r=i+1}^{j-1}[(\mathbf{R}_k^{(l)})_{rj}^2 - (\mathbf{R}_k^{(l)})_{ri}^2]\}, && (6.41)\\
b_{k4}(\beta) &= (\tan^2\beta-1)\,[(\mathbf{R}_k^{(l)})_{ij}(\mathbf{R}_k^{(l)})_{ji} + (\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{jj}] \\
&\quad +2\tan\beta\,[(\mathbf{R}_k^{(l)})_{ij}(\mathbf{R}_k^{(l)})_{jj} - (\mathbf{R}_k^{(l)})_{ii}(\mathbf{R}_k^{(l)})_{ji}], && (6.42)\\
b_{k5}(\beta) &= (\tan^2\beta-1)\,[(\mathbf{R}_k^{(l)})_{ji}(\mathbf{R}_k^{(l)})_{jj} + \sum_{r=i+1}^{j-1}(\mathbf{R}_k^{(l)})_{ri}(\mathbf{R}_k^{(l)})_{rj}] \\
&\quad +\tan\beta\,\{(\mathbf{R}_k^{(l)})_{jj}^2 - (\mathbf{R}_k^{(l)})_{ji}^2 + \sum_{r=i+1}^{j-1}[(\mathbf{R}_k^{(l)})_{rj}^2 - (\mathbf{R}_k^{(l)})_{ri}^2]\}. && (6.43)
\end{aligned}$$

The global minimum of $h(\alpha,\beta)$ can be determined by computing the various solutions of Eqs. (6.37-6.38), and selecting the one corresponding to the smallest value in Eq. (6.31).

For the solution of the set of biquadratic equations, let us first consider the special case where Eq. (6.37) is linear in $\tan\alpha$: $b_1(\beta) = 0$. The only ways in which a root $\beta_0$ of $b_1$ can lead to a solution of Eqs. (6.37-6.38), are (a) $\tan\alpha = 0$ and additionally $b_5(\beta_0) = 0$, and (b) $\alpha$ is a solution of Eq. (6.38), for $\beta = \beta_0$, which additionally satisfies $b_2(\beta_0) = 0$. Taking into account the definitions of $b_1(\beta)$, $b_2(\beta)$ and $b_5(\beta)$, neither case occurs, generically.

Now let us investigate the general case, i.e. $b_1(\beta) \neq 0$. Substitution of the square roots of Eq. (6.37) in Eq. (6.38) (considered as quadratic expressions

in the unknown $\tan\alpha$) then leads, after tedious formula manipulations, to the following polynomial of degree 8 in $\tan\beta$:

$$b_1^2(\beta)b_3^2(\beta) + b_1^2(\beta)b_5^2(\beta) - b_1^2(\beta)b_4^2(\beta) + 2b_1^2(\beta)b_3(\beta)b_5(\beta)$$
$$+b_2^2(\beta)b_3(\beta)b_5(\beta) - b_1(\beta)b_2(\beta)b_4(\beta)b_5(\beta) + b_1(\beta)b_2(\beta)b_3(\beta)b_4(\beta) = 0 \quad (6.44)$$

For the roots of this polynomial, the corresponding values of $\tan\alpha$ that give a solution to Eqs. (6.37-6.38), can be found by rooting the $S$-polynomial [236] (in the unknown $\tan\alpha$) of the set of Eqs. (6.37-6.38). This polynomial is linear; for a given value of $\beta$, $\tan\alpha$ can be found from:

$$(b_2(\beta)b_3(\beta) - b_1(\beta)b_4(\beta))\tan\alpha - b_1(\beta)(b_3(\beta) + b_5(\beta)) = 0. \quad (6.45)$$

**Remark 6.4.2** If the simultaneous equivalence transformation of Eqs. (6.8-6.10) is not exactly satisfied, different permutations of the CANDECOMP components may cause the corresponding orthogonal factors $\mathbf{Q}$ and $\mathbf{Z}$ to yield values of the upper-diagonal norm function $g$ (Eq. (6.20)) that are somewhat different. There is no guarantee that the two-sided Jacobi-algorithm will converge to the solution with that specific column ordering that leads to the global optimum. As far as the global optimization as such is concerned, simulations have lead to the heuristic rule that a necessary and sufficient condition for a global optimum is that the function value cannot be further increased by reinitializing the algorithm with the best outer rotation among the solutions of Eqs. (6.44-6.45), for any of the possible rotation pairs $(i, j)$ (with an "outer rotation" we mean that $\alpha \in (-\pi/2, -\pi/4] \cup (\pi/4, \pi/2]$ or $\beta \in (-\pi/2, -\pi/4] \cup (\pi/4, \pi/2]$). In accordance with this rule, we have observed in all our simulations that the global optimum is found by extensively checking all rotation pairs, and restarting the procedure if a reinitialized algorithm allows to improve the current estimate of the global optimum.
Apart from the re-ordering of columns, there is no formal evidence that the two-sided Jacobi-algorithm cannot get stuck in a local optimum; local or global convergence is still an open problem for the computation of the simultaneous matrix decompositions mentioned in Sect. 6.3 as well. We have not observed convergence to a local optimum in any of our simulations for the unsymmetric CANDECOMP-problem. For the case where $\mathbf{U}^{(1)} = \mathbf{U}^{(2)}$ (see Chapter 10), a meaningless result has been obtained for one out of hundreds of simulations. In this odd case, the problem could be overcome by reinitializing the algorithm as described above. ◇

**Remark 6.4.3** The computation scheme described above, can also be interpreted as follows. Remind that the cost function $h$ has been defined as an homogeneous quadratic polynomial in $\sin\alpha$ and $\cos\alpha$, of which the coefficients are themselves homogeneous quadratic polynomials, in $\sin\beta$ and $\cos\beta$ (6.31), i.e. $h$ takes the following form:

$$h(\alpha, \beta) = X^T \mathbf{M}(\beta) X,$$



in which $X = (\sin\alpha \;\; \cos\alpha)^T$ and in which $\mathbf{M} \in \mathbb{R}^{2\times 2}$ is a symmetric positive (semi)definite matrix of which the entries are homogeneous quadratic polynomials in $\sin\beta$ and $\cos\beta$. Given a particular value of $\beta$, the optimal choice for $X$ is the eigenvector of $\mathbf{M}$ that corresponds to the smallest eigenvalue; this is equivalent to the computation of Eq. (6.45). In addition, $\beta$ should be determined such that the smallest eigenvalue of $\mathbf{M}(\beta)$ is as small as possible. Expressing the eigenvalues of $\mathbf{M}(\beta)$ as a function of its entries, and setting the derivative with respect to $\beta$ equal to zero, in order to find the critical points, leads to the formula given in Eq. (6.44). ◇

**Remark 6.4.4** For the related problem of a simultaneous Schur decomposition, in which $\mathbf{Q} = \mathbf{Z}$ by definition, the cost function $h$ in Eq. (6.30) is an homogeneous function in $\cos\alpha$ and $\sin\alpha$. Hence the set (6.37-6.38) reduces to the rooting of a single polynomial of degree four in $\tan\alpha$ [135]. ◇

### 6.4.4 Gradient-based optimization

In this section we provide an explicit expression for the gradient of the function $g$, over the manifolds of orthogonal matrices to which $\mathbf{Q}$ and $\mathbf{Z}$ belong. This gradient can be interpreted as the "velocity" of $\mathbf{Q}$ and $\mathbf{Z}$ in a continuous-time version of the computation of the simultaneous generalized Schur decomposition, or, after integration, in discrete-time implementations.

**Theorem 6.4.5**
*Consider the function $g(\mathbf{Q}, \mathbf{Z})$, explicited in Eq. (6.20), in which $\mathbf{Q}, \mathbf{Z} \in O(R)$; let the matrices $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K$ be defined by Eqs. (6.17-6.19). The gradients of $g$, with respect to $\mathbf{Q}$ and $\mathbf{Z}$, over the manifold of orthogonal matrices, are given by:*

$$\nabla_{\mathbf{Q}} g = 2\,\mathtt{skew}\{\sum_k \mathtt{upp}(\mathbf{R}_k)\mathbf{R}_k^T\} \cdot \mathbf{Q}, \quad (6.46)$$

$$\nabla_{\mathbf{Z}} g = 2\mathbf{Z} \cdot \mathtt{skew}\{\sum_k \mathbf{R}_k \mathtt{upp}(\mathbf{R}_k)\}. \quad (6.47)$$

(For a definition of `skew` and `upp`, and $\mathrm{O}(R)$, we refer to App. A.)

**Proof:** The proof is similar in spirit to the proof of Theor. 5.3.9: the effect of infinitesimal changes of $\mathbf{Q}$ and $\mathbf{Z}$ on the value of $g$ is investigated, and expressed in terms of the gradients of $g$. The derivation involves a number of technicalities of the calculation of scalar products, and is therefore deferred to App. C. □

## 6.5 A matrix representation

In this section we propose an alternative method for the reconstruction of the CANDECOMP-model, given the orthogonal matrices $\mathbf{Q}$ and $\mathbf{Z}$, when the matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ in Eq. (6.6) are the same. In this case the tensor $\mathcal{A}$ is necessarily invariant for interchanging its first two indices. The situation can e.g. arise for (super-symmetric) moment and cumulant tensors; an application will be further developed in Chapter 10.

First let us remark that, even if $\mathbf{U}^{(1)} = \mathbf{U}^{(2)}$, the transition to the new unknowns $\mathbf{Q}$ and $\mathbf{Z}$ in Eqs. (6.17-6.19) is not an overparametrization. As a matter of fact, since $\mathbf{Q}$ and $\mathbf{Z}$ are orthogonal, the total number of degrees of freedom ($R(R-1)$) exactly corresponds to the number of parameters to be determined in $\mathbf{U}^{(1)}$ (taking into account that the norms of the columns of $\mathbf{U}^{(1)}$ are essentially unidentifiable). Hence the procedure to reconstruct $\mathbf{U}^{(1)}$, described in Sect. 6.3, can still be applied. On the other hand, we will show now that there is no further need to resort to Eqs. (6.17-6.19), once $\mathbf{Q}$ and $\mathbf{Z}$ have been determined. In other words, we prove that matrices that can only be identified up to a scaling of their columns, have a unique alternative representation in terms of a pair of orthogonal matrices ($\mathbf{Q}$,$\mathbf{Z}$); this representation is interesting as such, and it is attractive from a numerical point of view. We state:

**Theorem 6.5.1** *For a regular matrix $\mathbf{U} \in \mathbb{R}^{R\times R}$, consider the factorizations $\mathbf{U} = \mathbf{Q}^T\mathbf{R}'$ and $\mathbf{U}^T = \mathbf{R}''\mathbf{Z}^T$, in which $\mathbf{Q},\mathbf{Z} \in \mathbb{R}^{R\times R}$ are orthogonal and $\mathbf{R}',\mathbf{R}'' \in \mathbb{R}^{R\times R}$ are upper triangular. The rest class of the quotient set, defined by the equivalence relation $\mathbf{U} \sim \tilde{\mathbf{U}} \Leftrightarrow \tilde{\mathbf{U}} = \mathbf{UD}$, in which $\mathbf{D} \in \mathbb{R}^{R\times R}$ is arbitrary diagonal, is uniquely represented by the orthogonal pair ($\mathbf{Q}$,$\mathbf{Z}$).*

**Proof:** Any regular matrix in the rest class yields the same orthogonal pair, by construction. We now prove that, on the other hand, ($\mathbf{Q}$,$\mathbf{Z}$) uniquely defines the rest class of $\mathbf{U}$. From the definition of $\mathbf{Q}$ and $\mathbf{Z}$ we have:

$$(\mathbf{Q}\cdot\mathbf{Z})\cdot{\mathbf{R}''}^T = \mathbf{R}'. \tag{6.48}$$

Let $\mathbf{V}$ be equal to the orthogonal matrix $\mathbf{Q}\cdot\mathbf{Z}$. The lower triangular part of Eq. (6.48) can then be considered as a system of linear equations in the unknown coefficients of $\mathbf{R}''$:

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

$$\begin{bmatrix} v_{R-1,R-2} & v_{R-1,R-1} & v_{R-1,R} \\ v_{R,R-2} & v_{R,R-1} & v_{R,R} \end{bmatrix} \begin{bmatrix} r''_{R-2,R-2} \\ r''_{R-2,R-1} \\ r''_{R-2,R} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{6.49}$$



and so on. We remark that:

- a scaling of the rows of $\mathbf{R}''$ does not affect the homogeneous sets of equations (6.49), which is consistent with the fact that the columns of the matrices in the rest class are only determined up to a scalar multiple.
- none of the sets of equations in (6.49) can have a solution set of dimension higher than 1. This would mean that one or more columns of $\mathbf{R}''$ were underdetermined, which conflicts with the assumption that $\mathbf{U}$ is regular.

□

Remind that we have claimed, for identifiability (see Sect. 6.1), that $\mathbf{U}^{(1)}$ is regular.

We now propose an efficient scheme for the solution of the equations in (6.49). Let us fix the norms of the columns of $\mathbf{U}$ by setting the diagonal elements of $\mathbf{R}'$ equal to one. Then the homogeneous set (6.49) yields

$$\mathbf{V}_{R-r:R,R-r:R}\cdot({\mathbf{R}''}_{R-r,R-r:R})^T = [1\,0\,\ldots\,0]^T, \tag{6.50}$$

for $r = R-1, R-2, \ldots, 0$, in which we have used MATLAB-notation. For $r = R-1$, the first row of $\mathbf{R}''$ is found by bringing the orthogonal matrix $\mathbf{V}$ to the other side of the equation. The inverse for $r = R-2$ can relatively easily be derived from the inverse $\mathbf{V}^T$ at $r = R-1$. This procedure will be repeated for $r = R-2, R-3, \ldots, 0$. Let us briefly indicate the matrix $\mathbf{V}_{R-r:R,R-r:R}$, for a particular choice of $r$, by the matrix $\mathbf{W} \in \mathbb{R}^{I+1\times I+1}$, with inverse $\mathbf{W}^{-1}$. It can be verified that:

$$\mathbf{W}_{2:I+1,1}(\mathbf{W}^{-1})_{1,2:I+1} + \mathbf{W}_{2:I+1,2:I+1}(\mathbf{W}^{-1})_{2:I+1,2:I+1} = \mathbf{I}, \tag{6.51}$$

$$\mathbf{W}_{2:I+1,1}(\mathbf{W}^{-1})_{11} + \mathbf{W}_{2:I+1,2:I+1}(\mathbf{W}^{-1})_{2:I+1,1} = [0\,0\,\ldots\,0]^T. \tag{6.52}$$

From Eq. (6.52) we see that if the scalar $(\mathbf{W}^{-1})_{11}$ is close to zero, then $\mathbf{W}_{2:I+1,2:I+1}$ is close to singular, and hence the determination of the corresponding row of $\mathbf{R}''$ is ill-conditioned. If $(\mathbf{W}^{-1})_{11}$ is sufficiently large, then combination of Eqs. (6.51) and (6.52) shows that the inverse of $\mathbf{W}_{2:I,2:I}$ can be obtained from

$$\begin{aligned}(\mathbf{W}_{2:I+1,2:I+1})^{-1} &= (\mathbf{W}^{-1})_{2:I+1,2:I+1} \\ -\mathbf{W}_{2:I+1,2:I+1}&(\mathbf{W}^{-1})_{2:I+1,1}(\mathbf{W}^{-1})_{1,2:I+1}/(\mathbf{W}^{-1})_{11}.\end{aligned} \tag{6.53}$$

This transition formula can be used to solve the sets of Eq. (6.50) subsequently. The operations in Eq. (6.53) require $5I^2$ flops (a flop is any floating point operation). The multiplication of the inverse with the right-hand side of the equation additionally requires $2I^2$ flops. Hence the complete construction of $\mathbf{R}''$ requires $O(\sum_{i=1}^{R-1} 7i^2) = O(7R^3/6)$ flops.

Finally, we remark that counting the degrees of freedom at the beginning of this section revealed that in the unsymmetric case the transition to Eqs. (6.17-6.19) in fact reduces the number of unknowns by a half. One could also perform a simultaneous reduction to lower triangularity, by starting from a $QL$-factorization $\mathbf{U}^{(1)} = \mathbf{Q}'^T\mathbf{L}'$ and an $LQ$-factorization $\mathbf{U}^{(2)} = \mathbf{L}''\mathbf{Z}'^T$. For the reconstruction of $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$, the matrices $\mathbf{Q}'^T$ and $\mathbf{Z}'^T$ then play the same role for $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ as the matrix $\mathbf{Z}$ did for $\mathbf{U}^{(1)} = \mathbf{U}^{(2)}$ in the symmetric case. However, such an approach would double the computation time, as two sets of simultaneous generalized Schur decompositions would have to be solved.

## 6.6 Least-squares model fitting

The least-squares criterion associated to the upper triangularity in Eq. (6.20) and the least-squares criterion associated to the fit in Eq. (6.2) are only equivalent if the CANDECOMP-model is exactly satisfied. The simultaneous generalized Schur solution could be used as an interesting starting value for an additional minimization algorithm that matches both sides of Eq. (6.2) (see also Sect. 6.7). First we briefly describe an ALS descent algorithm [51]; as an alternative, we will additionally derive a closed-form gradient expression, with a view to gradient-based optimization algorithms.

### 6.6.1 ALS descent

Consider the following matrix representation of Eq. (6.2):

$$f = \|\mathbf{A}_{(n)}^T - \hat{\mathbf{U}}^{(n)} \cdot \hat{\Lambda} \cdot \tilde{\mathbf{U}}^{(n)}\|^2, \tag{6.54}$$

in which:

$$\begin{aligned}
\hat{\mathbf{U}}^{(n)} &\stackrel{\text{def}}{=} [\hat{U}_1^{(n)}\hat{U}_2^{(n)}\ldots\hat{U}_{I_n}^{(n)}], \\
\hat{\Lambda} &\stackrel{\text{def}}{=} \mathtt{diag}\{(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_R)\}, \\
\tilde{\mathbf{U}}^{(n)} &\stackrel{\text{def}}{=} \hat{\mathbf{U}}^{(1)} \odot \ldots \odot \hat{\mathbf{U}}^{(n-1)} \odot \ldots \odot \hat{\mathbf{U}}^{(n+1)} \odot \ldots \odot \hat{\mathbf{U}}^{(N)},
\end{aligned}$$

where $\odot$ denotes the Kathri-Rao product (App. A).

If we imagine that the matrices $\hat{\mathbf{U}}^{(1)}, \ldots, \hat{\mathbf{U}}^{(n-1)}, \hat{\mathbf{U}}^{(n+1)}, \ldots, \hat{\mathbf{U}}^{(N)}$ are fixed, then Eq. (6.54) is merely a quadratic expression in the components of the matrix $\hat{\mathbf{U}}^{(n)} \cdot \hat{\Lambda}$; its components can be estimated with a classical Linear Least Squares algorithm. Repeating this procedure for different mode numbers leads to an ALS algorithm for the (local) maximization of $f(\hat{\mathcal{A}})$: in each step the estimate of one of the matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ is optimized, while the other matrix estimates are kept constant. Over- and underflow can be avoided by normalizing the estimates of the columns $U_r^{(n)}$ $(1 \leqslant r \leqslant R; 1 \leqslant n \leqslant N)$ to unit-length.



### 6.6.2 Gradient

We now investigate which form the gradient of the cost function in Eq. (6.2) takes:

**Theorem 6.6.1** *Consider the least-squares cost function $f$, defined by Eq. (6.2), in which $\hat{\mathcal{A}}$ is defined in accordance with Eqs. (6.3). We assume that the vectors $\{\hat{U}_r^{(n)}\}_{1\leqslant r\leqslant R; 1\leqslant n\leqslant N}$ are normalized to unit-length. Then the gradient of $f$, with respect to $\hat{U}_r^{(n)}$, over the unit sphere, is given by*

$$\nabla_{\hat{U}_r^{(n)}} f = -2\hat{\lambda}_r\,(\mathbf{I} - \hat{U}_r^{(n)}\hat{U}_r^{(n)^T}). \tag{6.55}$$

**Proof:** Like the proofs of Theors. 5.3.9 and 6.4.5, the derivation involves a number of technicalities of the calculation of scalar products, and is therefore deferred to App. C. □

Given the estimates $\{\hat{U}_r^{(n)}\}_{1\leqslant r\leqslant R; 1\leqslant n\leqslant N}$ and the tensor $\mathcal{A}$ to be approximated, the optimal estimation of $\{\hat{\lambda}_r\}_{1\leqslant r\leqslant R}$ from Eqs. (6.2-6.3) is a classical linear least-squares problem. Hence the evolution of the coefficients $\{\hat{\lambda}_r\}_{1\leqslant r\leqslant R}$ can be implemented as a subspace tracking algorithm (e.g. [77]).

## 6.7 Simulations

In this section we illustrate the performance of the CANDECOMP-algorithms discussed in Sect. 6.4 with a number of simulations.

In a first series of experiments we generate tensors $\mathcal{A} \in \mathbb{R}^{3\times3\times3}$, of which the CANDECOMP-components will afterwards be estimated, in the following way:

$$\mathcal{A} = \tilde{\mathcal{A}}/\|\tilde{\mathcal{A}}\| + \sigma_N\tilde{\mathcal{N}}/\|\tilde{\mathcal{N}}\|, \tag{6.56}$$

in which $\tilde{\mathcal{A}}$ exactly satisfies the CANDECOMP-model:

$$\tilde{\mathcal{A}} = U_1^{(1)} \circ U_1^{(2)} \circ U_1^{(3)} + U_2^{(1)} \circ U_2^{(2)} \circ U_2^{(3)} + U_3^{(1)} \circ U_3^{(2)} \circ U_3^{(3)}. \tag{6.57}$$

The components in Eqs. (6.56-6.57) are generated as follows. First consider the $(3 \times 3)$-matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$, defined by Eq. (6.6). $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ are

derived from matrices of which the elements have been taken from a uniform distribution on the interval $[0,1)$, by setting the singular values equal to 3,2,1. For $\mathbf{U}^{(1)}$, three different sets of singular values will be considered: 3,2,1; 30,15,1; 100,50,1. The entries of $\tilde{\mathcal{N}}$ are drawn from a zero-mean unit-variance Gaussian distribution. For each particular choice of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$ and $\tilde{\mathcal{N}}$, the scalar $\sigma_N$ is varied between $1e-3$ and 1.

For each of the sets of singular values of $\mathbf{U}^{(1)}$, 50 independent samples of $\tilde{\mathcal{A}}$ are realized; for each of them 7 logarithmically equidistant values of $\sigma_N$ are considered. In each Monte Carlo simulation the following algorithms are run: (a) the two-sided Jacobi-algorithm, derived in Subsect. 6.4.3, (b) the two-sided Jacobi-algorithm, with the restart procedure described in Remark 6.4.2, (c) a least-squares matching of both sides of Eq. (6.2), for which the `leastsq` command of the Optimization Toolbox 1.0 of MATLAB 4.2 [186] has been used (the algorithm is an implementation of a Levenberg-Marquardt method [196]), initialized with the result of (b), and (d) the extended $QZ$-iteration, described in Subsect. 6.4.1. The algorithm (a) (which is also used in (b)) is terminated if a full sweep does not longer allow to reduce the cost function

$$h(\mathbf{Q},\mathbf{Z}) = \|\mathbf{Q}\cdot\mathbf{V}_1\cdot\mathbf{Z}\|_{LF}^2 + \|\mathbf{Q}\cdot\mathbf{V}_2\cdot\mathbf{Z}\|_{LF}^2 + \ldots + \|\mathbf{Q}\cdot\mathbf{V}_K\cdot\mathbf{Z}\|_{LF}^2 \quad (6.58)$$

with at least 0.01 %. The same termination criterion is used for a $Q$-step followed by a $Z$-step in the extended $QZ$-iteration. For the least-squares matching (c) a minimal precision of $1e-5$ for the optimal values of $h$ and the corresponding components is premised; the MATLAB routine maximally performs 2100 iteration steps.

To evaluate the accuracy of the different algorithms we will consider the quality of the estimate $\hat{\mathbf{U}}^{(1)}$ of $\mathbf{U}^{(1)}$; we assume that the columns of $\mathbf{U}^{(1)}$ have unit-length. In Fig. 6.7 the Root Mean Square Error (RMSE) between $\hat{\mathbf{U}}^{(1)}$ and $\mathbf{U}^{(1)}$ is plotted as a function of the noise level $\sigma_N$. For a given noise level and a given algorithm, the error measure has been computed as the average, over the different Monte Carlo simulations, of the Frobenius-norm $\|\hat{\mathbf{U}}^{(1)} - \mathbf{U}^{(1)}\|$; the ordering of the columns of $\hat{\mathbf{U}}^{(1)}$ that corresponds to the ordering of the columns of $\mathbf{U}^{(1)}$, has been determined as the ordering that minimizes the error. Algorithms (a), (b), (c) and (d) correspond to solid, dashed, dotted and dash-dot curves respectively. The upper, middle and lower curves correspond to a condition number of $\mathbf{U}^{(1)}$, equal to 100, 30 and 3 respectively.

The figure displays the expected performance degradation as the noise level and/or the condition number of $\mathbf{U}^{(1)}$ increases. The number of simulations is high enough to give a good picture: the efficiency (see Eq. (3.37)) of the estimated mean error ranges from at least $4e-6$ ($\sigma_N = 1e-3$) to typically $2.5e-2$ ($\sigma_N = 1$). We notice that on the average, the accuracy of methods (a), (b) and (d) is comparable. Moreover, in none of the simulations, the double-sided Jacobi-iteration technique ended up in a meaningless local optimum: the accuracy obtained by algorithms (a) and (b) was always in the same order of



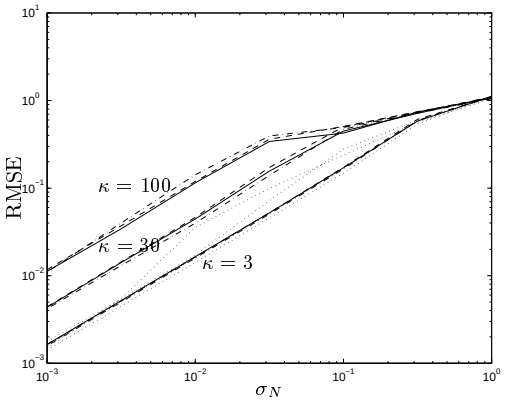



Figure 6.7: The RMSE for the estimation of $\mathbf{U}^{(1)}$, as a function of the noise level $\sigma_N$, for the two-sided Jacobi-algorithm (solid), the two-sided Jacobi-algorithm with restart procedure (dashed), an additional least-squares matching (dotted) and the extended $QZ$-algorithm (dashdot). The upper, middle and lower curves correspond to a condition number $\kappa$ of $\mathbf{U}^{(1)}$, equal to 100, 30 and 3 respectively.

magnitude. The figure also shows that an additional least-squares matching routine generally improves the accuracy, but that the marginal improvement becomes smaller as the CANDECOMP factors are better conditioned.

In Fig. 6.8 we have plotted the mean value of the cost function $h$ in Eq. (6.58) for the algorithms (a), (b) and (d). First, we notice that method (b) indeed shows a small marginal improvement of the overall upper triangular Frobenius-norm, with respect to method (a). Next, the figure shows that the extended $QZ$-iteration does not minimize the below-diagonal norm; this effect is more outspoken as the condition number of $\mathbf{U}^{(1)}$ is larger. On the other hand, it is clear from the discussion in Sect. 6.3 that, in the absence of noise, the theoretical solution is a stationary point of the extended $QZ$-algorithm; in Fig. 6.7 we see that the algorithm was still reliable in the presence of noise.

In a second series of experiments we illustrate the convergence behaviour of methods (a) and (d), and of methods (e1), (e2), (e3), (e4), consisting of the one-sided Givens-technique described in Subsect. 6.4.3, in which one alternates between sequences of respectively 1, 2, 3 or 4 cyclic sweeps on $\mathbf{Q}$ and $\mathbf{Z}$. For each of the sets of singular values of $\mathbf{U}^{(1)}$ (3,2,1; 30,15,1; 100,50,1), 100 independent samples of $\mathcal{A}$ are realized as before, with $\sigma_N = 0$. The different algorithms are now terminated if the instantaneous value of $h$ has been reduced below $1e-14$.

In Fig. 6.9 we have plotted the average evolution of the value of $h$ as a function

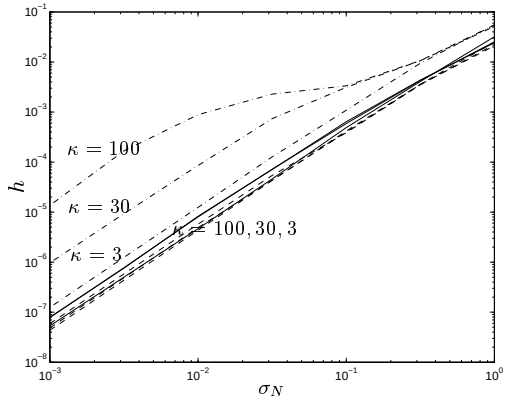



Figure 6.8: The mean value of $h$, defined in Eq. (6.58), as a function of the noise level $\sigma_N$, for the two-sided Jacobi-algorithm (solid), the two-sided Jacobi-algorithm with restart procedure (dashed) and the extended $QZ$-algorithm (dashdot). The upper, middle and lower curves correspond to a condition number $\kappa$ of $\mathbf{U}^{(1)}$, equal to 100, 30 and 3 respectively.

of the iteration step $l$ for algorithm (d) (for the scenario's with condition number $\kappa = 3, 30, 100$) and for algorithm (a) (only with condition number $\kappa = 3$, as will be motivated immediately). For these curves, the convergence speed is quasi-linear. The curves for the extended $QZ$-iteration have only been marginally affected by the chosen value of $\kappa$. On the other hand, it makes less sense to plot an average curve for the Jacobi-method in the cases where $\kappa = 30$ or 100, as the results can be strongly data-dependent. Namely, the convergence is still good in most cases, but for some particular instances of $\mathcal{A}$, the algorithm is observed to move through a swamp: apparently, like ALS iterations, Jacobi-iterations can be affected by swamps, though for well-conditioned problems they seem to form a minor issue. The extended $QZ$-algorithm appears to be less vulnerable, as the typical swamp behaviour has only been observed for one instance of $\mathcal{A}$ ($\kappa = 100$). Rather than plotting the remaining mean convergence curves, we show in a histogram how many iterations were needed to terminate the algorithm in the different Monte Carlo simulations, for the various set-ups (see Fig. 6.10, in which we have taken as a convention that experiments in which more than 100 iterations were needed, are added to the final histogram bin). Concerning Figs. 6.9 and 6.10, we finally remark that for 5 instances of $\mathcal{A}$, the extended $QZ$-algorithm was observed to start by increasing the value of $h$ to some extent, before actual convergence.

Fig. 6.11 is then the equivalent of Fig. 6.9, for methods (e1), (e2), (e3) and



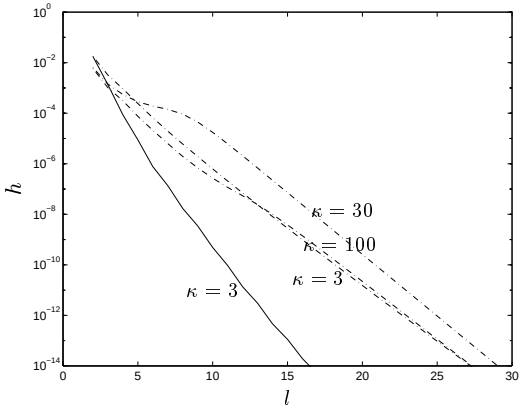



Figure 6.9: The evolution of $h$, defined in Eq. (6.58), as a function of the iteration step $l$, for the two-sided Jacobi-algorithm (solid) and the extended $QZ$-algorithm (dashdot). The parameter $\kappa$ is the condition number of $\mathbf{U}^{(1)}$.

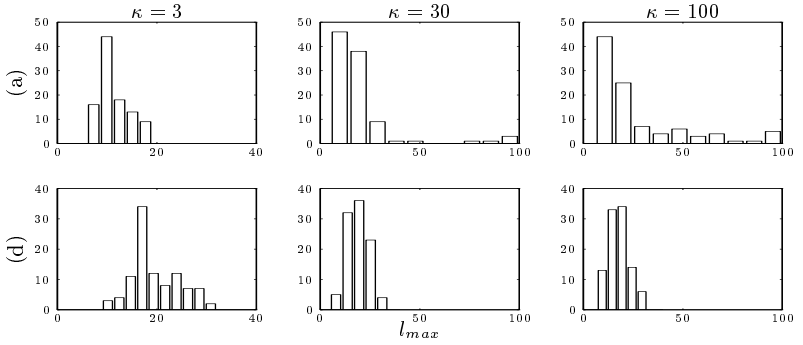



Figure 6.10: Histogram of 100 Monte Carlo simulations, showing the number of iterations required to reduce the value of $h$ (Eq. (6.58)) below $1e-14$, for methods (a) (two-sided Jacobi-algorithm) and (d)(extended $QZ$-algorithm), and a condition number $\kappa$ of $\mathbf{U}^{(1)}$, equal to 3, 30, or 100. Experiments in which more than 100 iterations were needed, have been added to the final histogram bin.

(e4) ($\kappa = 3$). For one-sided Givens iterations, swamps have been observed as well; Fig. 6.12 is the equivalent of Fig. 6.10. However, in both figures, for methods (e2), (e3) and (e4), the time scale was expanded with a factor 2, 3

and 4 respectively, to allow an easy comparison in terms of the actual number of Givens-iterations that have been calculated. It is clear that the simple cyclic scheme is on the average the most economical one; the extra cost of marginally improving the estimate of $\mathbf{Q}$, given $\mathbf{Z}$ or vice-versa, by performing extra sweeps, can in general not be motivated against the more significant improvement of the overall fit that can be achieved by switching between sweeps on $\mathbf{Q}$ and $\mathbf{Z}$. In all the Monte Carlo simulations, methods (a), (b) and (e1) to (e4) converged towards to global optimum, except for one case ($\kappa = 100$) in which methods (e2), (e3) and (e4) failed, because the convergence speed in a swamp became so low that the marginal improvement per iteration step was reduced to the order of magnitude of the numerical accuracy of MATLAB.

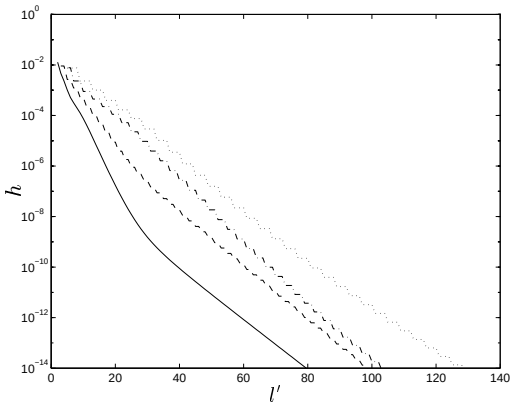



Figure 6.11: The evolution of $h$, defined in Eq. (6.58), as a function of the number of $\mathbf{Q}$-sweeps $l'$, for the one-sided Givens-technique in which one alternates between sequences of 1 (method (e1), solid), 2 (method (e2), dashed), 3 (method (e3), dashdot) or 4(method (e4), dotted) cyclic sweeps on $\mathbf{Q}$ and $\mathbf{Z}$. The condition number $\kappa$ of $\mathbf{U}^{(1)}$ is equal to 3.

## 6.8 Conclusions

In this chapter we have investigated the computation of the CANDECOMP, under the identifiability assumptions put forward in Sect. 6.1. Currently, the calculation of the factors mostly takes the form of an ALS descent algorithm, possibly initiated with an estimate obtained by matrix EVD.

We proposed to compute the CANDECOMP via simultaneous diagonalization, by equivalence or congruence, of a set of matrices. This is numerically more



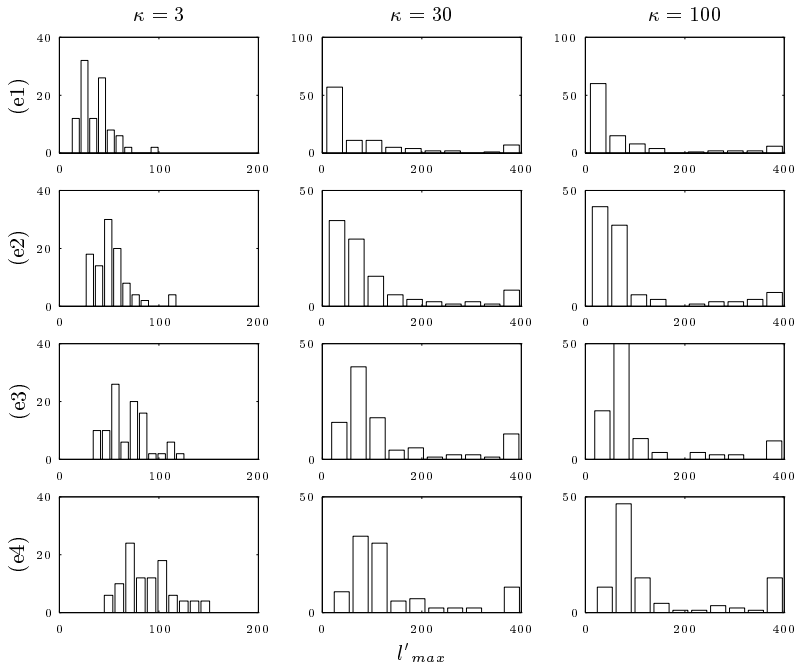



Figure 6.12: Histogram of 100 Monte Carlo simulations, showing the number of $\mathbf{Q}$-sweeps required to reduce the value of $h$ (Eq. (6.58)) below $1e-14$, for methods (e1), (e2), (e3) and (e4), and a condition number $\kappa$ of $\mathbf{U}^{(1)}$, equal to 3, 30, or 100. Experiments in which more than 400 iterations were needed, have been added to the final histogram bin.

reliable than the calculation of a single EVD, especially for moderate signal-to-noise ratios; actually the problem can be reformulated as a simultaneous matrix EVD.

Diagonalization by simultaneous congruence transformation was encountered as well in the derivation of an Analytical Constant Modulus Algorithm [266], where it was translated into a simultaneous generalized Schur decomposition and subsequently solved with an extended $QZ$-iteration scheme. In this chapter, we have focused on one-sided Givens-iterations, two-sided Jacobi-iterations and gradient-based optimization. It was shown that the computation of an elementary Givens rotation corresponds to the calculation of the dominant left singular vector of a matrix with two rows. The determination of an elementary Jacobi-rotation lead to a set of biquadratic equations; this problem has been

rewritten as the rooting of a polynomial of degree 8, followed by rooting a linear $S$-polynomial. As far as gradient-based optimization is concerned, we have derived an explicit expression for the gradient of the simultaneous Schur cost function, over the manifold of orthogonal matrices.

Gradient-based descent also forms an alternative for the ALS iterations in an additional least-squares model matching phase. With this respect, we have derived an explicit expression for the gradient of the quadratic error function, over the manifold of matrices with unit-norm columns. The marginal improvement of the fit that can be achieved with the extra least-squares matching, becomes smaller for better-conditioned problems, as indicated by simulations in which the performance of the different algorithms has been compared.

We have also proved that regular matrices of which the columns may be arbitrarily scaled (e.g. the CANDECOMP factor matrices, the mixing matrix in ICA, ... ), can be uniquely parametrized in terms of a pair of orthogonal matrices; this representation is interesting from a numerical point of view.

# Chapter 7

# Independent Component Analysis and (HO)SVD

*In this chapter we tackle a fundamental signal processing problem with tools from HOS and multilinear algebra, namely the problem of Independent Component Analysis (ICA), which is also known as Blind Source Separation (BSS).*

*In a nutshell, the goal of ICA is the decomposition of a multichannel dataset in an a priori unknown linear mixture of a priori unknown source signals, relying on the assumption that the source signals are mutually statistically independent. This concept is in fact a fine-tuning of the more well-known Principal Component Analysis (PCA), where one aims at the decomposition in a linear mixture of uncorrelated components - the weaker condition resulting in a rotational indeterminacy of the solution. ICA and PCA are not only related from a statistical point of view, but also from a computational perspective: where PCA involves the EVD of the covariance matrix, ICA (also) relies on tensorial generalizations of the EVD, applied to a higher-order cumulant.*

*First we discuss the topic of ICA in general terms and give some background material that is essential for the further developments. The concept of ICA itself is formally introduced in Sect. 7.1. In Sect. 7.2 we outline an important two-stage procedure that forms the basis of many algorithms. It is based on the diagonalization of the covariance matrix of the observations, like in PCA, followed by the tensorial diagonalization of the higher-order cumulant (in Chapter 10 we will follow an other strategy, resorting only to the cumulant). In Sect. 7.3 we discuss the uniqueness of the solution. The links between PCA and ICA are further explicited in Sect. 7.4. Sect. 7.5 gives a definition of contrast functions, as several ICA-techniques are based on, or can be interpreted as, the explicit optimization of such a function. Sect. 7.6 deals with the*

*performance of ICA-methods. First, we define some measures of performance; next, we explain that the quality of source separation can (for sufficiently low noise levels) be independent of the actual way in which the source signals were mixed, provided that the estimator exhibits a similar symmetry as the ICA-problem itself; finally, we pay attention to the influence of the quality of the PCA-stage on the overall performance - it appears that errors introduced at the PCA-level, cannot be eliminated by the higher-order step. Sect. 7.8 contains a literature overview.*

*Taking the discussion of Chapter 4 in mind, as well as the general introduction to ICA, we come to the conclusion that HOEVD of a cumulant tensor can be used to solve the ICA-problem. This new technique is analyzed in Sect. 7.7.*

*In Sect. 7.9 we wonder if the "source separations", obtained with a mere PCA in the past, are necessarily wrong. We prove that PCA can give a satisfactory quality of separation for uncorrelated sources with strongly different powers.*

## 7.1 Independent Component Analysis

Assume the following basic linear statistical model:

$$Y = \mathbf{M}X + N = \tilde{Y} + N, \tag{7.1}$$

in which $Y \in \mathbb{R}^I\,(\mathbb{C}^I)$ is called the *observation vector* or *output vector*, $X \in \mathbb{R}^J\,(\mathbb{C}^J)$ is the *source vector* and $N \in \mathbb{R}^I\,(\mathbb{C}^I)$ represents additive *noise*. $\tilde{Y} \in \mathbb{R}^I\,(\mathbb{C}^I)$ is the *signal part* of the observations. $\mathbf{M} \in \mathbb{R}^{I\times J}\,(\mathbb{C}^{I\times J})$ is the *transfer matrix* or *mixing matrix*. The columns $M_j$ of $\mathbf{M}$ are often called *steering vectors*, its range as the *signal subspace*.

The concept of Independent Component Analysis can be formulated as follows:

*The goal of Independent Component Analysis (ICA) consists of the estimation of the transfer matrix* $\mathbf{M}$ *and/or the corresponding realizations of the source vector* $X$*, given only realizations of the output vector* $Y$*, under the following assumptions:*

- *The columns of* $\mathbf{M}$ *are linearly independent.*
- *The components of* $X$ *are mutually statistically independent, as well as statistically independent from the noise components.*

The second assumption is the key ingredient for ICA. It is a very strong hypothesis, but also quite natural in lots of applications. The first assumption holds in a generical sense (regardless of the fact that an ill-conditioned mixing matrix can make the ICA-problem heavier from a numerical point of view, as will be explained later on). Merely resorting to these two assumptions it is impossible to distinguish between the signal and the noise term in Eq. (7.1). Hence the source signals will be estimated as $\hat{X}$, by a simple matrix multiplication:

$$\hat{X} = \mathbf{W}^H Y. \tag{7.2}$$

As an example $\mathbf{W}^H$ can take the form of $\hat{\mathbf{M}}^\dagger$, in which $\hat{\mathbf{M}}$ is an estimate of the transfer matrix. More generally various beamforming strategies [274] can be applied (see Sect. 7.6).

The ICA-problem is also addressed in literature under the labels *Blind Source Separation* (BSS), *Signal Copy*, *Waveform Preserving Estimation*, etc. However the assumptions on which the solution strategies are based, sometimes differ from paper to paper.



## 7.2 A two-stage procedure

The ICA-problem is usually solved by a two-stage algorithm, consisting of a second- and a higher-order step. In this section we will explain the technique in general terms; at this point the noise term in Eq. (7.1) is mostly neglected.

For the identification of the transfer matrix $\mathbf{M} \in \mathbb{C}^{I\times J}$ we will consider decompositions of the type:

$$\mathbf{M} = \mathbf{TQ}, \tag{7.3}$$

in which $\mathbf{T} \in \mathbb{C}^{I\times J}$ is regular and $\mathbf{Q} \in \mathbb{C}^{J\times J}$ is unitary. Examples are the LQ-decomposition $\mathbf{M} = \mathbf{LQ}$, the SVD $\mathbf{M} = (\mathbf{US})\mathbf{Q}$, etc.

In a first step $\mathbf{T}$ will be determined from the second-order statistics of the output $Y$. The remaining degree of freedom, the unitary factor $\mathbf{Q}$, will be recovered from the higher-order statistics of $Y$. An outline of the technique is presented as Alg. 7.2.

### 7.2.1 Step 1: prewhitening

The goal of the prewhitening step is to transform $Y$ into an other stochastic vector, $Z$, with unit covariance. This involves the multiplication of $Y$ with the inverse of a square root of its covariance matrix $\mathbf{C}^Y$. When $J < I$, a projection of $Y$ onto the signal subspace is carried out.

The covariance matrix $\mathbf{C}_2^Y$ takes the following form:

$$\mathbf{C}_2^Y = \mathbf{M}\mathbf{C}_2^X\mathbf{M}^H, \tag{7.4}$$

**Algorithm 7.2**
*PCA-based ICA*

Given: $N$ samples $\{Y_n\}_{1 \leqslant n \leqslant N}$ of $Y = \mathbf{M}X + N$, as in Eq. (7.1) ($Y, N \in \mathbb{C}^I$, $X \in \mathbb{C}^J$, $\mathbf{M} \in \mathbb{C}^{I \times J}$). Call $\mathbf{A}_Y = [Y_1 Y_2 \dots Y_N]$.

1. Prewhitening stage (PCA):
   - Compute sample mean $\hat{M}_Y$ from $\mathbf{A}_Y$.
     Define $\tilde{\mathbf{A}}_Y = [Y_1 - \hat{M}_Y Y_2 - \hat{M}_Y \dots Y_N - \hat{M}_Y]/\sqrt{N-1}$.
   - SVD of $\tilde{\mathbf{A}}_Y$: $\tilde{\mathbf{A}}_Y = \mathbf{E} \cdot \mathbf{D} \cdot \tilde{\mathbf{Q}}$, with $\mathbf{E} \in \mathrm{St}(J, I)$, $\tilde{\mathbf{Q}}^H \in \mathrm{St}(J, N)$, and $\mathbf{D} \in \mathbb{R}^{I \times J}$ positive (semi)definite pseudo-diagonal.
   - If $N$ is spatially white and Gaussian, with variance $\sigma_N^2$, replace the diagonal entries of $\mathbf{D}$ by $\sqrt{d_{jj}^2 - \sigma_N^2}$ $(1 \leqslant j \leqslant J)$.

   (Subsect. 7.2.1.)
2. Higher-order stage:
   - Compute sample cumulant $\hat{\mathcal{C}}_4^Y$ from $\mathbf{A}_Y$.
   - $\hat{\mathcal{C}}_4^Z = \hat{\mathcal{C}}_4^Y \times_1 (\mathbf{D}^\dagger \cdot \mathbf{E}^H) \times_2 (\mathbf{D}^\dagger \cdot \mathbf{E}^T) \times_3 (\mathbf{D}^\dagger \cdot \mathbf{E}^T) \times_4 (\mathbf{D}^\dagger \cdot \mathbf{E}^H)$.
   - (Approximate) diagonalization ($\mathbf{Q}$ is unitary):

$$\hat{\mathcal{C}}_4^Z = \hat{\mathcal{C}}_4^X \times_1 \mathbf{Q} \times_2 \mathbf{Q}^* \times_3 \mathbf{Q}^* \times_4 \mathbf{Q}.$$

     - HOEVD (Sect. 7.7).
     - Maximal Diagonality (MD) (Chapter 8).
     - Simultaneous Third-Order Tensor Diagonalization (STOTD) (Chapter 9).
     - Joint Approximate Diagonalization of Eigenmatrices (JADE) (Sect. 9.3).
   - Mixing matrix estimate: $\hat{\mathbf{M}} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{Q}$.
   - Source separation by means of MVDR, LCMV, ... beamforming.

in which the covariance $\mathbf{C}_2^X$ of $X$ is diagonal, since the source signals are uncorrelated. Assuming that the source signals have unit variance, possibly by rescaling the steering vectors, we have:

$$\mathbf{C}_2^Y = \mathbf{M}\mathbf{M}^H. \tag{7.5}$$

From this equation it is clear that prewhitening allows to determine the matrix $\mathbf{T}$, while the factor $\mathbf{Q}$ remains unknown. The computation takes the form of



diagonalizing the output covariance by congruence transformation:

$$\mathbf{C}_2^Y = \mathbf{M}\mathbf{M}^H = (\mathbf{T}\mathbf{Q})(\mathbf{T}\mathbf{Q})^H = \mathbf{T}\mathbf{T}^H. \tag{7.6}$$

The fact that any square-root decomposition can be used for the prewhitening, corresponds to the degree of freedom in the factorization of the transfer matrix proposed in Eq. (7.3). E.g. an LQ-decomposition leads to a Cholesky factorization of $\mathbf{C}_2^Y$, an SVD of $\mathbf{M}$ corresponds to an EVD of $\mathbf{C}_2^Y$, etc.

The diagonalization of $\mathbf{C}_2^Y$ by a symmetric EVD, and the corresponding determination of $Z$, is known as *Principal Component Analysis* (PCA):

$$\mathbf{C}_2^Y = \mathbf{E}\mathbf{D}^2\mathbf{E}^H = (\mathbf{E}\mathbf{D})(\mathbf{E}\mathbf{D})^H. \tag{7.7}$$

This type of whitening is preferable, since it allows to handle easily and accurately the case of singular or ill-conditioned covariance matrices, met when $J < I$, by projection on the space spanned by the $J$ dominant eigenvectors.

The effect of the additive noise term $N$ can be neutralized by replacing $\mathbf{C}_2^Y$ by the noise-free covariance $\mathbf{C}_2^Y - \mathbf{C}_2^N$. In the case of white noise, $\mathbf{C}_2^N$ takes the form of $\sigma_N^2\mathbf{I}$, in which $\sigma_N^2$ is the noise variance on each data channel. In a more-sensors-than-sources setup, $\sigma_N^2$ can be estimated as the mean of the "noise-eigenvalues", i.e. the smallest $I - J$ eigenvalues, of $\mathbf{C}_2^Y$.

When the output covariance is found as

$$\mathbf{C}_2^Y = \tilde{\mathbf{A}}_Y \tilde{\mathbf{A}}_Y^H, \tag{7.8}$$

in which $\tilde{\mathbf{A}}_Y$ is an $(I \times N)$-dimensional dataset containing $N$ realizations of $Y$, divided by $\sqrt{N}$ after subtraction of the sample mean, and possibly rescaled in accordance with Sect. 3.5, then it is preferable from a numerical point of view to compute the factor $\mathbf{T}$ without explicit calculation of the product (7.8). Instead $\mathbf{T}$ should be estimated from a direct factorization of $\tilde{\mathbf{A}}_Y$ of the form $\tilde{\mathbf{A}}_Y = \mathbf{T}\tilde{\mathbf{Q}}$, in which $\tilde{\mathbf{Q}} \in \mathbb{C}^{I \times N}$ has mutually orthonormal rows. E.g. the PCA-factor ($\mathbf{E}\mathbf{D}$) can be obtained in a numerically reliable way from the SVD of $\tilde{\mathbf{A}}_Y$ [130]:

$$\tilde{\mathbf{A}}_Y = \mathbf{E}\mathbf{D}\tilde{\mathbf{Q}}. \tag{7.9}$$

### 7.2.2 Step 2: determination of Q from the higher-order statistics of $Y$

In this section we will explain how $\mathbf{Q}$ can be estimated by decomposing a higher-order cumulant tensor of $Y$. As an example the fourth-order case will be addressed.

The fourth-order cumulant of the standardized random vector $Z$ is given by

$$\mathcal{C}_4^Z = \mathcal{C}_4^X \times_1 \mathbf{Q} \times_2 \mathbf{Q}^* \times_3 \mathbf{Q}^* \times_4 \mathbf{Q}. \tag{7.10}$$

This tensor is related to the fourth-order output cumulant by the multilinearity property 3.2.3:

$$\mathcal{C}_4^Z = \mathcal{C}_4^Y \times_1 \mathbf{T}^\dagger \times_2 (\mathbf{T}^*)^\dagger \times_3 (\mathbf{T}^*)^\dagger \times_4 \mathbf{T}^\dagger. \tag{7.11}$$

The key observation is that the source cumulant $\mathcal{C}_4^X$ is theoretically a diagonal tensor, since the source signals are also higher-order independent (Prop. 3.2.7), such that Eq. (7.10) is in fact a symmetric EVD-like tensor decomposition. This decomposition is unique if at most one diagonal element equals 0, as will be explained in Sect. 7.3. However, as mentioned in Sect. 4.2, it is in general impossible to diagonalize a higher-order tensor by means of unitary transformations. The way in which the estimation error (finite datalength, additive noise, etc.) of *sample* cumulants is dealt with, allows to distinguish different solution strategies. New techniques are presented in Chapters 7, 8 and 9; Sect. 7.8 contains a literature overview.

## 7.3 Identifiability

It is impossible to determine the norm of the steering vectors in Eq. (7.1), since a rescaling of these vectors can be compensated by the inverse scaling of the source signal values. Similarly the ordering of the source signals, having no physical meaning, cannot be identified. In general, these indeterminacies are the only way in which an ICA-solution is not unique, as is clear from the following theorem:

**Theorem 7.3.1** *Let $X$ be a $J$-dimensional vector with independent components, of which at most one is Gaussian, and whose densities are not reduced to a point-like mass. Let $\mathbf{Q} \in \mathbb{C}^{J\times J}$ be unitary, and consider $Z = \mathbf{Q}X$. Then the components $z_j$ are mutually independent iff $\mathbf{Q} = \Lambda\mathbf{P}$, in which $\Lambda$ is a diagonal matrix of unit-norm elements and $\mathbf{P}$ is a permutation.*

**Proof:** The theorem is a direct consequence of Darmois's Theorem [75] and can be found in [64]. □

The higher-order cumulants of Gaussian components vanish (Prop. 3.2.9), and hence these components cannot be separated in an essentially unique way. We have:



**Theorem 7.3.2** *Let $X$ be a $J$-dimensional non-deterministic Gaussian random vector. Let $\mathbf{M} \in \mathbb{C}^{J\times J}$ have linearly independent columns, and consider $Y = \mathbf{M}X$. Then the components $y_i$ are mutually independent iff $\mathbf{M} = \Lambda_\mathbf{1}\mathbf{Q}\Lambda_\mathbf{2}$, with $\Lambda_\mathbf{1}$, $\Lambda_\mathbf{2}$ diagonal and $\mathbf{Q}$ unitary.*

**Proof:** The proof is easily established by claiming that both the covariance matrix of $Y$ and $X$ are diagonal [64]. □

Analysis of Theor. 7.3.1 from a tensor algebraic point of view reveals that its conditions can be weakened to statistical independence at some order $N > 2$. This means that the matrix $\mathbf{Q}$ can be obtained in an essentially unique way by the diagonalization of an $N$th order output cumulant, iff the corresponding source cumulant has at most one 0 on the diagonal:

**Theorem 7.3.3** *Let the $N$th order super-symmetric tensor $\mathcal{C} \in \mathbb{C}^{J\times J\times\ldots\times J}$ be given by*

$$\mathcal{C} = \mathcal{D} \times_1 \tilde{\mathbf{Q}}_1 \times_2 \tilde{\mathbf{Q}}_2 \ldots \times_N \tilde{\mathbf{Q}}_N, \tag{7.12}$$

*in which $\mathcal{D} \in \mathbb{C}^{J\times J\times\ldots\times J}$ is diagonal, containing at most one 0 on the diagonal, and $\tilde{\mathbf{Q}}_n$ equals either $\mathbf{Q}$ or $\mathbf{Q}^*$, with $\mathbf{Q} \in \mathbb{C}^{J\times J}$ unitary. $\mathcal{C}$ can be decomposed by the same model in terms of $\mathcal{D}'$ and $\mathbf{Q}'$ iff:*

- *$\mathbf{Q}' = \mathbf{Q}\Lambda\mathbf{P}$, in which $\Lambda$ is a diagonal matrix of unit-norm elements and $\mathbf{P}$ is a permutation, and*
- *$\mathcal{D}'$ is related to $\mathcal{D}$ in the inverse way:*

$$\mathcal{D}' = \mathcal{D} \times_1 \tilde{\mathbf{U}}_1^{-1} \times_2 \tilde{\mathbf{U}}_2^{-1} \ldots \times_N \tilde{\mathbf{U}}_N^{-1}, \tag{7.13}$$

*where $\tilde{\mathbf{U}}_n$ represents $\mathbf{U}$ or $\mathbf{U}^*$, in accordance to Eq. (7.12), with $\mathbf{U} = \Lambda\mathbf{P}$.*

**Proof:** It is clear that if $\mathcal{D}$ and $\mathbf{Q}$ are replaced by $\mathcal{D}'$ and $\mathbf{Q}'$ as specified by the theorem, Eq. (7.12) is still valid.
We prove that the factor $\mathbf{U}$ is the only way in which decomposition (7.12) is not unique. Associate to $\mathcal{C}$ the linear mapping obtained by summation over the $N-2$ last indices (Sect. 2.8). Consider the $(N-2)$th order tensor $\mathcal{A} \in \mathbb{C}^{J\times J\times\ldots\times J}$, defined by

$$\mathcal{A} = \mathcal{S} \times_1 \tilde{\mathbf{Q}}_3^H \times_2 \tilde{\mathbf{Q}}_4^H \ldots \times_{N-2} \tilde{\mathbf{Q}}_N^H, \tag{7.14}$$

in which the $(N-2)$th order tensor $\mathcal{S} \in \mathbb{C}^{J\times J\times\ldots\times J}$ is a diagonal tensor (or a vector, for $N = 3$), chosen such that the products $d_{jj\ldots j}s_{jj\ldots j}$ are all different. $\mathcal{A}$ is mapped to the matrix $\mathbf{B}$, given by

$$\mathbf{B} = \tilde{\mathbf{Q}}_1\check{\mathbf{D}}\tilde{\mathbf{Q}}_2, \tag{7.15}$$

in which $\tilde{\mathbf{D}}$ is a diagonal matrix containing the elements $d_{jj\ldots j}s_{jj\ldots j}$ on the diagonal. Possibly up to the sign of the elements $\tilde{d}_{jj}$, Eq. (7.15) is an SVD of $\mathbf{B}$. Hence $\mathbf{Q}$ can be at most undetermined up to a factor $\Lambda\mathbf{P}$, as specified by the theorem. □

As a summary, we can state that ICA does not allow to estimate the transfer matrix $\mathbf{M}$ as such, but that under the conditions of Theor. 7.3.1 it determines the appropriate rest class of the quotient set, defined by the equivalence relation $\mathbf{M} \sim \mathbf{M}' \Leftrightarrow \mathbf{M}' = \mathbf{M}\Lambda\mathbf{P}$. Generically a unique representative of this rest class can be obtained, assuming that:

- the source components have unit variance,
- the steering vectors are ordered by decreasing Frobenius-norm,
- the entry of largest modulus in each steering vector is positive real.

**Remark 7.3.4** Remark that Theor. 7.3.1, 7.3.2 and 7.3.3 are established under noise-free conditions and assume a perfect knowledge of the output statistics. It remains an open question in the literature under which assumptions the results are still valid for sample cumulants, affected by additive noise. However no examples of non-uniquess have been reported so far. ◇

## 7.4 PCA versus ICA

From Sect. 7.2 it is clear that ICA is in fact the natural way to fine-tune PCA. Both statistical techniques are linked with the algebraic concepts of EVD and SVD in a remarkable way.

In the "classical" *second*-order statistical problem of PCA, the problem of interest is to remove the correlation from data obtained as a linear mixture of independent source signals. The key tool to realize this, comes from "classical" *linear* algebra: it is the EVD of the observations covariance or, numerically, the SVD of the dataset.

In this way the signal subspace, or more precisely the factors $\mathbf{U}$ and $\mathbf{S}$ in the SVD of the mixing matrix $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, can be identified. The fact that the steering vectors and the source signals can only be found up to an orthogonal transformation, is sometimes called the *rotation invariance* of PCA. Uniqueness is usually obtained by adding (often artificial) constraints, like mutual orthogonality of the columns of the mixing matrix estimate.

In the more recent problem of ICA, one also aims at the removal of higher-order dependence, which additionally involves the use of *Higher*-Order Statistics. It



appears that the unknown orthogonal factor can be estimated by a *multilinear* EVD. By inverting the transfer matrix, the original source signals can be estimated.

Although any dataset can be decorrelated by a linear transformation, it is not always possible to express it as a linear combination of independent contributions. This corresponds to the fact that Eq. (7.10) is overdetermined for an arbitrary cumulant tensor $\mathcal{C}$. Whether ICA is useful, depends on the context. On the other hand the extent to which $\mathcal{C}$ can be diagonalized, gives an indication of how independent the estimated source signals are (see also Sect. 3.2).

## 7.5 Contrast functions

In several ICA-techniques the transfer matrix is estimated through optimization of a so-called *contrast function*. The concept and the term have been introduced by Donoho and Gassiat in the context of scalar blind deconvolution [111, 128]. In a consistent way, the statistical framework of contrast functions has been further developed by Comon [64].

**Definition 7.5.1 (Contrast)** *A contrast over a set* $\mathtt{M}$ *of matrices is a mapping* $\Psi$ *from the set of probability densities* $\{p_X | X \in \mathbb{C}^I \text{ or } \mathbb{R}^I\}$ *to* $\mathbb{R}$ *that satisfies the following requirements:*

1. $\Psi(p_X)$ *does not change if the components of* $X$ *are permuted or scaled:*

$$\Psi(p_X) = \Psi(p_{\Lambda\mathbf{P}X}), \tag{7.16}$$

   *in which* $\Lambda$ *is a diagonal matrix and* $\mathbf{P}$ *a permutation matrix.*

2. *If* $X$ *has mutually independent components, then* $\Psi(p_{\mathbf{M}X}) \leqslant \Psi(p_X)$ *for every matrix* $\mathbf{M} \in \mathtt{M}$.

3. *If* $X$ *has mutually independent components, then* $\Psi(p_{\mathbf{M}X}) = \Psi(p_X)$ *only if* $\mathbf{M} = \Lambda\mathbf{P}$, *with* $\Lambda$ *a diagonal matrix and* $\mathbf{P}$ *a permutation matrix.*

The first condition reflects that contrast optimization should show the same indeterminacies as ICA itself. The basis of contrast optimization is established by condition 2; uniqueness is obtained by the third condition. Property 3 is referred to as the *discrimination property* of the contrast function.

The notation $\Psi(p_X)$ will often be abbreviated as $\Psi(X)$.

**Example 7.5.2** The function $\Psi(X) = -\mathtt{I}(X)$, in which $\mathtt{I}(X)$ is the average mutual information of the components of the random vector $X \in \mathbb{C}^J$,

$$\mathtt{I}(X) = \int p_X(x) \log \frac{p_X(x)}{\prod p_{x_j}(x_j)} dx_1 dx_2 \dots dx_J, \tag{7.17}$$

is a contrast over all unitary matrices, for all $X$ with unit covariance [64]. Hence ICA could take the form of whitening, followed by optimization of $\Psi(X)$. However, $\Psi(X)$ involves all the higher-order statistics of $X$; further we will discuss contrasts that are more interesting from a computational point of view. □

## 7.6 Performance of ICA

### 7.6.1 Measures of performance

The quality of source separation and reconstruction is naturally formulated in terms of beamforming performance. In terms of Eq. (7.1) and (7.2), and assuming that the sources are estimated in the right order (for notational convenience), we define the following indices of performance:

$$\mathrm{SNR}_i \stackrel{\mathrm{def}}{=} \frac{\sigma_{x_i}^2 |W_i^H M_i|^2}{W_i^H \mathbf{C}_2^N W_i}, \tag{7.18}$$

$$\mathrm{SIR}_{ij} \stackrel{\mathrm{def}}{=} \frac{\sigma_{x_i}^2 |W_i^H M_i|^2}{\sigma_{x_j}^2 |W_i^H M_j|^2}, \tag{7.19}$$

$$\mathrm{SINR}_i \stackrel{\mathrm{def}}{=} \frac{\sigma_{x_i}^2 |W_i^H M_i|^2}{W_i^H (\mathbf{C}_2^Y - \sigma_{x_i}^2 M_i M_i^H) W_i} \tag{7.20}$$

(formulated for a single ICA-problem; in a series of Monte Carlo runs, averaged values are considered). The first index is the Signal to Noise Ratio (SNR) of the estimate of the $i$th source. The second index is a Signal to Interference Ratio (SIR), quantifying the contamination by the $j$th source of the $i$th source estimate. Finally, the third index is the Signal to Interference-plus-Noise Ratio (SINR) of the $i$th source estimate. Often we will only consider the numerator of $\mathrm{SIR}_{ij}$, indicating to which extent $\mathbf{W}^H$ approximates $\mathbf{M}^\dagger$. Note that for a good estimate the denominator approximates 1, which allows us to define an approximate Interference to Signal Ratio (ISR), as $\mathrm{ISR}_{ij} \stackrel{\mathrm{def}}{=} \sigma_{x_j}^2 |W_i^H M_j|^2$.

The SINR is optimized by a *minimum variance distortionless response* (MVDR) filter, given by:

$$\mathbf{W} = (\mathbf{C}_2^Y)^\dagger \mathbf{M}. \tag{7.21}$$



This implements a matched filter. On the other hand the mutual interference of the sources can be cancelled by implementing a *linear constrained minimum variance* (LCMV) filter:

$$\mathbf{W} = (\mathbf{C}_2^{\tilde{Y}})^\dagger \mathbf{M}. \tag{7.22}$$

In practice these filters will be approximated in terms of sample statistics and the estimate of the mixing matrix.

In case the estimation of the transfer matrix is of primary importance, it is natural to express the ICA-performance as the least-squares difference between $\mathbf{M}$ and its estimate $\hat{\mathbf{M}}$. We implicitly assume that the columns of $\hat{\mathbf{M}}$ are optimally ordered and scaled. In a Monte Carlo experiment, the Root Mean Square Error (RMSE) $\sqrt{\mathtt{E}\|\mathbf{M} - \hat{\mathbf{M}}\|^2}$ is considered.

### 7.6.2 Invariance and uniform performance

For a fixed distribution of the source signals, and sufficiently low noise level, the probability distribution of the observations depends only on the mixing matrix $\mathbf{M}$. Multiplication of the observations with a regular matrix $\mathbf{A}$ corresponds to a similar transformation of the unknown parameters, now given by the matrix $\mathbf{AM}$. The group structure induced by left matrix multiplication can be interpreted as a symmetry, inherent to the blind source separation problem. Estimation procedures that show a consistent symmetry are qualified as invariant. For more details on the general theory of invariance we refer to [179], Chapter 6.

For an ICA-estimator to be invariant its estimates should be transformed in the same way as the mixing matrix (up to the indeterminacies mentioned in Sect. 7.3). Hence invariant estimators do not extract the transfer matrix in prespecified, normalized way. A natural way to obtain an invariant ICA-estimator, is by expressing the estimation procedure in terms of the output of the separator (i.e. in terms of the source estimates). Examples of invariant estimators are the Maximum Likelihood estimator, the Constant Modulus estimator, and several contrast optimization methods [42].

Call $\hat{\mathbf{M}}$ the estimate of the transfer matrix, obtained from a dataset $\mathbf{Y}$ by an invariant ICA algorithm. For convenience we assume that the original transfer matrix $\mathbf{M}$ is square. The estimated transfer matrix, obtained from the transformed dataset $\mathbf{AY}$ is called $\mathbf{B}$. We have:

$$\mathbf{B}^{-1}\mathbf{AM} = (\mathbf{A}\hat{\mathbf{M}})^{-1}\mathbf{AM} = \hat{\mathbf{M}}^{-1}\mathbf{M}, \tag{7.23}$$

which does not depend on $\mathbf{A}$. Hence, the ISR of invariant ICA estimators does not depend on the mixing matrix; it can even be computed using the unmixed

dataset (still under the no-noise assumption). This property is called *uniform performance* of invariant estimators [42].

### 7.6.3 Effect of prewhitening

Estimation errors can have different reasons, such as a model misfit, finite datalength effects, perturbation by noise, etc. The second-order step in the scheme of Sect. 7.2 has the disadvantage w.r.t. the higher-order step that, at a conceptual level, its partial results can be affected by additive Gaussian noise. The error introduced at this stage cannot be compensated by the higher-order step, and introduces an upper-bound to the performance of the ICA-algorithm. We will discuss this effect in terms of the ISR and in terms of the least-squares error in the transfer matrix estimate. The derivations apply to any ICA-procedure of the type of Sect. 7.2. For notational convenience we disregard the indeterminacies described in Sect. 7.3.

First we discuss the effect of the prewhitening on the ISR. The derivation is based on the following lemma:

**Lemma 7.6.1** *For any matrix* $\mathbf{A} \in \mathbb{C}^{J\times J}$ *we have:*

$$\left(\|\mathbf{A}\|_{LFs}^2 + \|\mathbf{A}^H\|_{LFs}^2\right)\|\mathbf{A}\|^2 \geqslant \|\mathbf{A}^H\mathbf{A}\|_{LFs}^2. \tag{7.24}$$

*The equality sign holds iff* $\mathbf{A}$ *is diagonal, in which case both sides vanish.*

(For a definition of $\|\mathbf{A}\|_{LFs}$, we refer to App. A.)

**Proof:** The proof is given in App. C. The derivation yields a closed-form expression for the difference between the two sides of Eq. (7.24). □

Let the estimate of the factor $\mathbf{T}$ in Sect. 7.2, obtained by prewhitening, be given by $\hat{\mathbf{T}}$. The matrix $\hat{\mathbf{T}}^\dagger\mathbf{M}$ is theoretically unitary; however due to imprecisions in the PCA-stage this might not be the case in practice. Consider the SVD

$$\hat{\mathbf{T}}^\dagger\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^H, \tag{7.25}$$

in which the factor $\mathbf{S}\mathbf{V}^H$ describes the effect of the prewhitening error.

Now let $\hat{\mathbf{Q}}$ represent the estimate of the factor $\mathbf{Q}$ in Sect. 7.2. Adopting the convention that the sources have unit variance, we can express the ISR as:

$$\mathrm{ISR}_{ij} = |(\hat{\mathbf{Q}}^H\hat{\mathbf{T}}^\dagger\mathbf{M})_{ij}|^2 = |(\tilde{\mathbf{U}}\mathbf{S}\mathbf{V}^H)_{ij}|^2, \tag{7.26}$$

in which $\tilde{\mathbf{U}} = \hat{\mathbf{Q}}^H\mathbf{U}$ is unitary. Now we prove that the quality of ICA is bounded by the quality of the prewhitening in the following way:



**Theorem 7.6.2 (ISR lower-bound)** *In the notation specified above, we have:*

$$\sum_{i\neq j} ISR_{ij} \geqslant \frac{\|(\hat{\mathbf{T}}^\dagger\mathbf{M})^H(\hat{\mathbf{T}}^\dagger\mathbf{M})\|_{LF}^2}{\|\hat{\mathbf{T}}^\dagger\mathbf{M}\|^2}. \tag{7.27}$$

*The equality sign holds only for a perfect reconstruction of the transfer matrix, in which case both sides vanish.*

**Proof:** The theorem can be translated in terms of Lemma 7.6.1, by defining $\mathbf{A} \stackrel{\text{def}}{=} \tilde{\mathbf{U}}\mathbf{S}\mathbf{V}^H$, in accordance with Eq. (7.26). □

Starting from the polar decomposition [130] of $\hat{\mathbf{T}}^\dagger\mathbf{M}$, a similar theorem was proposed in [42]. In this paper the asymptotical (i.e. the number of samples $T \to \infty$) lower-bound is explicitly calculated for the case of two sources, in spatially and temporally white Gaussian noise, when $\mathbf{C}^{\tilde{Y}}$ is estimated and decomposed according to the PCA-scheme described in Sect. 7.2:

$$\lim_{T\to\infty} T\,\mathrm{E}\{\mathrm{ISR}_{21} + \mathrm{ISR}_{12}\} \geqslant \frac{1}{2}\left\{(1+\sigma_N^2\rho_{11})(1+\sigma_N^2\rho_{22}) + \frac{\sigma_N^4\rho_{12}\rho_{21}}{I-J}\right\}, \tag{7.28}$$

in which $\rho_{pq} = \left((\mathbf{M}^H\mathbf{M})^{-1}\right)_{pq}$. This bound appears to be independent of the distribution of the source signals. In the low noise limit, it even reduces to a numerical constant:

$$\lim_{T\to\infty,\sigma_N\to 0} T\,\mathrm{E}\{\mathrm{ISR}_{21} + \mathrm{ISR}_{12}\} \geqslant \frac{1}{2}. \tag{7.29}$$

This bound is tight; it is e.g. approached by several techniques to be discussed in this thesis. This implies e.g. that an average rejection of 26 dB can be obtained from 100 independent samples.

Now let us discuss the effect of the prewhitening on the reconstruction of the transfer matrix. The derivation is based on the following lemma:

**Lemma 7.6.3** *Let the SVD's of* $\mathbf{A} \in \mathbb{C}^{I\times I}$, $\mathbf{B} \in \mathbb{C}^{I\times I}$ *and the product* $\mathbf{AB}$ *be given by:*

$$\mathbf{A} = \mathbf{U}_\mathbf{A}\mathbf{S}_\mathbf{A}\mathbf{V}_\mathbf{A}^H, \qquad \mathbf{B} = \mathbf{U}_\mathbf{B}\mathbf{S}_\mathbf{B}\mathbf{V}_\mathbf{B}^H, \qquad \mathbf{AB} = \mathbf{U}_\mathbf{AB}\mathbf{S}_\mathbf{AB}\mathbf{V}_\mathbf{AB}^H, \tag{7.30}$$

*and let their respective singular values be given by* $\sigma_i(\mathbf{A})$, $\sigma_i(\mathbf{B})$ *and* $\sigma_i(\mathbf{AB})$ *$(1 \leqslant i \leqslant I)$. Then we have:*

$$\sum_i \sigma_i(\mathbf{AB}) \leqslant \sum_i \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \tag{7.31}$$

*The equality sign holds iff the matrix* $\mathbf{V}_\mathbf{A}^H\mathbf{U}_\mathbf{B}$ *is diagonal, containing only unit-norm scalars, in the case that all the pairs* $(\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B}))$ *are mutually different; in the case that* $(\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B})) = (\sigma_{i+1}(\mathbf{A}), \sigma_{i+1}(\mathbf{B})) = \ldots = (\sigma_j(\mathbf{A}), \sigma_j(\mathbf{B}))$, *the equality sign still holds if the corresponding diagonal block of* $\mathbf{V}_\mathbf{A}^H\mathbf{U}_\mathbf{B}$ *is a unitary matrix.*

**Proof:** Although the lemma looks like a classical result, we have not been able to find it in the standard literature on matrix algebra. In particular it is stronger than what could be expected from the inequality in [146], Chapter 7, Sect. 3, Problem 18. We give a proof in App. C. □

Let the sample estimate of $\mathbf{C}_2^{\check{Y}}$ be given by $\hat{\mathbf{C}}_2^{\check{Y}}$, and let the symmetric EVD's of these matrices be given by

$$\mathbf{C}_2^{\check{Y}} = \mathbf{E}\mathbf{D}^2\mathbf{E}^H, \qquad \hat{\mathbf{C}}_2^{\check{Y}} = \hat{\mathbf{E}}\hat{\mathbf{D}}^2\hat{\mathbf{E}}^H. \tag{7.32}$$

The diagonal entries of $\mathbf{D}$ and $\hat{\mathbf{D}}$ are supposed to be positive; for convenience we will also assume that for each matrix all the elements are mutually different and not equal to zero. The derivation for the general case is analogous, but more cumbersome. Finally we write the SVDs of $\mathbf{M}$ and its estimate $\hat{\mathbf{M}}$ as:

$$\mathbf{M} = \mathbf{E}\mathbf{D}\mathbf{Q}, \qquad \hat{\mathbf{M}} = \hat{\mathbf{E}}\hat{\mathbf{D}}\hat{\mathbf{Q}}. \tag{7.33}$$

Now we show that the quality of the estimation is bounded by the quality of the prewhitening in the following way:

**Theorem 7.6.4 (Least-squares estimation error lower-bound)** *In the notation specified above, we have:*

$$\|\mathbf{M} - \hat{\mathbf{M}}\|^2 \geqslant \sum_i (d_{ii}^2 + \hat{d}_{ii}^2 - 2s_{ii}) \geqslant \sum_i (d_{ii} - \hat{d}_{ii})^2, \tag{7.34}$$

*in which* $s_{ii}$ *is the ith singular value of* $\hat{\mathbf{D}}\hat{\mathbf{E}}^H\mathbf{E}\mathbf{D}$. *The first inequality reduces to an equality for an optimal choice of* $\hat{\mathbf{Q}}$ *in the higher-order ICA-step; the second inequality vanishes iff* $\mathbf{E} = \hat{\mathbf{E}}\mathbf{P}$, *in which* $\mathbf{P}$ *is a diagonal matrix, containing only unit-modulus entries.*

**Proof:** The minimization of $\|\mathbf{M} - \hat{\mathbf{M}}\|^2$ in terms of $\hat{\mathbf{Q}}$ is a unitary Procrustes problem ([130], p. 582). We have:

$$\begin{aligned} \|\mathbf{M} - \hat{\mathbf{M}}\|^2 &= \|\mathbf{M}\|^2 + \|\hat{\mathbf{M}}\|^2 - 2\mathtt{Re}(\langle \mathbf{M}, \hat{\mathbf{M}} \rangle) && (7.35) \\ &= \|\mathbf{D}\|^2 + \|\hat{\mathbf{D}}\|^2 - 2\mathtt{Re}(\langle \hat{\mathbf{D}}\hat{\mathbf{E}}^H\mathbf{E}\mathbf{D}, \tilde{\mathbf{Q}} \rangle), && (7.36) \end{aligned}$$



in which $\tilde{\mathbf{Q}} \stackrel{\text{def}}{=} \hat{\mathbf{Q}}^H\mathbf{Q}$. If the SVD of $\hat{\mathbf{D}}\hat{\mathbf{E}}^H\mathbf{E}\mathbf{D}$ is given by $\mathbf{U}\mathbf{S}\mathbf{V}^H$, then the optimal $\hat{\mathbf{Q}}$ takes the form of $\mathbf{U}\mathbf{V}^H$. This proves the first inequality.
The second inequality,

$$\sum_i s_{ii} \leqslant \sum_i d_{ii}\hat{d}_{ii}, \tag{7.37}$$

is established by resorting to Lemma 7.6.3, in which $\mathbf{A} = \hat{\mathbf{D}}\hat{\mathbf{E}}^H$ and $\mathbf{B} = \mathbf{E}\mathbf{D}$. □

## 7.7 ICA by means of HOSVD

### 7.7.1 Principle

As tensor diagonality is a special case of all-orthogonality, Eq. (7.10) shows a HOEVD of the cumulant tensor $\mathcal{C}_4^Z$. Hence the higher-order step in the scheme of Sect. 7.2 can be interpreted as a HOEVD. This technique is summarized as Alg. 7.7.1.

**Algorithm 7.7.1**
*ICA by means of HOEVD*

Given: $N$ samples $\{Y_n\}_{1 \leqslant n \leqslant N}$ of $Y = \mathbf{M}X + N$, as in Eq. (7.1). Let $\hat{\mathbf{M}} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{Q}$ be an SVD of the mixing matrix estimate.

1. Determine $\mathbf{E}$, $\mathbf{D}$ from a prewhitening (Subsect. 7.2.1).
2. Call $\hat{\mathcal{C}}_4^Z$ the sample cumulant of the whitened observations. Determine $\mathbf{Q}$ from the HOEVD

$$\hat{\mathcal{C}}_4^Z = \hat{\mathcal{C}}_4^X \times_1 \mathbf{Q} \times_2 \mathbf{Q}^* \times_3 \mathbf{Q}^* \times_4 \mathbf{Q}.$$

When dealing with sample cumulants, the fourth-order output cumulant will only be approximately diagonalizible. For high S/N-ratios, the perturbation of the estimate can be analyzed with the expressions presented in Sect. 4.7.

### 7.7.2 Properties

In this section we will study some properties of the HOEVD-technique:

**Property 7.7.1** *Let the HOEVD of an Nth order cumulant tensor $\mathcal{C}_N^X$ be given by*

$$\mathcal{C}_N^X = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)}, \tag{7.38}$$

*in which either $\mathbf{U}^{(n)} = \mathbf{U}$ or $\mathbf{U}^{(n)} = \mathbf{U}^*$ $(1 \leqslant n \leqslant N)$, in accordance with the exact way $\mathcal{C}_N^X$ has been defined. The calculation of this decomposition is mathematically equivalent to the optimization of the following function:*

$$f = \sum_\alpha \|\mathcal{S}_{i=\alpha}\|^4, \tag{7.39}$$

*in which $\mathcal{S}_{i=\alpha}$ is the subtensor of $\mathcal{S}$, obtained by setting its first index to $\alpha$.*

**Proof:** Recall the link between HOSVD and matrix EVD (Prop. 4.4.7). Let the matrices $\mathbf{H}^{(n)}$ and $\mathbf{D}^{(n)}$ be defined in analogy with Eq. (4.15), with the difference however that $\mathbf{U}^{(n)}$ is an arbitrary unitary (orthogonal) matrix. It is easy to verify that the function $f$ in Eq. (7.39) equals:

$$f = \sum_\alpha |\mathbf{d}_{\alpha\alpha}^{(n)}|^2, \tag{7.40}$$

in which the components of $\mathbf{D}^{(n)}$ depend on $\mathbf{U}^{(n)}$. In other words, the optimization in Eq. (7.39) is equivalent to the diagonalization, in least-squares sense, of the matrix $\mathbf{H}^{(n)}$. □

The following theorem shows that a higher-order ICA-stage, based on the algebraic HOEVD-technique, can be reformulated in the statistical framework of contrast optimization.

**Property 7.7.2 (Contrast optimization)** *Let $\mathcal{C}_N^Z$ be the Nth order cumulant of a white stochastic vector $Z$, and let $\mathcal{C}_{N;i=\alpha}^Z$ represent the subtensor of $\mathcal{C}_N^Z$, obtained by setting its first index to $\alpha$. In this notation, we have that the function*

$$f(Z) = \sum_\alpha \|\mathcal{C}_{N;i=\alpha}^Z\|^4 \tag{7.41}$$

*is a contrast function over the manifold of unitary (orthogonal) matrices. This contrast is discriminating if the Nth order cumulants of the components of $Z$ are mutually different ($N$ is even), or mutually different in absolute value ($N$ is odd).*

**Proof:** We refer to the formulation of the contrast concept in Sect. 7.5. As $Z$ is supposed to be white (since we are only dealing with the higher-order



ICA-step, and not with the prewhitening), the only relevant diagonal matrix $\Lambda$ contains unit-norm elements on the diagonal. Hence, $\Lambda$ nor $\mathbf{P}$ affect the cost function of Eq. (7.41), such that the first condition is satisfied.
If $Z$ has mutually independent components, then $f(Z)$ takes the value $\sum_i \|c_N^{z_i}\|^4$. Now recall the proof of Prop. 7.7.1. The case in which $Z$ has mutually independent components, corresponds to the case where $\mathbf{D}^{(n)}$ is fully diagonal:

$$\mathbf{D}^{(n)} = \mathtt{diag}(|c_N^{z_1}|^2, |c_N^{z_2}|^2, \dots).$$

Multiplication of $Z$ with a unitary $\mathbf{Q}$ changes $\mathbf{D}^{(n)}$ to a matrix $\tilde{\mathbf{D}}^{(n)}$, as follows:

$$\tilde{\mathbf{D}}^{(n)} = \mathbf{Q}\mathbf{D}^{(n)}\mathbf{Q}^H.$$

Hence multiplication with $\mathbf{Q}$ cannot further increase the diagonal norm. This proves the second condition.
The conditions under which $f(Z)$ is discriminating, correspond to the conditions under which the HOEVD of $\mathcal{C}_N^Z$ is unique. This proves the third condition. The distinction between even and odd orders, arises from the fact that for odd orders, the sign of $c_N^{z_n}$ can be incorporated in the corresponding higher-order eigenvector. □

**Property 7.7.3 (Invariant estimator)** *The two-stage ICA-procedure, based on prewhitening and HOEVD of the standardized cumulant tensor, is an invariant estimator of the mixing matrix.*

**Proof:** In [46] it is explained that "orthogonal contrast optimization methods" are invariant if the contrast function is expressed in terms of the output of the separator. The HOEVD-technique belongs to this class. Call $\hat{X}$ the estimate of the source vector $X$, with estimated covariance $\hat{\mathbf{C}}^{\hat{X}}$ and higher-order tensor $\hat{\mathcal{C}}_N^{\hat{X}}$. We have indeed that the prewhitening takes the form of $\hat{\mathbf{C}}^{\hat{X}} = \mathbf{I}$, and that the higher-order step is based on the optimization of $\sum_\alpha \|\mathcal{C}_{N;i=\alpha}^{\hat{X}}\|^4$. It is easy to understand why the fact that the technique is based on the separator outputs, implies invariance: if the observed data are premultiplied with a matrix, say $\mathbf{A}$, then the separator outputs corresponding to the global optimum will now be observed when the estimated transfer matrix is premultiplied by $\mathbf{A}$ as well. □

**Property 7.7.4 (Matched filter)** *Assume an ICA set-up with more sensors than sources, subject to additive white Gaussian noise. The asymptotical estimate of the transfer matrix, obtained by prewhitening and HOEVD of the standardized cumulant tensor, is such that the corresponding MVDR beamformer (Eq. (7.21)) implements a matched filter, if the standardized source cumulants are mutually different.*

**Proof:** See App. C. □

### 7.7.3 Simulations

We illustrate the HOEVD-technique for ICA with two simulations.

In a first experiment we consider a set-up of two observation channels, listening to two sources, in which the signals are real-valued. Three different types of zero-mean unit-variance source distributions are used: (a) a binary distribution, with $\texttt{Prob}(x = 1) = \texttt{Prob}(x = -1) = 1/2$, (b) a uniform distribution over the interval $[-\sqrt{3}, +\sqrt{3}]$, and (c) a double-sided exponential distribution $p_X(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|)$. The three possible configurations (binary / uniform, binary / double-sided exponential , uniform / double-sided exponential) are considered. Since the three distributions are even, the higher-order stage of the algorithm is based on the HOEVD of a fourth-order cumulant. In each simulation a new mixing matrix is generated by taking its elements from a zero-mean Gaussian distribution. Each mixing matrix is normalized: the first column is scaled to unit-length and the Frobenius-norm of the second column is set equal to 2. The additive noise term $N$ can be neglected. Several performance parameters are plotted as a function of the length of the datasets $T$. All the curves are obtained by averaging over 500 Monte Carlo simulations.

In Fig. 7.1 we plot the mean ISR of the estimated LCMV filters (Eq. (7.22)), for the three source configurations. The dotted line is the theoretical upper-bound of performance, given by Eq. (7.29). The dashdotted line shows the mean ISR for the filter defined by Eq. (7.22), in which the true value for $\mathbf{M}$ has been used (note that in this case the estimate of $\mathbf{C}_2^{\hat{Y}}$ does not equal $\mathbf{M}\mathbf{M}^T$, which causes a separation error).

In Fig. 7.2 we plot the Root Mean Square Error (RMSE) between the true mixing matrix and the one estimated by HOEVD, both normalized in the same way.

The second experiment is inspired by the simulations described in [41], to allow an easy comparison. We consider two zero-mean complex-valued source signals. The first source has a uniform distribution over the circle with radius $\sigma_1$; the second source is uniformly distributed over the disc with radius $\sqrt{2}\,\sigma_2$. Hence the source powers equal $\sigma_1^2$ and $\sigma_2^2$ respectively. Both signals impinge on a linear $\lambda/2$ equispaced array of 10 unit-gain omnidirectional sensors in the far field of the emitters. In the simulations the lengths of the columns of the estimated mixing matrix $\hat{\mathbf{M}}$ are normalized in the sense that the source estimates have unit power. Under this assumption, the theoretical values of the elements of the transfer matrix are given by $m_{pq} = \sigma_q e^{2j\pi p\phi_q}$, where $\phi_q$ equals the electrical angle of source $q$. The noise is Gaussian, with power $\sigma_N^2$. In each experiment the datalength $T = 100$ and the angle $\phi_1 = 0$. All curves are obtained by averaging over 500 Monte Carlo simulations. Since the source distributions are point-symmetric around the origin, the higher-order stage of the algorithm is



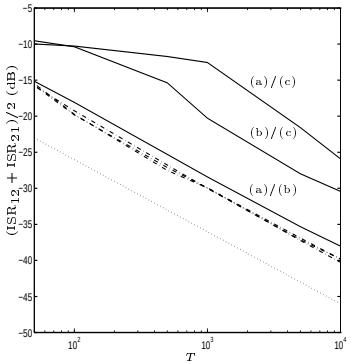



Figure 7.1: The mean ISR of the LCMV-beamformer in the first simulation of Sect. 7.7.3, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated by HOEVD. Dashdotted: the true mixing matrix is used in Eq. (7.22). Dotted: upper-bound of performance.

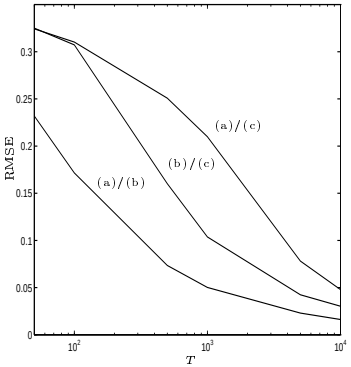



Figure 7.2: The RMSE for the HOEVD-technique in the first simulation of Sect. 7.7.3, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential.

based on the HOEVD of a fourth-order cumulant.

In Fig. 7.3 we plot the mean ISR of the LCMV filter estimates. The left figure shows the effect of the SNR, for three different mixing configurations $\phi_2 = 0.02$, 0.05 and 0.1 (note that in the latter case the steering vectors are

mutually orthogonal). The right figure shows the effect of the difference in Direction-of-Arrival (DOA), for three different noise levels $\sigma_N^2 = -15$ dB, $-5$ dB and 5 dB. The dotted line is the theoretical upper-bound of performance of Eq. (7.29). The dashdotted lines show the mean ISR for the filter defined by Eq. (7.22), in which the true value for $\mathbf{M}$ has been used.

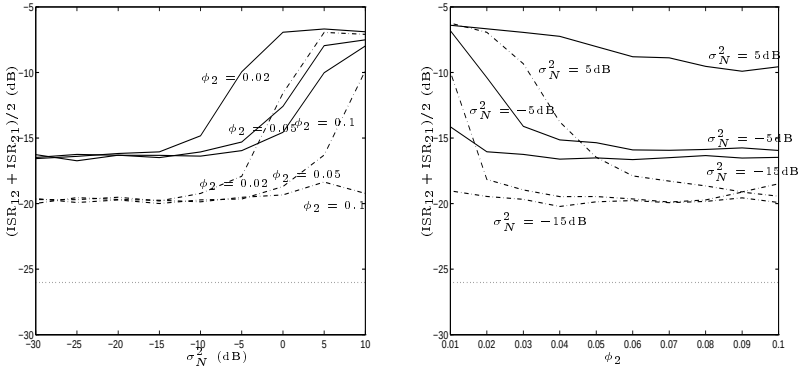



Figure 7.3: The mean ISR of the LCMV-beamformer in the second simulation of Sect. 7.7.3. Solid: the mixing matrix is estimated by HOEVD. Dashdotted: the true mixing matrix is used in Eq. (7.22). Dotted: low-noise upper-bound of performance. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of separation. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of separation. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

Fig. 7.4 is a similar plot for the INSR of the first source estimate, obtained by the MVDR filter estimates (Eq. (7.21)). The dashdotted lines now show the INSR for the filter defined by Eq. (7.21), in which the true value for $\mathbf{M}$ has been used. The dotted line is the optimal value for the INSR, given by Eq. (7.21) with a perfect knowledge of the covariance ánd the mixing matrix.

Finally, Fig. 7.5 is the analogous plot for the RMSE of the reconstruction of the mixing matrix.

## 7.8 Literature overview

Originally the problem of ICA was formulated by Hérault and Jutten, around 1983; for the solution they proposed an iterative algorithm on a neuromimetic architecture [143]. However, a problem of the same type had already been studied by Bar-Ness et al. a few years before, perhaps without recognizing its



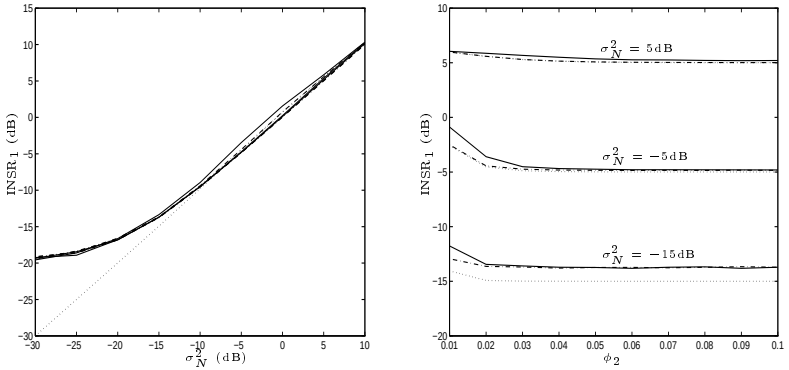



Figure 7.4: The INSR of the first source for the MVDR-beamformer in the second simulation of Sect. 7.7.3. Solid: the mixing matrix is estimated by HOEVD. Dashdotted: the true mixing matrix is used in Eq. (7.21). Dotted: optimal matched filter. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

full implications [11]. It took several years before the Jutten-Hérault algorithm

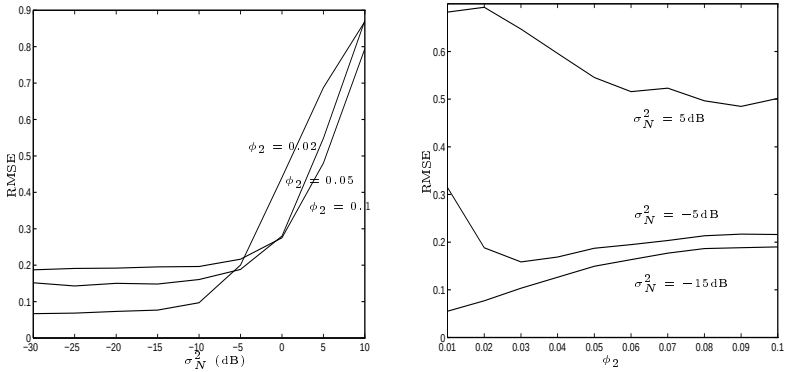



Figure 7.5: The RMSE between the true mixing matrix and its estimate, in the second simulation of Sect. 7.7.3. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

could be explained in theoretical terms [153, 68, 247]. Since the seminal paper, hundreds of papers have been published on the issue of blind source separation. E.g. there is a wide class of ICA-techniques based on Artificial Neural Networks (ANN's) [58]. In this section we will mainly concentrate (in a non-chronological way) on the contributions that are of direct importance to this thesis.

In [35] a matrix-to-matrix mapping is associated to $\mathcal{C}_4^Z$ in Eq. (7.10). It is observed that in the noise-free case the EVD of this mapping, in the sense of Sect. 2.8, allows a straightforward identification of the standardized transfer matrix $\mathbf{Q}$: every eigenmatrix corresponds to the outer product of a column of $\mathbf{Q}$ with itself, when $\mathcal{C}_4^X$ is diagonal with mutually different entries. The paper further discusses cheap implementations of this idea, among which a so-called "contracted scheme": it is explained that linear combinations of the eigenmatrices yield a Hermitean matrix that has, at the matrix level, $\mathbf{Q}$ as eigenmatrix.

In [172] Lacoume and Ruiz take into account that it is generally impossible to diagonalize the sample cumulant. They propose to minimize the sum of squared fourth-order cross-cumulants. In [124, 125] Gaeta and Lacoume show that, under the condition that all the third-order cumulants are zero, this optimization corresponds to a maximum likelihood approach, based on fourth-order approximations of the Gram-Charlier expansions of the probability density functions. From a practical point of view however, the technique is only realistic for a two-sources scenario, as it is based on an exhaustive search of the global minimum.

In [64], preceded by [59, 60, 61, 62], Comon shows that, in the real-valued two-dimensional case, the least-squares diagonalization of the fourth-order cumulant can take the form of rooting a polynomial of degree four (degree two, if dealing with third-order cumulants). In the general complex case the solution can be obtained by a more involved elimination scheme. Explicit expressions are derived for the noiseless case [61]. For the general ICA set-up, a Jacobi-type iteration scheme is proposed. This algorithm will be further studied in Chapter 8, where the link between the computation of a Jacobi-rotation and computation of the best rank-1 approximation of a tensor is investigated.

[64] not only explores the algebraic aspects of ICA, but also contains a profound discussion of the subject from a statistical point of view. As far as the core algorithm itself is concerned, its derivation is embedded in the development of the framework of contrast functions, consistent with [111, 128]. It is explained that the mutual information (see Ex. 7.5.2) corresponds to a contrast function; an approximation of this contrast in terms of third- and fourth-order cumulants is obtained by truncation of the Edgeworth expansions after the terms with fourth-order cumulants. In addition it is shown that the least-squares diagonalization of the third- resp. fourth-order cumulant can be considered as an approximative minimization of the mutual information, for large resp. small



third-order cumulants.

[64] also contains identifiability conditions for ICA. Independently, Lacoume and Ruiz came to the conclusion that ICA requires the processing of HOS [172]. The identifiability conditions proposed in [62] (see Theor. 7.3.1,7.3.2) are based on pair-wise statistical independence of the source estimates; as such, they are weaker than the conditions that are derived in [257]. Remark that our Theor. 7.3.3, derived from an algebraic point of view, contains even slightly weaker conditions, as it requires only independence for some order $N > 2$.

In [41] Cardoso and Souloumiac follow an other track to improve [35]. They propose to exploit all the available information, while still retaining a matrix-like method, by a simultaneous least-squares diagonalization of a set of matrices, spanning the range of $\mathcal{C}_4^Z$. The solution is obtained by a Jacobi-iteration scheme. The resulting technique is called *JADE* (Joint Approximate Diagonalization of Eigen-Matrices), and will briefly be recapitulated in Chapter 9. JADE is relatively cheap, particularly in the complex case: the computation of an elementary rotation is formulated as the best rank-1 approximation of a real symmetric $(2 \times 2)$-matrix (real data) or $(3 \times 3)$-matrix (complex data). On the other hand, the performance of the techniques developed in [64] and [41] are asymptotically equivalent, as proved in [245] and experimentally verified in [54]. [41] also contains an interpretation of JADE in terms of contrast optimization.

The scheme of Sect. 7.2 is not the only possible way to solve the ICA-problem. In particular, it is possible to identify the mixing matrix resorting only to the decomposition of the higher-order cumulant in rank-1 terms. This approach has the advantage that it is conceptually blind to the effect of additive Gaussian noise. Another interesting point is that the decomposition of the higher-order tensor allows the identification of a transfer matrix with more columns than rows. In Chapter 10 we will derive a new "higher-order-only" ICA-algorithm. Now we discuss some related publications.

[67] contains an historical survey on the decomposition of super-symmetric tensors in rank-1 components. The paper contains a.o. generic rank values, rank upper-bounds and a complete discussion of the $(2 \times 2 \times \ldots \times 2)$-case. In [34] Cardoso shows that under mild conditions the matrices in the intersection of the range of the cumulant tensor and the manifold of rank-1 matrices take the form of an outer product of a steering vector with itself; consequently MUSIC-like algorithms are devised. In [37] the same author investigates the link between symmetry of the cumulant tensor and the rank-1 property of its components. Several remarkable subresults are derived. A "rank-1 detection device", based on the symmetry properties of rank-1 tensors, is proposed to distinguish between rank-1 and general matrices in the further developments. This device is also used to obtain the number of components that can be identified. By relating the decomposition of the cumulant in rank-1 components to the EVD

of its associated matrix-to-matrix mapping, the problem is reformulated as the estimation of an orthogonal matrix, at least when all the sources have a cumulant with the same sign. Combination of the subresults leads to an algorithm, based on matrix EVD. To cope with the fact that sample cumulants can only be expanded as a sum of $R$ rank-1 components in an approximate sense, the author advises a preceding cumulant enhancement stage [36, 39], based on alternating projections of the cumulant on the rank-$R$ manifold and the subspace of super-symmetric tensors, in the sense of Cadzow [23].

It is important to notice that this thesis is primarily dealing with the *blind* source separation problem. In several situations however, some kind of a priori information is available, sometimes making it no longer necessary to resort to HOS. As an example, the second-order technique discussed in Sect. 7.9 is based on the fact that there is a large gap between the powers of the different sources. In [72] an accurate and efficient version (based on the $l_1$-norm [146] instead of the Frobenius-norm) of the technique in [64] is proposed, for the case in which all the sources have a kurtosis of the same sign. Important other examples are the separation of time-dependent sources (where one can also minimize the cross-correlation between samples at different sample times), discrete sources, or sources with known distributions. For an overview we refer to [152]. Finally we want to draw the attention to [44], in which Cardoso derived fourth-order-only contrast functions that correspond to the optimal statistic matching, for sufficiently low noise levels. The required a priori knowledge consists of the source cumulants up to order 8. The exercise has been repeated in [47], to establish optimal contrast functions based on both second-order and fourth-order cumulants.

## 7.9 ICA by mere PCA?

In Section 7.4 we stressed that in general the classical PCA does not allow to identify the transfer matrix and the source signals of Eq. (7.1) in a blind way. However in some cases PCA does allow to obtain a fairly good source separation. An example is the separation of uncorrelated source signals of which the variances show a large gap (under the convention that the norms of the corresponding steering vectors are comparable - see Sect. 7.3), which is the topic of this section.

The technique of separating sources by mere PCA has been widely used. Surprisingly, to our knowledge we were among the first in the ICA-community to analyze the quality of separation in a formal way. At the same time, and independently, similar results have been developed by Thirion [256]. While writing this thesis, we discovered that comparable results were already present in earlier work [264].



For readibility we will restrict the derivation to the real-valued case, with a set-up consisting of two observation channels and two zero-mean source signals.

Let $\mathbf{A}_Y$ be a $(2 \times T)$-dimensional dataset, containing $T$ independent realizations of $Y$. Fig. 7.9 shows a typical polar plot of the oriented energy of $\mathbf{A}_Y$ (for the definition of oriented energy, we refer to App. A). The directions of extremal oriented energy correspond to the left singular vectors of $\mathbf{A}_Y$, represented by $U_1$ and $U_2$; the extremal values are given by the squared singular values ($\sigma_1^2$ and $\sigma_2^2$). $M_1$ and $M_2$ represent the steering vectors, normalized as

$$\mathbf{M} = \begin{pmatrix} \cos\alpha & \cos\beta \\ \sin\alpha & \sin\beta \end{pmatrix} = (M_1 \; M_2)\,. \tag{7.42}$$

$\theta$ is the angle between $U_1$ and $M_1$. Now recall that the $t$th realization of $Y$, represented by $Y(t)$, has the following form:

$$Y(t) = M_1 x_1(t) + M_2 x_2(t) + N(t), \tag{7.43}$$

in which $x_1(t)$, $x_2(t)$, $N(t)$ are resp. the $t$th value of the source signals and the noise. We will adopt the convention that $\sigma_{x_1}^2 \geqslant \sigma_{x_2}^2$. Finally, we will assume that the noise is white - in this way, the oriented energy plot of the noise terms is theoretically a circle, and hence does not affect the plot for $\mathbf{A}_Y$ $(T \to \infty)$.

If $\sigma_{x_1}^2 \gg \sigma_{x_2}^2$, then large values of oriented energy occur in the direction of $M_1$. Hence, $U_1$ is a good approximation of $M_1$. As a consequence, the sequence $U_1^T \mathbf{A}_Y$ is a good approximation of $x_1(t)$: the interference from the second source signal is just the projection of $M_2 x_2(t)$ on $U_1$, such that the ISR is proportional to $\sigma_{x_2}^2 / \sigma_{x_1}^2$.

On the other hand, if $U_1$ is a very good approximation of $M_1$, then the sequence $U_2^T \mathbf{A}_Y$ mainly contains the second source signal: through the multiplication with $U_2^H$, the first source signal is to a large extent projected away. It remains to be proved that, under the condition that $\sigma_{x_1}^2 \gg \sigma_{x_2}^2$, the difference between the directions of $U_1$ and $M_1$ is indeed so small that the projection of the large excursions of $Y(t)$ (in the direction of $M_1$), over a small angle ($\theta$), can be neglected with respect to the values of the weak source signal. This is guaranteed by the following theorem, which states that the estimation error of $M_1$ is not a first- but a second-order effect as $\sigma_{x_2} / \sigma_{x_1} \to 0$:

**Theorem 7.9.1 (Source separation by PCA)**
*Let the real-valued two-dimensional case of model (7.1) be represented by:*

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\alpha & \cos\beta \\ \sin\alpha & \sin\beta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}, \tag{7.44}$$

*in which the source components are ordered such that $\sigma_{x_1}^2 \geqslant \sigma_{x_2}^2$, and in which the noise term is white with variance $\sigma_n^2$. Let the symmetric EVD of the covariance $\mathbf{C}_2^Y$ be given by*

$$\mathbf{C}_2^Y = \mathbf{U}\mathbf{S}^2\mathbf{U}^T, \tag{7.45}$$

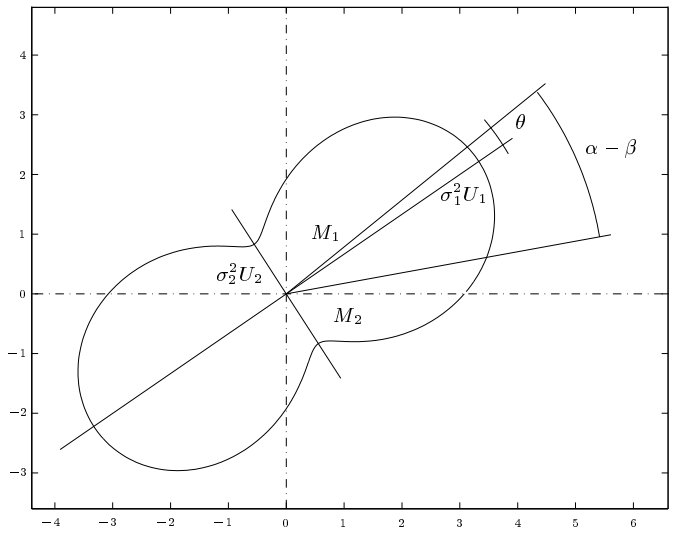



Figure 7.6: Schematic representation of a polar plot of the oriented energy of an ICA-dataset. The directions of extremal oriented energy are given by $U_1$ and $U_2$; the extremal values are given by $\sigma_1^2$ and $\sigma_2^2$. $M_1$ and $M_2$ are unit-norm steering vectors. $\theta$ is the angle between $U_1$ and $M_1$. $\alpha - \beta$ is the angle between $M_1$ and $M_2$.

*in which* $\mathbf{U} = (U_1\, U_2)$ *and* $\mathbf{S} = \mathtt{diag}(\sigma_1, \sigma_2)$. *Under the condition that* $x_1$ *and* $x_2$ *are uncorrelated,* $U_1$ *is proportional to a vector* $\hat{M}_1$*, related to the first steering vector* $M_1$ *as follows:*

$$\hat{M}_1 = M_1 + \Delta_1, \tag{7.46}$$

*in which*

$$\lim_{(\sigma_{x_2}/\sigma_{x_1})\to 0} \frac{\|\Delta_1\|}{(\sigma_{x_2}/\sigma_{x_1})^2} = f(\alpha, \beta), \tag{7.47}$$

*in which* $f(\alpha, \beta)$ *does not depend on* $\sigma_{x_1}$ *or* $\sigma_{x_2}$.

**Proof:** See App. C. □

The theorem is illustrated by Fig. 7.7. In this figure, the angle $\theta$ between the estimated and the true steering vector is plotted as a function of $(\sigma_{x_2}/\sigma_{x_1})^2$ and $\alpha - \beta$. In the neighbourhood of $\sigma_{x_2}/\sigma_{x_1} = 0$ this function is indeed very



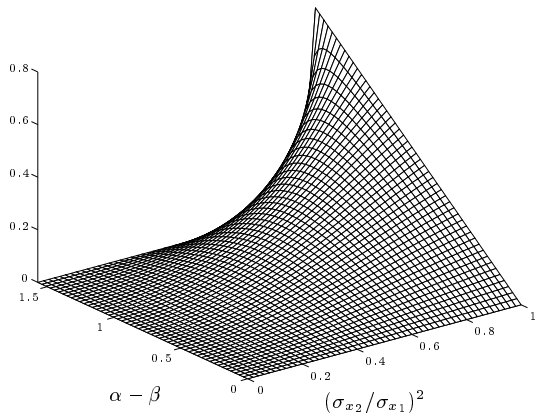



Figure 7.7: The angle (in radians) between the true dominant steering vector and the estimate obtained by PCA. The parameters are on one hand, the ratio $(\sigma_{x_2}/\sigma_{x_1})^2$ between the powers over the source signals, and on the other hand the angle $\alpha - \beta$ (in radians) between the steering vectors of these signals.

flat. Note also that, if the transfer matrix $\mathbf{M}$ is (close to) orthogonal, then (a good approximation of) it can be obtained from Eq. (7.45).

The preceding derivation is valid under the assumption that the noise term $N$ is white. If $N$ has a substantial coloured component, then the shape of the oriented energy plot (Fig. 7.9) is influenced by $N$, and the estimation of $M_1$ by $U_1$ is generally biased.

Even if there is a large difference in the variances of the various source components, PCA does not allow to extract the same amount of information from the dataset as ICA:

- PCA is still subject to rotation invariance, and hence it is impossible to identify the transfer matrix.
- The estimate of the strong source still contains the weak component and vice-versa. In both estimates the ISR is proportional to $(\sigma_{x_2}/\sigma_{x_1})^2$.
- The reconstruction of the weak source is only based on the component of the observations that is orthogonal to $U_1$.

**Example 7.9.2** We illustrate the method by the separation of the source signals of Fig. 7.8. In contrast to our usual way of presentation, these signals

are deterministic instead of stochastic; however, there is no difference from a computational point of view if one deals with a covariance matrix and a higher-order cumulant tensor obtained by time averaging, instead of sample averaging. Both sources are zero-mean, the covariance matrix is the identity, and the fourth-order cumulant tensor is diagonal. As the example is intended as a visualization of the effect of the power ratio $(\sigma_{x_2}/\sigma_{x_1})^2$, we will not consider an additive noise term.

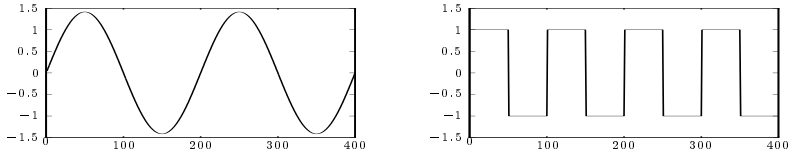


Figure 7.8: The source signals, considered in Ex. 7.9.2.

We will consider the following transfer matrix:

$$\mathbf{M} = \begin{pmatrix} 1/\sqrt{2} & 2/\sqrt{5} \\ 1/\sqrt{2} & 1/\sqrt{5} \end{pmatrix}, \tag{7.48}$$

which has unit-norm steering vectors. The corresponding observation signals are displayed on top of Fig. 7.9. The signals below are obtained by pre-multiplication of the sine wave source by resp. the factors 3, 5, 10 and 20. The source signals estimated by PCA, are displayed in Fig. 7.10. The figure clearly shows how the source separation becomes better, as the gap between $\sigma^2_{x_1}$ and $\sigma^2_{x_2}$ increases. This does not imply that the estimation of $\mathbf{M}$ becomes more accurate; instead the SVD-technique makes the steering vector estimates converge to $(1/\sqrt{2} \quad 1/\sqrt{2})^T$ resp. $(1/\sqrt{2} \quad -1/\sqrt{2})^T$. On the other hand, the ICA-technique of Sect. 7.7 allows a perfect reconstruction of the source signals in all cases, within the limits of identifiability discussed in Sect. 7.3 (Fig. 7.11). □

## 7.10 Conclusion

In this chapter we have discussed some basic aspects of ICA, such as a general two-stage solution strategy, conditions for identifiability, the link with PCA, the definition of contrast functions, a number of performance indicators, the concept of uniform performance and the effect of a prewhitening error on the overall performance. We have also given an overview of the publications that are most relevant to our work.

The general discussion already contained some new material. We have given a new, algebraic, proof of the identifiability of the ICA-solution, in which we



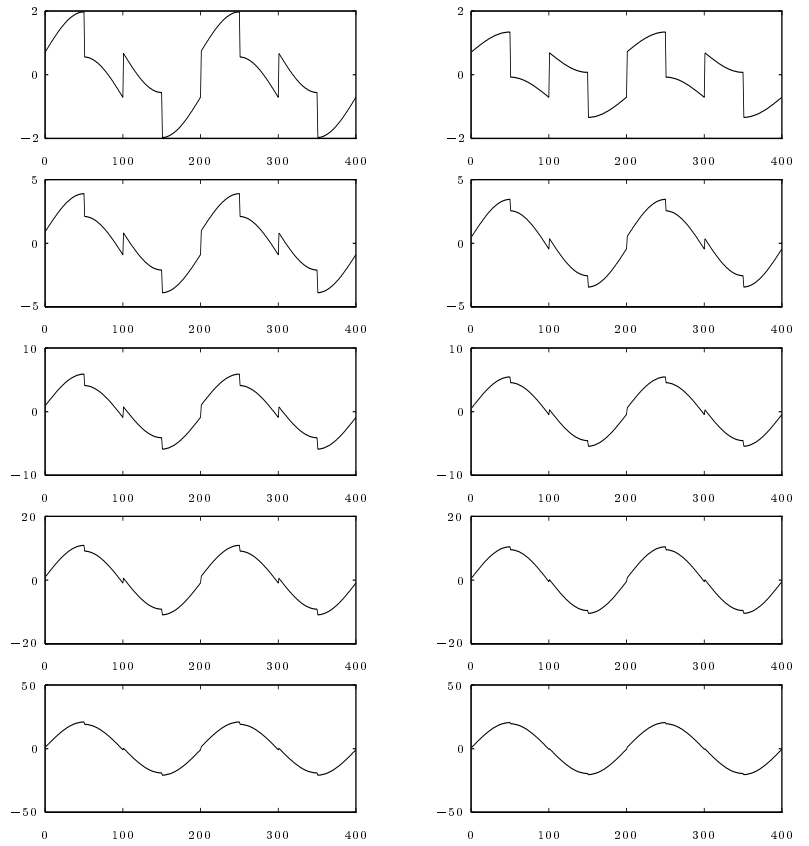


Figure 7.9: Top: the observation signals, obtained by multiplying the sources of Fig. 7.8 with the matrix of Eq. (7.48). Below: observations obtained when the sine wave source was premultiplied with the factors 3, resp. 5, 10 and 20.

rely on the assumptions that the sources are uncorrelated and independent "at some specific order $N > 2$". We have also presented an original proof for a lower-bound on the ISR's of the source estimates, induced by imperfections in the PCA-stage. In addition, we have derived a lower-bound on the Frobenius-norm of the difference between the original mixing matrix and its estimate. The latter derivation was based on a lemma, stating that the sum of singular values of a matrix product cannot be larger than the sum of the products of

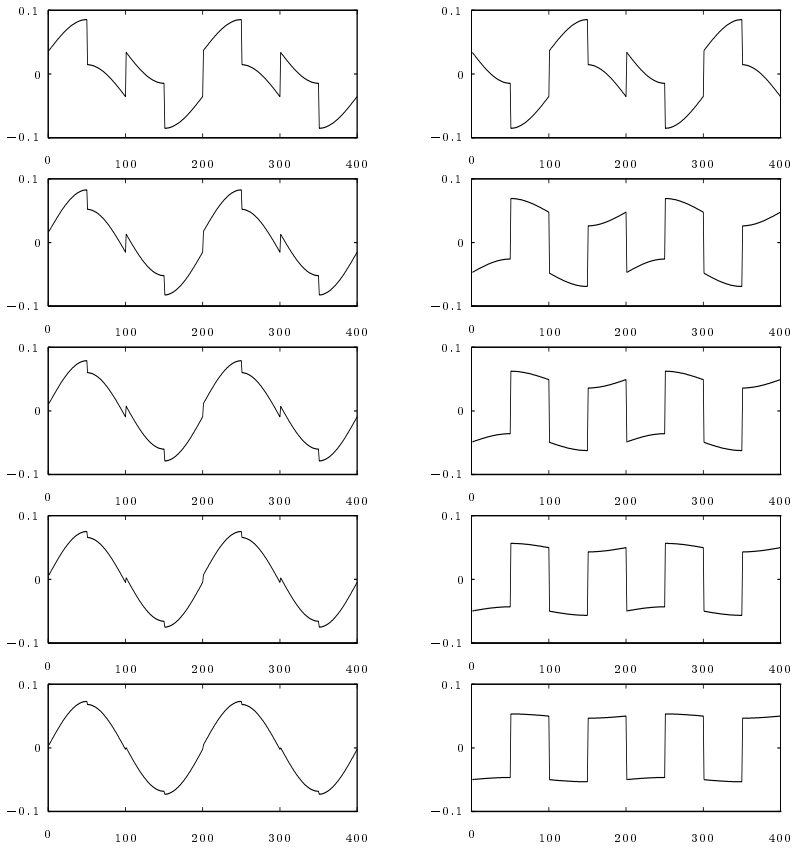


Figure 7.10: Source estimates, obtained by PCA from the datasets in Fig. 7.9.

the singular values of the distinct matrices.

In Sect. 7.7 we have proposed a new ICA-algorithm, based on the HOEVD of a higher-order cumulant, and analyzed its properties. The computational load of the technique is very low. In the next chapters, we will show that the price has to be paid in terms of numerical accuracy, compared to methods that are more computationally demanding. A second drawback of the HOEVD-technique is that the required conditions for identifiability are somewhat stronger than the ones imposed by the ICA-problem itself.



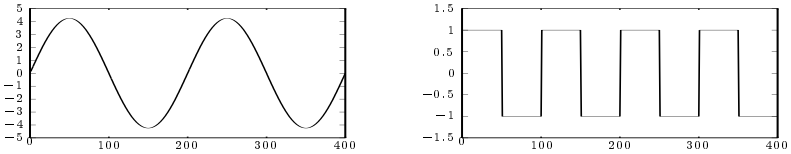


Figure 7.11: The source estimates, obtained with the ICA-technique of Sect. 7.7, for any dataset of Fig. 7.9 (different datasets only resulting in a different scaling and/or permutation of the estimated source signals).

Finally, we have derived a theoretical foundation for the fact that a mere PCA allows to obtain a fairly good source separation, provided that the sources are mutually uncorrelated and have a large difference in variance.

# Chapter 8

# ICA and Higher-Order Best Rank-1 Approximation

*In this chapter we discuss a second algebraic technique to solve the ICA-problem. The main difference with the HOEVD-method is the way in which we deal with the fact that a general super-symmetric tensor cannot be diagonalized by orthogonal (unitary) transformations. Instead of imposing the condition of all-orthogonality on the estimate of the source cumulant tensor, we now claim that this tensor should be "maximally diagonal" (in a quadratic sense). The formulation of this concept in literature, and the derivation of a reliable algorithm, have meant an important progress in the field of HOS-based signal processing [172, 124, 125, 64]. In this chapter we establish a link with the best rank-1 approximation problem of Chapter 5, which leads to a new line of approach.*

*In Sect. 8.1 we summarize the available literature. Subsect. 8.1.1 deals with the concept of maximal diagonality itself. The ICA-method is interpreted in terms of contrast functions, maximum likelihood estimation, and minimization of mutual information. In Subsect. 8.1.2 a Jacobi-type algorithm is discussed.*

*In Sect. 8.2 we show that in the fourth-order case an elementary Jacobi-rotation can also be found via the best rank-1 approximation of a* $(2 \times 2 \times 2 \times 2)$*-tensor (real-valued signals) or a* $(3 \times 3 \times 3 \times 3)$*-tensor (complex-valued signals). (The third-order case is analyzed in the next chapter.) In Subsect. 8.2.2 the technique is compared to the method of Subsect. 8.1.2. Subsect. 8.2.3 contains some simulations.*

# 8.1 The Maximal Diagonality approach to ICA

In this section we recapitulate the ICA-algorithm presented by Comon in [64]. In this technique, the higher-order stage is based on the optimal (in least-squares sense) diagonalization of the higher-order cumulant tensor, by unitary transformations. Algorithmic issues are discussed in Subsect. 8.1.2; first we develop the essential idea behind the method against the background of contrast functions.

## 8.1.1 Basic idea

In Chapter 4 it was stressed that a generic higher-order tensor cannot be diagonalized by unitary (orthogonal) transformations. As far as the development of a tensorial equivalent of the SVD is concerned, this problem was intercepted by replacing the condition of diagonality on the matrix of singular values by a condition of all-orthogonality on the corresponding tensor in the tensor decomposition. An interesting alternative definition of a tensorial SVD could involve the optimal diagonalization, in least-squares sense, of the higher-order tensor. This criterion will be called *Maximal Diagonality* (MD), or the *Comon-Lacoume* (CL) criterion.

Applied to Eq. (7.10), for sample statistics, this idea defines a new ICA-algorithm. The CL-approach can be interpreted in terms of the maximization of a contrast function:

**Theorem 8.1.1 (Contrast)** *Let the $N$th order cumulant of a white stochastic vector $X$, with values in $\mathbb{C}^J$ ($\mathbb{R}^J$), be given by $\mathcal{C}_N^X$. The function*

$$f(X) = \sum_{j=1}^{J} |(\mathcal{C}_N^X)_{jj\ldots j}|^2, \tag{8.1}$$

*with $N \geqslant 2$, is a contrast function over the manifold of unitary (orthogonal) matrices. For $N \geqslant 3$, this contrast is discriminating for the set of random vectors of which at most one component has a zero cumulant.*

**Proof:** For $N = 2$, the result is trivial. For $N \geqslant 3$, the theorem is a consequence of Theor. 7.3.3, taking into account that $n$-mode multiplication of $\mathcal{C}_N^X$ with a unitary (orthogonal) matrix does not alter the Frobenius-norm.
[64] contains an other proof, which is based on properties of the entries of an arbitrary unitary (orthogonal) matrix $\mathbf{Q}$. To be more specific, the theorem can be proved by noticing that the matrix, derived from $\mathbf{Q}$ by element-wise squaring of the moduli of the entries, is bistochastic (i.e. the sum of its entries in any row or column is equal to 1). □



**Remark 8.1.2** The original proof of Theor. 8.1.1 implies that it is not necessary to perform a full prewhitening to obtain the mixing matrix by optimization of contrast (8.1). A normalization to a mixing matrix with unit-norm rows and columns is sufficient. In this thesis we will not further investigate ICA-techniques based on this "weak" form of normalization. ◇

**Remark 8.1.3** In the CL-approach it is possible to solve the ICA-problem in its full generality (Theor. 7.3.3). The identifiability conditions of Theor. 8.1.1 are weaker than those of the HOEVD-technique (Prop. 7.7.2). ◇

**Remark 8.1.4** The fact that the contrast is discriminating, even if the cumulants of all the components of $X$ are the same (but different from zero), corresponds to the fact that the tensor with entries $\delta_{j_1 j_2 \ldots j_J}$ is non-isotropic for $J > 2$ (see Sect. 2.8). ◇

The optimization of contrast (8.1) can be interpreted as the approximative optimization of the maximum likelihood function for the ICA-problem. Assume the case in which $X$ is real-valued with null third-order cumulants (e.g. $X$ has an even distribution). Gaeta and Lacoume have shown that the function in (8.1) is obtained as the maximum likelihood function, if the Taylor series expansion of the second characteristic function of $X$ is truncated after the fourth-order terms [124, 125].

There is also a link between the optimization of contrast (8.1) and the minimization of the average mutual information of the components of a random vector (see Ex. 7.5.2). Assume that $X$ is a real-valued random vector. By substituting the Edgeworth series expansion of the probability density function $p_X(x)$ in Eq. (7.17), and truncating the expansion after the terms of order $O(P^{-1})$ in Eq. (3.34), the contrast can be approximated as [64]:

$$\hat{\Psi}(X) = \sum_{i=1}^{I} 4(\mathcal{C}_3^X)_{iii}^2 + (\mathcal{C}_4^X)_{iiii}^2 + 7(\mathcal{C}_3^X)_{iii}^4 - 6(\mathcal{C}_3^X)_{iii}^2 (\mathcal{C}_4^X)_{iiii}.$$

It turns out that for large values of $(\mathcal{C}_3^X)_{iii}$, the optimization of this function can be approximated by the optimization of the contrast in Eq. (8.1), for $N = 3$. On the other hand, if the third-order cumulants are null, then $\hat{\Psi}(X)$ equals $f(X)$, for $N = 4$.

## 8.1.2 Algorithm

### Computation by pair-wise processing

It was Comon [64] who came to the idea to base the optimal diagonalization of the standardized higher-order cumulant $\mathcal{C}_N^Z$ in Eq. (7.10), on a multilinear

generalization of the Jacobi-technique for the computation of the symmetric EVD of an Hermitean (symmetric) matrix. Hence the unknown factor $\mathbf{Q}$ in Eq. (7.10) will be estimated as a product of elementary Jacobi rotations, where each elementary rotation diagonalizes as far as possible the $(2 \times 2 \times \ldots \times 2)$-subtensor, associated with the marginal cumulants of the estimates of two different source components.

From an algebraic point of view, it is obvious to motivate the Jacobi-idea: $\mathcal{C}_N^Z$ is made diagonal iff all the elementary subtensors are made diagonal. Interestingly, the Jacobi-approach is consistent with a statistical consideration:

**Theorem 8.1.5** *Let $X$ be a $J$-dimensional vector with independent components, of which at most one is Gaussian, and whose densities are not reduced to a point-like mass. Let $\tilde{\mathbf{Q}} \in \mathbb{C}^{J\times J}$ ($\mathbb{R}^{J\times J}$) be an arbitrary unitary (orthogonal) matrix, and consider $\tilde{Z} = \tilde{\mathbf{Q}}X$. Then mutual independence of the components of $\tilde{Z}$ is equivalent with pair-wise independence.*

**Proof:** See [64]. □

The determination of the optimum of contrast (8.1) can even be formulated in terms of pair-wise conditions, when $X$ does not necessarily have mutually independent components, as is clear from the following more general theorem:

**Theorem 8.1.6** *For a random vector $X$ with values in $\mathbb{C}^J$ ($\mathbb{R}^J$), consider $\tilde{Z} = \tilde{\mathbf{Q}}X$, in which $\tilde{\mathbf{Q}} \in \mathbb{C}^{J\times J}$ ($\mathbb{R}^{J\times J}$) is unitary (orthogonal). Consider also a function $\Psi$, depending on $\tilde{\mathbf{Q}}$, of the form $\Psi = \sum_i \Psi_i$, in which $\Psi_i$ is a polynomial expression in cumulants of $\tilde{z}_i$. Then the equation $d\Psi = 0$ only imposes conditions on entries with two distinct indices, of the corresponding cumulants of $\tilde{Z}$. The same holds for the condition $d^2\Psi < 0$. By extension, $\Psi_i$ can be any analytic expression in cumulants of $z_i$.*

**Proof:** See [64]. □

This theorem allows us to express the optimal diagonality of sample cumulants in pair-wise terms, even if they cannot be fully diagonalized.

### The Jacobi-technique

In this section we will show how a Jacobi-rotation for the maximal diagonalization of an elementary subtensor of $\mathcal{C}_N^Z$ can be computed, for $N = 3$ or 4. Although the derivation is reminiscent of [64], the formulas from which the



rotation angles can be obtained, are given in quite some details: for completeness, and to allow an easy comparison with other methods discussed in this thesis. In addition, we touch upon the existence of spurious local optima, at the end of the section.

First let us consider the *third-order real-valued case*. Assume that the $(2\times2\times2)$-tensor to be diagonalized, is given by $\tilde{\mathcal{C}}$, and define $\mathcal{C} = \tilde{\mathcal{C}} \times_1 \mathbf{J} \times_2 \mathbf{J} \times_3 \mathbf{J}$, in which $\mathbf{J}$ is a Jacobi rotation matrix. $\mathbf{J}$ can be parametrized as

$$\mathbf{J} = \frac{1}{\sqrt{1+\theta^2}} \begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}, \tag{8.2}$$

where $\theta$ symbolizes the tangent of the rotation angle: $\theta = \tan\alpha$. The squared norm of the diagonal of $\mathcal{C}$ can then be written in the form

$$f(\theta, \tilde{\mathcal{C}}) = (\theta + \frac{1}{\theta})^{-3} \sum_{k=1}^{3} a_k (\theta^i - (-\theta)^{-i}). \tag{8.3}$$

For explicit expressions of the coefficients $a_k$, in terms of the entries of $\tilde{\mathcal{C}}$, we refer to [66], p.97. In Chapter 9 we will present a full expression for $f$, for a different parametrization of $\mathbf{J}$. The optimal rotation can be found by setting the derivative of $f$ w.r.t. $\theta$ equal to zero, and selecting the solution that corresponds to the global optimum of the contrast function. In terms of an auxiliary variable $\xi = \theta - 1/\theta$, the condition on the derivative is given by

$$b_2\xi^2 + b_1\xi - 4b_2 = 0, \tag{8.4}$$

in which

$$\begin{aligned} b_1 &= 3(\tilde{c}_{122}^2 + \tilde{c}_{112}^2) + 2(\tilde{c}_{112}\tilde{c}_{222} + \tilde{c}_{111}\tilde{c}_{122}) - \tilde{c}_{111}^2 - \tilde{c}_{222}^2, \\ b_2 &= \tilde{c}_{122}\tilde{c}_{222} - \tilde{c}_{111}\tilde{c}_{112}. \end{aligned}$$

After computation of the roots of the quadratic function in Eq. (8.4), the corresponding values of $\theta$ can be computed from $\theta^2 - \xi\theta - 1 = 0$. The latter equation always shows precisely 1 solution in the interval $(-1, +1]$, corresponding to an inner rotation. If Eq. (8.4) has more than 1 real root, then the one corresponding to the optimal contrast value is selected.

The *fourth-order real-valued case* can be treated in a similar way. We use the same notation as in the third-order case. The contrast function can now be written in the form

$$f(\xi, \tilde{\mathcal{C}}) = \frac{1}{(\xi^2+4)^2} \sum_{k=0}^{4} c_k\, \xi^k. \tag{8.5}$$

For explicit expressions of the coefficients $c_k$, in terms of the entries of $\tilde{\mathcal{C}}$, we refer to [64]. In Sect. 8.2 we will present a full expression for $f$, for a different

parametrization of $\mathbf{J}$. Setting the derivative with respect to $\xi$ equal to zero, gives:

$$\sum_{k=0}^{4} d_k \xi^k = 0, \tag{8.6}$$

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

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

The roots of the polynomial of degree 4 in Eq. (8.6) can be found in a non-iterative way [2].

For the *fourth-order complex-valued case* we keep using a similar notation. We assume that the complex tensor $\tilde{\mathcal{C}}$ has been defined in accordance with the conventions of Sect. 3.2. The complex Jacobi rotation $\mathbf{J}$ is now parametrized as

$$\mathbf{J} = c \begin{pmatrix} 1 & \theta \\ -\theta^* & 1 \end{pmatrix}, \tag{8.7}$$

with $c \in \mathbb{R}^+$ and $c = 1/\sqrt{1+|\theta|^2}$. $c$ symbolizes the real cosine of the rotation, and $\theta$ corresponds to the tangent of the complex rotation: $\theta = s/c$, with $c^2 + |s|^2 = 1$. The contrast function now takes the form

$$f(\xi, \tilde{\mathcal{C}}) = \sum_{\substack{-2 \leqslant k,l \leqslant 2 \\ 0 \leqslant k+l \leqslant 4}} e_{kl}\, \xi^k \xi^{l*} / (\xi\xi^* + 4)^2 \tag{8.8}$$

in terms of the auxiliary variable $\xi = \theta - 1/\theta^*$; the coefficients satisfy $e_{kl} = e_{lk}^*$. Explicit formulas for the coefficients are given in Eqs. (8.40-8.48) in Subsect. 8.2.1, where we will present a full expression for $f$ for a different parametrization of $\mathbf{J}$. Differentiation of the contrast function with respect to the modulus $\rho$ and the argument $\phi$ of $\xi$ yields a system of two polynomials of degree 3 and 4 in $\rho$, where the coefficients are polynomials in $e^{i\phi}$. Admissible



values of $\xi$ can be obtained by performing a formal elimination procedure on $\rho$, followed by rooting the resulting polynomial in $e^{i\phi}$.

Instead of solving the problem in an analytic way, one could also explicitly compute the contrast values for the complex points $c\theta$ on a grid inside the unit disk. In addition, the noise-free solution can be computed (this basically involves the rooting of a quadratic polynomial [61]). Consequently, the best solution is retained. It can be refined by a gradient-based algorithm, but this does not seem to be necessary in practice.

Finally, we remark that for higher-order tensors the cost function of the MD-technique can have spurious local maxima, contrarily to the matrix case. As a matter of fact, in the discussion concerning an elementary rotation in the real-valued fourth-order case, any cumulant for which Eq. (8.6) has exactly 4 distinct solutions, leads to a local maximum. The super-symmetric tensor defined by

$$\tilde{c}_{1111} = -5, \quad \tilde{c}_{1112} = -5, \quad \tilde{c}_{1122} = 2, \quad \tilde{c}_{1222} = 8, \quad \tilde{c}_{2222} = 4, \tag{8.9}$$

is an example. In Fig. 8.1, we have plotted the contrast value for $\mathcal{C} = \tilde{\mathcal{C}} \times_1 \mathbf{J} \times_2 \mathbf{J} \times_3 \mathbf{J} \times_4 \mathbf{J}$, as a function of the rotation angle $\alpha$.

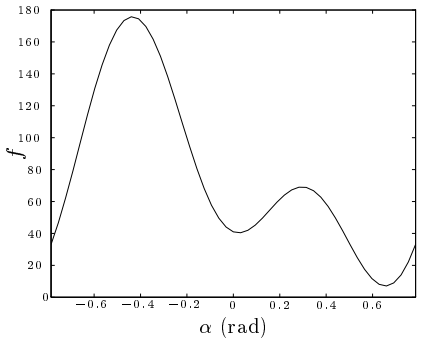



Figure 8.1: The squared diagonal norm, after Jacobi-rotation, of the super-symmetric tensor defined by Eq. (8.9), as a function of the rotation angle.

Theor. 8.1.6 does not imply that the Jacobi-technique itself will converge to the global optimum for an $I$-dimensional ICA-problem, with $I > 2$. However, in practice it has not been observed that the algorithm gets stuck in a local optimum.

**Algorithm 8.2**
*ICA by Maximal Diagonality - Best rank-1 Approach*

Given: $N$ samples $\{Y_n\}_{1 \leqslant n \leqslant N}$ of $Y = \mathbf{M}X + N$, as in Eq. (7.1). Let $\hat{\mathbf{M}} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{Q}$ be an SVD of the mixing matrix estimate.

1. Determine $\mathbf{E}$, $\mathbf{D}$ from a prewhitening (Subsect. 7.2.1).
2. Call $\hat{\mathcal{C}}_4^Z$ the sample cumulant of the whitened observations. Initialize $\mathbf{Q} \leftarrow \mathbf{I}$ and $\mathcal{C}' \leftarrow \hat{\mathcal{C}}_4^Z$. Sweep the $J(J-1)$ pairs $(i,j)$, according to a fixed ordering. Iterate until convergence ($|\alpha| < \epsilon$ for all rotation pairs). For each pair:
   (a) Call $\tilde{\mathcal{C}}$ the $(2 \times 2 \times 2 \times 2)$-subtensor of $\mathcal{C}'$, corresponding to sources $i$ and $j$. Construct super-symmetric $\mathcal{T} \in \mathbb{R}^{2\times2\times2\times2}$ in accordance with Eqs. (8.11 - 8.21) (real case), or super-symmetric $\mathcal{T} \in \mathbb{R}^{3\times3\times3\times3}$ in accordance with Eqs. (8.25 - 8.48) (complex case).
   (b) Maximize $h(U) = \mathcal{T} \times_1 U^T \times_2 U^T \times_3 U^T \times_4 U^T$ over $U \in \mathbb{R}^2$ ($\mathbb{R}^3$), with $\|U\| = 1$ and $u_1 \in \mathbb{R}^+$ (Chapter 5). $\mathtt{max}(h(U)) = h(U_{\max})$.
   (c) Construct $\mathbf{J}_{ij}$ from $U_{\max}$, in accordance with Theor. 8.2.2 and Eq. (8.10) (real case), or in accordance with Theor. 8.2.4 and Eq. (8.24) (complex case).
   (d) Update $\mathcal{C}' \leftarrow \mathcal{C}' \times_1 \mathbf{J}_{ij} \times_2 \mathbf{J}_{ij}^* \times_3 \mathbf{J}_{ij}^* \times_4 \mathbf{J}_{ij}$.
   (e) Accumulate $\mathbf{Q} \leftarrow \mathbf{Q} \cdot \mathbf{J}_{ij}^H$.

## 8.2 Fourth-order Jacobi-rotations and best rank-1 approximation

In this section we show that, in the fourth-order case, the computation of a Jacobi-rotation is equivalent to the calculation of the best rank-1 approximation of a super-symmetric fourth-order tensor. The link is derived in Subsect. 8.2.1. Subsect. 8.2.2 contains a brief comparison with the method described in Subsect. 8.1.2. Subsect. 8.2.3 contains a number of simulations, to illustrate the performance of the ICA-algorithm based on maximal diagonality. The algorithm is outlined in Alg. 8.2.

The computation of Jacobi-rotations in the third-order case will be further discussed in Chapter 9.



### 8.2.1 Determination of a Jacobi-rotation

Let us first consider the *real-valued case*. Like in Subsect. 8.1.2, the $(2 \times 2 \times 2 \times 2)$-subtensor to be diagonalized is called $\tilde{\mathcal{C}}$, and we define $\mathcal{C} = \tilde{\mathcal{C}} \times_1 \mathbf{J} \times_2 \mathbf{J} \times_3 \mathbf{J} \times_4 \mathbf{J}$, in which $\mathbf{J}$ is a Jacobi rotation matrix. Here we will use the standard representation of Jacobi rotations (note that the rotation angle $\alpha$ has the opposite sign as in Eq. (8.2)):

$$\mathbf{J} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{8.10}$$

Now we construct the super-symmetric $(2 \times 2 \times 2 \times 2)$-tensor $\mathcal{T}$, defined by

$$\begin{aligned}
t_{1111} &= a, && (8.11)\\
t_{1112} &= d, && (8.12)\\
t_{1122} &= 2b/3 + v/2 + a/6, && (8.13)\\
t_{1222} &= 3d/4 - e/4 - 3u/2, && (8.14)\\
t_{2222} &= 2b + 3v/2 + 9a/8 + 9\,\tilde{c}_{1122}^2/2 \\
&\quad + \tilde{c}_{1111}\tilde{c}_{2222}/4 + 4\,\tilde{c}_{1112}\tilde{c}_{1222}, && (8.15)
\end{aligned}$$

in which the auxiliary variables $a, b, d, e, u, v$ are given by

$$\begin{aligned}
a &= \tilde{c}_{1111}^2 + \tilde{c}_{2222}^2, && (8.16)\\
b &= \tilde{c}_{1112}^2 + \tilde{c}_{1222}^2, && (8.17)\\
d &= \tilde{c}_{1222}\tilde{c}_{2222} - \tilde{c}_{1111}\tilde{c}_{1112}, && (8.18)\\
e &= \tilde{c}_{1111}\tilde{c}_{1222} - \tilde{c}_{1112}\tilde{c}_{2222}, && (8.19)\\
u &= \tilde{c}_{1122}\,(\tilde{c}_{1112} - \tilde{c}_{1222}), && (8.20)\\
v &= \tilde{c}_{1122}\,(\tilde{c}_{1111} + \tilde{c}_{2222}). && (8.21)
\end{aligned}$$

Using these notations, we state:

**Theorem 8.2.1** *Assume a super-symmetric tensor $\tilde{\mathcal{C}} \in \mathbb{R}^{2\times2\times2\times2}$ and consider the tensors $\mathcal{C}$ and $\mathcal{T}$, defined above. The squared norm of the diagonal of $\mathcal{C}$ is given by*

$$\sum_{i=1}^{2} c_{iiii}^2 = \mathcal{T} \times_1 U^T \times_2 U^T \times_3 U^T \times_4 U^T, \tag{8.22}$$

*in which $U \stackrel{\text{def}}{=} (\cos(2\alpha) \;\; \sin(2\alpha))^T$.*

**Proof:** The theorem can be proved with straightforward calculus. The tedious goniometric manipulations are omitted. □

We remark that Eq. (8.22) clearly shows that the contrast function is periodic in the rotation angle, with period $\pi/2$, as the sign of $U$ is of no importance. By limitation of $\alpha$ to the interval $(-\pi/4, +\pi/4]$, $\mathbf{J}$ can be restricted to the set of inner rotations.

Furthermore Eq. (8.22) establishes the link with the best rank-1 approximation:

**Theorem 8.2.2** *Assume a cumulant tensor $\tilde{\mathcal{C}} \in \mathbb{R}^{2\times2\times2\times2}$ and consider the tensors $\mathcal{C}$ and $\mathcal{T}$, defined above. The best rank-1 approximation of $\mathcal{T}$ is given by*

$$\hat{\mathcal{T}} = \lambda\, U \circ U \circ U \circ U, \tag{8.23}$$

*in which $U = (\cos(2\alpha) \;\; \sin(2\alpha))^T$ corresponds to the rotation angle for the maximization of contrast $\sum_i c_{iiii}^2$, and in which $\lambda$ is the global contrast maximum.*

**Proof:** The proof follows from comparison of Theor. 5.2.2 and Theor. 8.2.1, taking into account that the contrast function itself is positive by definition. □

Hence the techniques described in Chapter 5, Sect. 5.2, can be used for the computation of the optimal $U$. Next, the actual elements of the optimal inner Jacobi rotation can be obtained from the basic goniometric relations $\cos\alpha = (1 + \cos(2\alpha)/2)^{1/2}$ and $\sin\alpha = \sin(2\alpha)/(2\cos\alpha)$.

Now let us consider the *complex-valued case*. Like before, the $(2 \times 2 \times 2 \times 2)$-subtensor to be diagonalized is called $\tilde{\mathcal{C}}$, and we define $\mathcal{C} = \tilde{\mathcal{C}} \times_1 \mathbf{J} \times_2 \mathbf{J}^* \times_3 \mathbf{J}^* \times_4 \mathbf{J}$, in which $\mathbf{J}$ is a complex Jacobi rotation matrix. Here we will use the standard representation of complex Jacobi rotations (note that the rotation angle $\alpha$ has the opposite sign as in Eq. (8.7)):

$$\mathbf{J} = \begin{pmatrix} \cos\alpha & -\sin\alpha\, e^{i\phi} \\ \sin\alpha\, e^{-i\phi} & \cos\alpha \end{pmatrix}. \tag{8.24}$$

Now we construct the real-valued super-symmetric $(3 \times 3 \times 3 \times 3)$-tensor $\mathcal{T}$, defined by

$$\begin{aligned}
t_{1111} &= a_{22}, && (8.25)\\
t_{1112} &= -\mathtt{Im}(a_{21}), && (8.26)\\
t_{1113} &= \mathtt{Re}(a_{21}), && (8.27)\\
t_{1122} &= a_{11} - \mathtt{Re}(a_{20}), && (8.28)\\
t_{1123} &= -\mathtt{Im}(a_{20}), && (8.29)\\
t_{1133} &= a_{11} + \mathtt{Re}(a_{20}), && (8.30)
\end{aligned}$$



$$\begin{aligned}
t_{1222} &= \mathtt{Im}(a_{2-1}) - \mathtt{Im}(a_{10}), && (8.31)\\
t_{1223} &= -\mathtt{Re}(a_{2-1}) + \mathtt{Re}(a_{10})/3, && (8.32)\\
t_{1233} &= -\mathtt{Im}(a_{2-1}) - \mathtt{Im}(a_{10})/3, && (8.33)\\
t_{1333} &= \mathtt{Re}(a_{2-1}) + \mathtt{Re}(a_{10}), && (8.34)\\
t_{2222} &= \mathtt{Re}(a_{2-2}) - \mathtt{Re}(a_{1-1}) + a_{00}, && (8.35)\\
t_{2223} &= -\mathtt{Im}(a_{2-2}) - \mathtt{Im}(a_{1-1})/2, && (8.36)\\
t_{2233} &= \mathtt{Re}(a_{2-2})/6 + a_{00}/3, && (8.37)\\
t_{2333} &= \mathtt{Im}(a_{2-2}) - \mathtt{Im}(a_{1-1})/2, && (8.38)\\
t_{3333} &= 8\mathtt{Re}(a_{2-2}) + \mathtt{Re}(a_{1-1}) + a_{00}, && (8.39)
\end{aligned}$$

in which the auxiliary variables are given by

$$\begin{aligned}
a_{22} &= \tilde{c}_{1111}^2 + \tilde{c}_{2222}^2 = e_{22}, && (8.40)\\
a_{21} &= \tilde{c}_{2222}\tilde{c}_{1222}^* - \tilde{c}_{1111}\tilde{c}_{1112} = e_{21}/4, && (8.41)\\
a_{20} &= (\tilde{c}_{1112}^2 + \tilde{c}_{1222}^{*2})/3 + \tilde{c}_{2112}(\tilde{c}_{1111} + \tilde{c}_{2222})/6 = e_{20}/12, && (8.42)\\
a_{2-1} &= \tilde{c}_{2112}(\tilde{c}_{1222}^* - \tilde{c}_{1112})/4 = e_{2-1}/16, && (8.43)\\
a_{2-2} &= \tilde{c}_{2112}^2/4 = e_{2-2}/8, && (8.44)\\
a_{11} &= a_{22}/6 + (|\tilde{c}_{1112}|^2 + |\tilde{c}_{1222}|^2)/3 + \tilde{c}_{1122}(\tilde{c}_{1111} + \tilde{c}_{2222})/12 \\
&\quad = e_{11}/24, && (8.45)\\
a_{10} &= (\tilde{c}_{1122}(\tilde{c}_{1222}^* - \tilde{c}_{1112}) + \tilde{c}_{2112}(\tilde{c}_{1222} - \tilde{c}_{1112}^*) \\
&\quad +3(\tilde{c}_{2222}\tilde{c}_{1222}^* - \tilde{c}_{1111}\tilde{c}_{1112}) + \tilde{c}_{2222}\tilde{c}_{1112} - \tilde{c}_{1111}\tilde{c}_{1222}^*)/4 \\
&\quad = e_{10}/16, && (8.46)\\
a_{1-1} &= \tilde{c}_{1222}^2 + \tilde{c}_{1112}^2 + 2(\tilde{c}_{1112}\tilde{c}_{1222}^*) + \tilde{c}_{2112}(\tilde{c}_{1111} + \tilde{c}_{1122} + \tilde{c}_{2222})/2 \\
&\quad = e_{1-1}/8, && (8.47)\\
a_{00} &= (\tilde{c}_{1111} + \tilde{c}_{1122} + \tilde{c}_{2222})^2/8 + \tilde{c}_{1112}\tilde{c}_{1222} + \tilde{c}_{1112}^*\tilde{c}_{1222}^* \\
&\quad +|\tilde{c}_{1112}|^2 + |\tilde{c}_{1222}|^2 + |\tilde{c}_{2112}|^2/4 = e_{00}/16, && (8.48)
\end{aligned}$$

where $e_{kl}$ correspond to the coefficients in Eq. (8.8).

Using these notations, we state:

**Theorem 8.2.3** *Assume a complex-valued $(2\times2\times2\times2)$-cumulant $\tilde{\mathcal{C}}$, defined in accordance with the complex conjugation convention of Sect. 3.2, and consider the tensors $\mathcal{C}$ and $\mathcal{T}$, defined above. The squared norm of the diagonal of $\mathcal{C}$ is given by*

$$\sum_{i=1}^{2} c_{iiii}^2 = \mathcal{T} \times_1 U^T \times_2 U^T \times_3 U^T \times_4 U^T, \tag{8.49}$$

*in which* $U \stackrel{\text{def}}{=} (\cos(2\alpha) \quad \sin(2\alpha)\sin\phi \quad \sin(2\alpha)\cos\phi)^T$.

**Proof:** The theorem can be proved with straightforward calculus. The tedious formula manipulations are omitted. □

Eq. (8.49) establishes the link with the best rank-1 approximation:

**Theorem 8.2.4** *Assume a cumulant tensor* $\tilde{\mathcal{C}} \in \mathbb{C}^{2\times2\times2\times2}$ *and consider the tensors* $\mathcal{C}$ *and* $\mathcal{T}$*, defined above. The best rank-1 approximation of* $\mathcal{T}$ *is given by*

$$\hat{\mathcal{T}} = \lambda\, U \circ U \circ U \circ U, \tag{8.50}$$

*in which* $U = (\cos(2\alpha) \quad \sin(2\alpha)\sin\phi \quad \sin(2\alpha)\cos\phi)^T$ *corresponds to the Jacobi-rotation for the maximization of contrast* $\sum_i c_{iiii}^2$*, and in which* $\lambda$ *is the global contrast maximum.*

**Proof:** The proof follows from comparison of Theor. 5.2.2 and Theor. 8.2.3, taking into account that the contrast function itself is positive by definition. □

Hence the techniques described in Chapter 5, Sect. 5.2, can be used for the computation of the optimal $U$ in the complex-valued case as well. Next, the actual elements of the optimal inner Jacobi rotation can be obtained from the basic relations $\cos\alpha = \sqrt{1+\cos(2\alpha)/2}$ and $\sin\alpha\, e^{i\phi} = (\sin(2\alpha)\cos\phi + i\sin(2\alpha)\sin\phi)/(2\cos\alpha)$.

### 8.2.2 Comparison

As a brief comparison between the technique proposed in [64] and the method described in the previous subsection, we remark the following:

- In Subsect. 8.1.2 the computation of the Jacobi-rotation basically involves the rooting of a polynomial of degree four, in the real-valued case. The roots can be found via analytical expressions. However, as [2] illustrates, these formulas are complicated. In the complex case, the computation of the optimal rotation leads to the calculation of the roots of a polynomial with degree higher than four. Here the rooting step is necessarily based on iterative techniques.
  In the formulation of Subsect. 8.2.1, the polynomial rooting is replaced by e.g. a higher-order power iteration. This iteration is essentially based on a sequence of simple sums and products.



- In Subsect. 8.1.2, all the roots have to be computed before the global optimum can be selected, or an extensive search for the global optimum is carried out from the beginning (as an alternative in the complex case). The best rank-1 procedure has the advantage that it immediately aims at the global optimum. The price consists of the computation of the dominant left singular vector of a $(2 \times 8)$-matrix (in the real case) or $(3 \times 27)$-matrix (in the complex case), to obtain the initial value corresponding to the appropriate HOEVD-truncate.
  As Ex. 5.2.10 indicated, there is not an absolute guarantee that initialization of an optimization algorithm for the computation of a best rank-1 approximation will lead to the global optimum (the defective cases in Ex. 5.2.10 would in the present context correspond to a contrast for which the position of the global maximum is not so clear). However, this did not cause a problem in any of the simulations discussed in the next section.

- Combining Theor. 5.2.2 and 8.2.1 (or 8.2.3) shows that, during the best rank-1 approximation procedure, the optimization of the contrast function can be monitored by the evolution of the estimate of $\lambda$. E.g. as far as the higher-order power iteration is concerned, this corresponds to inspecting the values $\lambda_k^{(1)}$, $\lambda_k^{(2)}$, ... in Alg. 5.2.2. This property allows to implement the termination criterion easily in terms of the value of the contrast. This is interesting, since in Jacobi-type methods it is seldom necessary to compute the rotation angles with very high accuracy: for the global optimization it is more important that the cost function is optimized in a monotonous way at a sufficient convergence speed.

### 8.2.3 Simulations

We illustrate the ICA-technique based on maximal diagonality with two simulations, similar to those for the HOEVD-technique.

As far as the first experiment is concerned, the only difference with Subsect. 7.7.3 is, that now both columns of the the mixing matrix are normalized to unit-length.

In Fig. 8.2 we plot the mean ISR of the estimated LCMV filters (Eq. (7.22)), for the three source configurations. The dotted line is the theoretical upper-bound of performance, given by Eq. (7.29).

In Fig. 8.3 we plot the RMSE between the true mixing matrix and the one estimated with the maximal diagonality approach, both normalized in the same way. The dashdotted and dotted lines give the two upper-bounds of performance specified in Theor. 7.6.4. The dashdotted lines image the performance that could be obtained with an optimal higher-order step; the dotted lines additionally take the quality of the prewhitening step into account.

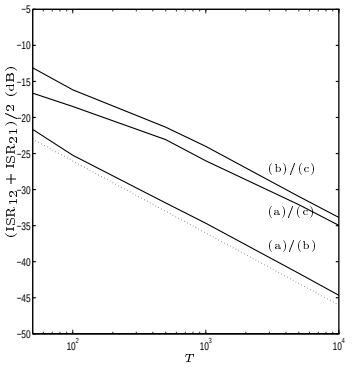



Figure 8.2: The mean ISR of the LCMV-beamformer in the first simulation of Sect. 8.2.3, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated with the maximal diagonality approach. Dotted: upper-bound of performance.

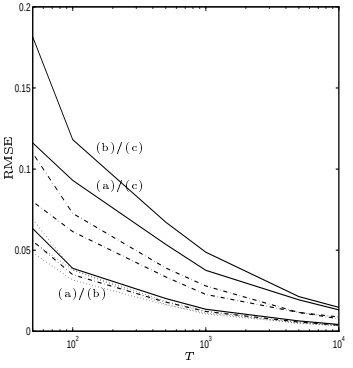



Figure 8.3: The RMSE in the first simulation of Sect. 8.2.3, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated with the maximal diagonality approach. Dashdotted: the first upper-bound of performance in Theor. 7.6.4. Dotted: the second upper-bound of performance in Theor. 7.6.4.



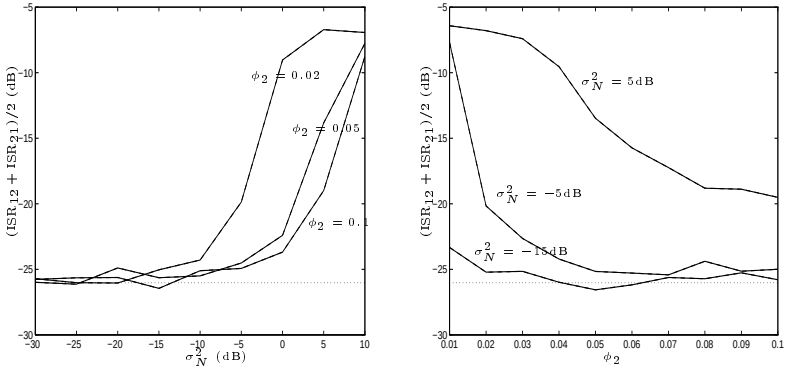



Figure 8.4: The mean ISR of the LCMV-beamformer in the second simulation of Sect. 8.2.3. Solid: the mixing matrix is estimated with the MD-technique. Dotted: low-noise upper-bound of performance. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of separation. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of separation. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

In the second experiment, the only difference with Subsect. 7.7.3 is, that now both sources are taken from a uniform distribution over a circle. 100 Monte Carlo runs were carried out. For the estimation of the best rank-1 approximation, Alg. 5.2.2 was used. The algorithm was initialized with the HOEVD-truncate, and terminated if the sine of the angle between $U_k^{(n)}$ and $U_{k-1}^{(n)}$ was below $1e-7$ in absolute value, for $n = 1$, 2, 3 or 4. In addition, the higher-order power algorithm was initialized with the vectors $U_0 = (\cos(2\alpha)\ \ \sin(2\alpha)\sin\beta\ \ \sin(2\alpha)\cos\beta)^T$, for $\alpha$ uniformly varying between $-\pi/4$ and $\pi/4$ in increments equal to $\pi/12$, and $\beta$ independently varying between $-\pi$ and $\pi$ in increments of $\pi/6$. In none of the cases a better rank-1 approximation was found than with the HOEVD-initialization.

In Fig. 8.4 the mean ISR of the LCMV filter estimates is plotted. The worst efficiency values for the different curves are in the order of magnitude of $1e-2$. The dotted line is the theoretical upper-bound of performance of Eq. (7.29).

In Fig. 8.5 the mean INSR of the MVDR filter estimates is plotted. The worst efficiency values for the different curves are in the order of magnitude of $5e-4$ to $5e-3$. The dotted line is the optimal value for the INSR, given by Eq. (7.21) with a perfect knowledge of the covariance and the mixing matrix.

In Fig. 8.2.3 the RMSE between the true mixing matrix and its estimate is shown. Here, the worst efficiency value is in the order of magnitude of $5e-3$. Like in Fig. 8.3, the dashdotted and dotted lines give the two upper-bounds of

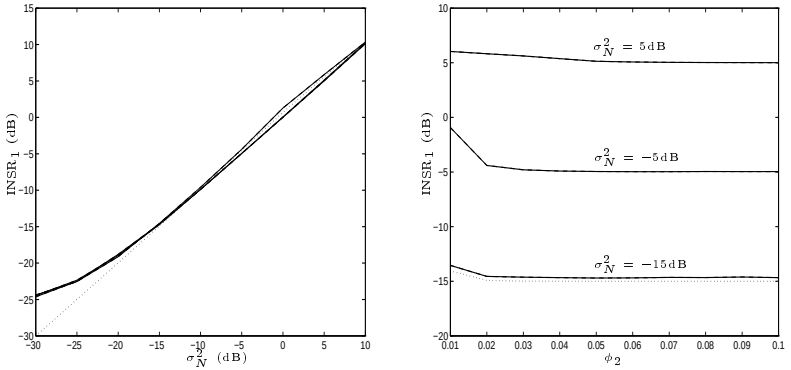



Figure 8.5: The INSR of the first source for the MVDR-beamformer in the second simulation of Sect. 8.2.3. Solid: the mixing matrix is estimated with the MD-technique. Dotted: optimal matched filter. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

performance specified in Theor. 7.6.4.

Fig. 8.7 shows the number of higher-order power iteration cycles until termination. It turns out that if the algorithm is initialized with the HOEVD-truncate, only a very limited number of iterations is required in most cases.

From both experiments it can be seen that the results obtained with the maximal diagonality technique are more accurate than their HOEVD-counterparts. An intuitive interpretation is that in the former approach, the off-diagonal cumulant entries are forced to zero, thereby imposing some degree of statistical independence, whereas in the HOEVD-method the degree of diagonality is merely determined by the perturbation properties in Sect. 4.7.

## 8.3 Conclusion

In this chapter we have studied a basic ICA-technique, in this thesis referred to as the "maximal diagonality" method [64]. We have given a simple proof of identifiability, and we have indicated the existence of local optima for the maximal diagonality cost function.

We have demonstrated that, for a higher-order ICA-stage involving fourth-order



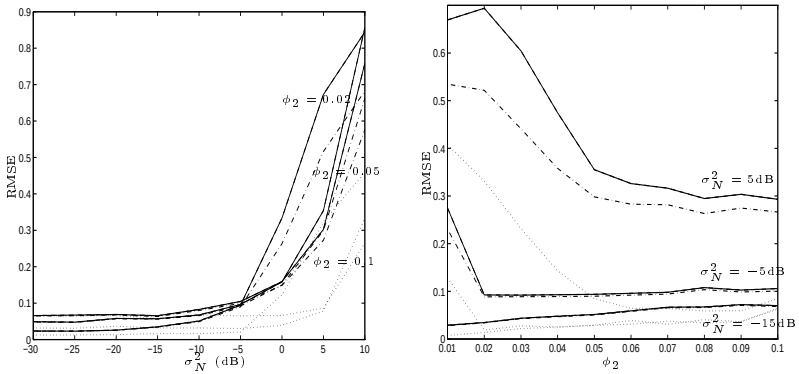



Figure 8.6: The RMSE between the true mixing matrix and its estimate, in the second simulation of Sect. 8.2.3. Solid: the mixing matrix is estimated with the MD-technique. Dashdotted: the first upper-bound of performance in Theor. 7.6.4. Dotted: the second upper-bound of performance in Theor. 7.6.4. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

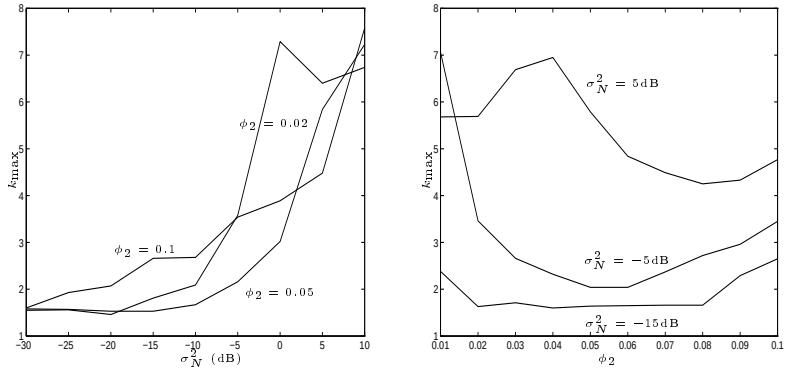



Figure 8.7: The averaged number of higher-order power iterations, starting from the HOEVD-truncate, in the second simulation of Sect. 8.2.3. *Left:* Varying SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB). $\phi_1 = 0$. *Right:* Varying difference in DOA ($\phi_1 = 0$). $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

cumulants, the computation of an elementary Jacobi-rotation amounts to the best rank-1 approximation of a super-symmetric real-valued $(2\times2\times2\times2)$-tensor (real case) or a super-symmetric real-valued $(3\times3\times3\times3)$-tensor (complex case). This approach has three advantages: (1) it can take the form of a sequence of simple operations (sums and products), (2) instead of computing all the points where the derivative of the cost function is null, the method aims directly at the global optimum, and (3) during the computation of the rotation, the optimization of the cost function is easily monitored. A disadvantage is that the technique is iterative, whereas in the real-valued case, the solution can be obtained by rooting a polynomial of degree 4; as such, the technique is especially interesting in the complex case.

# Chapter 9

# ICA and (Simultaneous) Third-Order Tensor Diagonalization

*Where the previous chapter dealt with MD ICA based on a fourth-order cumulant tensor, the third-order case is investigated in this chapter. It turns out that the computation of an elementary Jacobi-rotation is again equivalent to a best rank-1 approximation. For third-order tensors however, it is a matrix that has to be approximated; the fourth-order case lead to the approximation of a fourth-order tensor. This favourable computational load makes it attractive to do "something third-order-like" for fourth-order cumulant tensors as well. We show that simultaneous optimal diagonalization of "third-order tensor slices" of the fourth-order cumulant, is a suitable strategy. This "Simultaneous Third-Order Tensor Diagonalization" approach (STOTD) is similar in spirit to the efficient JADE-algorithm [41], that will briefly be discussed.*

*Sect. 9.1 is the third-order equivalent of Sect. 8.2. The step to optimal simultaneous diagonalization of a set of third-order tensors is set in Sect. 9.2. The JADE-algorithm is outlined in Sect. 9.3. In Sect. 9.4 it is noticed that the HOEVD-method, the JADE-method, the STOTD-method and the fourth-order MD-method are in fact elements of an algebraic framework for prewhitening-based ICA involving the fourth-order cumulant.*

## 9.1 The MD technique for third-order cumulants

In this section we show that the computation of the Jacobi-rotation that maximally diagonalizes a given $(2 \times 2 \times 2)$-tensor, is equivalent to the best rank-1 approximation of a symmetric matrix. Both the real and complex case are studied.

Let us first consider the *real-valued case*. Similar to Sect. 8.2, the $(2 \times 2 \times 2)$-tensor to be diagonalized is called $\tilde{\mathcal{C}}$, and we define $\mathcal{C} = \tilde{\mathcal{C}} \times_1 \mathbf{J} \times_2 \mathbf{J} \times_3 \mathbf{J}$, in which $\mathbf{J}$ is a Givens rotation matrix. We use the standard representation of Givens rotations (see Eq. (8.10)). We construct a real symmetric $(2 \times 2)$-matrix $\mathbf{B}$, as follows:

$$b_{11} = 9a_2/4 + 3a_3/2 + a_1/4, \tag{9.1}$$
$$b_{12} = 3a_4/2, \tag{9.2}$$
$$b_{22} = a_1, \tag{9.3}$$

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

$$a_1 = \tilde{c}_{111}^2 + \tilde{c}_{222}^2, \tag{9.4}$$
$$a_2 = \tilde{c}_{112}^2 + \tilde{c}_{122}^2, \tag{9.5}$$
$$a_3 = \tilde{c}_{111}\tilde{c}_{122} + \tilde{c}_{112}\tilde{c}_{222}, \tag{9.6}$$
$$a_4 = \tilde{c}_{122}\tilde{c}_{222} - \tilde{c}_{111}\tilde{c}_{112}. \tag{9.7}$$

Using these notations, we state:

**Theorem 9.1.1** *Assume a super-symmetric tensor $\tilde{\mathcal{C}} \in \mathbb{R}^{2\times2\times2}$ and consider the tensor $\mathcal{C}$ and the matrix $\mathbf{B}$, defined above. The squared norm of the diagonal of $\mathcal{C}$ is given by*

$$\sum_{i=1}^{2} c_{iii}^2 = U^T \cdot \mathbf{B} \cdot U, \tag{9.8}$$

*in which $U \stackrel{\text{def}}{=} (\cos(2\alpha) \;\; \sin(2\alpha))^T$.*

**Proof:** The theorem can be proved with straightforward calculus. The main subresults are given in App. C. □

Hence 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



inner Givens rotation can be obtained from the entries of $U$ in the same way as in Subsect. 8.2.1. Like in the fourth-order case, it is clear that the contrast function is periodic in the rotation angle, with period $\pi/2$, as the sign of $U$ is of no importance. The sign of the dominant eigenvector can be fixed to restrict $\mathbf{J}$ to the set of inner rotations ($\alpha \in (-\pi/4, +\pi/4]$).

Now let us consider the *complex-valued case*. We assume that the complex tensor $\tilde{\mathcal{C}}$ has been defined by the element-wise equation $\tilde{c}_{ijk} = \texttt{Cum}(\tilde{z}_i, \tilde{z}_j^*, \tilde{z}_k^*)$, for some random vector $\tilde{Z}$. This exception to the conjugation convention of Sect. 3.2 will make the formulae more analogous to those of the fourth-order technique, to be developed in Sect. 9.2. One can easily translate the expressions below, to the diagonalization of the conventionally defined tensor $\overline{\mathcal{C}}$ (of which the elements are given by $\overline{c}_{ijk} = \texttt{Cum}(\tilde{z}_i, \tilde{z}_j^*, \tilde{z}_k)$), by noticing that $\tilde{c}_{ijk} = \overline{c}_{jik}^*$.

We define $\mathcal{C} = \tilde{\mathcal{C}} \times_1 \mathbf{J} \times_2 \mathbf{J}^* \times_3 \mathbf{J}^*$, in which $\mathbf{J}$ is a complex Givens rotation matrix. For $\mathbf{J}$, we use the standard representation of Eq. (8.24). We construct a real symmetric $(3 \times 3)$-matrix $\mathbf{B}$, as follows:

$$b_{11} = a_1, \tag{9.9}$$
$$b_{12} = \texttt{Im}(v_1) + \texttt{Im}(v_2), \tag{9.10}$$
$$b_{13} = \texttt{Re}(v_1) - \texttt{Re}(v_2), \tag{9.11}$$
$$b_{22} = v_4 - \texttt{Re}(v_3), \tag{9.12}$$
$$b_{23} = \texttt{Im}(v_3), \tag{9.13}$$
$$b_{33} = v_4 + \texttt{Re}(v_3), \tag{9.14}$$

in which the auxiliary variables are given by

$$a_1 = |\tilde{c}_{111}|^2 + |\tilde{c}_{222}|^2, \tag{9.15}$$
$$a_2 = |\tilde{c}_{112}|^2 + |\tilde{c}_{212}|^2, \tag{9.16}$$
$$a_3 = |\tilde{c}_{211}|^2 + |\tilde{c}_{112}|^2, \tag{9.17}$$
$$a_4 = \tilde{c}_{111}\tilde{c}_{112}^*, \tag{9.18}$$
$$a_5 = \tilde{c}_{111}\tilde{c}_{211}^*, \tag{9.19}$$
$$a_6 = \tilde{c}_{222}\tilde{c}_{112}^*, \tag{9.20}$$
$$a_7 = \tilde{c}_{222}\tilde{c}_{212}^*, \tag{9.21}$$
$$a_8 = \tilde{c}_{111}\tilde{c}_{212}^* + \tilde{c}_{222}\tilde{c}_{112}^*, \tag{9.22}$$
$$a_9 = \tilde{c}_{111}^*\tilde{c}_{122} + \tilde{c}_{222}\tilde{c}_{211}^*, \tag{9.23}$$
$$a_{10} = \tilde{c}_{211}^*\tilde{c}_{112} + \tilde{c}_{122}\tilde{c}_{212}^*, \tag{9.24}$$
$$v_1 = a_7 - a_5/2, \tag{9.25}$$
$$v_2 = a_4 - a_6/2, \tag{9.26}$$
$$v_3 = a_9/2 + a_{10}, \tag{9.27}$$

$$v_4 \quad = \quad (a_1 + a_3)/4 + a_2 + \mathtt{Re}(a_8). \tag{9.28}$$

Using these notations, we state:

**Theorem 9.1.2** *Consider the complex-valued* $(2 \times 2 \times 2)$*-cumulant tensor* $\tilde{\mathcal{C}}$*, the tensor* $\mathcal{C}$ *and the matrix* $\mathbf{B}$*, defined above. The squared norm of the diagonal of* $\mathcal{C}$ *is given by*

$$\sum_{i=1}^{2} |c_{iii}|^2 = U^T \cdot \mathbf{B} \cdot U, \tag{9.29}$$

*in which* $U \stackrel{\text{def}}{=} (\cos(2\alpha) \quad \sin(2\alpha)\sin\phi \quad \sin(2\alpha)\cos\phi)^T$.

**Proof:** The theorem can be proved with straightforward calculus. The main subresults are given in App. C. □

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 inner Givens rotation can be obtained from the entries of $U$ in the same way as in Subsect. 8.2.1. The sign of the dominant eigenvector can be fixed to restrict $\mathbf{J}$ to the set of inner rotations ($\alpha \in (-\pi/4, +\pi/4]$).

## 9.2 ICA by Simultaneous Third-Order Tensor Diagonalization

In this section we show that the results of the previous section, involving third-order cumulants, can still be useful for ICA in which the higher-order stage is based on fourth-order cumulants. This leads to a new technique, which will be referred to as "ICA by Simultaneous Third-Order Tensor Diagonalization" (STOTD). In the first subsection we explain the basics of this method. Subsect. 9.2.2 contains the algorithmic details. In Subsect. 9.2.3 we discuss some properties of the technique. Finally, Subsect. 9.2.4 illustrates the performance with some simulations. An outline of the algorithm is presented as Alg. 9.2.

### 9.2.1 Basic idea

Consider the fourth-order cumulant $\mathcal{C}_4^Z$ of the standardized random vector $Z$ in the higher-order ICA-stage, as in Eq. (7.10). Let us associate to $\mathcal{C}_4^Z$ a linear



**Algorithm 9.2**
*ICA by means of STOTD*

Given: $N$ samples $\{Y_n\}_{1 \leqslant n \leqslant N}$ of $Y = \mathbf{M}X + N$, as in Eq. (7.1). Let $\hat{\mathbf{M}} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{Q}$ be an SVD of the mixing matrix estimate.

1. Determine $\mathbf{E}$, $\mathbf{D}$ from a prewhitening (Subsect. 7.2.1).
2. Call $\hat{\mathcal{C}}_4^Z$ the sample cumulant of the whitened observations. Associate to $\hat{\mathcal{C}}_4^Z$ a linear mapping from $\mathbb{R}^J$ ($\mathbb{C}^J$) to $\mathbb{R}^{J\times J\times J}$ ($\mathbb{C}^{J\times J\times J}$) (Eq. (9.30)). Determine a basis $\{\mathcal{T}^{(l)}\}_{(1 \leqslant l \leqslant J)}$ of $\mathtt{R}\{\hat{\mathcal{C}}_4^Z\}$ (Subsect. 9.2.1).
3. Initialize $\mathbf{Q} \leftarrow \mathbf{I}$ and $\tilde{\mathcal{T}}^{(l)} \leftarrow \mathcal{T}^{(l)}$ $(1 \leqslant l \leqslant J)$. Sweep the $J(J-1)$ pairs $(i, j)$, according to a fixed ordering. Iterate until convergence ($|\alpha| < \epsilon$ for all rotation pairs). For each pair:
   (a) Call $\tilde{\mathcal{C}}^{(l)}$ the $(2 \times 2 \times 2)$-subtensor of $\tilde{\mathcal{T}}^{(l)}$ $(1 \leqslant l \leqslant J)$, corresponding to sources $i$ and $j$. Construct symmetric matrix $\mathbf{B} \in \mathbb{R}^{2\times 2}$ in accordance with Eqs. (9.1 - 9.3, 9.35 - 9.38) (real case), or symmetric matrix $\mathbf{B} \in \mathbb{R}^{3\times 3}$ in accordance with Eqs. (9.9 - 9.14, 9.39 - 9.48) (complex case).
   (b) Determine $U \in \mathbb{R}^2$ ($\mathbb{R}^3$), with $\|U\| = 1$ and $u_1 \in \mathbb{R}^+$, as dominant eigenvector of $\mathbf{B}$.
   (c) Construct $\mathbf{J}_{ij}$ from $U$, in accordance with Theor. 9.1.1 and Eq. (8.10) (real case), or in accordance with Theor. 9.1.2 and Eq. (8.24) (complex case).
   (d) Update $\tilde{\mathcal{T}}^{(l)} \leftarrow \tilde{\mathcal{T}}^{(l)} \times_1 \mathbf{J}_{ij} \times_2 \mathbf{J}_{ij}^* \times_3 \mathbf{J}_{ij}^*$ $(1 \leqslant l \leqslant J)$.
   (e) Accumulate $\mathbf{Q} \leftarrow \mathbf{Q} \cdot \mathbf{J}_{ij}^H$.

transformation of $\mathbb{C}^J$ to the vector space of third-order tensors $\mathbb{C}^{J\times J\times J}$, in the following way:

$$\mathcal{V}' = \mathcal{C}_4^Z(V) \qquad \Longleftrightarrow \qquad {v'}_{ijk} = \sum_l (\mathcal{C}_4^Z)_{ijkl} v_l \tag{9.30}$$

for all index values. This linear mapping has a very special structure if we neglect the additive noise term in the ICA-model. In terms of Eq. (7.10), the SVD of the mapping (Theor. 2.8.5) is given by

$$\mathcal{C}_4^Z = \sum_j^J \sigma_j \, \mathcal{Q}_j \circ Q_j, \tag{9.31}$$

in which:

- the singular values are given by $\text{sign}(\kappa_j^X)\,\kappa_j^X$ $(1 \leqslant j \leqslant J)$, where $\kappa_j^X$ symbolizes the kurtosis of the $j$th source.
- the corresponding right singular vectors $Q_j^*$ are the columns of $\mathbf{Q}$, represented by $Q_j$ $(1 \leqslant j \leqslant J)$, and complex conjugated by convention.
- the corresponding "left singular tensors" $\mathcal{Q}_j$ $(1 \leqslant j \leqslant J)$ are given by

$$(\mathcal{Q}_j)_{pqr} = \text{sign}(\kappa_j^X)\,(Q_j)_p (Q_j)_q^* (Q_j)_r^*. \tag{9.32}$$

So the SVD of the linear mapping is strongly related to the standardized transfer matrix $\mathbf{Q}$. Moreover, we remark that all the third-order tensors in the range space of $\mathcal{C}_4^Z$ can be written as a linear combination of the left singular tensors, such that they can be diagonalized by $\mathbf{Q}$.

When noise is present and/or when the statistics of $Z$ are only available with limited accuracy, the derivation above is only approximately valid. We propose to estimate $\mathbf{Q}$ as the unitary (orthogonal) matrix that simultaneously diagonalizes as far as possible (in least-squares sense) a set of third-order tensors that form a basis for the range of $\mathcal{C}_4^Z$. Formally, if the set to be diagonalized is given by $\{\mathcal{T}^{(l)}\}_{(1 \leqslant l \leqslant J)}$, $\mathbf{Q}$ is estimated as the unitary (orthogonal) matrix $\mathbf{U}$ that maximizes the function

$$f(\mathbf{U}) = \sum_l (\sum_n |t'^{(l)}_{nnn}|^2), \tag{9.33}$$

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

$$\mathcal{T}'^{(l)} = \mathcal{T}^{(l)} \times_1 \mathbf{U} \times_2 \mathbf{U} \times_3 \mathbf{U}. \tag{9.34}$$

By simultaneously diagonalizing a full basis of the span of $\mathcal{C}_4^Z$, the information contained in all the entries of $\mathcal{C}_4^Z$ can be exploited. An orthogonal basis for the range of the linear mapping can be obtained from SVD in Eq. (9.31), together with a first estimate of $\mathbf{Q}$. It is also possible to resort to an ordinary basis by simple transformation under $\mathcal{C}_4^Z$ of $J$ linearly independent vectors. E.g. transformation of the canonical unit vectors corresponds to choosing the "third-order tensor slices" $(\mathcal{C}_4^Z)^{(l)}$, obtained by keeping the index $l$ in $(\mathcal{C}_4^Z)_{ijkl}$ fixed.

### 9.2.2 Computation

Like for the optimal diagonalization of a single third-order tensor, the optimal simultaneous diagonalization of a set of third-order tensors will be computed



by Jacobi-iteration. In each step the set of $J$ $(2 \times 2 \times 2)$-tensors associated with the elementary rotation, will be diagonalized as far as possible. This set will be represented by $\{\tilde{\mathcal{C}}^{(l)}\}_{(1 \leqslant l \leqslant J)}$. By Jacobi-rotation these tensors are transformed into the set $\{\mathcal{C}^{(l)}\}_{(1 \leqslant l \leqslant J)}$.

The function that will be maximized, is the sum of the squared norms of the diagonals of the tensors $\mathcal{C}^{(l)}$. Theor. 9.1.1 and 9.1.2 show that the optimal rotation can still be estimated via the dominant eigenvector of a real symmetric $(2 \times 2)$-matrix (real case) or a real symmetric $(3 \times 3)$-matrix (complex case); however, the matrix itself is computed as a sum of contributions of the individual tensors.

As far as *real-valued tensors* are concerned, this boils down to replacing the definitions of $a_1$ to $a_4$ in Eqs. (9.4 - 9.7) by the following summations:

$$a_1 = \sum_l (\tilde{c}^{(l)}_{111})^2 + (\tilde{c}^{(l)}_{222})^2, \tag{9.35}$$

$$a_2 = \sum_l (\tilde{c}^{(l)}_{112})^2 + (\tilde{c}^{(l)}_{122})^2, \tag{9.36}$$

$$a_3 = \sum_l \tilde{c}^{(l)}_{111}\tilde{c}^{(l)}_{122} + \tilde{c}^{(l)}_{112}\tilde{c}^{(l)}_{222}, \tag{9.37}$$

$$a_4 = \sum_l \tilde{c}^{(l)}_{122}\tilde{c}^{(l)}_{222} - \tilde{c}^{(l)}_{111}\tilde{c}^{(l)}_{112}. \tag{9.38}$$

In the *complex case*, the definitions of $a_1$ to $a_{10}$ in Eqs. (9.15 - 9.24) have to modified to:

$$a_1 = \sum_l |\tilde{c}^{(l)}_{111}|^2 + |\tilde{c}^{(l)}_{222}|^2, \tag{9.39}$$

$$a_2 = \sum_l |\tilde{c}^{(l)}_{112}|^2 + |\tilde{c}^{(l)}_{212}|^2, \tag{9.40}$$

$$a_3 = \sum_l |\tilde{c}^{(l)}_{211}|^2 + |\tilde{c}^{(l)}_{112}|^2, \tag{9.41}$$

$$a_4 = \sum_l \tilde{c}^{(l)}_{111}(\tilde{c}^{(l)}_{112})^*, \tag{9.42}$$

$$a_5 = \sum_l \tilde{c}^{(l)}_{111}(\tilde{c}^{(l)}_{211})^*, \tag{9.43}$$

$$a_6 = \sum_l \tilde{c}^{(l)}_{222}(\tilde{c}^{(l)}_{112})^*, \tag{9.44}$$

$$a_7 = \sum_l \tilde{c}^{(l)}_{222}(\tilde{c}^{(l)}_{212})^*, \tag{9.45}$$

$$a_8 = \sum_l \tilde{c}^{(l)}_{111}(\tilde{c}^{(l)}_{212})^* + \tilde{c}^{(l)}_{222}(\tilde{c}^{(l)}_{112})^*, \tag{9.46}$$

$$a_9 = \sum_l (\tilde{c}^{(l)}_{111})^* \tilde{c}^{(l)}_{122} + \tilde{c}^{(l)}_{222} (\tilde{c}^{(l)}_{211})^*, \quad (9.47)$$

$$a_{10} = \sum_l (\tilde{c}^{(l)}_{211})^* \tilde{c}^{(l)}_{112} + \tilde{c}^{(l)}_{122} (\tilde{c}^{(l)}_{212})^*. \quad (9.48)$$

### 9.2.3 Properties

In this subsection, we will study some properties of the STOTD-technique.

**Theorem 9.2.1 (Contrast)** *Let $\mathcal{C}_4^X$ be the 4th order cumulant of a white stochastic vector $X$, with values in $\mathbb{C}^J$ ($\mathbb{R}^J$). The function*

$$f(X) = \sum_{l=1}^{J} \sum_{j=1}^{J} |(\mathcal{C}_4^X)_{jjjl}|^2 \quad (9.49)$$

*is a contrast function over the manifold of unitary (orthogonal) matrices. It is discriminating for the set of random vectors of which at most one component has zero kurtosis.*

**Proof:** First let us assume that all the components of $X$ are kurtic. Define $\tilde{X} = \mathbf{Q}X$, in which $X$ has mutually independent components and $\mathbf{Q}$ is unitary (orthogonal). We will prove that $f(\tilde{X}) \leqslant f(X)$, and that the equality sign only holds if $\mathbf{Q} = \Lambda\mathbf{P}$, in which $\Lambda$ is a diagonal matrix of unit-norm elements and $\mathbf{P}$ is a permutation.
Since the components of $X$ are mutually independent, its fourth-order cumulant is diagonal and $f(X)$ is the squared Frobenius-norm of this cumulant:

$$f(X) = \sum_{j=1}^{J} |(\mathcal{C}_4^X)_{jjjj}|^2 = \|\mathcal{C}_4^X\|^2.$$

On the other hand, we remind that the cumulant of $\tilde{X}$ is given by:

$$\mathcal{C}_4^{\tilde{X}} = \mathcal{C}_4^X \times_1 \mathbf{Q} \times_2 \mathbf{Q}^* \times_3 \mathbf{Q}^* \times_4 \mathbf{Q}.$$

$f(\tilde{X})$ can be bounded by the squared Frobenius-norm of $\mathcal{C}_4^{\tilde{X}}$, to prove that $f(\tilde{X}) \leqslant f(X)$:

$$f(\tilde{X}) = \sum_{j,l} |(\mathcal{C}_4^{\tilde{X}})_{jjjl}|^2 \leqslant \|\mathcal{C}_4^{\tilde{X}}\|^2 = \|\mathcal{C}_4^X\|^2 = f(X),$$

in which we used that the fact that $n$-mode multiplication with a unitary matrix does not alter the Frobenius-norm of a tensor.



$f(\tilde{X}) = f(X)$, or $f(\tilde{X}) = \|\mathcal{C}_4^{\tilde{X}}\|^2$, implies that the third-order cumulant slices of $\mathcal{C}_4^{\tilde{X}}$, represented by $(\mathcal{C}_4^{\tilde{X}})^{(l)}$, are strictly diagonal. E.g. the first tensor slice is given by

$$(\mathcal{C}_4^{\tilde{X}})^{(1)} = \sum_j (q_{1j}\kappa^{x_j})\, Q_j \circ Q_j^* \circ Q_j^*,$$

in which $Q_j$ represents the $j$th column of $\mathbf{Q}$. Now assume that exactly $K > 1$ of these terms do not vanish (say: $q_{11}\kappa^{x_1} \neq 0$, $q_{12}\kappa^{x_2} \neq 0$, ..., $q_{KK}\kappa^{x_K} \neq 0$ ). According to a reasoning similar to the one behind Theor. 7.3.3, claiming that $(\mathcal{C}_4^{\tilde{X}})^{(1)}$ is strictly diagonal means that the matrix $\mathbf{Q}_{1:K} = [Q_1 Q_2 \dots Q_K]$ is in the form

$$\mathbf{Q}_{1:K} = \tilde{\Lambda}\tilde{\mathbf{P}},$$

in which $\tilde{\Lambda} \in \mathbb{C}^{J \times K}$ contains $K$ unit-norm elements on the diagonal, and $\tilde{\mathbf{P}} \in \mathbb{C}^{K \times K}$ is a permutation. But this implies that only one of the coefficients $q_{1k}\kappa^{x_k}$ $(1 \leqslant k \leqslant K)$ could have been different from zero. Hence we conclude that diagonality of $(\mathcal{C}_4^{\tilde{X}})^{(1)}$ involves that $K = 1$. In other words, the first row of $\mathbf{Q}$ only contains one non-zero entry; this entry has unit-norm. Repeating the derivation for the other third-order cumulant slices, and taking into account that $\mathbf{Q}$ is unitary, shows that $\mathbf{Q}$ contains exactly one unit-modulus entry in each column and each row.
Finally, if e.g. $x_J$ is non-kurtic, then the reasoning above can be repeated for the random vector formed by the first $J-1$ components of $X$. The fact that the original $\mathbf{Q}$ has to be orthogonal, then induces the required form of the remaining column $Q_J$. □

**Remark 9.2.2** Remark that the STOTD-technique, like the CL-technique, allows to solve the ICA-problem in its full generality. ◇

**Property 9.2.1 (Invariant estimator)** *The two-stage ICA-procedure, based on prewhitening and simultaneous diagonalization of the third-order tensor slices of the standardized fourth-order cumulant tensor, is an invariant estimator of the mixing matrix.*

**Proof:** Call $\hat{X}$ the estimate of the source vector $X$, with estimated covariance $\hat{\mathbf{C}}^X$ and cumulant $\hat{\mathcal{C}}_4^{\hat{X}}$. The prewhitening takes the form of $\hat{\mathbf{C}}^{\hat{X}} = \mathbf{I}$, and the higher-order step is based on the optimization of $\sum_{l,j} |(\mathcal{C}_4^{\hat{X}})_{jjjl}|^2$. Hence, the algorithm is based on the output of the separator, which is a sufficient condition for invariance [46] (see also the proof of Prop. 7.7.3). □

**Property 9.2.2 (Matched filter)** *Assume an ICA set-up with more sensors than sources, subject to additive white Gaussian noise. The asymptotical esti-*

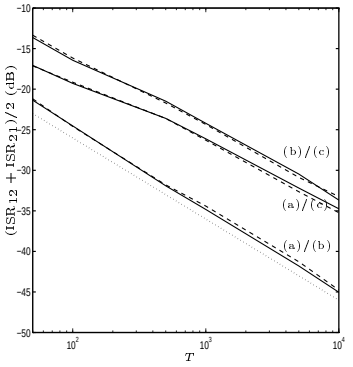



Figure 9.1: The mean ISR of the LCMV-beamformer in the first simulation of Sect. 9.2.4, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated by means of the STOTD-technique. Dashed: the mixing matrix is estimated by means of the JADE-technique. Dotted: upper-bound of performance.

*mate of the transfer matrix, obtained by prewhitening and simultaneous diagonalization of the third-order tensor slices of the standardized fourth-order cumulant tensor, is such that the corresponding MVDR beamformer (Eq. (7.21)) implements a matched filter, if at most one source component is non-kurtic.*

**Proof:** Analogous to Prop. 7.7.4. □

### 9.2.4 Simulations

We illustrate the STOTD-technique by means of the same experiments as for the MD-approach (see Subsect. 8.2.3). In each of the figures, the solid lines are obtained by STOTD. The dashed lines are obtained by means of the JADE-algorithm, to be discussed in the next section. (For the second experiment, 500 Monte Carlo runs have been conducted. The worst efficiency value for each STOTD- or JADE-curve is then in the order of magnitude of $1e-5$ to $1e-3$.)

The results obtained by STOTD turn out to be nearly the same as the results of the JADE-algorithm. Comparison to Subsect. 8.2.3, shows that the MD-method behaves similarly as well. The latter observation has also been reported in [54].



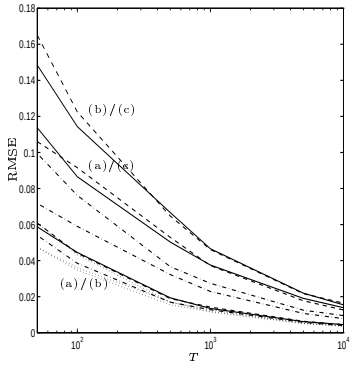



Figure 9.2: The RMSE in the first simulation of Sect. 9.2.4, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated by means of the STOTD-technique. Dashed: the mixing matrix is estimated by means of the JADE-technique. Dashdotted: the first upper-bound of performance in Theor. 7.6.4. Dotted: the second upper-bound of performance in Theor. 7.6.4.

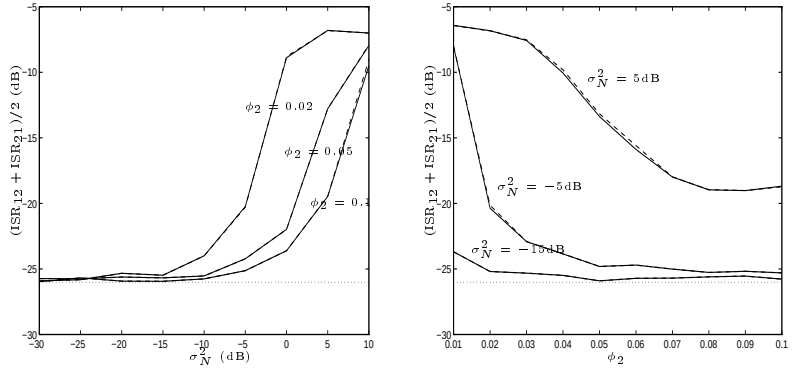



Figure 9.3: The mean ISR of the LCMV-beamformer in the second simulation of Sect. 9.2.4. Solid: the mixing matrix is estimated by means of the STOTD-technique. Dashed: the mixing matrix is estimated by means of the JADE-technique. Dotted: low-noise upper-bound of performance. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of separation. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of separation. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

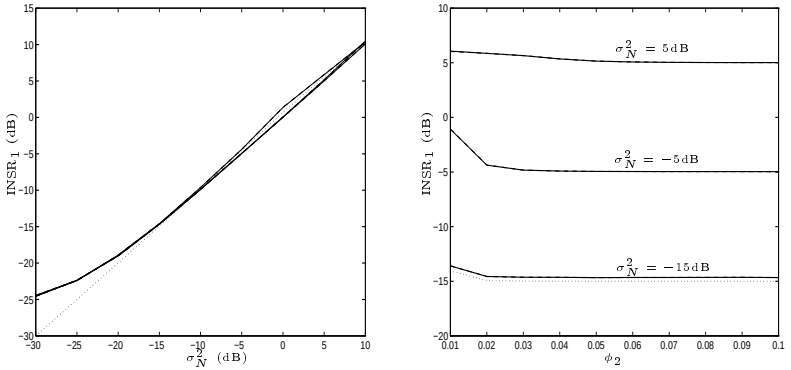



Figure 9.4: The INSR of the first source for the MVDR-beamformer in the second simulation of Sect. 9.2.4. Solid: the mixing matrix is estimated by means of the STOTD-technique. Dashed: the mixing matrix is estimated by means of the JADE-technique. Dotted: optimal matched filter. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

## 9.3 The JADE-algorithm

Besides the derivation by Comon [64] of the ICA-method based on maximal diagonality, an other major breakthrough in algebraic ICA-techniques was the development of the JADE-algorithm by Cardoso and Souloumiac [41]. "JADE" stands for "Joint Approximate Diagonalization of Eigen-Matrices" and is very similar in spirit to STOTD; as a matter of fact, the idea of simultaneous manipulation of "slices" of the tensor, rather than addressing the tensor as a whole, with the purpose to reduce the computational load while retaining an excellent performance, is reminiscent of [41]. In this section we briefly describe the JADE-concept, for completeness, and to allow an easy comparison with the algebraic methods that are developed in this thesis.

Consider the fourth-order cumulant $\mathcal{C}_4^Z$ of the standardized random vector $Z$ in the higher-order ICA-stage, as in Eq. (7.10). We associate a matrix-to-matrix mapping to $\mathcal{C}_4^Z$, in the following way:

$$\mathbf{V}' = \mathcal{C}_4^Z(\mathbf{V}) \qquad \Longleftrightarrow \qquad {v'}_{ij} = \sum_{kl} (\mathcal{C}_4^Z)_{ijkl} v_{kl} \tag{9.50}$$

for all index values. If we neglect the additive noise term in the ICA-model,



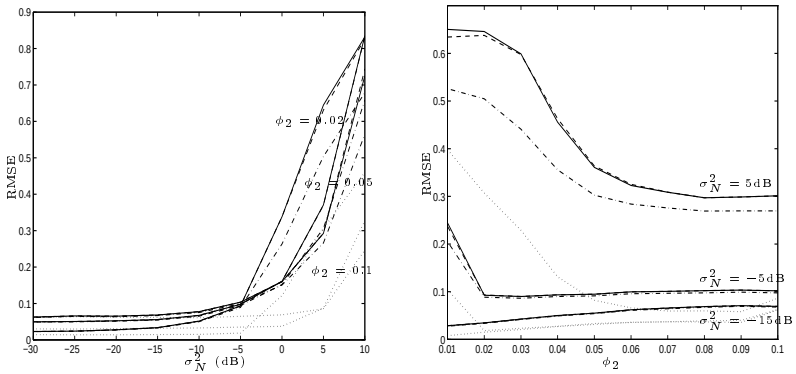



Figure 9.5: The RMSE between the true mixing matrix and its estimate, in the second simulation of Sect. 9.2.4. Solid: the mixing matrix is estimated by means of the STOTD-technique. Dashed: the mixing matrix is estimated by means of the JADE-technique. Dashdotted: the first upper-bound of performance in Theor. 7.6.4. Dotted: the second upper-bound of performance in Theor. 7.6.4. *Left:* Effect of the SNR ($\sigma_1^2 = \sigma_2^2 = 0$dB) on the quality of the reconstruction. $\phi_1 = 0$. *Right:* Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. $\sigma_1^2 = \sigma_2^2 = 0$ (dB).

then the EVD of the mapping (Theor. 2.8.4) is given by:

$$\mathcal{C}_4^Z = \sum_j^J \kappa_j^X (Q_j \circ Q_j^*) \circ (Q_j \circ Q_j^*)^*. \tag{9.51}$$

In the JADE-algorithm $\mathbf{Q}$ is estimated as the unitary (orthogonal) matrix that simultaneously diagonalizes, as far as possible in least-squares sense, the estimates of the eigenmatrices in Eq. (9.51). A variant is the simultaneous EVD of the "matrix slices" $(\mathbf{C}_4^Z)^{(k,l)}$, obtained by keeping the indices $k, l$ in $(\mathcal{C}_4^Z)_{ijkl}$ fixed. In the latter case, $J^2$ matrices are jointly diagonalized (of course, an estimate of $\mathbf{Q}$ can be obtained from a limited set of matrices in the span of the mapping, without exploiting all the information available in $\mathcal{C}_4^Z$). In the former case, the a priori computation of the EVD in Eq. (9.51) allows to restrict the set of matrices for the diagonalization procedure, to the $J$ dominant eigen-matrices.

The simultaneous EVD is computed by Jacobi-iteration. This iteration can e.g. be initialized with an estimate of $\mathbf{Q}$, obtained from the EVD of a single matrix in the span of the matrix-to-matrix mapping. Let the set of matrices to be diagonalized be given by $\{\tilde{\mathbf{C}}^{(p)}\}$. We also define $\mathbf{C}^{(p)} = \mathbf{J} \cdot \tilde{\mathbf{C}}^{(p)} \cdot \mathbf{J}^H$, in which we use the standard representation for the Givens rotation $\mathbf{J}$ (Eqs.

(8.10) and (8.24)).

The way in which an optimal Jacobi-rotation can be computed, is detailed in [246]. In the *real-valued case*, the function to be maximized has the following form, in terms of a matrix $\mathbf{G}$ of which the columns are given by $(\tilde{c}_{11}^{(p)} - \tilde{c}_{22}^{(p)}, -\tilde{c}_{12}^{(p)} - \tilde{c}_{21}^{(p)})^T$:

$$\sum_p (c_{11}^{(p)})^2 + (c_{22}^{(p)})^2 = U^T \cdot \mathbf{G}\mathbf{G}^T \cdot U, \tag{9.52}$$

in which $U = (\cos(2\alpha) \;\; \sin(2\alpha))^T$. Hence the optimal rotation can be found via the dominant left singular vector of $\mathbf{G}$.

In the *complex case*, the corresponding formula takes the form, in terms of a matrix $\mathbf{G}$ of which the columns are given by $(\tilde{c}_{11}^{(p)} - \tilde{c}_{22}^{(p)}, -\tilde{c}_{12}^{(p)} - \tilde{c}_{21}^{(p)}, i(-\tilde{c}_{12}^{(p)} + \tilde{c}_{21}^{(p)}))^T$:

$$\sum_p |c_{11}^{(p)}|^2 + |c_{22}^{(p)}|^2 = U^T \cdot \mathtt{Re}(\mathbf{G}\mathbf{G}^H) \cdot U, \tag{9.53}$$

in which $U = (\cos(2\alpha) \;\; \sin(2\alpha)\sin\phi \;\; \sin(2\alpha)\cos\phi)^T$. Hence the optimal rotation can be found via the dominant eigenvector of the real symmetric $(3 \times 3)$-matrix $\mathbf{G}\mathbf{G}^H$.

**Remark 9.3.1** Although this result might seem contra-intuitive, it turns out now that the simultaneous diagonalization of $J$ $(2 \times 2 \times 2)$-tensors is mathematically equivalent to the simultaneous diagonalization of $J$ Hermitean (symmetric) $(2 \times 2)$-matrices, that can be constructed from the third-order tensors. ◇

**Remark 9.3.2** The computational load of JADE and STOTD are comparable. In the real case, the computation of the optimal vector $U$ leads to the rooting of a quadratic polynomial in both methods. In the complex case, both methods lead to the rooting of a polynomial of degree 3. The only differences consist in the computation of the matrices of which the dominant eigenvector (or singular vector) has to be determined, and in the implementation of the Jacobi-rotation; these calculations only involve a limited number of sums and products. With this respect, the number of flops per third-order tensor slice in the STOTD-method is higher than the number of flops per matrix slice in JADE. On the other hand, JADE involves the manipulation of $J^2$ matrix slices (we neglect the possible a priori computation of the EVD in Eq. (9.51) at this point, to make a comparison possible). This suggests that the STOTD-method is slightly more efficient if the number of sources is large; for small numbers of sources, JADE is somewhat faster. ◇



**Remark 9.3.3** Like in other sections, the Jacobi-technique is not guaranteed to converge to the global optimum. [21] contains, for the simultaneous EVD-problem, an example of a stationary point that is different from the global optimum. However, problems of this kind have not been met in ICA-practice. ◇

Simultaneous diagonalization of the full set of matrix slices is equivalent to the maximization of the function $f(X) = \sum_{jkl}^{J} |(\mathcal{C}_4^{\hat{X}})_{jjkl}|^2$, in which $\hat{X}$ represents the source estimate. This function is a contrast; it is discriminating for the set of random vectors of which at most one component is non-kurtic.

In [245] it has been proved that the asymptotical performance (i.e. for infinitesimal errors in the cumulant estimate) of the MD-approach and the JADE-approach are the same. Moreover, it was indicated that for sources with kurtosis equal to $-1$, and in the absence of additive noise, both methods reach the ISR performance bound specified by Eq. (7.28).

## 9.4 An algebraic framework for ICA

From the previous section, it is clear that the simultaneous diagonalization of the third-order cumulant slices in the STOTD-technique is analogous to the simultaneous diagonalization of matrix slices in the JADE-technique. Actually, the MD-technique of Chapter 8 and the HOEVD-technique of Chapter 7 also fit in this algebraic framework. Obviously, as far as the maximal diagonality-approach is concerned, the standardized fourth-order cumulant $\mathcal{C}_4^Z$ is considered as a single "fourth-order tensor slice", that has to be diagonalized. In the HOEVD-method, $\mathcal{C}_4^Z$ is regarded as a set of "vector slices". The SVD of the matrix unfolding $\mathbf{C}_{4\,(n)}^Z$, of which the columns are the $n$-mode vectors of $\mathcal{C}_4^Z$, is computed - this is equivalent to the diagonalization of $\mathbf{C}_{4\,(n)}^Z \mathbf{C}_{4\,(n)}^{Z\,H}$.

Some important properties of the four approaches are brought together in Table 9.1.

## 9.5 Conclusion

In this chapter we have proved that the computation of an elementary Jacobi-rotation, in the optimal diagonalization of a third-order cumulant, can be formulated as the best rank-1 approximation of a real symmetric $(2 \times 2)$-matrix (real case) or $(3 \times 3)$-matrix (complex case). This is still true for the optimal diagonalization of a set of third-order cumulants; the matrix is obtained as a sum of contributions of the individual tensors in the set. Comparison to formulas

| aspect / technique | HOEVD | JADE |
|---|---|---|
| $\mathcal{C}_4^{\hat{X}}$ | $\mathcal{C}_4^{\hat{X}}$ all-orthogonal | $(\mathcal{C}_4^{\hat{X}})^{(k,l)}$ max. diagonal |
| contrast | $\sum_l^J \|(\mathcal{C}_4^{\hat{X}})^{(l)}\|^4$ | $\sum_{jkl}^J \|(\mathcal{C}_4^{\hat{X}})_{jjkl}\|^2$ |
| comp. $U$ (real) | linear | quadratic |
| comp. $U$ (complex) | linear | degree 3 |
| aspect / technique | STOTD | MD |
| $\mathcal{C}_4^{\hat{X}}$ | $(\mathcal{C}_4^{\hat{X}})^{(l)}$ max. diagonal | $\mathcal{C}_4^{\hat{X}}$ max. diagonal |
| contrast | $\sum_{jl}^J \|(\mathcal{C}_4^{\hat{X}})_{jjjl}\|^2$ | $\sum_j^J ((\mathcal{C}_4^{\hat{X}})_{jjjj})^2$ |
| comp. $U$ (real) | quadratic | degree 4 |
| comp. $U$ (complex) | degree 3 | (iterative) |

Table 9.1: A comparison of some important aspects of the four prewhitening-based ICA-techniques discussed in this thesis: (1) the criterion imposed on the source cumulant estimate, (2) the form of the corresponding contrast function, (3) the degree of the polynomial that has to be rooted for the computation of $U$, representing an optimal Jacobi-rotation, in the real case, and (4) in the complex case.

for simultaneous matrix diagonalization, leads to the result that simultaneous diagonalization of $N$ complex (real) super-symmetric $(2 \times 2 \times 2)$-tensors is mathematically equivalent to the simultaneous diagonalization of $N$ Hermitean (symmetric) $(2 \times 2)$-matrices, constructed from the third-order tensors.

We have shown that ICA can be realized by simultaneous diagonalization of third-order tensors, spanning the range of the standardized fourth-order cumulant. Computational requirements and performance are similar to those of the JADE-algorithm.

Actually the new STOTD-technique fills the gap in an algebraic framework for fourth-order cumulant-based ICA with prewhitening. In the HOEVD-technique of Chapter 7, the JADE-technique in Sect. 9.3 and the MD-technique of Chapter 8, the fourth-order cumulant is regarded as a set of vector slices, as a set of matrix slices, or a single fourth-order cumulant slice respectively, that are "diagonalized" in some sense.

# Chapter 10

# ICA and CANDECOMP

*In Sect. 7.2, we have explained that it is generally impossible to solve the ICA-problem resorting only to second-order statistics, although a reasonable source separation might be obtained under specific conditions (Sect. 7.9). Consequently, Sect. 7.2 proposed a general ICA-scheme in which an initial PCA is followed by a higher-order tensorial EVD-type step. The ICA-algorithms studied in Chapters 7-9 follow this scheme. In Sect. 7.6 however, we have stressed that estimation errors, introduced in the prewhitening, cannot be compensated in the higher-order stage. In this chapter we will show that, on the other hand, the ICA-problem can be solved resorting only to the higher-order cumulant. In this way, the new technique is conceptually insensitive to the presence of additive Gaussian noise.*

*The chapter is organized as follows. In Sect. 10.1 we will demonstrate that the cumulants of the observations and of the sources are related in terms of a CANDECOMP, such that the ICA-problem can be solved with the techniques discussed in Chapter 6. Some properties of the resulting algorithm will be investigated. The technique is illustrated in Sect. 10.2 with a number of simulations.*

## 10.1 Higher-order-only ICA

### 10.1.1 Concept

Let us use the same notation for the basic ICA-model as in Eq. (7.1). As a consequence of the multilinearity of higher-order cumulant tensors, the fourth-

order cumulant of the observations is given by

$$\mathcal{C}_4^Y = \mathcal{C}_4^X \times_1 \mathbf{M} \times_2 \mathbf{M}^* \times_3 \mathbf{M}^* \times_4 \mathbf{M} + \mathcal{C}_4^N. \tag{10.1}$$

As in Sect. 7.2, the algebraic key observation is that the source cumulant $\mathcal{C}_4^X$ is theoretically a diagonal tensor, since the source signals are assumed to be statistically independent (Prop. 3.2.7). On the other hand, the noise cumulant $\mathcal{C}_4^N$ vanishes if the noise is Gaussian (Prop. 3.2.9). Eq. (10.1) is in fact a tensorial equivalent of the diagonalization of a symmetric matrix by congruence. It can be rewritten as

$$\mathcal{C}_4^Y \simeq \sum_{j=1}^{J} \kappa_{X_j} M_j \circ M_j^* \circ M_j^* \circ M_j, \tag{10.2}$$

in which $\kappa_{X_j}$ represents the marginal cumulant of the $j$th source, and in which $M_j$ represents the $j$th column of $\mathbf{M}$. Eq. (10.2) is a decomposition of $\mathcal{C}_4^Y$ in a minimal number of rank-1 terms, as the columns of $\mathbf{M}$ are assumed to be linearly independent; it is the CANDECOMP of $\mathcal{C}_4^Y$ (see Chapter 6). This decomposition is unique if all the sources are kurtic, as will be reminded in Subsect. 10.1.2. However, as mentioned in Sect. 2.6, it is in general impossible to expand an $(I \times I \times \ldots \times I)$-tensor in a sum of $J \leqslant I$ rank-1 terms. The way in which the estimation error (finite datalength, additive noise, model misfit, etc.) of *sample* cumulants is dealt with, allows to distinguish different solution strategies. One could e.g. resort to the techniques discussed in Chapter 6. In the context of ICA, a limited number of other algebraic techniques have been published; these methods are overviewed together with the two-stage algorithms in Sect. 7.8. A general outline of the ICA-technique based on the CANDECOMP, is presented as Alg. 10.1.1.

### 10.1.2 Identifiability

The following uniqueness theorem is the counterpart of Theor. 7.3.3:

**Theorem 10.1.1** *Let the $N$th order super-symmetric tensor $\mathcal{C} \in \mathbb{C}^{J \times J \times \ldots \times J}$ be given by*

$$\mathcal{C} = \mathcal{D} \times_1 \tilde{\mathbf{Q}}_1 \times_2 \tilde{\mathbf{Q}}_2 \ldots \times_N \tilde{\mathbf{Q}}_N, \tag{10.3}$$

*in which $\mathcal{D} \in \mathbb{C}^{J \times J \times \ldots \times J}$ is diagonal, containing no null entries on the diagonal, and $\tilde{\mathbf{Q}}_n$ equals either $\mathbf{Q}$ or $\mathbf{Q}^*$, with $\mathbf{Q} \in \mathbb{C}^{J \times J}$ regular. $\mathcal{C}$ can be decomposed by the same model in terms of $\mathcal{D}'$ and $\mathbf{Q}'$ iff:*

- *$\mathbf{Q}' = \mathbf{Q}\Lambda\mathbf{P}$, in which $\Lambda$ is a diagonal matrix and $\mathbf{P}$ a permutation, and*



**Algorithm 10.1.1**
*ICA by means of CANDECOMP*

Given: $N$ samples $\{Y_n\}_{1 \leqslant n \leqslant N}$ of $Y = \mathbf{M}X + N$, as in Eq. (7.1) ($Y, N \in \mathbb{C}^I$, $X \in \mathbb{C}^J$, $\mathbf{M} \in \mathbb{C}^{I \times J}$). Call $\mathbf{A}_Y = [Y_1 Y_2 \ldots Y_N]$.

1. Compute sample cumulant $\hat{\mathcal{C}}_4^Y$ from $\mathbf{A}_Y$.
2. Obtain mixing matrix estimate $\hat{\mathbf{M}} = [\hat{M}_1 \hat{M}_2 \ldots \hat{M}_J]$ from the CANDECOMP of $\hat{\mathcal{C}}_4^Y$:

$$\hat{\mathcal{C}}_4^Y \simeq \sum_{j=1}^{J} \hat{\kappa}_{X_j} \hat{M}_j \circ \hat{M}_j^* \circ \hat{M}_j^* \circ \hat{M}_j$$

   (Chapter 6).
3. Source separation by means of MVDR, LCMV, ... beamforming.

- *$\mathcal{D}'$ is related to $\mathcal{D}$ in the inverse way:*

$$\mathcal{D}' = \mathcal{D} \times_1 \tilde{\mathbf{U}}_1^{-1} \times_2 \tilde{\mathbf{U}}_2^{-1} \ldots \times_N \tilde{\mathbf{U}}_N^{-1}, \tag{10.4}$$

  *where $\tilde{\mathbf{U}}_n$ represents $\mathbf{U}$ or $\mathbf{U}^*$, in accordance to Eq. (10.3), with $\mathbf{U} = \Lambda\mathbf{P}$.*

**Proof:** The theorem follows from the derivation in Sect. 6.2. □

**Remark 10.1.2** It is possible to smooth the conditions even further, to allow the identification of a transfer matrix with more columns than rows in a more sources-than-sensors set-up [37, 165] (in this situation, the sources themselves remain necessarily unknown, as the mixing matrix cannot be inverted). This topic falls outside the scope of this thesis. ◇

### 10.1.3 Simultaneous generalized Schur decomposition versus JADE

Solving the higher-order-only ICA-problem of Eq. (10.2) via the simultaneous generalized Schur decomposition of Sect. 6.3, is in fact the higher-order-only equivalent of the JADE-algorithm, discussed in Sect. 9.3. *Unsymmetric* and *symmetric* simultaneous EVD play exactly the same role. In the JADE-algorithm prewhitening leads to a simultaneous matrix decomposition from which an *orthogonal* matrix has to be computed; in the current approach the absence of prewhitening makes that a general *regular* matrix has to be determined. In other words: without prewhitening, *a pair* of orthogonal matrices

has to be determined, instead of a *single* orthogonal simultaneous eigenmatrix. The equivalent of simultaneous *diagonalization* turns out to be simultaneous *triangularization*.

In principle, it should be possible to generalize the MD- and STOTD-techniques as well. However, this will not be further investigated in this thesis.

### 10.1.4 The equivalent of PCA in higher-order-only ICA

Let us remind that the PCA-step in prewhitening-based ICA serves a three-fold goal:

- reduction of the parameter set of unknowns to the manifold of orthogonal matrices,
- standardization of the unknown source signals to mutually uncorrelated unit-variance signals, and
- determination of the number of sources.

In higher-order-only ICA it makes still sense to reduce the dimensionality of the ICA-model to the true number of sources, as this may significantly reduce the computational efforts in the actual algorithm. As already indicated in Remark 6.2.2, the preprocessing can be carried out with the techniques of Chapter 5. The way in which this dimensionality reduction can be achieved, shows some analogy with PCA, but at the same time it is also remarkably different.

According to the definition of $n$-rank, every tensor satisfying Eq. (10.2) for kurtic sources, is a rank-$(J, J, J, J)$ tensor. To deal with the situation in which $I > J$, we propose to perform an orthogonal projection of the sample cumulant $\hat{\mathcal{C}}_4^Y$ on the manifold of super-symmetric rank-$(J, J, J, J)$ tensors, before addressing the harder problem of the further projection on the submanifold of rank-$J$ tensors and the actual computation of decomposition (10.2). Notice that in the second-order case both projections coincide, as $n$-rank and rank are necessarily the same, which corresponds to the single best rank-$R$ approximation problem, associated with PCA.

## 10.2 Simulations

A first simulation is similar to the first experiments in Subsect. 7.7.3, 8.2.3 and 9.2.4. The only difference is the way in which the mixing matrix has been



generated: after drawing the entries from a zero-mean unit-variance Gaussian distribution, the singular values are set equal to $\{2, 1\}$ (see Sect. 6.7), and the resulting columns are rescaled to unit-length. The mixing matrix is estimated by means of three procedures: (1) direct two-sided Jacobi-rotation, (2) followed by least-squares matching, and (3) extended $QZ$-iteration. These methods have been implemented in the same way as in Sect. 6.7; only, for the least-squares matching procedure, a maximum of 1600 steps has been put forward.

The results are visualized in Figs. 10.1 and 10.2. With respect to Fig. 10.1, we remark that for small data lengths, the algorithms tended to identify two times the same steering vector for a number of datasets, with as a result much higher ISR values than in the majority of cases. To keep the plot meaningful, these Monte Carlo runs have been discarded. For $T = 50$, this concerned 16, 17, 18 cases in the (b)/(c)-experiment, and 1, 1, 2 cases in the (a)/(c)-experiment, for methods (1), (2) and (3) respectively. For $T = 100$, a single $QZ$-iteration had to be neglected in the (a)/(c)-experiment and two $QZ$-iterations in the (b)/(c)-experiment. In all these cases, the condition number of the estimated mixing matrix was at least outside the interval $(m_{\text{rem}} - 25\sigma_{\text{rem}}, m_{\text{rem}} + 25\sigma_{\text{rem}})$, in which $m_{\text{rem}}$ and $\sigma_{\text{rem}}$ represent the mean and standard deviation of the condition numbers obtained for the remaining samples. The efficiency (Eq. (3.37)) of the displayed values ranges from at least 0.0027 ($T = 50$) to at least $1.2e-8$ ($T = 10000$). For Fig. 10.2 all 500 Monte Carlo runs have been taken into account. As convenient, the dB-values in Fig. 10.1 have been obtained by a logarithmic transformation *after* averaging over the various Monte Carlo runs; on the other hand, averaging the dB-values obtained in each run, would lead to curves that e.g. start some 5 dB lower at $T = 50$. With respect to Fig. 10.2, the Frobenius-norm errors have been determined as the minimal error, over the possible column reorderings and rescalings of the mixing matrix estimates.

Figs. 10.1 and 10.2 show similar curves as their counterparts in e.g. Subsect. 8.2.3 and 9.2.4. The accuracy is a bit lower than for the two-stage ICA-algorithms, though: due to the absence of (Gaussian) noise in the experiment, the attained precision is mainly determined by the accuracy with which the higher-order cumulant has been estimated. As indicated in Sect. 3.5 this cumulant is somewhat harder to estimate than the covariance, for a given probability distribution and data length.

The effect of additive Gaussian noise is investigated in a second experiment, of which the results are shown in Figs. 10.3 and 10.4. The mixing matrix is generated as in the first experiment. The sources have both a binary distribution, with equal probability for the values $\pm 1$. Additive Gaussian noise with variance $\sigma_N^2$ is added to the dataset. The ICA-algorithms are the higher-order-only algorithm, based on direct two-sided Jacobi-rotation (solid), and the prewhitening-based JADE-algorithm (dashed) (Sect. 9.3). 100 Monte Carlo runs have been carried out. For the same reason as in the first experiment, a number of runs have been discarded in the ISR-plot of the CANDECOMP-method: 5 runs at

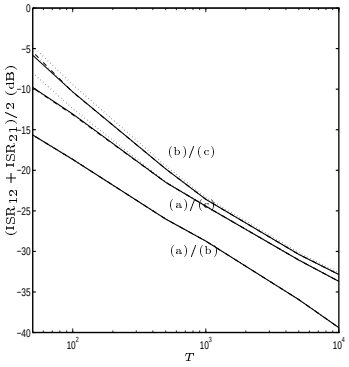



Figure 10.1: The mean ISR of the LCMV-beamformer in the first simulation of Sect. 10.2, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated by direct two-sided Jacobi-rotation. Dashdotted: the mixing matrix is estimated by extended $QZ$-iteration. Dotted: the mixing matrix is estimated by Jacobi-rotation, followed by least-squares matching.

$\sigma_N^2 = 0$dB ($T = 100$), and 8, 9, 1 runs at $\sigma_N^2 = 5$dB ($T = 100, 1000, 10000$). In the RMSE-plot, all runs have been taken into account. The minimal efficiency value is 0.0025.

For high SNR-values we observe a situation like in the first experiment, namely, prewhitening-based ICA is advantageous, as smaller datasets are required to estimate the covariance with the same precision as the fourth-order cumulant. However, as the noise power increases, it directly affects the estimation of the covariance, while it is to some extent, depending on the length of the dataset, cancelled out in the estimation of the cumulant. In a wide range the CANDECOMP-approach yields more accurate results than the JADE-algorithm. Data length can be traded off for accuracy: roughly, recording a ten times longer data sequence corresponds to a decrease of the ISR with 10 dB. This is consistent with the theorems in Sect. 3.5. For poor SNR-values and short datasets, the interest of higher-order-only ICA decreases again, as it becomes hard to estimate the higher-order cumulant with a sufficient accuracy. We also mention that in this simulation the accuracy of the CANDECOMP-estimate could not be further improved with least squares matching.

Non-Gaussian noise components are not attenuated in the estimate of the higher-order cumulant. For the influence of perturbations of the higher-order tensor, and other more general effects, we refer to the simulations in Sect. 6.7.



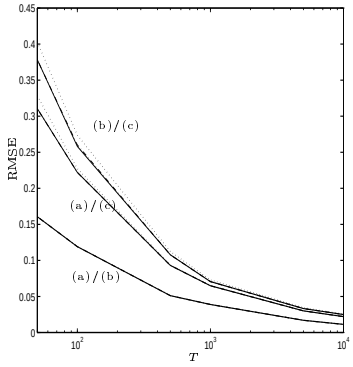



Figure 10.2: The RMSE between the mixing matrix and its estimate, for the higher-order-only ICA-techniques in the first simulation of Sect. 10.2, as a function of the length of the dataset $T$. Configurations of the source distributions: (a) binary, (b) uniform, (c) double-sided exponential. Solid: the mixing matrix is estimated by direct two-sided Jacobi-rotation. Dashdotted: the mixing matrix is estimated by extended $QZ$-iteration. Dotted: the mixing matrix is estimated by Jacobi-rotation, followed by least-squares matching.

## 10.3 Conclusions

We have presented a new ICA-technique that resorts only to the higher-order cumulants of the observations. We have indicated the uniqueness properties of the transfer matrix estimate in our approach. Higher-order-only BSS has the advantage, w.r.t. prewhitening-based ICA, 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 the estimation of higher-order statistics generally requires larger datasets than the estimation of second-order statistics. In a more sensors-than-sources scenario, the dimensionality of the problem can be reduced to the number of sources without resorting to second-order statistics, by exploiting the difference between the concepts of $n$-rank and rank of higher-order tensors.

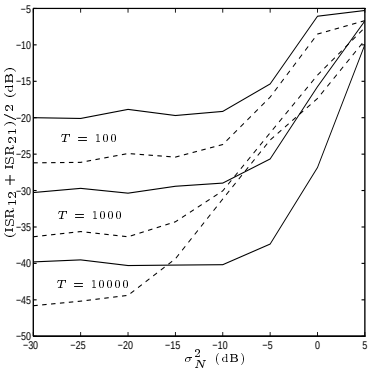



Figure 10.3: The mean ISR of the LCMV-beamformer in the second simulation of Sect. 10.2, as a function of SNR. The parameter $T$ is the length of the dataset. Solid: the mixing matrix is estimated by means of the CANDECOMP-approach based on direct two-sided Jacobi-rotation. Dashed: the mixing matrix is estimated by means of the JADE-algorithm.

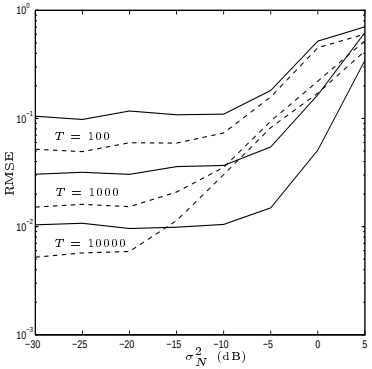



Figure 10.4: The RMSE between the true mixing matrix and its estimate, in the second simulation of Sect. 10.2. The parameter $T$ is the length of the dataset. Solid: the mixing matrix is estimated by means of the CANDECOMP-approach based on direct two-sided Jacobi-rotation. Dashed: the mixing matrix is estimated by means of the JADE-algorithm.

# Chapter 11

# Applications

*In Chapter 6, it was explained that the CANDECOMP-model combines a relative simplicity with interesting uniqueness properties and a high interpretability. What this means in practice, is illustrated in the current chapter by using the model to describe the form of the vocal tract during vowel production; as will be clarified, this forms an important issue in Linguistics (Sect. 11.1).*

*A second part of this chapter deals with the topic of ICA; the assets of this fundamental concept are illustrated with two examples taken from different contexts. The most-quoted application is blind beamforming in telecommunications. The background of this application is briefly sketched in Sect. 11.2. In Sect. 11.3 we propose a variant of ICA, called Blind Source Subspace Estimation (BSSS), as an innovating way to solve a classical problem in biomedical engineering: the extraction of the fetal electrocardiogram (FECG) from potential measurements recorded on the mother's skin.*

## 11.1 Factor analysis of tongue shapes

### 11.1.1 Modelling the vocal tract during vowel production

The first application example is situated in Linguistics. Maybe surprisingly at first sight, it turns out that the CANDECOMP is a highly suitable model for the description of tongue positions that occur in speech. This is a very active field because of the great interest in the neural control of speech motor movements, all part of the overall desire to understand how ideas become language.

In the literature a number of techniques have been proposed to parametrize

tongue shapes in a compact but yet accurate way. Among the features that have been discussed, we mention the coordinates of the highest point of the tongue, the width of the vocal tract at its narrowest point together with the distance of this point from the glottis, the coordinates of the center of a circle an arc of which reflects the position of the body of the tongue, the degree of distortion of the tongue shape from a neutral reference together with the position of the point of maximal displacement, and truncation of the Fourier series describing the profile of the tongue after a coordinate transformation that straightens out the vocal tract (for bibliographic pointers, we refer to [141]).

It were Harshman, Ladefoged and Goldstein [141] who proposed the use of a multilinear factor-analytic technique (CANDECOMP), for the description of tongue shapes. In the next subsection, we will analyze their dataset with the methods discussed in Chapter 6. Subsect. 11.1.3 contains a discussion on the assets of the CANDECOMP-approach.

### 11.1.2 Analysis

The dataset is a real-valued $(5 \times 10 \times 13)$-tensor; basically, it was obtained as follows. High-quality audio recordings and cine-fluorograms were made of five English-speaking test persons while saying sentences of the form "Say h(*vowel*)d again". The vowels included /i, ɪ, eɪ, ɛ, æ, a, ɔ, o, ɷ, u/ as in "heed, hid, hayed, head, had, hod, hawed, hoed, hood, who'd"; these vowels cover the range of articulatory movement that occurs in the pronunciation of English vowels. For each of these 10 vowels an acoustic reference moment was defined and the corresponding 5 (corresponding to the different speakers) frames in the film located. Next, speaker-dependent reference grids, taking into account the anatomy of each test person, were defined and superimposed on the remaining x-ray images. The grids consisted of 13 equidistant lines, in the region from epiglottis to tongue tip, more or less perpendicular to the midline of the vocal tract. The tensor entries now consist of the distance along the grid lines between the surface of the tongue and the harder upper surface of the vocal tract, after subtraction of the mean distance, averaged over all vowels. The obtained values are referred to as "pseudosagittal dimensions"; they display the displacement of the tongue surface from that speaker's reference position. The dataset can be downloaded from the Web [105]. The pseudosagittal dimensions are given in centimeters and have been measured to the nearest 0.5 mm.

First we observed that the two dominant 1-mode, 2-mode and 3-mode singular values explain 94.5 %, 95.4 % and 96.0 % respectively of the "energy" in the dataset. Therefore we performed a dimensionality reduction by calculating the best rank-$(2,2,2)$ approximation of the dataset, before starting the actual CANDECOMP computations (see Remark 6.2.2). The approximation was obtained with higher-order orthogonal iterations, initialized with the truncated



HOSVD. The stop criterion consisted of checking if the adjustment of each of the component matrices in an iteration step was below $1e-4$ (Frobenius-norm). For programming convenience, a full dominant subspace estimation was carried out in each step. The algorithm converged in 5 steps. The approximation contained 92.6 % of the energy.

Next we looked for the least-squares approximation of the $(2 \times 2 \times 2)$-core tensor by a sum of two rank-1 components. The solution could be found by rooting a polynomial of degree 8, as explained in Sect. 6.4. The core tensor belongs to the class of $(2 \times 2 \times 2)$-tensors of which the rank equals two (and not three, see Sect. 2.6), as the error of the fit was in the order of the numerical accuracy of MATLAB. Backtransformation to the original dimensionality by multiplication with the best rank-$(2,2,2)$ components yielded the best rank-2 approximation of the original dataset. Further enhancement by e.g. least-squares matching was not possible; otherwise, the rank-$(2,2,2)$ approximation would not have been optimal.

The cosine of the angle in $\mathbb{R}^{5\times10\times13}$ between the original data tensor and its rank-2 approximation was equal to 0.962, which was even slightly better than the result of [141] (0.956); the latter result had been obtained by repeating ALS iterations for different rank estimates and different starting values, and cross-examining the results. On a SUN Ultra 2 Sparc and using MATLAB 4.2c, our computations took $0.2+0.04s$ of CPU-time, which was a drastic improvement[1].

### 11.1.3 Discussion

The CANDECOMP factors can be interpreted as follows:

- The 3-mode factors represent the components of a normalized speaker-independent displacement of the tongue along reference gridlines. It appears that two factors are enough to describe the tongue positions, occurring during the pronunciation of English vowels, with a sufficient accuracy (the results have been validated for other datasets). These components are displayed in Fig. 11.1. In [207] it has been shown that the modelling process is equally successful if the tongue positions are measured in an unconstrained $(x, y)$-coordinate system.
- The 2-mode factors give the contribution of each mode of tongue displacement to each vowel and allow to compare the articulatory features of different vowels in a speaker-independent way.
- The 1-mode factors show the speaker-specific contribution to each mode of tongue displacement; this modulation is modelled as uniform over the

[1] Personal communication with Prof. Peter Ladefoged (UCLA, Linguistics Dept., Los Angeles).

production of the different vowels. It has been shown that the factor entries are correlated with some gross anatomical features of the vocal tract of the different speakers [141].

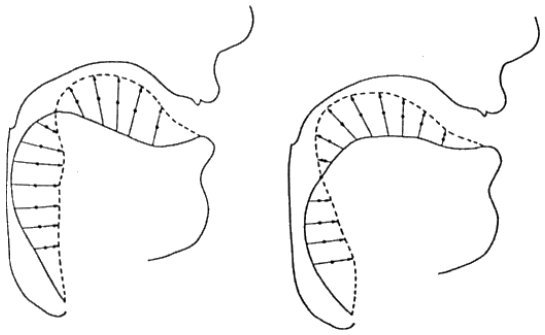


Figure 11.1: Visualization of the 3-mode factors of the dataset described in Sect. 11.1.2. The points show the mean position of the vocal tract. The solid and dashed lines indicate the tongue positions corresponding to the CANDECOMP-factors, with a plus- and minus-sign.

With respect to the parametrization techniques mentioned in Subsect. 11.1.1, the CANDECOMP-model has a number of advantages: (a) it is less intuitive, (b) it is generally easier to quantify the quality of the fit, (c) like for the spectral analysis method, the features are chosen in such a way that they insure an optimal fit, (d) individual differences among speakers are integrated in the model, and (e) the techniques in Sect. 11.1.1 are less suitable for interpretation as "real" features.

With this respect, we also remind that a classical second-order factor analysis would be less appropriate than the multiway technique, as it does not allow to incorporate individual differences either, and since its components are not unique.

Finally we mention that the CANDECOMP-technique has served as the starting point for several further studies, e.g. in the analysis of cross-linguistic effects underlying tongue shapes for vowels [207], and in the development of an algorithm for the computation of vocal tract shapes, given acoustic properties (the three basic "formant frequencies") of a particular speech sound [175].



## 11.2 Blind beamforming in wireless communications

The most intensively studied application area of ICA-type problems is spatial multiplexing in wireless communications. The simulations in Chapters 7-10 have been chosen in this context. In this section, we will briefly sketch the background of the application.

Assume a number of "source signals", emitted at the same time and frequency, but at different locations. After propagation, the signals are captured by an array of antenna elements. If the sources are narrow-band signals [222], modulated onto a sinusoidal carrier, then the complex envelope representation of the antenna outputs is given by

$$Y(t) = \mathbf{M}X(t) + N(t). \tag{11.1}$$

In this equation, $Y(t)$ contains the antenna outputs, $X(t)$ represents the sources and $N(t)$ is an additive noise term at time instant $t$. The matrix $\mathbf{M}$ depends on the propagation and the antenna characteristics; more in particular, its entries are influenced by the angles of arrival. The goal of a beamformer is to linearly combine the antenna outputs to obtain an estimate of the original source signals, free from interference (for a number of beamforming strategies, we refer to Sect. 7.6). The beamforming problem is visualized in Fig. 11.2.

The data model (11.1) is standard in array processing [160]. It arises e.g. in radar, radio astronomy, satellite communications [149]. The current interest is largely related to wireless terrestrial communication systems, such as cellular mobile telephony, and wireless data networks [6, 8, 251]. In this context, beamforming may be used to increase the capacity of the communications system by "spatial division multiple access" (SDMA), complementarily to the techniques of time and frequency division multiple access (TDMA,FDMA). Directional transmitting has the additional advantage that point-to-point links require significantly less power than omnidirectional broadcasting; at the same time, this lowers the probability of interference with distant users, allocated to the same time/frequency bin. A good introduction to beamforming in the context of mobile communications is the first chapter of [273].

A number of important algorithms have been developed to solve the multi-user beamforming problem by exploiting a priori knowledge on the mixing mechanism. In particular, it is assumed that each direction-of-arrival (DOA) corresponds to a unique steering vector, and that the set of all possible steering vectors (the "array manifold") is known. This implies the use of calibrated antenna arrays (e.g. MUSIC [229]) or special array geometries of identical antenna elements (e.g. ESPRIT [222]). However, it is known that these techniques are very vulnerable to errors in the array manifold model. An other disadvantage is related to multipath: the fact that the signal, emitted by a

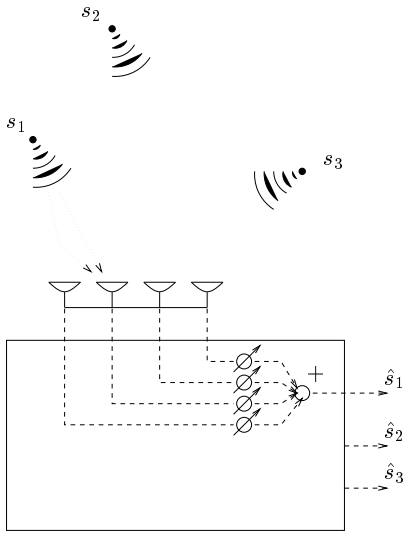



Figure 11.2: The beamforming problem.

single source, may reach the antenna array in different ways, and from different DOA's. For short-delay or diffuse multipath, it might be more accurate to model the corresponding steering vector as the sum of two or more vectors on the array manifold, which is problematic from a computational point of view.

With this respect it is preferable to solve the beamforming problem in a blind way. This approach is by definition insensitive to calibration errors; steering vectors consisting of a linear combination of vectors on the array manifold will be found directly in the form of the sum. It is natural to assume that different sources are statistically independent; hence, Eq. (11.1) may be considered as an ICA-problem. Other approaches are to resort to signal properties, besides mutual statistical independence, as constant modulus [266], finite alphabet [253], spectral self-coherence [4] or cyclostationarity [187, 258, 199, 244].

It should be noted that the ICA-model involves a *scalar* mixture, i.e. it is assumed that different multipath components reach the antenna array within the same sampling interval. On the other hand, multipath delays of 10-15 $\mu$s are observed in practice; comparison with e.g. a bit rate of 270.3 kbit/s in GSM [200] shows that in the context of mobile communications the problem rather consists of the equalization of a set of linear time-invariant filters; by performing this equalization in a blind way, instead of resorting to training sequences, inefficient use of the available spectrum could be avoided. Van der



Veen et al. have shown that such problems can often be separated into a separation and an equalization step [267], such that blind beamforming techniques are also important in the more general context.

Finally, we remark that the high throughput rates in wireless communication systems are extremely demanding in terms of required computational power. Because of the fast fading, due to the interference of multipath components, propagation channels should be re-estimated in a period in the order of magnitude of 5 ms [234]. This imposes hard constraints on the development of fast algorithms, application specific processors and strategies for parallel computing. With this respect, the integrated development of algorithms and architectures, often called "algorithmic engineering" [189, 190, 218], is important.

## 11.3 FECG extraction by BSSS

### 11.3.1 Introduction

The mechanical action of the heart muscle is caused by the propagation of an electrical wavefront, known as the *depolarization wave*, which is initiated by the nerve system, and followed by *repolarization* (recovery). This quasi-periodical stimulus involves an electrical current, propagating through the body and resulting in potential differences. The potential differences can be measured between electrodes on the skin (*cutaneous* recordings). The registration of these potential signals, visualized as a function of time, is called the *electrocardiogram* (ECG). (We remark that in medical practice the term additionally implies that the electrodes are placed at standard locations on the person's chest and/or limbs.)

The ECG is a common and fairly effective diagnostic tool, and in addition it is cheap and painless. It should be noted that, since the electrical activity is the driving mechanism behind the mechanical activity of the heart, the ECG reveals important medical information that cannot be extracted from audiorecordings (e.g. [193]). On the other hand, it is possible to obtain important information, not merely about the electrical system that controls the mechanical function of the heart, but also about the health of the heart muscle itself, from potential measurements: mechanical events can influence the electrical context to some extent. A recent evolution in the field is *cardiac electrical imaging*, which is a generalization of the ECG, in which more information is acquired by using a larger array of (e.g. 200) electrodes to record a sequence of "electrical images" of the body [20]. This technique can be seen as an emerging modality for medical imaging, complementary to e.g. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI); it is worth mentioning that in Japan the technique is already common practice (*Body Surface Potential Mapping* (BSPM)).

Like for adults, it should be possible to visualize the electrical activity of a fetal heart: the *fetal electrocardiogram* (FECG) contains important indications about the health and condition of the fetus. With this respect, analysis of the *fetal heart rate* (FHR) has become a routine procedure for the evaluation of the well-being of the fetus [270] (variations of the FHR (*fetal heart rate variability* (FHRV)) are a result of a continuous interaction with the central nervous system - a loss of beat-to-beat variation may indicate that the fetus is under stress). The cardiac waveform reveals important diagnostic information as well, e.g. for the diagnosis of *arrhytmia* (abnormal heart rates). For an overview of some clinical applications of FECG-analyses, we refer to [26], Sect. 2.2.2.

During delivery accurate recordings can be made by placing an electrode on the fetal scalp. As long as the membranes protecting the child have not been broken (*antepartum*), one should look for non-invasive techniques. Currently the FHR is mostly measured using ultrasound techniques, based on a Doppler-shifted ultrasonic heart echo [109, 243]. The main advantage of this approach is that it is almost guaranteed that a FHR recording will be obtained. On the other hand, the method requires that the ultrasound beam is pointed at the fetal heart, which may be hard to achieve; an experiment may involve continuous replacement of the transducer, due to fetal and maternal movements. A second disadvantage is the poor accuracy, for a number of reasons: (a) the mechanical movement of the heart may introduce spurious FHRV, (b) most Doppler systems rely on a number of heart beats to compute an averaged instantaneous FHR, so they don't record the real beat-to-beat variability, and (c) monitoring units are often only able to process FHR's up to 220 bpm - in case of higher rates, the number is divided by two, resulting in an apparently normal rate, which hides however the fact that the fetus is in *tachycardia* (FHR $\geqslant$ 200 bpm). These disadvantages are not shared by the alternative technique of examining the FECG from ECG-recordings measured on the mother's skin. As indicated before, ECG-data also allow to analyze the cardiac waveform. The main disadvantage is the poor SNR of most recordings, and the fact that the approach is less robust. We also mention that the amplitude of the FECG is observed to decrease around the 30th week of pregnancy - although this is considered to be an important period for fetal monitoring - due to the isolating effect of the *vernix caseosa*, which surrounds the fetus at that time [208]. Other means to obtain FHR-recordings are processing of the *fetal magnetocardiogram* (FMCG) and analysis of the audio sound emitted by the heart (*phonocardiography*). In [74] these different techniques are compared with respect to monitoring over larger periods of time (e.g. 24 hours) and long term ambulant use, and it was concluded that processing of the ECG can be qualified as the most appropriate approach.

The aim of this section is to show that ICA is a promising tool for the estimation of the FECG from recordings on the mother's skin. In Subsect. 11.3.2 we motivate that cutaneous recordings contain instantaneous linear mixtures of MECG and FECG, and we briefly situate the ICA-approach against other methods for the extraction of the FECG from cutaneous potential measurements. The ICA-method itself is further discussed in Sect. 11.3.3. Subsect. 11.3.4 contains a real-life example.



### 11.3.2 Extraction of the FECG from abdominal recordings

Potential measurements on the mother's skin contain contributions from several bioelectric phenomena (maternal and fetal heart activity, potential distributions generated by respiration and stomach activity, ... ) and are affected by various kinds of noise (thermal noise, noise from electrode-skin contact, ... ). Two aspects have to be discussed here: first, the nature of the occurring signals and secondly, the characteristics of the propagation from bioelectric source to electrode.

In [216] it is shown that, at a considerable distance from the mother heart, its activity as a bioelectric current source can be represented in first order approximation by a three-dimensional vector signal, that can be imagined as the effect of a rotating current dipole in the chest (the model has further been refined with BSPM [20]). The three-dimensional vector space, described by the discrete-time evolution of the maternal ECG (MECG) after sampling, will be called the *MECG-subspace*. On the other hand [208] states that the "dimension" of the fetal heart, i.e. the number of independent signals describing its electrical activity, is not necessarily equal to three, but subject to changes during the period of pregnancy. In this paper the term *FECG-subspace* will be used. In comparison with the low-voltage range of the FECG other electrical signals can play an important role too: *electromyographic* activity (electrical potentials generated by the muscles, the uterus, etc.), 50 Hz net-interference, etc.

Moreover the ECG-recordings are corrupted by noise that is not necessarily generated by bioelectric mechanisms. A first cause of disturbance is thermal noise in the electronic recording equipment. A second example consists of noise generated by the electrode-skin contact: due to pressure, movements, etc., there exist electrode-specific contributions to the junction potential. ([26], Sect. 2.3.3).

The transfer from bioelectric current source to body surface electrode can be assumed to be linear and resistive [216]. On the other hand the frequency at which the cutaneous potential distribution is sampled (typically 500 Hz) can be considered as low, taking into account the high propagation velocity of the electrical signals. We can conclude that, in first approximation, cutaneous potential measurements can be considered as instantaneous linear mixtures of

potential signals generated by underlying bioelectric phenomena; noise can be taken into account as an additive perturbation.

The FECG and MECG in cutaneous potential recordings are overlapping in time as well as in frequency. As a consequence, separation of these signals cannot be realized by simple windowing or linear filtering. For a first overview of techniques that aim at the extraction of the FECG from the available measurements, we refer to [26], Sect. 2.2.3. Briefly, the earliest attempts involved a scaled subtraction of an approximate MECG-recording, obtained by measuring at a distance from the fetus (e.g. on the *thorax* (chest)), from a linear superposition of MECG and FECG, recorded on the mother's *abdomen* (belly). Among the more involved approaches we mention the use of matched filters, adaptive noise cancellers, and spatial filtering (in which the FECG is estimated as a linear combination of a number of electrode signals - the method to be discussed further, belongs to this class). Most of these techniques have one or more of the following drawbacks:

- together with the elimination of the MECG, important contributions of the FECG are eliminated as well, especially when the fetal and maternal heartbeat coincide,
- the FECG is only described by one signal,
- only the fetal *QRS-complex* (the central part of the cardiac waveform, with high potential values) is extracted.

Recently, a number of methods have been proposed to extract the FECG from a single-lead recording. The mathematics behind these approaches are diverse: the interpretation of the fetal cardiac waveform as the orbit of a non-linear stationary system [232], wavelet analysis [116], pattern matching (e.g. [3]), ... These techniques have the obvious advantage of a simple data acquisition set-up, but on the other hand, they all rely on some kind of pattern averaging to detect and isolate the FECG, making it unlikely that unforeseen FECG-patterns can be succesfully dealt with.

### 11.3.3 Extraction of the FECG with BSSS

The propagation of $q$ bioelectric sources to an array of $p$ body surface electrodes ($p \geqslant q$), can be formulated as

$$Y(t) = \mathbf{M}X(t) + N(t), \tag{11.2}$$

where $Y(t) = (y_1(t) \dots y_p(t))^T$ contains the potential recordings, $X(t) = (x_1(t) \dots x_q(t))^T$ contains the signal values of the bioelectric sources, and the



noise on each channel is represented by $N(t) = (n_1(t) \dots n_p(t))^T$. The matrix $\mathbf{M}$ describes the propagation from source to electrode. It is natural to assume that the different bioelectric sources - since they originate at different locations, correspond to different mechanisms, etc. - can be approximately modelled as statistically independent. The noise components $N_i(t)$ $(1 \leqslant i \leqslant q)$ are assumed to be Gaussian, with variance $\sigma_N^2$, mutually independent as well as independent from the source signals.

As a conclusion, the derivation of the antepartum FECG from multilead cutaneous recordings can be considered as an example of BSS (Chapter 7), in which however the sources have a multidimensional nature. With this respect, we will use the term *Blind Source Subspace Separation* (BSSS). The fact that only the different source subspaces have to be separated, instead of all the source components, limits the number of Jacobi-rotation angles that have to be identified, in comparison to conventional ICA, and allows to reduce the computational cost without loss of medical information. Since there is a large gap between the amplitudes of the MECG and the FECG, a good separation can be expected from a mere PCA (Sect. 7.9); this is the philosophy behind the important class of SVD-techniques for the extraction of the FECG [24, 26, 29].

At a conceptual level, the higher-order processing step in ICA may add the following advantages to the second-order approach:

- It is possible to enhance the quality of separation, as explained in Sect. 7.9; Subsect. 11.3.4 contains an illustration. In case the higher-order ICA-step would fail, one can still resort to the results of the PCA-step.
- The propagation of the electrical signals can be characterized in an essentially unique way. We mention three important implications:
  - The transfer vectors indicate how strongly the different electrodes capture each source signal; from this information, better measurement positions might be deduced. It should be mentioned that the positioning of the electrodes is still the most crucial factor for the success of the PCA-method [29].
  - An important aspect in the evaluation of the fetal well-being is the quantification of fetal movements [26]. At this moment the required information can only be obtained by echography or, simply, by asking the mother. The number of significant changes in the FECG-subspace, which could be obtained from an on-line adaptive ICA-implementation, could be very useful information here.
  - The properties of the human body as a conducting medium are, in their own, subject of medical research [216]. The study of the propagation of the fetal heart signal to the mother's skin is an important subaspect [208]. The transfer matrix can provide more understand-

ing with respect to the propagation of electrical signals through the body.

- The physician can resort to a more intuitive interpretation of the results: the separation of the measured signals into statistically independent source signals with a physical meaning, is easier to interpret than a decomposition in time-orthogonal principal components.

We stress the fact that the FECG-extraction is formulated as a *blind identification* problem, since it is less meaningful in practice to resort to a more parametric approach:

- The transfer coefficients are subject to a large uncertainty: the development of propagation models is still in its infancy. Moreover it is clear that length, weight, contour, etc. are significantly different from patient to patient.
- The geometrical and resistivity parameters of the body of a single patient are not constant in time. Fetal growth, a different position of the fetus in the uterus, the variation in the characteristics of the amniotic fluid and the placenta during pregnancy, the changing geometry, ... imply important changes of the transfer matrix.
- For the application in medical diagnosis and treatment it is crucial that *un*expected ECG-patterns can be detected and examined. E.g. the parametric formulation of the quasi-periodicity of a regular heart rate pattern would hamper the detection of *extrasystoles* (extra heart beats between the regular beat-to-beat pattern).

We may conclude that conceptually BSSS is a very promising technique to tackle the problem of FECG-extraction. Subsect. 11.3.4 contains an example. It is clear however, that hard conclusions on the merits and drawbacks of the method can only be drawn, after intensive testing in medical practice to which extent the assumptions, underlying the ICA-model, are valid.

### 11.3.4 Example

Fig. 11.3 shows a set of potential signals measured in an 8-channel experiment. The dataset is available from the Web [105]. The horizontal axis displays the time in seconds; concerning the vertical axes only the relative values are important. The sampling frequency was 500 Hz. For details about the data acquisition we refer to [29]. Channels 1 to 5 show abdominal signals; for channels 6 to 8 the electrodes have been placed further away from the fetus, e.g. on the thorax. Channels 1 and 3 clearly contain weak fetal contributions. Due



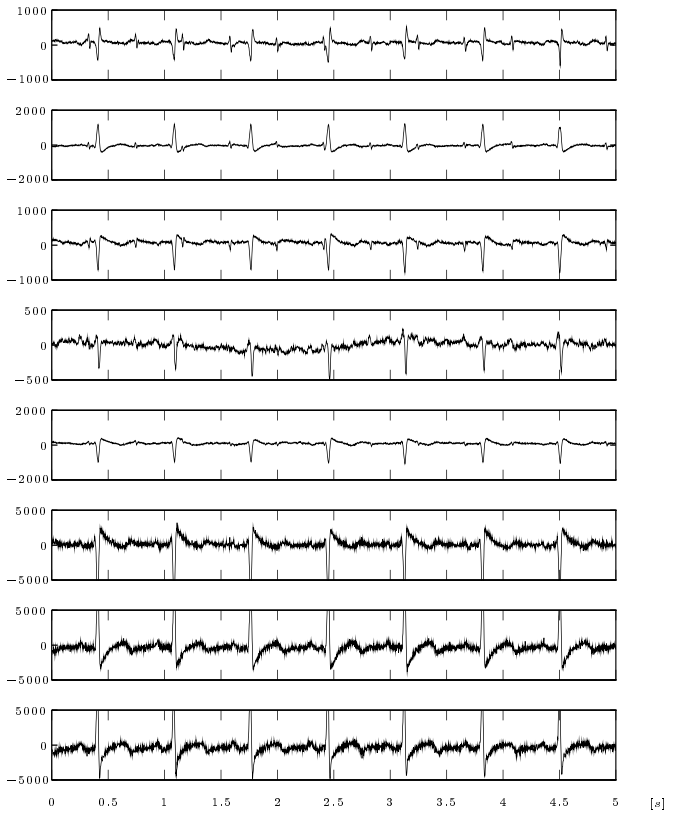



Figure 11.3: 8-channel set of cutaneous data recordings.

to the large amplitudes of the MECG in the thoracic signals, the FECG is less visible. However the strong pick-up of MECG allows an accurate modelling of the electrical activity of the mother heart. From the figure it is clear that there is a large gap between the oriented energy in the MECG-subspace and the FECG-subspace.

The source estimates after PCA are displayed in Fig. 11.4. Two MECG-free FECG-components have been obtained as resp. the 6th and the 7th right singular vector of the data-matrix. The signals 1 and 2 partly describe the MECG-subspace; the MECG also appears in signals 3 and 5. Channels 4 and 8 mainly show noise contributions.

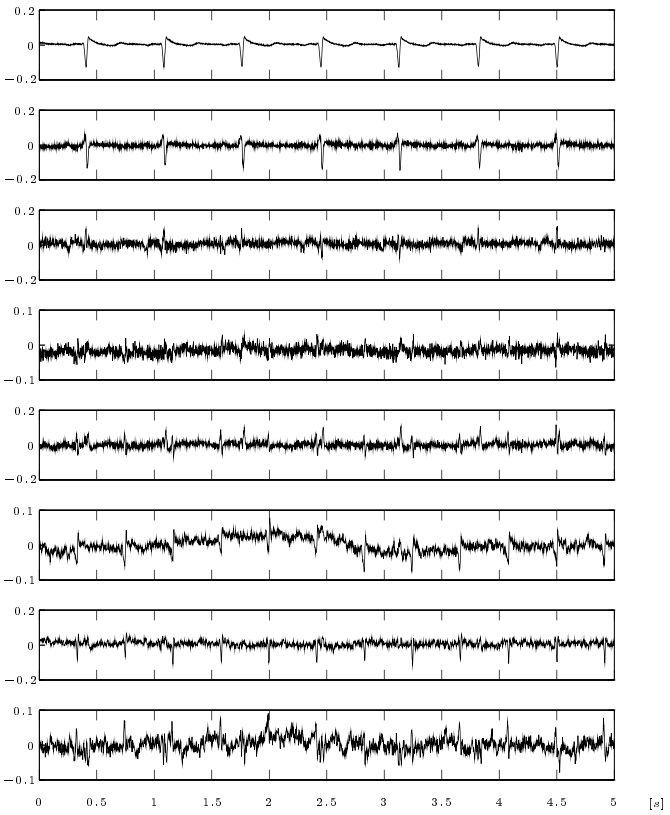



Figure 11.4: Source estimates obtained via PCA.

The result after BSSS is shown in Fig. 11.5 (we used the JADE-algorithm (Sect. 9.3)). Although the statistics of the non-stationary signals have been estimated by time-averaging, the result is an excellent source separation. Where the PCA-method obtains only two clear MECG-components (the 3rd signal is heavily perturbated by noise and the fifth principal component contains important FECG-contributions), BSSS accurately reconstructed the full three-dimensional MECG-subspace (signals 1 to 3 in Fig. 11.5). As far as the FECG is concerned, the quality of the 7th principal component and the 8th BSSS-signal are comparable, but in the 6th BSSS-signal the SNR is somewhat higher than in the 6th PCA-estimate. The off-set in the 6th PCA-signal is found back as an extra source signal (the 7th signal in Fig. 11.5; this sequence continues



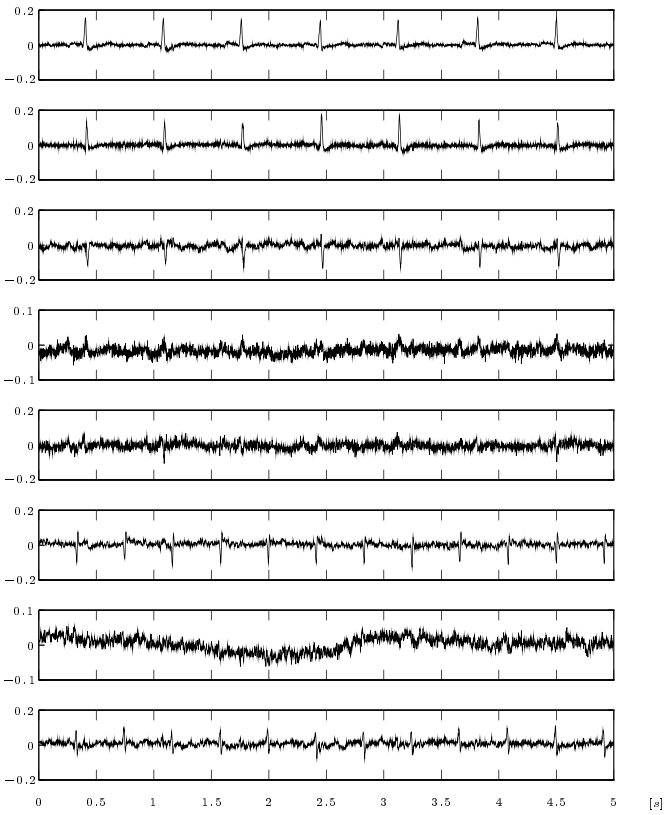



Figure 11.5: Source estimates obtained via BSSS.

as a low-periodic signal and deserves further medical interpretation). The 5th BSSS-signal mainly shows noise contributions.

## 11.4 Conclusion

In this chapter, we have illustrated some typical assets of the CANDECOMP-concept by means of a key problem in Linguistics, namely the analysis of multi-way datasets describing vocal tract shapes occurring during the pronunciation of different vowels by different speakers. Compared to the methods currently

used, the techniques discussed in Chapter 6 can be considered as interesting numerical tools to carry out the analysis.

We have also illustrated the power of the ICA-concept with two applications.

First, we have briefly sketched the background of the application of ICA for blind beamforming in wireless telecommunications. We remind that, even if a model of the array manifold is available, blind methods are still preferable in many cases, due to their inherent insensitivity to calibration errors and the more convenient way to deal with multipath propagation.

Secondly, we have proposed ICA as a new way to address the problem of extracting the FECG from multilead potential recordings on the mother's skin. With respect to the previously proposed SVD-based method, the higher-order ICA-step additionally requires the estimation and the (partial) diagonalization of the fourth-order cumulant tensor of the data. From a conceptual point of view, ICA is a very ambitious approach: it aims at the direct reconstruction of the different statistically independent bioelectric source signals, as well as the characteristics of their propagation to the electrodes, each revealing important medical information.

# Chapter 12

# Conclusion

## 12.1 Contributions

One of the main themes of this thesis is the generalization of the SVD of matrices to higher-order tensors.

As a starting-point could serve that the multilinear decomposition should address the $n$-mode vectors in a similar way as the SVD does for columns and rows. This implies the orthogonal transformation of the original tensor to all-orthogonality. The resulting HOSVD indeed exhibits very clear counterparts of SVD-properties related to the characteristics of column and row space. In particular, the numerical $n$-rank values can be read from the $n$-mode singular values. However, the SVD-link between a matrix and its best rank-$R$ approximation in least-squares sense does not persist. If the aim is to approximate a given $N$th order tensor by another one of which the $n$-mode vectors span at most an $R_n$-dimensional space ($1 \leqslant n \leqslant N$), then a tree of best rank-$(R_1, R_2, \ldots, R_N)$ approximations can be defined; the different parts of this tree are only in special cases related by truncation of the HOSVD. Other approaches are possible here, as rank and $n$-mode rank are essentially different concepts for higher-order tensors. For the best rank-$R$ approximation of a given tensor, non-orthogonal rank-1 terms may be required (CANDECOMP); with this respect, the orthogonality of SVD-components should be considered as some kind of flexibility, inherent to the matrix case .

Multilinear algebraic techniques find an important application area in HOS, of which the basic quantities are super-symmetric higher-order tensors. In this thesis we have mainly concentrated on the fundamental problem of ICA.

For sufficiently low noise levels, the source estimates can be decorrelated by

an SVD of the matrix dataset, and further made mutually independent by a HOEVD of the higher-order cumulant. A computationally harder but more accurate technique consists of replacing the higher-order step by the maximal diagonalization, in least squares sense, of the higher-order cumulant (MD); this approach can be interpreted in terms of maximum likelihood and minimal mutual information. Efficient schemes are obtained by considering the simultaneous diagonalization of the matrix slices (JADE) or third-order tensor slices (STOTD) of the cumulant tensor. Because of the super-symmetry, transformation of the cumulant by means of the orthogonal factor found in this way, still reduces the cross-cumulants to a large extent; actually, it has been proved that the JADE- and MD-approach are asymptotically equivalent, for sufficiently low noise levels and sufficiently large datasets [245]. As an alternative for two-stage ICA, the prewhitening step can be skipped by directly solving the ICA-problem via the CANDECOMP of the cumulant tensor.

Now let us, chapter by chapter, go through the main contributions of the thesis.

In the literature of Psychometrics, the HOSVD of real-valued tensors is known as the Tucker model [259, 260], originally developed for the analysis of three-way data tables. In Chapter 4, we have presented a square-root version of the algorithm, valid for real- and complex-valued tensors of arbitrary order, and shown that the result is a convincing generalization of the SVD-concept itself. The properties of the decomposition have thorougly been analyzed and interpreted in terms of their matrix counterparts. We explained that the HOSVD always exists for real or complex $N$th-order tensors; when applied to matrices, it reduces - up to some trivial indeterminacies - to the matrix SVD. From the fact that $n$-mode vectors and column (row) vectors play the same role in HOSVD resp. SVD, it followed that the matrices of singular vectors can be computed from the sets of $n$-mode vectors in the same way as in the second-order case. In other words, the HOSVD of an $N$th-order tensor boils down to $N$ matrix SVD's. Uniqueness, link with EVD, directions and values of extremal oriented energy, etc. have been investigated and explicit expressions for a first-order perturbation analysis have been presented. The link between tensor symmetry and HOSVD was also investigated: if a higher-order tensor is transformed into itself or into its complex conjugate by a permutation of the indices, then this symmetry is reflected by the HOSVD-components. It turned out that the most adequate way to analyze this relation is the decomposition of the permutation in a sequence of permutation cycles. When every pair of indices is connected by some cycle, a property which was called pair-wise symmetry, then the different matrices of higher-order singular vectors are either equal or complex conjugated. In analogy to the EVD of a real symmetric or complex Hermitean matrix the Higher-Order EVD (HOEVD) of a pair-wise symmetric higher-order tensor was defined.

The least-squares approximation of a three-way dataset by a dataset with reduced core tensor is known as three-mode factor analysis, in the literature of



Psychometrics [161, 162, 163, 254]. In Chapter 5, we have further contributed to a linear/multilinear algebraic framework for the least-squares approximation of tensors of arbitrary order. In the square-root algorithm that we presented, an iteration step was based on a multilinear mapping, followed by the estimation of a dominant subspace; the technique was interpreted as a higher-order generalization of the method of orthogonal iterations. Aiming at a higher efficiency, we have indicated how two modes can be updated at the same time. We have demonstrated that starting the iteration from a truncated HOSVD does not lead to the global optimum in all possible cases. We have paid special attention to the case of super-symmetric tensors, in view of applications in HOS. Explicit expressions for the gradient of the cost function over the Stiefel manifold have been provided. We have extensively discussed the best rank-1 approximation problem. For super-symmetric $(2 \times 2 \times \ldots \times 2)$-tensors, the determination of the best rank-1 approximation, and the calculation of the super-symmetric stationary points and of the other stationary points of the higher-order power algorithm, have been reformulated as a polynomial rooting problems. To improve the efficiency of higher-order power and orthogonal iterations, we proposed a preprocessing step, based on a higher-order generalization of the Hessenberg Decomposition.

In Chapter 6 we have investigated the computation of the CANDECOMP, under the identifiability assumptions put forward in Sect. 6.1. Currently, the calculation of the factors mostly takes the form of an ALS descent algorithm, possibly initiated with an estimate obtained by matrix EVD. We proposed to compute the CANDECOMP via simultaneous diagonalization, by equivalence or congruence, of a set of matrices. This is numerically more reliable than the calculation of a single EVD, especially for moderate signal-to-noise ratios; actually the problem can be reformulated as a simultaneous matrix EVD. Diagonalization by simultaneous congruence transformation can be realized by translating the problem into a simultaneous generalized Schur decomposition, which can subsequently be solved by means of an extended $QZ$-iteration scheme [266]. In Chapter 6, we have focused on one-sided Givens-iterations, two-sided Jacobi-iterations and gradient-based optimization. It was shown that the computation of an elementary Givens rotation corresponds to the calculation of the dominant left singular vector of a matrix with two rows. The determination of an elementary Jacobi-rotation lead to a set of biquadratic equations; this problem has been rewritten as the rooting of a polynomial of degree 8. As far as gradient-based optimization is concerned, we have derived an explicit expression for the gradient of the simultaneous Schur cost function, over the manifold of orthogonal matrices. Gradient-based descent also forms an alternative for the ALS iterations in an additional least-squares model matching phase. With this respect, we have derived an explicit expression for the gradient of the quadratic error function, over the manifold of matrices with unit-norm columns. The marginal improvement of the fit that can be achieved with the extra least-squares matching, becomes smaller for better-conditioned

problems, as indicated by simulations in which the performance of the different algorithms has been compared. We have also proved that regular matrices of which the columns may be arbitrarily scaled (such as the CANDECOMP factor matrices, the mixing matrix in ICA, ... ), can be uniquely parametrized in terms of a pair of orthogonal matrices; this representation is interesting from a numerical point of view.

In Chapter 7, we have given a new, algebraic, proof of the identifiability of the solution of ICA, in which we rely on the assumptions that the sources are uncorrelated and independent "at some specific order $N > 2$". We have also presented an original proof for a lower-bound on the ISR's of the source estimates, induced by imperfections in the PCA-stage. In addition, we have derived a lower-bound on the Frobenius-norm of the difference between the original mixing matrix and its estimate. The latter derivation was based on a lemma, stating that the sum of singular values of a matrix product cannot be larger than the sum of the products of the singular values of the distinct matrices. In Sect. 7.7 we have proposed a new ICA-algorithm, based on the HOEVD of a higher-order cumulant, and analyzed its properties. The computational load of the technique is very low; the price has to be paid in terms of numerical accuracy, compared to methods that are more computationally demanding, like the algorithms discussed in Chapter 8 and 9. A second drawback of the HOEVD-technique is that the required conditions for identifiability are somewhat stronger than the ones imposed by the ICA-problem itself. We have also derived a theoretical foundation for the fact that a mere PCA allows to obtain a fairly good source separation, provided that the sources are mutually uncorrelated and have a large difference in variance.

In Chapter 8 we have demonstrated that, for a higher-order MD ICA-stage involving fourth-order cumulants, the computation of an elementary Jacobi-rotation amounts to the best rank-1 approximation of a super-symmetric real-valued $(2 \times 2 \times 2 \times 2)$-tensor (real case) or a super-symmetric real-valued $(3 \times 3 \times 3 \times 3)$-tensor (complex case). This approach has three advantages: (1) it can take the form of a sequence of simple operations (sums and products), (2) instead of computing all the points where the derivative of the cost function is null, the method aims directly at the global optimum, and (3) during the computation of the rotation, the optimization of the cost function is easily monitored. A disadvantage is that the technique is iterative, whereas in the real-valued case, the solution can be obtained by rooting a polynomial of degree 4; as such, the technique is especially interesting in the complex case.

In Chapter 9, we have proved that, as far as the optimal diagonalization of a third-order cumulant is concerned, the computation of an elementary Jacobi-rotation can be formulated as the best rank-1 approximation of a real symmetric $(2 \times 2)$-matrix (real case) or $(3 \times 3)$-matrix (complex case). This is still true for the optimal diagonalization of a set of third-order cumulants; the matrix is obtained as a sum of contributions of the individual tensors in the set.



Comparison to formulas for simultaneous matrix diagonalization, leads to the result that simultaneous diagonalization of $N$ complex (real) super-symmetric $(2 \times 2 \times 2)$-tensors is mathematically equivalent to the simultaneous diagonalization of $N$ Hermitean (symmetric) $(2 \times 2)$-matrices, constructed from the third-order tensors. We have shown that ICA can be realized by simultaneous diagonalization of third-order tensors, spanning the range of the standardized fourth-order cumulant. Computational requirements and performance are similar to those of the JADE-algorithm.

In Chapter 10, we have presented a new ICA-technique that resorts only to the higher-order cumulants of the observations. We have indicated the uniqueness properties of the transfer matrix estimate in our approach. Higher-order-only BSS has the advantage, w.r.t. prewhitening-based ICA, 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 the estimation of higher-order statistics generally requires larger datasets than the estimation of second-order statistics. We also explained that, in a more sensors-than-sources scenario, the dimensionality of the problem can be reduced to the number of sources without resorting to second-order statistics, by exploiting the difference between the concepts of $n$-rank and rank of higher-order tensors.

In Chapter 11, we have illustrated some typical assets of the CANDECOMP-concept by means of a key problem in Linguistics, namely the analysis of multi-way datasets describing vocal tract shapes occurring during the pronunciation of different vowels by different speakers. We have also proposed ICA as a new way to address the problem of extracting the FECG from multilead potential recordings on the mother's skin. With respect to the previously proposed SVD-based method, the higher-order ICA-step additionally requires the estimation and the (partial) diagonalization of the fourth-order cumulant tensor of the data. From a conceptual point of view, ICA is a very ambitious approach: it aims at the direct reconstruction of the different statistically independent bioelectric source signals, as well as the characteristics of their propagation to the electrodes, each revealing important medical information.

## 12.2 Suggestions for further research

Numerical multilinear algebra is a very young discipline; several new results, extensions and improvements should be expected. In this section, we will touch on some open problems and indicate suggestions for further research.

The conditions ensuring uniqueness of the CANDECOMP may be relaxed, in comparison to the assumptions put forward in Sect. 6.1, to include more general patterns of linear independence of the components [165, 37, 67]. A

nice feature is that for an $(I_1 \times I_2 \times \ldots \times I_N)$-tensor, with $N > 2$, more than $\mathtt{max}\{I_1, I_2, \ldots, I_N\}$ rank-1 terms may be identified. (However, it cannot be excluded a priori that, from a numerical point of view, such problems are badly conditioned [137].) Important work has to be done with respect to the development of numerical algorithms that can successfully deal with these cases; only the super-symmetric $(2 \times 2 \times \ldots \times 2)$-case can be solved in its full generality at present [67]. In the context of Factor Analysis, these algorithms could extract more factors from multiway datasets than indicated by the $n$-mode ranks. In higher-order-only ICA, more steering vectors could be estimated than there are sensors available. As a matter of fact, the same issue deserves investigation for prewhitening-based ICA as well; this leads to the decomposition of the cumulant of the whitened I-dimensional observations in a sum of $J > I$ rank-1 components $Q_j \circ Q_j \circ \ldots \circ Q_j$, under the constraint that $\mathbf{Q} = [Q_1 Q_2 \ldots Q_J]$ has mutually orthonormal rows.

An important remark concerns the convergence behaviour of the algorithms that have been discussed. Currently, explicit formulas describing the convergence speed are only available for a very limited number of cases (for the simultaneous diagonalization of a set of mutually commuting normal matrices by unitary similarity transformations, under no-noise conditions, it has been proved that the convergence is quadratric in a neighbourhood of the global optimum [21]). Also, global convergence is often assumed in the literature, but formal proofs are lacking. To a large extent, this is due to the fact that tensor decompositions have a more complicated structure than their matrix equivalents. As an example, the convergence of the power algorithm in the matrix case can easily be analyzed in terms of the EVD-components; for higher-order tensors however, the residu after best rank-1 approximation is much harder to parametrize and can a priori still have full rank and full $n$-mode ranks. In addition, the derivation of algorithms is often established under low-noise assumptions, which can considerably simplify the exposition. Usually, the overall performance of algorithms is analyzed by means of "proofs by MATLAB", but more general theoretical derivations could increase the insight in tensorial mechanisms. To investigate global convergence for the computation of the best rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order cumulants, one could possibly exploit the inequality relations that exist among the entries [65]. We also mention the need for further analysis with respect to the nature of swamps in CANDECOMP-algorithms, possibly followed by the development of regularization tools to enhance the performance.

The techniques developed in this thesis are mainly based on multilinear extensions of power iteration and Givens/Jacobi iteration schemes. Additional results and fast algorithms may be expected from the generalization of more involved matrix techniques for EVD/SVD-computations, such as e.g. the $QR$-iteration [130]. A second important discipline from which numerical multilinear algebra may benefit, is the research with respect to the solution of sets of polynomial equations. The best rank-1 approximation problem, the computation of the CANDECOMP, etc., are easily formulated in this format, and can therefore theoretically be solved by e.g. homotopy methods [126, 276]. However, the computational cost of this approach can be prohibitive in practice, even for moderate problem sizes. Nevertheless, the intensive study of relatively simple types of sets of polynomial equations may form an important input for the formulation of e.g. new Givens/Jacobi-type methods.



The discussion on Blind Source Separation in this thesis was restricted to the identification of scalar mixtures of mutually independent signals. It has been indicated in Sect. 11.2 that the results are also relevant for the blind deconvolution of a set of linear time-invariant filters. In addition, tensor decompositions can possibly be considered as building blocks for the (blind) identification of Volterra-type transfers [241], like their matrix equivalents play an important role in the identification of the linear state space model [271, 272].

With respect to gradient-based optimization strategies for the computation of tensor decompositions, we remark that the derivation of discrete-time implementations of continuous-time algorithms is not a priori straightforward. A discussion and examples can be found in [77].

Finally, we indicate two real-life applications that deserve further analysis by means of multilinear techniques.

The first problem can be situated in the area of Chemometrics [225, 226]. In laboratoria for chemical analysis, one makes frequently use of bilinear equipment (tandem massa spectrometry, 2D-NMR, ... ): the output of the device is a matrix, formed by rank-1 contributions of the various chemical components. E.g. for emission-excitation fluorescence data, each row is an emission spectrum and each column an excitation scan or vice-versa. Theoretically these row- and column-vectors are linear combinations of the emission and excitation spectra of the components, which have to be determined. At a matrix-level however, the chemical analysis cannot be realized without a priori knowledge, as an infinite number of matrix decompositions is possible. On the other hand, it is often necessary to examine several samples; sometimes, multiway datasets are readily available (e.g. fluorescence data). With this respect, the CANDECOMP can be considered as an important technique to realize the chemical analysis: if its identifiability conditions are satisfied, then the CANDECOMP of the data tensor could allow to determine the components and their concentrations in a blind way. Non-negativity constraints (e.g. for spectra) could be imposed in a least-squares matching step (Sect. 6.6).

As a second application field, we mention the domain of operating response analysis in vibro-acoustics [211]. Here the dynamical behaviour of a mechanical structure is studied from observations under operating conditions; in many cases this provides enough insight to understand e.g. the nature of undesired

vibrations, without having to resort to an expensive full identication of the system in a laboratorium set-up. A key technique is the computation, as a function of frequency, of the principal components of the signals measured by a number of reference transducers under specific stationary conditions, and the assessment of their relative importance. Further insight can be obtained by the animation, as so-called Principal Operating Field Shapes, of the contribution of the different principal components to the signals recorded by a battery of "target" sensors, distributed over the structure. It is natural to refine this PCA-technique to ICA, with the aim to extract as much information from the data as possible. This could lead to the visualization of "Independent Operating Field Shapes", showing the contribution of the distinct mutually statistically independent contributions to the vibration problem. Of course, whether the dataset can succesfully be modelled as a linear combination of statistically independent source signals, depends on the specific context, but the measure to which the higher-order spectrum of the data can be diagonalized, offers an indication for the validity of this assumption.

# Appendix A

# Matrix Definitions

In this appendix we briefly describe some matrix-related definitions that are used in this thesis. For a detailed exposition on matrix decompositions, and computation issues, we refer to [146, 130]. First we elucidate some notations.

The sets of $(I \times J)$-matrices with real entries or complex entries are represented by $\mathbb{R}^{I\times J}$ and $\mathbb{C}^{I\times J}$ respectively. The *range* and *null space* of a matrix $\mathbf{A}$ are represented by $\mathtt{R}(\mathbf{A})$ and $\mathtt{N}(\mathbf{A})$. The *determinant* and the *trace* of a square matrix $\mathbf{A}$ are represented by $\mathtt{det}(\mathbf{A})$ and $\mathtt{trace}(\mathbf{A})$. $\mathtt{diag}(\mathbf{A})$, $\mathtt{upp}(\mathbf{A})$ and $\mathtt{skew}(\mathbf{A})$ represent the diagonal, upper triangular, and skew-symmetric part of $\mathbf{A}$:

$$\begin{aligned}
\mathbf{B} = \mathtt{diag}(\mathbf{A}) \quad &\Leftrightarrow \quad \begin{cases} b_{ii} = & a_{ii}, \\ b_{ij} = & 0 \quad \text{if} \quad i \neq j, \end{cases} \\
\mathbf{B} = \mathtt{upp}(\mathbf{A}) \quad &\Leftrightarrow \quad \begin{cases} b_{ij} = & a_{ij} \quad \text{if} \quad i \leqslant j, \\ b_{ij} = & 0 \quad \text{if} \quad i > j, \end{cases} \\
\mathbf{B} = \mathtt{skew}(\mathbf{A}) \quad &\Leftrightarrow \quad \mathbf{B} = \frac{\mathbf{A} - \mathbf{A}^H}{2}.
\end{aligned}$$

On the other hand, $\mathtt{diag}(a_1, a_2, \ldots, a_N)$ defines a diagonal matrix with diagonal elements equal to $a_1, a_2, \ldots, a_N$; the dimensions of this matrix are $N$ by $N$, unless specified otherwise. Some higher-order tensor concepts can be formulated in a matrix language by resorting to the Kronecker or Kathri-Rao product [221, 17]. For matrices $\mathbf{A} \in \mathbb{C}^{I_1\times I_2}$ and $\mathbf{B} \in \mathbb{C}^{J_1\times J_2}$, the *Kronecker product* is defined by

$$(\mathbf{A} \otimes \mathbf{B})_{(i_1-1)J_1+1:i_1J_1,(i_2-1)J_2+1:i_2J_2} = a_{i_1i_2}\mathbf{B} \qquad (1 \leqslant i_1 \leqslant I_1; 1 \leqslant i_2 \leqslant I_2).$$

For matrices $\mathbf{F} = [F_1 F_2 \ldots F_K] \in \mathbb{C}^{I\times K}$ and $\mathbf{G} = [G_1 G_2 \ldots G_K] \in \mathbb{C}^{J\times K}$, the *Kathri-Rao product* is defined by

$$\mathbf{F} \odot \mathbf{G} = [F_1 \otimes G_1 \; F_2 \otimes G_2 \; \ldots \; F_K \otimes G_K]. \tag{A.1}$$

The sets of orthogonal matrices, unitary matrices, and rectangular matrices with mutually orthonormal columns are called the *orthogonal manifold*, *unitary manifold*, *Stiefel manifold*, and represented by

$$\begin{aligned} \mathrm{O}(I) &= \{\mathbf{A} \in \mathbb{R}^{I\times I} | \mathbf{A}^T\mathbf{A} = \mathbf{I}\}, \\ \mathrm{U}(I) &= \{\mathbf{A} \in \mathbb{C}^{I\times I} | \mathbf{A}^H\mathbf{A} = \mathbf{I}\}, \\ \mathrm{St}(J,I) &= \{\mathbf{A} \in \mathbb{R}^{I\times J} | \mathbf{A}^T\mathbf{A} = \mathbf{I}\} \end{aligned}$$

respectively. The Frobenius-norm of a matrix $\mathbf{A}$ is denoted as $\|\mathbf{A}\|$. To indicate the Frobenius-norms of the upper-triangular and strictly upper-triangular part of $\mathbf{A}$, $\|\mathbf{A}\|_{UF} = \|\mathtt{upp}(\mathbf{A})\|$ and $\|\mathbf{A}\|_{UFs} = \|\mathtt{upp}(\mathbf{A}) - \mathtt{diag}(\mathbf{A})\|$ are used. The Frobenius-norms of the lower-triangular and strictly lower-triangular part of $\mathbf{A}$ are represented in a similar way.

Now let us pay some attention to a number of basic matrix decompositions.

**Theorem A.1 ($QR$-Decomposition)** *Any matrix $\mathbf{A} \in \mathbb{R}^{I\times J}$ ($\mathbb{C}^{I\times J}$) can be decomposed as*

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{R}, \tag{A.2}$$

*in which $\mathbf{Q}$ is a square orthogonal (unitary) matrix and $\mathbf{R}$ is upper triangular. This decomposition is called the QR-Decomposition (QRD).*

Similar decompositions can be obtained by reversing the order of $\mathbf{Q}$ and $\mathbf{R}$, or by putting a condition of lower-triangularity (involving a matrix $\mathbf{L}$) instead of upper-triangularity. The resulting decompositions are referred to as RQ-, QL-, and LQ-factorizations.

**Definition A.2 (Unsymmetric Eigenvalue Decomposition)** *An Eigenvalue Decomposition (EVD) of a matrix $\mathbf{A} \in \mathbb{R}^{I\times I}$ ($\mathbb{C}^{I\times I}$) has the form*

$$\mathbf{A} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{E}^{-1}, \tag{A.3}$$

*in which:*

- $\mathbf{E} = [E_1 E_2 \dots E_I]$ *is a regular matrix in* $\mathbb{C}^{I\times I}$,
- $\mathbf{D} = \mathtt{diag}(\lambda_1, \lambda_2, \dots, \lambda_I)$ *is a diagonal matrix in* $\mathbb{C}^{I\times I}$.

*$\lambda_i$ and $E_i$ are the ith eigenvalue resp. eigenvector of $\mathbf{A}$.*

In a dense subset of $\mathbb{R}^{I\times I}$ and $\mathbb{C}^{I\times I}$ every matrix has an EVD. By the EVD of a matrix $\mathbf{A}$ the associated linear mapping is decoupled in the following way:



each eigenvector is mapped in its own direction, where the amplification factor is given by the corresponding eigenvalue: $\mathbf{A}E_i = \lambda_i E_i$. Vector spaces spanned by a set of eigenvectors are called *invariant subspaces*, since they are mapped onto themselves through multiplication with $\mathbf{A}$.

The EVD of a real symmetric matrix and a complex Hermitean matrix always exists, and shows a specific structure:

**Theorem A.3 (Symmetric Eigenvalue Decomposition)** *Any real symmetric (complex Hermitean) matrix $\mathbf{A}$ can be decomposed as*

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{D} \cdot \mathbf{Q}^T \tag{A.4}$$

*(or $\mathbf{A} = \mathbf{Q} \cdot \mathbf{D} \cdot \mathbf{Q}^H$), in which $\mathbf{Q}$ is a square orthogonal (unitary) matrix and $\mathbf{D}$ is real diagonal.*

Similarly, the Singular Value Decomposition (SVD) shows how an arbitrary matrix can be diagonalized by orthogonal (unitary) transformations:

**Theorem A.4 (Singular Value Decomposition)** *Any matrix $\mathbf{A} \in \mathbb{R}^{I\times J}$ ($\mathbb{C}^{I\times J}$) can be decomposed as*

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T \tag{A.5}$$

*(or $\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^H$), in which*

- $\mathbf{U} = [U_1 U_2 \dots U_I] \in \mathbb{R}^{I\times I}$ ($\mathbb{C}^{J\times J}$) *and* $\mathbf{V} = [V_1 V_2 \dots V_J] \in \mathbb{R}^{I\times I}$ ($\mathbb{C}^{J\times J}$) *are square orthogonal (unitary) matrices,*
- $\mathbf{S} = \mathtt{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\mathtt{min}(I,J)})$ *is pseudo-diagonal, with positive real values on the diagonal, in decreasing order:*

$$\sigma_1 \geqslant \sigma_2 \geqslant \dots \geqslant \sigma_{\mathtt{min}(I,J)} \geqslant 0. \tag{A.6}$$

*The $\sigma_r$ are the singular values of $\mathbf{A}$ and the vectors $U_r$ and $V_r$ are respectively the rth left and the rth right singular vector.*

The SVD plays a crucial role in this thesis. We briefly enumerate the matrix equivalents of the tensor properties derived in Sect. 4.4:

- The singular values of a real matrix are unique. For a singular value with multiplicity 1, only the sign of the corresponding singular vectors is undetermined. When the sign of the left singular vector is changed,

the sign of the right singular vector should be altered as well. Singular vectors corresponding to a multiple singular value are determined up to an orthogonal transformation. If the left singular vectors are multiplied from the right with an orthogonal matrix $\mathbf{Q}$, the right singular vectors should be multiplied from the right with $\mathbf{Q}^T$. For complex-valued matrices uniqueness is up to multiplication with a unit-norm scalar resp. a unitary matrix.

- The rank $R$ of a matrix equals its number of non-zero singular values. This also holds in a numerical sense: the best rank-$k$ approximation of a given matrix, in least-squares sense, is obtained by discarding the smallest singular values.
- The column space $\mathtt{R}(\mathbf{A}) = \mathtt{span}\{U_1, U_2, \ldots, U_R\}$; the orthogonal complement of $\mathtt{R}(A)$, the left null space $\mathtt{N}(\mathbf{A}^H) = \mathtt{span}\{U_{R+1}, U_{R+2}, \ldots, U_I\}$; the complex conjugated row space $\mathtt{R}(\mathbf{A}^H) = \mathtt{span}\{V_1, V_2, \ldots, V_R\}$; and the orthogonal complement of $\mathtt{R}(\mathbf{A}^H)$, the null space $\mathtt{N}(\mathbf{A}) = \mathtt{span}\{V_{R+1}, V_{R+2}, \ldots, V_J\}$.
- The Frobenius-norm of $\mathbf{A}$ equals the sum of its squared singular values.
- Considering $\mathbf{A}$ as a collection of $J$ $I$-dimensional vectors, the oriented energy of $\mathbf{A}$ in the direction of a unit-norm vector $X$, denoted as $\mathtt{OE}_{\mathtt{n}}(X, \mathbf{A})$, is defined by $\mathtt{OE}(X, \mathbf{A}) = \|X^H \mathbf{A}_{(n)}\|^2$. The left singular vectors give the directions of extremal oriented energy, with extremal energy value equal to the corresponding squared singular value.
- An orthonormal set of eigenvectors of $\mathbf{A}\mathbf{A}^H$ corresponds to the left singular vectors of $\mathbf{A}$. The eigenvalues are the squared singular values of $\mathbf{A}$. Via $\mathbf{A}^H\mathbf{A}$ the right singular vectors are obtained.

Another generalization of the symmetric EVD, to arbitrary square matrices, is the Schur Decomposition:

**Theorem A.5 (Schur Decomposition)** *Any matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$ can be decomposed as*

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{R} \cdot \mathbf{Q}^H, \tag{A.7}$$

*in which*

- $\mathbf{Q} = [Q_1 Q_2 \ldots Q_I] \in \mathbb{C}^{I \times I}$ *is unitary,*
- $\mathbf{R} \in \mathbb{C}^{I \times I}$ *is upper triangular.*

*The columns $Q_i$ are Schur vectors and the diagonal entries of $\mathbf{R}$ are the eigenvalues of $\mathbf{A}$.*



The first $i$ Schur vectors span an invariant subspace of $\mathbf{A}$ ($1 \leqslant i \leqslant I$). $\mathbf{Q}$ can be chosen so that the eigenvalues appear in any order along the diagonal.

For the computation of tensor decompositions, the technique of Givens or Jacobi-iteration [130] plays a prominent role in this thesis. A Givens rotation matrix that affects rows $i$ and $j$ (assume that $j > i$), is symbolized by $\mathbf{G}_{ij}$. $\mathbf{G}_{ij} \in \mathbb{C}^{I \times I}$ equals the identity matrix, except for the entries

$$\begin{aligned} (\mathbf{G}_{ij})_{ii} &= (\mathbf{G}_{ij})_{jj} = \cos\alpha, \\ (\mathbf{G}_{ij})_{ji} &= -(\mathbf{G}_{ij})^*_{ij} = \sin\alpha e^{j\phi}, \end{aligned}$$

in which $\alpha$ and $\phi$ are the rotation angles. For $\mathbf{G}_{ij} \in \mathbb{R}^{I \times I}$, $\phi = 0$. An alternative parametrization, assuming that $\alpha \in (-\pi/2, \pi/2)$, is

$$\begin{aligned} (\mathbf{G}_{ij})_{ii} &= (\mathbf{G}_{ij})_{jj} = \frac{1}{\sqrt{1+|t|^2}}, \\ (\mathbf{G}_{ij})_{ji} &= -(\mathbf{G}_{ij})^*_{ij} = \frac{t}{\sqrt{1+|t|^2}}, \end{aligned}$$

where $t$ corresponds to the tangent of the complex rotation: $t = \sin\alpha e^{j\phi}/\cos\alpha$. $(2 \times 2)$ Givens rotations will briefly be denoted as $\mathbf{G}$.

A Jacobi rotation matrix that affects rows $i$ and $j$ is denoted as $\mathbf{J}_{ij}$ or $\mathbf{J}$. There is no difference with Givens rotations, but the term is used in this thesis to indicate that application of a Jacobi rotation to a matrix $\mathbf{A}$ involves both a left and a right multiplication $\mathbf{J}_{ij} \cdot \mathbf{A} \cdot \mathbf{J}_{ij}^H$ (or all possible $n$-mode multiplications in the higher-order case), whereas Givens rotations are only applied to columns or rows separately (or to a single mode of $n$-mode vectors in the higher-order case). With this respect, we also use the terms "two-sided" and "one-sided" rotation.

# Appendix B

# The Multivariate Gaussian Distribution

One of the major assets of higher-order cumulants is that they are conceptually insensitive to additive Gaussian noise. Since we are primarily dealing with higher-order cumulant tensors, it is important to have a clear understanding of what is meant with a *multivariate Gaussian distribution*.

**Definition B.1** *A set of real random variables $X_1$, $X_2$, ... , $X_I$ is said to be jointly Gaussian if their joint probability density function is given by*

$$p_X(x_1, x_2, \ldots, x_I) = \frac{(\mathtt{det}(\mathbf{C}))^{-1/2}}{(2\pi)^{I/2}} \exp(-\frac{1}{2}(X - M)^T \mathbf{C}^{-1}(X - M)), \quad \text{(B.1)}$$

*with $X = (x_1, x_2, \ldots, x_N)^T$, $M \in \mathbb{R}^I$, and $\mathbf{C} \in \mathbb{R}^{I \times I}$ positive definite and symmetric. The function $p_X$ is called a multivariate Gaussian distribution [198].*

**Theorem B.2** *Let $X = (X_1, X_2, \ldots, X_N)^T$ be a real random vector with an $I$-dimensional Gaussian distribution, given by Eq. (B.1). Then the mean of $X$ equals $M$, and its covariance is given by $\mathbf{C}$:*

$$\mathtt{E}\{X\} = M, \qquad \mathtt{E}\{(X - M) \cdot (X - M)^T\} = \mathbf{C}. \quad \text{(B.2)}$$

**Proof:** See [9] (Section 8.7). □

A complex random variable $Z = X + jY$, in which $X$ and $Y$ are real random variables, will be called Gaussian if $X$ and $Y$ are jointly Gaussian. For brevity

we often call a complex random variable $Z$ "Gaussian, with variance $\sigma^2$". By convention this implies that $X$ and $Y$ are mutually independent, that $m_X = m_Y = 0$ and that $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Hence, the probability density function of $Z$, in terms of the unknowns $x$ and $y$, is given by

$$p_Z(x,y) = \frac{1}{2\pi\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{B.3}$$

**Remark B.3** Consider a complex random variable $Z = Re^{j\Theta}$, in which the radial part $R$ has a zero-mean Gaussian distribution with variance $\sigma_R^2$ and the angular part $\Theta$ is independently distributed, uniform over $(-\pi, +\pi]$. The probability density function of $Z$, in terms of the unknowns $r$ and $\theta$, is given by

$$\tilde{p}_Z(r,\theta) = \frac{1}{\pi\sqrt{2\pi\sigma_R}} e^{-\frac{r^2}{2\sigma_R^2}}. \tag{B.4}$$

Note that this distribution is not Gaussian, and that its even higher-order cumulants do not vanish. Let us e.g. compare the expressions for the marginal probability that $Z$ has a radius between $r_0$ and $r_0 + dr$, for the distributions of Eq. (B.3) and Eq. (B.4). In the former case we end up with a marginal probability given by

$$\frac{r}{\sigma} e^{-\frac{r^2}{2\sigma^2}} dr,$$

with $r^2 = x^2 + y^2$; the latter case yields

$$\frac{1}{\sqrt{2\pi\sigma_R}} e^{-\frac{r^2}{2\sigma_R^2}} dr.$$

The fact that a Gaussian distribution is more prohibitive for a small radius, can intuitively be explained by observing that a small radius corresponds to small realizations of *both* $X$ and $Y$. Hence the marginal probability equals 0, for $r_0 = 0$. ◇

# Appendix C

# Derivations

*Theor. 5.3.9:*

**Proof:** The gradient of $g(\mathbf{U})$ can be determined by assuming that $\mathbf{U}$ has a velocity $\dot{\mathbf{U}}$ on the Stiefel manifold, and expliciting the evolution of $g$:

$$\dot{g} = \langle \nabla_{\mathbf{U}} g, \dot{\mathbf{U}} \rangle \tag{C.1}$$

(see e.g. [77], p. 48; the formula corresponds to a chain rule for the derivation).

In the theorem we have:

$$g = \langle \mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \ldots \times_N \mathbf{U}^T, \mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \ldots \times_N \mathbf{U}^T \rangle.$$

Assuming that $\mathbf{U}$ is time-dependent, the derivative with respect to the time coordinate is given by

$$\dot{g} = 2N \langle \mathcal{A} \times_1 \dot{\mathbf{U}}^T \times_2 \mathbf{U}^T \ldots \times_N \mathbf{U}^T, \mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \ldots \times_N \mathbf{U}^T \rangle,$$

in which we took into account that $\langle \mathcal{A} \times_1 \dot{\mathbf{U}}^T \ldots \times_N \mathbf{U}^T, \mathcal{A} \times_1 \mathbf{U}^T \ldots \times_N \mathbf{U}^T \rangle = \langle \mathcal{A} \times_1 \mathbf{U}^T \ldots \times_N \mathbf{U}^T, \mathcal{A} \times_1 \dot{\mathbf{U}}^T \ldots \times_N \mathbf{U}^T \rangle$ (symmetry of the inner product) and $\langle \mathcal{A} \times_1 \dot{\mathbf{U}}^T \ldots \times_N \mathbf{U}^T, \mathcal{A} \times_1 \mathbf{U}^T \ldots \times_N \mathbf{U}^T \rangle = \langle \mathcal{A} \times_1 \mathbf{U}^T \ldots \times_n \dot{\mathbf{U}}^T \ldots \times_N \mathbf{U}^T, \mathcal{A} \times_1 \dot{\mathbf{U}}^T \ldots \times_N \mathbf{U}^T \rangle$ (super-symmetry of $\mathcal{A}$). Using the definition of Eq. (5.41), we can write:

$$\dot{g} = 2N \langle \dot{\mathbf{U}}^T \cdot \tilde{\mathbf{U}}, \mathbf{U}^T \cdot \tilde{\mathbf{U}} \rangle.$$

With the properties of a scalar product, we obtain:

$$\dot{g} = 2N \langle \dot{\mathbf{U}}^T, \mathbf{U}^T \cdot \tilde{\mathbf{U}} \cdot \tilde{\mathbf{U}}^T \rangle = 2N \langle \dot{\mathbf{U}}, \tilde{\mathbf{U}} \cdot \tilde{\mathbf{U}}^T \cdot \mathbf{U} \rangle.$$

The right term is proportional to the gradient of $g$ over $\mathbb{R}^{I\times R}$. To ensure that $\mathbf{U}$ stays on the Stiefel manifold, we claim additionally that

$$\dot{\mathbf{U}} = \mathbf{U}\cdot\Omega + \mathbf{U}^{\perp}\cdot\Psi,$$

in which $\Omega\in\mathbb{R}^{R\times R}$ is skew-symmetric, $\Psi\in\mathbb{R}^{(I-R)\times R}$ and $\mathbf{U}^{\perp}\in\mathbb{R}^{I\times(I-R)}$ is the orthogonal complement of $\mathbf{U}$ (i.e. $\mathbf{U}^T\cdot\mathbf{U}^{\perp}=\mathbf{0}$ and ${\mathbf{U}^{\perp}}^T\cdot\mathbf{U}^{\perp}=\mathbf{I}$) ([77], p. 22). Now the inner product can be decomposed as:

$$\dot{g} = 2N\,\langle\Omega,\mathbf{U}^T\cdot\tilde{\mathbf{U}}\cdot\tilde{\mathbf{U}}^T\cdot\mathbf{U}\rangle + 2N\,\langle\Psi,{\mathbf{U}^{\perp}}^T\cdot\tilde{\mathbf{U}}\cdot\tilde{\mathbf{U}}^T\cdot\mathbf{U}\rangle,$$

in which the first term vanishes, since its right component is symmetric. We now rewrite the expression for $\dot{g}$ in the form of Eq. (C.1):

$$\begin{aligned}\dot{g} &= 2N\,\langle\mathbf{U}\cdot\Omega,\mathbf{0}\rangle + 2N\,\langle\mathbf{U}^{\perp}\cdot\Psi,\mathbf{U}^{\perp}\cdot{\mathbf{U}^{\perp}}^T\cdot\tilde{\mathbf{U}}\cdot\tilde{\mathbf{U}}^T\cdot\mathbf{U}\rangle\\ &= 2N\,\langle\mathbf{U}\cdot\Omega+\mathbf{U}^{\perp}\cdot\Psi,\mathbf{U}^{\perp}\cdot{\mathbf{U}^{\perp}}^T\cdot\tilde{\mathbf{U}}\cdot\tilde{\mathbf{U}}^T\cdot\mathbf{U}\rangle\\ &= \langle 2N\,(\mathbf{I}-\mathbf{U}\mathbf{U}^T)\cdot\tilde{\mathbf{U}}\cdot\tilde{\mathbf{U}}^T\cdot\mathbf{U},\mathbf{U}\cdot\Omega+\mathbf{U}^{\perp}\cdot\Psi\rangle,\end{aligned}$$

in which the last equation was obtained by substitution of $\mathbf{U}^{\perp}\cdot{\mathbf{U}^{\perp}}^T=\mathbf{I}-\mathbf{U}\mathbf{U}^T$. □

*Theor. 6.4.5:*

**Proof:** The gradient of $g$ to $\mathbf{Q}$ can be determined by assuming that $\mathbf{Q}$ has a velocity $\dot{\mathbf{Q}}$ on the manifold of orthogonal matrices, and expliciting the evolution of $g$:

$$\dot{g} = \langle\nabla_{\mathbf{Q}}\,g,\dot{\mathbf{Q}}\rangle \tag{C.2}$$

(see e.g. [77], p. 48; the formula corresponds to a chain rule for the derivation).

First we express the function $g$ as

$$g(\mathbf{Q},\mathbf{Z}) = \sum_{k=1}^{K}\langle\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z},\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\rangle.$$

Assuming that $\mathbf{Q}$ is time-dependent, the derivative with respect to the time coordinate is given by (taking into account that $\mathtt{upp}(\cdot)$ is a linear operation)

$$\begin{aligned}\dot{g} &= \sum_{k=1}^{K}[\langle\dot{\mathbf{Q}}\cdot\mathbf{V}_k\cdot\mathbf{Z},\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\rangle + \langle\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z},\mathtt{upp}(\dot{\mathbf{Q}}\cdot\mathbf{V}_k\cdot\mathbf{Z})\rangle]\\ &= 2\sum_{k=1}^{K}\langle\dot{\mathbf{Q}}\cdot\mathbf{V}_k\cdot\mathbf{Z},\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\rangle.\end{aligned}$$



With a property of the scalar product, we obtain:

$$\dot{g} = 2\sum_{k=1}^{K}\langle\dot{\mathbf{Q}},\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\cdot\mathbf{Z}^T\cdot\mathbf{V}_k^T\rangle.$$

The right term is proportional to the gradient of $g$ over $\mathbb{R}^{R\times R}$. To ensure that $\mathbf{Q}$ stays on the manifold of orthogonal matrices, we claim additionally that

$$\dot{\mathbf{Q}} = \Omega\cdot\mathbf{Q},$$

in which $\Omega\in\mathbb{R}^{R\times R}$ is skew-symmetric ([77], p. 17). Now the inner product can be written in the form of Eq. (C.2):

$$\begin{aligned}\dot{g} &= 2\sum_{k=1}^{K}\langle\Omega,\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\cdot\mathbf{Z}^T\cdot\mathbf{V}_k^T\cdot\mathbf{Q}^T\rangle\\ &= \langle\Omega,2\sum_{k=1}^{K}\mathtt{skew}\{\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\cdot\mathbf{Z}^T\cdot\mathbf{V}_k^T\cdot\mathbf{Q}^T\}\rangle\\ &= \langle\Omega\cdot\mathbf{Q},2\sum_{k=1}^{K}\mathtt{skew}\{\mathtt{upp}(\mathbf{Q}\cdot\mathbf{V}_k\cdot\mathbf{Z})\cdot\mathbf{Z}^T\cdot\mathbf{V}_k^T\cdot\mathbf{Q}^T\}\cdot\mathbf{Q}\rangle,\end{aligned}$$

which proves Eq. (6.46). The gradient with respect to $\mathbf{Z}$ can be found in an analogous way. □

*Theor. 6.6.1:*

**Proof:** For clarity, we omit the caps on the symbols $\{\hat{U}_r^{(n)}\}_{1\leqslant r\leqslant R;1\leqslant n\leqslant N}$ and $\{\hat{\lambda}_r\}_{1\leqslant r\leqslant R}$. We will first determine the gradient of $f$ to $U_r^{(n)}$ without taking into account the normalization constraint. This gradient will be represented by $\tilde{\nabla}_{U_r^{(n)}}f$. Afterwards, the gradient $\nabla_{U_r^{(n)}}f$ over the unit sphere will be obtained by projecting $\tilde{\nabla}_{U_r^{(n)}}f$ on the unit sphere.

First, the unconstrained gradient can be determined by assuming that $U_r^{(n)}$ has a velocity $\dot{U}_r^{(n)}$ as a function of a time coordinate $t$, and expliciting the evolution of $f$:

$$\dot{f} = \langle\tilde{\nabla}_{U_r^{(n)}}f,\dot{U}_r^{(n)}\rangle \tag{C.3}$$

(see e.g. [77], p. 48; the formula corresponds to a chain rule for the derivation).

In the theorem we have:

$$f = \langle\mathcal{A}-\hat{\mathcal{A}},\mathcal{A}-\hat{\mathcal{A}}\rangle.$$

Assuming that $U_r^{(n)}$ is time-dependent, the derivative with respect to the time coordinate is given by

$$\begin{aligned}\dot{f} &= \langle\frac{d}{dt}(\mathcal{A}-\hat{\mathcal{A}}),\mathcal{A}-\hat{\mathcal{A}}\rangle+\langle\mathcal{A}-\hat{\mathcal{A}},\frac{d}{dt}(\mathcal{A}-\hat{\mathcal{A}})\rangle = 2\langle\frac{d}{dt}(\mathcal{A}-\hat{\mathcal{A}}),\mathcal{A}-\hat{\mathcal{A}}\rangle \\ &= 2\langle-\lambda_r U_r^{(1)}\circ\ldots\circ U_r^{(n-1)}\circ\dot{U}_r^{(n)}\circ U_r^{(n+1)}\circ\ldots\circ U_r^{(N)},\mathcal{A}-\hat{\mathcal{A}}\rangle.\end{aligned}$$

With a property of the scalar product, we can write this expression in the form of Eq. (C.3):

$$\begin{aligned}\dot{f} &= \langle\dot{U}_r^{(n)},-2\lambda_r(\mathcal{A}-\hat{\mathcal{A}})\times_1 {U_r^{(1)}}^T\ldots\times_{n-1}{U_r^{(n-1)}}^T \\ &\qquad \times_{n+1}{U_r^{(n+1)}}^T\ldots\times_N{U_r^{(N)}}^T\rangle.\end{aligned}$$

Projection of the unconstrained gradient leads to the following expression for $\nabla_{U_r^{(n)}} f$:

$$\nabla_{U_r^{(n)}} f = -2\lambda_r(\mathbf{I}-U_r^{(n)}{U_r^{(n)}}^T).$$

□

*Lemma 7.6.1:*

**Proof:** In terms of the elements of $\mathbf{A}$, Eq. (7.24) takes the form:

$$(\sum_{\substack{i,j\\i\neq j}}|a_{ij}|^2)(\sum_{k,l}|a_{kl}|^2) \geqslant \sum_{\substack{i,j\\i>j}}|\sum_k a_{ki}a_{kj}|^2, \tag{C.4}$$

$$(\sum_{\substack{i,j\\i\neq j}}|a_{ij}|^2)(\sum_{k,l}|a_{kl}|^2) \geqslant \sum_{\substack{i,j\\i>j}}(\sum_k|a_{ki}a_{kj}|^2+\sum_{\substack{k,l\\k\neq l}}a_{ki}a_{kj}a_{li}^*a_{lj}^*). \tag{C.5}$$

We will prove that the square-terms on the right-hand side of Eq. (C.5) also appear on the left-hand side, and that the product-terms on the right-hand side can be combined with terms on the left-hand side to form squares of differences. First we rewrite the square-terms:

$$\sum_{\substack{i,j,k\\i>j}}|a_{ki}a_{kj}|^2=\sum_{\substack{i,j\\i\neq j}}|a_{ii}a_{ij}|^2+\sum_{\substack{i,j,k\\i>j\\i\neq k\neq j}}|a_{ki}a_{kj}|^2. \tag{C.6}$$

The second part of Eq. (C.5) can be written as:

$$\sum_{\substack{i,j,k,l\\i>j\\k\neq l}}a_{ki}a_{kj}a_{li}^*a_{lj}^* =$$



$$2\mathtt{Re}(\sum_{\substack{i,j\\i>j}}a_{ii}a_{ij}a_{ji}^*a_{jj}^*+\sum_{\substack{i,j,k\\j\neq i\neq k\neq j}}a_{ii}a_{ij}a_{ki}^*a_{kj}^*+\sum_{\substack{i,j,k,l\\i>j,k>l\\i\neq k\neq j,i\neq l\neq j}}a_{ki}a_{kj}a_{li}^*a_{lj}^*). \tag{C.7}$$

On the other hand, the left-hand side of Eq. (C.5) can be split up in 4 parts:

$$(\sum_{\substack{i,j\\i\neq j}}|a_{ij}|^2)(\sum_{k,l}|a_{kl}|^2) =$$

$$\sum_{\substack{i,j\\i\neq j}}|a_{ij}|^2|a_{ii}|^2+\sum_{\substack{i,j,l\\l\neq i\neq j}}|a_{ij}|^2|a_{il}|^2+\sum_{\substack{i,j,l\\k\neq i\neq j}}|a_{ij}|^2|a_{ki}|^2+\sum_{\substack{i,j,k,l\\i\notin\{j,k,l\}}}|a_{ij}|^2|a_{kl}|^2. \tag{C.8}$$

The second, third and fourth term can further be decomposed as:

$$\sum_{\substack{i,j,l\\l\neq i\neq j}}|a_{ij}|^2|a_{il}|^2 = \sum_{\substack{i,j\\i\neq j}}|a_{ij}|^4+2\sum_{\substack{i,j,k\\i>j\\i\neq k\neq j}}|a_{ki}|^2|a_{kj}|^2, \tag{C.9}$$

$$\sum_{\substack{i,j,k\\k\neq i\neq j}}|a_{ij}|^2|a_{ki}|^2 = \sum_{\substack{i,j,k\\j\neq i\neq k\neq j}}|a_{ij}|^2|a_{ki}|^2+\sum_{\substack{i,j\\i\neq j}}|a_{ij}|^2|a_{ji}|^2, \tag{C.10}$$

$$\begin{aligned}\sum_{\substack{i,j,k,l\\i\notin\{j,k,l\}}}|a_{ij}|^2|a_{kl}|^2 &= \sum_{\substack{i,j\\i\neq j}}|a_{ji}|^2|a_{ii}|^2+\sum_{\substack{i,j,k\\j\neq i\neq k\neq j}}|a_{ji}|^2|a_{kk}|^2 \\ &\quad+\sum_{\substack{i,j,k,l\\k\neq l\\j\notin\{i,k,l\}}}|a_{ji}|^2|a_{kl}|^2 \qquad\text{(C.11)}\\ &= \sum_{\substack{i,j\\i\neq j}}|a_{ji}|^2|a_{ii}|^2+\sum_{\substack{i,j,k\\k\neq j\neq i\neq k}}|a_{ii}|^2|a_{kj}|^2 \\ &\quad+\sum_{\substack{i,j,k,l\\j\neq l\\k\notin\{i,j,l\}}}|a_{ki}|^2|a_{lj}|^2. \qquad\text{(C.12)}\end{aligned}$$

Finally, the last sum of Eq. (C.12) can be decomposed as:

$$\sum_{\substack{i,j,k,l\\j\neq l\\k\notin\{i,j,l\}}}|a_{ki}|^2|a_{lj}|^2 =$$

$$\sum_{\substack{i,j,k,l\\i>j,i\neq l\neq j\\k\notin\{i,j,l\}}}|a_{ki}|^2|a_{lj}|^2+\sum_{\substack{i,j,k\\i>j\\j\neq k\neq i}}|a_{ki}|^2|a_{lj}|^2+\sum_{\substack{i,j,k,l\\i\leqslant j,l\neq j\\k\notin\{i,j,l\}}}|a_{ki}|^2|a_{lj}|^2. \tag{C.13}$$

Combination of the preceding equations shows that inequality (C.4) is equivalent to

$$\sum_{\substack{i,j\\ i>j}} |a_{ii}a_{ji}^* - a_{ij}a_{jj}^*|^2 + \sum_{\substack{i,j,k\\ j\neq i\neq k\neq j}} |a_{ii}a_{kj}^* - a_{ij}a_{ki}^*|^2 + \sum_{\substack{i,j,k,l\\ i>j,k>l\\ \{i,j\}\cap\{k,l\}=\varnothing}} |a_{ki}a_{lj}^* - a_{kj}a_{li}^*|^2 +$$

$$\sum_{\substack{i,j\\ i\neq j}} |a_{ij}|^4 + \sum_{\substack{i,j\\ i\neq j}} |a_{ij}|^2|a_{ji}|^2 + \sum_{\substack{i,j,k\\ i>j\\ j\neq k\neq i}} |a_{ki}|^2|a_{lj}|^2 + \sum_{\substack{i,j,k,l\\ i\leqslant j,l\neq j\\ k\notin\{i,j,l\}}} |a_{ki}|^2|a_{lj}|^2 \geqslant 0, \quad \text{(C.14)}$$

which is always satisfied; an equality holds iff $\mathbf{A}$ is diagonal. □

*Lemma 7.6.3:*

**Proof:** First we observe that the sum $\sum_i a_i b_i$, in which $a_i$ and $b_i$ are positive real numbers, is maximal if $a_i$ and $b_i$ are ordered in the same way as a function of magnitude. This is easily verified in the case of two terms. The general case follows immediately, as any permutation (of e.g. the elements $b_i$) is a composition of transpositions (i.e. permutations of two elements).
Next, observe that the right-hand side of Eq. (7.31) can be written as the inner product $\langle \mathbf{S_A}, \mathbf{S_B}\rangle$. The left-hand side takes the form

$$\langle \mathbf{AB}, \mathbf{U_{AB}}\mathbf{V_{AB}}^H\rangle = \langle \mathbf{S_A}(\mathbf{V_A}^H\mathbf{U_B}), (\mathbf{U_A}^H\mathbf{U_{AB}}\mathbf{V_{AB}}^H\mathbf{V_B})\mathbf{S_B}\rangle \quad \text{(C.15)}$$

$$\stackrel{\text{def}}{=} \langle \mathbf{S_A}\mathbf{P_1}, \mathbf{P_2}\mathbf{S_B}\rangle. \quad \text{(C.16)}$$

We will consider this expression as a function $f$ of the unitary matrices $\mathbf{P_1}$ and $\mathbf{P_2}$. It will be differentiated over the manifolds of unitary matrices, and the global optimum will be found among the "critical points" (i.e. where an infinitesimal variation of $\mathbf{P_1}$, or $\mathbf{P_2}$, does not influence the value of $f(\mathbf{P_1}, \mathbf{P_2})$). For more details on the mathematical technique we refer to [77].
The effect on $f$ of a variation of e.g. $\mathbf{P_1}$ is investigated by giving $\mathbf{P_1}$ a "velocity" as a function of a "time" coordinate $t$. The derivative with respect to $t$ is indicated by a dot: $\dot{\mathbf{P}}_\mathbf{1}$. The condition for $\mathbf{P_1}(t)$ to remain unitary, is that $\dot{\mathbf{P}}_\mathbf{1} = \Omega\mathbf{P_1}$, in which $\Omega$ is skew-Hermitean. We have:

$$\dot{f}(\mathbf{P_1}, \mathbf{P_2}) = \langle \mathbf{S_A}\dot{\mathbf{P}}_\mathbf{1}, \mathbf{P_2}\mathbf{S_B}\rangle \quad \text{(C.17)}$$

$$= \langle \Omega, \mathbf{S_A}\mathbf{P_2}\mathbf{S_B}\mathbf{P_1}^H\rangle. \quad \text{(C.18)}$$

This inner product vanishes iff

$$\mathbf{S_A}\mathbf{P_2}\mathbf{S_B}\mathbf{P_1}^H = \mathbf{P_1}\mathbf{S_B}\mathbf{P_2}^H\mathbf{S_A}, \quad \text{(C.19)}$$

for which we define $\mathbf{C} = \mathbf{P_2}\mathbf{S_B}\mathbf{P_1}^H$. A similar derivation in terms of $\mathbf{P_2}$ yields:

$$\mathbf{S_A}\mathbf{C}^H = \mathbf{C}\mathbf{S_A}. \quad \text{(C.20)}$$



Eqs.(C.19,C.20) can be combined to yield:

$$\mathbf{S_A}^2(\mathbf{C}\mathbf{C}^H) = (\mathbf{C}\mathbf{C}^H)\mathbf{S_A}^2, \qquad \mathbf{S_A}^2(\mathbf{C}^H\mathbf{C}) = (\mathbf{C}^H\mathbf{C})\mathbf{S_A}^2, \quad \text{(C.21)}$$

$$\mathbf{S_B}^2(\mathbf{C}\mathbf{C}^H) = (\mathbf{C}\mathbf{C}^H)\mathbf{S_B}^2, \qquad \mathbf{S_B}^2(\mathbf{C}^H\mathbf{C}) = (\mathbf{C}^H\mathbf{C})\mathbf{S_B}^2. \quad \text{(C.22)}$$

Assume that all the pairs $(\sigma_i(\mathbf{A}), \sigma_i(\mathbf{B}))$ are mutually different. Then the preceding commutation equations can only hold, if $\mathbf{P_1}$ is a matrix that contains exactly one, unit-modulus, entry in each column and row. Recalling the definition of $\mathbf{P_2}$, the equivalent condition for $\mathbf{P_2}$ is then automatically satisfied. Recalling the preliminary observation, the global optimum of $f$ is reached when $\mathbf{P_1}$ is diagonal. In case that not all the singular value pairs are mutually different, the derivation can be generalized as specified by the lemma. □

*Prop. 7.7.4:*

**Proof:** Let the SVD of the transfer matrix be given by $\mathbf{M} = \mathbf{EDQ}$. The conditions of the theorem are such that the prewhitening stage allows a perfect reconstruction of $\mathbf{ED}$ (see Sect. 7.2) and the higher-order stage allows a perfect reconstruction of $\mathbf{Q}$ (see e.g. Prop. 7.7.2). Hence Eq. (7.21) is the implementation of a matched filter. □

*Theor. 7.9.1:*

**Proof:** Under the conditions specified by the theorem, the covariance $\mathbf{C}_2^Y$ has the form:

$$\mathbf{C}_2^Y = \begin{pmatrix} \sigma_{x_1}^2\cos^2(\alpha) + \sigma_{x_2}^2\cos^2(\beta) + \sigma_n^2 & \sigma_{x_1}^2\cos\alpha\sin\alpha + \sigma_{x_2}^2\cos\beta\sin\beta \\ \sigma_{x_1}^2\cos\alpha\sin\alpha + \sigma_{x_2}^2\cos\beta\sin\beta & \sigma_{x_1}^2\sin^2(\alpha) + \sigma_{x_2}^2\sin^2(\beta) + \sigma_n^2 \end{pmatrix}. \quad \text{(C.23)}$$

It can be verified that the eigenvalues of $\mathbf{C}_2^Y$ are given by

$$\sigma_1^2 = \sigma_n^2 + \sigma_{x_1}^2 + \gamma, \qquad \sigma_2^2 = \sigma_n^2 + \sigma_{x_2}^2 - \gamma, \quad \text{(C.24)}$$

in which

$$\gamma = \frac{1}{2}\left[\sqrt{(\sigma_{x_1}^2 - \sigma_{x_2}^2)^2 + 4\sigma_{x_1}^2\sigma_{x_2}^2)^2\cos^2(\alpha-\beta)} - (\sigma_{x_1}^2 - \sigma_{x_2}^2)\right]. \quad \text{(C.25)}$$

The first eigenvector $U_1$ is proportional to:

$$U_1 \sim \begin{pmatrix} \sigma_{x_1}^2\cos\alpha\sin\alpha + \sigma_{x_2}^2\cos\beta\sin\beta \\ \sigma_{x_1}^2\sin^2(\alpha) - \sigma_{x_2}^2\cos^2(\beta) + \gamma \end{pmatrix}. \quad \text{(C.26)}$$

Computation of the limit for $\frac{\sigma_{x_2}}{\sigma_{x_1}} \to 0$ proves the theorem. □

*Theor. 9.1.1:*

**Proof:** The diagonal entries of $\mathcal{C}$ have the form

$$\begin{aligned} c_{111} &= \cos^3\alpha\tilde{c}_{111} - 3\cos^2\alpha\sin\alpha\tilde{c}_{112} + 3\cos\alpha\sin^2\alpha\tilde{c}_{122} - \sin^3\alpha\tilde{c}_{222}, \\ c_{222} &= \sin^3\alpha\tilde{c}_{111} + 3\sin^2\alpha\cos\alpha\tilde{c}_{112} + 3\sin\alpha\cos^2\alpha\tilde{c}_{122} + \cos^3\alpha\tilde{c}_{222}. \end{aligned}$$

The squared diagonal norm of $\mathcal{C}$ can then be written as:

$$\begin{aligned} c_{111}^2 + c_{222}^2 &= (\frac{9}{4}(\tilde{c}_{112}^2 + \tilde{c}_{122}^2) + \frac{3}{2}(\tilde{c}_{111}\tilde{c}_{122} + \tilde{c}_{112}\tilde{c}_{222}))\sin^2(2\alpha) \\ &\quad +(\tilde{c}_{111}^2 + \tilde{c}_{222}^2)(1 - \frac{3}{4}\sin^2(2\alpha)) \\ &\quad +(\tilde{c}_{122}\tilde{c}_{222} - \tilde{c}_{111}\tilde{c}_{112})(3\sin(2\alpha)\cos(2\alpha)). \end{aligned}$$

In Eq. (9.8), this expression is given in a matrix format. □

*Theor. 9.1.2:*

**Proof:** The diagonal entries of $\mathcal{C}$ have the form

$$\begin{aligned} c_{111} &= \cos^3\alpha\tilde{c}_{111} - \cos^2\alpha\sin\alpha(2\tilde{c}_{112}e^{-i\phi} + \tilde{c}_{211}e^{i\phi}) \\ &\quad + \cos\alpha\sin^2\alpha(\tilde{c}_{122}e^{-2i\phi} + 2\tilde{c}_{212}) - \sin^3\alpha\tilde{c}_{222}e^{-i\phi}, \\ c_{222} &= \sin^3\alpha\tilde{c}_{111}e^{i\phi} + \sin^2\alpha\cos\alpha(2\tilde{c}_{112} + \tilde{c}_{211}e^{2i\phi}) \\ &\quad + \sin\alpha\cos^2\alpha(\tilde{c}_{122}e^{-i\phi} + 2\tilde{c}_{212}e^{i\phi}) + \cos^3\alpha\tilde{c}_{222}. \end{aligned}$$

The squared diagonal norm of $\mathcal{C}$ can then be written as:

$$\begin{aligned} |c_{111}|^2 + |c_{222}|^2 &= \sin^2(2\alpha)\,[-\frac{3}{4}(|\tilde{c}_{111}|^2 + |\tilde{c}_{222}|^2) \\ &\quad +\frac{1}{4}(|2\tilde{c}_{112} + \tilde{c}_{211}e^{2i\phi}|^2 + |2\tilde{c}_{212} + \tilde{c}_{122}e^{-2i\phi}|^2) \\ &\quad +\frac{1}{2}\mathtt{Re}(\tilde{c}_{111}(\tilde{c}_{122}e^{-2i\phi} + 2\tilde{c}_{212})^* + \tilde{c}_{222}(\tilde{c}_{211}e^{2i\phi} + 2\tilde{c}_{112})^*)] \\ &\quad -\sin(2\alpha)\cos(2\alpha)\mathtt{Re}(\tilde{c}_{111}e^{i\phi}(\tilde{c}_{211}e^{2i\phi} + 2\tilde{c}_{112})^* \\ &\qquad\qquad - \tilde{c}_{222}^*e^{i\phi}(\tilde{c}_{122}e^{-2i\phi} + 2\tilde{c}_{212})). \end{aligned}$$

In Eq. (9.29), this expression is reorganized and presented in a matrix format. □

# Bibliography


[1] R. Abraham, *Linear and Multilinear Algebra*, Benjamin, N.Y., 1966.

[2] M. Abramowitz, I. Segun (Eds.), *Handbook of Mathematical Functions*, Dover Publ., N.Y., 1968, pp. 17-18.

[3] S. Abboud et al., "Real-Time Abdominal Fetal ECG Recording Using a Hardware Correlator", *Comp. in Biology and Med.*, No. 22, 1995, pp. 325-335.

[4] B.G. Agee, S.V. Schell, W.A. Gardner, "Spectral Self-Coherence Restoral: A New Approach to Blind Adaptive Signal Extraction Using Antenna Arrays", *Proc. IEEE*, Vol. 78, April 1990, pp. 753-767.

[5] M.A. Akivis, V.V. Goldberg, R.A. Silverman (Ed.), *An Introduction to Linear Algebra and Tensors*, Dover, N.Y., 1977.

[6] S. Anderson, M. Millnert, M. Viberg, B. Wahlberg, "An Adaptive Array for Mobile Communication Systems", *IEEE Trans. VT*, Vol. 40, No. 1, Feb. 1991, pp. 230-236.

[7] C.J. Appellof, E.R. Davidson, "Strategies for Analyzing Data from Video Fluoromatric Monitoring of Liquid Chromatographic Effluents", *Analytical Chemistry*, Vol. 53, No. 13, Nov. 1981, pp. 2053-2056.

[8] J. Arnbak, "The European (R)evolution of Wireless Personal Communications," *IEEE Comm. Mag.*, Sept. 1993, pp. 74-82.

[9] R.B. Ash, *Basic Probability Theory*, Wiley, N.Y., 1970.

[10] T.A. Bak, J. Lichtenberg, *Vectors, Tensors, and Groups*, Benjamin, N.Y., 1967.

[11] Y. Bar-Ness, J.W. Carlin, M.L. Steinberger, "Bootstrapping Adaptive Interference Cancellers: Some Practical Limitations", *Proc. Globecom. Conf.*, Miami, Nov. 1982, pp. 1251-1255.

[12] B. Bastijns, L. De Lathauwer, *Real-time Extractie van Medische Informatie uit Foetale Hartsignalen*, Master's Thesis, K.U.Leuven, E.E. Dept., 1992 (in Dutch).

[13] J.H. Bennett (Ed.), *Collected Papers of R.A. Fisher*, Vol. 2, Paper 74, University of Adelaide Press, 1972, pp. 351-354.

[14] Bloxom B., "Tucker's Three-Mode Factor Analysis Model", in: H.G. Law, C.W. Snyder, J.A. Hattie, R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis*, Praeger, N.Y., 1984, pp. 104-121.

[15] C. Bourin, P. Bondon, "Efficiency of High-Order Moment Estimates", *Proc. IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics*, Girona, Spain, June 12-14, 1995, pp. 186-190.

[16] R.M. Bowen, C.-C. Wang, *Introduction to Vectors and Tensors. Linear and Multilinear Algebra*, Vol. 1, Plenum Press, N.Y., 1980.

[17] J. Brewer, "Kronecker Products and Matrix Calculus in System Theory", *IEEE Trans. CAS*, Vol. 25, Sept. 1978, pp. 772-780.

[18] D.R. Brillinger, "Basic Aspects of Higher-order Spectra and Some of their Uses", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 9-18.

[19] R. Bro, "PARAFAC. Tutorial & Applications", to appear in: *Chemom. Intell. Lab. Syst.*, Special Issue *2nd Internet Conf. in Chemometrics (INCINC'96)*.

[20] D.H. Brooks, R.S. McLeod, "Electrical Imaging of the Heart", *IEEE Signal Processing Mag.*, Jan. 1997, pp. 24-42.

[21] A. Bunse-Gerstner, R. Byers and V. Mehrmann, "Numerical Methods for Simultaneous Diagonalization", *SIAM J. Matrix Anal. Appl.*, Vol. 14, No. 4, Oct. 1993, pp. 927-949.

[22] D. Burdick, X. Tu, L. McGown, D. Millican, "Resolution of Multicomponent Fluorescent Mixtures by Analysis of the Excitation-Emision-Frequency Array", *Journal of Chemometrics*, Vol. 4, 1990, pp. 15-28.

[23] J.A. Cadzow, "Signal Enhancement - A Composite Property Mapping Algorithm", *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 36, No. 1, Jan. 1988, pp. 49-62.

[24] D. Callaerts, J. Vanderschoot, J. Vandewalle, W. Sansen, G. Vantrappen, J. Janssens, "Fetal Electrocardiogram Measuring Method and Equipment (FEMME)", *Journal of Perinatal Medicine*, Vol. 15, Suppl. 1, World Symp. on Computers in the Care of Mother, Fetus and Newborn, Wien, March 1987, p. 33.



[25] D. Callaerts, J. Vandewalle, W. Sansen, M. Moonen, "On-line Algorithm for Signal Separation based on SVD", in: E. Deprettere (Ed.), *SVD and Signal Processing. Algorithms, Applications and Architectures*, Elsevier Science Publishers B.V., North-Holland, 1988, pp. 269-276.

[26] D. Callaerts, *Signal Separation based on Singular Value Decomposition and their Application to the Real-time Extraction of the Fetal Electrocardiogram from Cutaneous Recordings*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., Dec. 1989.

[27] D. Callaerts, J. Vanderschoot, J. Vandewalle, W. Sansen, "An On-Line Adaptive Algorithm for Signal Processing using SVD", *Signal Processing III. Proc. EURASIP 86*, Den Haag, The Netherlands, Sept. 1986, pp. 953-956.

[28] D. Callaerts, J. Vandewalle, "Adaptive Singular Value Decomposition and its Use for Adaptive Signal Separation", *Proc. Third Int. Symp. on Mathematical Theory of Networks and Systems, MTNS-89*, Amsterdam, June 1989, Vol. 3, pp. 255-262.

[29] D. Callaerts, J. Vandewalle, W. Sansen, J. Janssens, G. Vantrappen, "Acquisition and Processing of the Antepartum FECG", in: H.P. van Geijn, F.J.A. Copray (Eds.), *A Critical Appraisal of Fetal Surveillance*, Elsevier Science B.V., 1994, pp. 371-380.

[30] M. Callendar, "Standards for Global Communications Services", *Proc. Mobile Radio Conf.*, Nice, France, Nov. 1991, pp. 229-234.

[31] J.-F. Cardoso, "Higher-Order Narrow-Band Array Processing", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 39-48.

[32] J.-F. Cardoso, "Source Separation using Higher-Order Moments", *Proc. ICASSP-89*, 1989, pp. 2109-2112.

[33] J.-F. Cardoso, "Eigen-Structure of the Fourth-Order Cumulant Tensor with Application to the Blind Source Separation Problem", *Proc. ICASSP-90*, 1990, pp. 2655-2658.

[34] J.-F. Cardoso, "Iterative Techniques for Blind Source Separation using only Fourth-Order Cumulants", *Signal Processing VI: Theories and Applications, Proc. EUSIPCO-92*, Brussels, Belgium, Aug. 24-27, 1992, Vol. 2, pp. 739-742.

[35] J.-F. Cardoso, P. Comon, "Tensor-Based Independent Component Analysis", *Signal Processing V: Theories and Applications, Proc. EUSIPCO-90*, 1990, pp. 673-676.

[36] J.-F. Cardoso, "Fourth-Order Cumulant Structure Forcing. Application to Blind Array Processing." *Proc. SSAP-92*, pp. 136-139.

[37] J.-F. Cardoso, "Super-Symmetric Decomposition of the Fourth-Order Cumulant Tensor. Blind Identification of More Sources than Sensors." *Proc. ICASSP-91*, Vol. 5, pp. 3109-3112.

[38] J.-F. Cardoso, "Localisation and Identification with the Quadricovariance", *Traitement du Signal*, Special Issue *Non-linear and Non-Gaussian*, Vol. 7, No. 5, Autumn 1990, pp. 397-406.

[39] J.-F. Cardoso, "A Tetradic Decomposition of 4th-order Tensors. Application to the Source Separation Problem." in: B. De Moor, M. Moonen (Eds.), *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1995, pp. 375-382.

[40] J.-F. Cardoso, A. Souloumiac, "An Efficient Technique for Blind Separation of Complex Sources", *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, Lake Tahoe, U.S.A., 1993, pp. 275-279.

[41] J.-F. Cardoso, A. Souloumiac, "Blind Beamforming for Non-Gaussian Signals", *IEE Proc.-F*, Vol. 140, No. 6, 1994, pp. 362-370.

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

[43] J.-F. Cardoso, E. Moulines, "How Much More DOA Information in Higher-Order Statistics?", *Proc. SSAP-94*, Québec, pp. 199-202.

[44] J.-F. Cardoso, S. Bose, B. Friedlander, "Output Cumulant Matching for Source Separation", *Proc. IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics*, Girona, Spain, June 12-14, 1995, pp. 44-50.

[45] J.-F. Cardoso (Ed.), *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 36, No. 3, April 1994.

[46] J.-F. Cardoso, "The Invariant Approach to Source Separation", *Proc. NOLTA'95*, Las Vegas, U.S.A., Dec. 10-14, 1995, pp. 55-60.

[47] J.-F. Cardoso, "On Optimal Source Separation Based on Second and Fourth Order Cumulants", *Proc. SSAP-96*, June 22-24, 1996, Corfu, Greece, pp. 198-201.

[48] J.-F. Cardoso, A. Souloumiac, "Jacobi Angles for Simultaneous Diagonalization", *SIAM J. Matrix Anal. Appl.*, Vol. 17, No. 1, Jan. 1996, pp. 161-164.



[49] J.D. Carroll, S. Pruzansky, "The CANDECOMP-CANDELINC Family of Models and Methods for Multidimensional Data Analysis", in: H.G. Law, C.W. Snyder, J.A. Hattie, R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis*, Praeger, N.Y., 1984, pp. 372-402.

[50] J.D. Carroll, G. Desoete, S. Pruzansky, "An Evaluation of Five Algorithms for Generating an Initial Configuration for SINDSCAL", *Journal of Classification*, Vol. 6, 1989, pp. 105-119.

[51] J.D. Carroll, J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an $N$-Way Generalization of "Eckart-Young" Decompositions", *Psychometrika*, Vol. 35, 1970, pp. 283-319.

[52] L. Chambadal, J.-L. Ovaert, *Algèbre Linéaire et Algèbre Tensorielle*, Dunod Paris, 1968.

[53] C. Charles, D. Allouch, *Algèbre Générale*, PUF, Paris, 1984.

[54] P. Chevalier, "On the Performances of Higher Order Blind Source Separation Methods", *Proc. IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics*, Girona, Spain, June 12-14, 1995, pp. 30-34.

[55] F. Chorlton, *Vector and Tensor Methods*, Horwood Chichester, 1976.

[56] M.T. Chu, "A Continuous Jacobi-Like Approach to the Simultaneous Reduction of Real Matrices", *Lin. Alg. Appl.*, Vol. 147, 1991, pp. 75-96.

[57] E.K. Chung, *Electrocardiography: Practical Applications with Vectorial Principles*, Harper & Row Publ. Inc., Hagerstown, Maryland, U.S.A., 1980.

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

[59] P. Comon, "Blind Identification in Presence of Noise", *Signal Processing VI: Theories and Applications, Proc. EUSIPCO-92*, Brussels, Aug. 24-27, 1992, Vol. 2, pp. 835-838.

[60] P. Comon, "Independent Component Analysis and Blind Identification", *Traitement du Signal*, Special Issue *Non-linear and Non-Gaussian*, Vol. 7, No. 5, Autumn 1990, pp. 435-450.

[61] P. Comon, "Higher-Order Separation, Application to Detection and Localization", *Signal Processing V: Theories and Applications, Proc. EUSIPCO-90*, 1990, pp. 277-280.

[62] P. Comon, "Independent Component Analysis", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 29-38.

[63] P. Comon, J.-F. Cardoso, "Eigenvalue Decomposition of a Cumulant Tensor with Applications", *Advanced Algorithms and Architectures for Signal Processing IV, Proc. SPIE Conf. on Adv. Sig. Proc. Algorithms, Architectures and Implementations*, San Diego, CA, U.S.A., July 10-12, 1990, Vol.1348, pp. 361-372.

[64] P. Comon, "Independent Component Analysis, A New Concept?" *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 36, No. 3, April 1994, pp. 287-314.

[65] P. Comon, "Remarques sur la Diagonalisation Tensorielle par la Méthode de Jacobi", *Actes du XIV Colloque GRETSI*, Sept. 13-16, 1993, pp. 125-128.

[66] P. Comon, *Quelques Développements Récents en Traitement du Signal*, Habilitation à Diriger des Recherches, Université de Nice, 1995.

[67] P. Comon, B. Mourrain, "Decomposition of Quantics in Sums of Powers of Linear Forms", *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 53, Nos. 2-3, Sept. 1996, pp. 93-108.

[68] P. Comon, "Blind Separation of Sources, Part II: Problems Statement", *Signal Processing*, No. 24, 1991, pp. 11-20.

[69] P. Comon, "Tensor Diagonalization, a Useful Tool in Signal Processing", in: M. Blanke, T. Söderström (Eds.), *Proc. SYSID'94*, Copenhagen, Denmark, Vol. 1, pp. 77-82.

[70] P. Comon, L. De Lathauwer, B. De Moor, "A Contrast-Based Independent Component Analysis without Second-Order Moments", Tech. Report No. 95-67, ESAT/SISTA, K.U.Leuven; Tech. Report No. 96-39, I3S, C.N.R.S. - U.R.A. 1376, Univ. de Nice - Sophia Antipolis.

[71] P. Comon, G. Golub, "Tracking a Few Extreme Singular Values and Vectors in Signal Processing", *Proc. IEEE*, Vol. 78, No. 8, August 1990, pp. 1327-1343.

[72] P. Comon, E. Moreau, "Improved Contrast Dedicated to Blind Separation in Communications", *Proc. ICASSP'97*, Vol. 5, pp. 3453-3456.

[73] M.J. Crowe, *A History of Vector Analysis. The Evolution of the Idea of a Vectorial System.* Dover Publications Inc., N.Y., 1985.

[74] J.A. Crowe, B.H. Hayes-Gill, D.K. James, "Towards Fetal Heart Rate Monitoring at Home Using the Abdominal Fetal Electrocardiogram", *Proc. 18th Int. Conf. on Engineering in Medicine and Biology*, Amsterdam, The Netherlands, Oct. 31- Nov. 3, 1996, paper no. 57.

[75] G. Darmois, "Analyse Générale des Liaisons Stochastiques", *Rev. Inst. Internat. Stat.*, Vol. 21, 1953, pp. 2-8.



[76] C. Davis, W. Kahan, "The Rotation of Eigenvectors by a Perturbation", *SIAM J. Numer. Anal.*, Vol. 7, No. 1, March 1970, pp. 1-46.

[77] J. Dehaene, *Continuous-Time Matrix Algorithms, Systolic Algorithms and Adaptive Neural Networks*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., Oct. 1995.

[78] L. De Lathauwer, B. De Moor, J. Vandewalle, "A Singular Value Decomposition for Higher-Order Tensors", *Proc. ProRISC IEEE Workshop on Circuits, Systems and Signal Processing*, Houthalen, Belgium, March 24-25, 1993, pp. 37-43.

[79] L. De Lathauwer, B. De Moor, J. Vandewalle, "A Singular Value Decomposition for Higher-Order Tensors", *Proc. ATHOS Workshop on System Identification and High-Order Statistics*, Nice, France, Sept. 20-21, 1993.

[80] L. De Lathauwer, B. De Moor, J. Vandewalle, "The Application of Higher Order Singular Value Decomposition to Independent Component Analysis", in: B. De Moor, M. Moonen (Eds.), *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1995, pp. 383-390.

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

[82] L. De Lathauwer, B. De Moor, J. Vandewalle, "A Singular Value Decomposition for Higher-Order Tensors", Tech. Report No. 94-31, ESAT/SISTA, K.U.Leuven, submitted to: *SIAM J. Matrix Anal. Appl.*.

[83] L. De Lathauwer, D. Callaerts, B. De Moor, J. Vandewalle, "Fetal Electrocardiogram Extraction by Source Subspace Separation", *Proc. IEEE SP / ATHOS Workshop on HOS*, Girona, Spain, June 12-14, 1995, pp. 134-138.

[84] L. De Lathauwer, B. De Moor, J. Vandewalle, "Linear Transformations of Higher-Order Spaces: Geometrical Concepts and Decompositions", Tech. Report No. 94-86, ESAT/SISTA, K.U.Leuven (in Dutch).

[85] L. De Lathauwer, B. De Moor, J. Vandewalle, "Reconstruction of Multidimensional Harmonics by means of Higher-Order Matrices", Tech. Report No. 94-87, ESAT/SISTA, K.U.Leuven (in Dutch).

[86] L. De Lathauwer, B. De Moor, J. Vandewalle, "Geometrical Interpretation of Givens and Jacobi Rotations for Vectors, Matrices and Tensors", Tech. Report No. 94-89, ESAT/SISTA, K.U.Leuven (in Dutch).

[87] L. De Lathauwer, P. Comon, B. De Moor, J. Vandewalle, "Higher-Order Power Method - Application in Independent Component Analysis", *Proc. Int. Symp. on Nonlinear Theory and Its Applications (NOLTA'95)*, Las Vegas, U.S.A., Dec. 10-14, 1995, Vol. 1, pp. 91-96.

[88] L. De Lathauwer, B. De Moor, J. Vandewalle, "Blind Source Separation by Simultaneous Third-Order Tensor Diagonalization", *Proc. EUSIPCO-96*, Sept. 10-13, 1996, Trieste, Italy, Vol. 3, pp. 2089-2092.

[89] L. De Lathauwer, B. De Moor, J. Vandewalle, "Independent Component Analysis Based on Higher-Order Statistics Only", *Proc. 8th IEEE SP Workshop on Statistical Signal and Array Processing (SSAP-96)*, Corfu, Greece, June 24-26, 1996, pp. 356-359.

[90] L. De Lathauwer, B. De Moor, J. Vandewalle, "A Technique for Higher-Order-Only Independent Component Analysis", *Proc. Int. Conf. on Neural Information Processing ICONIP'96*, Sept. 24-27, 1996, Hong Kong, Vol. 2, pp. 1223-1228.

[91] L. De Lathauwer, B. De Moor, "From Matrix to Tensor: Multilinear Algebra and Signal Processing", Tech. Report No. 97-58, ESAT/SISTA, K.U.Leuven, to appear in: *Mathematics in Signal Processing IV*, IMA Conference Series, Oxford University Press, 1997. Papers presented at *4th IMA Int. Conf. on Mathematics in Signal Processing*, Dec. 17-19, 1996, University of Warwick, U.K. (15 p.)

[92] L. De Lathauwer, B. De Moor, J. Vandewalle, "Dimensionality Reduction in Higher-Order-Only ICA", *Proc. IEEE Signal Processing workshop on HOS*, July 21-23, 1997, Banff, Alberta, Canada, pp. 316-320.

[93] L. De Lathauwer, B. De Schutter, B. De Moor, "An SVD for All Seasons", Tech. Report No. 97-60, ESAT/SISTA, K.U.Leuven (in preparation). Abstract accepted for *SIAM Rev.*

[94] L. De Lathauwer, B. De Moor, J. Vandewalle, "Canonical Cumulant Decomposition and Blind Source Separation", Tech. Report No. 97-, ESAT/SISTA, K.U.Leuven (in preparation). To be submitted to: *Journal A*.

[95] L. De Lathauwer, B. De Moor, J. Vandewalle, "A Prewhitening-Induced Lower-Bound on the Quality of the Mixing Matrix Estimate in ICA", Tech. Report No. 97-, ESAT/SISTA, K.U.Leuven (in preparation). To be submitted to: EUSIPCO-98, Greece, Sept. 8-11, 1998.

[96] L. De Lathauwer, B. De Moor, J. Vandewalle, "The Equivalent of PCA in Higher-Order-Only ICA", Tech. Report No. 97-, ESAT/SISTA, K.U.Leuven (in preparation). To be submitted to: *Signal Processing*.



[97] B. De Moor, *Mathematical Concepts and Techniques for Modelling of Static and Dynamic Systems*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., June 1988.

[98] B. De Moor, "The Singular Value Decomposition and Long and Short Spaces of Noisy Matrices", *IEEE Trans. Signal Processing*, Vol. 41, No. 9, Sept. 1993, pp. 2826-2838.

[99] B. De Moor, S. Boyd, "Analytic properties of singular values and vectors", Tech. Report No. 1989-28, ESAT/SISTA, K.U.Leuven, 1989.

[100] B. De Moor, G. Golub, "Generalized Singular Value Decompositions: A Proposal for a Standardized Nomenclature", Tech. Report No. 1989-10, ESAT/SISTA, K.U.Leuven, 1989.

[101] B. De Moor, "On the Structure of Generalized Singular Value and QR Decompositions", *SIAM J. Matrix Anal. Appl.*, Vol. 15, No. 1, Jan. 1994, pp. 347-358.

[102] B. De Moor, "Total Least Squares for Affinely Structured Matrices and the Noisy Realization Problem", *IEEE Trans. Signal Processing*, Vol. 42, No. 11, Nov. 1994, pp. 3104-3113.

[103] B. De Moor, "Dynamical Total Least Squares", *Proc. 10th IFAC Symp. on System Identification*, Copenhagen, Denmark, July 4-6, 1994, Vol. 3, pp. 159-164.

[104] B. De Moor, "The Riemannian Singular Value Decomposition", in: M. Moonen, B. De Moor (Eds.), *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1995, pp. 61-78.

[105] B. De Moor (Ed.), *DAISY (Database for the Identification of Systems)*, E.E. Dept., ESAT/SISTA, K.U.Leuven, Belgium, URL: `http://www.esat.kuleuven.ac.be/sista/daisy/`.

[106] E.F. Deprettere (Ed.), *SVD and Signal Processing. Algorithms, Applications and Architectures*, North-Holland, Amsterdam, 1988.

[107] G. Desanghere, *Identification and Quantification of Noise and Vibration Sources Using Frequency Response and Coherence Functions*, Ph.D. Thesis, K.U.Leuven, Mech. Dept., 1987.

[108] B. De Schutter, *Max-Algebraic System Theory for Discrete Event Systems*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., Feb. 1996.

[109] M.Y. Divon et al., "Autocorrelation Techniques in Fetal Monitoring", *Am. J. Obste. Gynecol.*, Vol. 151, Jan. 1985, pp. 2-6.

[110] C.T.J. Dodson, T. Poston, *Tensor Geometry: the Geometric Viewpoint and its Uses*, Pitman London, 1977.

[111] D. Donoho, "On Minimum Entropy Deconvolution", *Applied Time-Series Analysis II*, Academic Press, 1981, pp. 565-609.

[112] O. Dossing, "Structural Stroboscopy - Measurements of Operational Deflection Shapes", *Sound and Vibration*, Aug. 1988, pp. 18-26.

[113] O. Dossing, C.H. Staker, "Operational Deflection Shapes: Background, Measurement and Applications", *Proc. 7th Int. Modal Analysis Conf.*, Las Vegas, U.S.A., 1989, pp. 1372-1378.

[114] P. Duvaut, "Principles of Source Separation Methods Based on Higher Order Statistics", *Traitement du Signal*, Special Issue *Non-linear and Non-Gaussian*, Vol. 7, No. 5, Autumn 1990, pp. 407-418.

[115] P. Duvaut, "Non Linear Filtering in Signal Processing", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 19-28.

[116] J.C. Echeverría et al., "Fetal QRS Extraction Based on Wavelet Analysis and Pattern Matching", *Proc. 18th Int. Conf. on Engineering in Medicine and Biology*, Amsterdam, The Netherlands, Oct. 31- Nov. 3, 1996.

[117] M. Farrashkhalvat, J.P. Miles, *Tensor Methods for Engineers*, Ellis Horwood, N.Y., 1990.

[118] B. Feinberg, H.B. Krebbs, "Intrapartum Fetal Heart Rate Patterns", in: J.A.D. Spencer (Ed.), *Fetal Monitoring: Physiology and Techniques of Antenatal and Intrapartum Assessment*, Castle House Publ. Ltd., Turnbridge Wells, 1989.

[119] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, N.Y., 1971.

[120] B. Flury, *Common Principal Components & Related Multivariate Models*, John Wiley & Sons, 1988.

[121] J. Fonollosa, "Sample Cumulants of Stationary Processes: Asymptotic Results", *IEEE Trans. Signal Processing*, Vol. 43, No. 4, April 1995, pp. 967-977.

[122] J. Fonollosa (Ed.), *Proc. IEEE SP/ATHOS Workshop on Higher-Order Statistics*, Girona, Spain, June 12-14, 1995, Pub. de la UPC.

[123] M. Gaeta, F. Harroy, J.L. Lacoume, "Utilisation de Statistiques d'Ordre Supérieur pour la Séparation de Sources Vibrantes", *Progrès Récents des Méthodes de Surveillance Acoustiques et Vibratoires*, Senlis, France, October 27-29, 1992, pp. 523-533.



[124] M. Gaeta, J.L. Lacoume, "Sources Separation without a Priori Knowledge: the Maximum Likelihood Solution", *Signal Processing V: Theories and Applications, Proc. EUSIPCO-90*, pp. 621-624.

[125] M. Gaeta, J.L. Lacoume, "Estimateurs du Maximum de Vraisemblance Etendus à la Séparation de Sources Non Gaussiennes", *Traitement du Signal*, Special Issue *Non-linear and Non-Gaussian*, Vol. 7, No. 5, Autumn 1990, pp. 419-434.

[126] C.B. Garcia, W.I. Zangwill, "An Approach to Homotopy and Degree Theory", *Mathematics of Operations Research*, Vol. 4, No. 4, Nov. 1979, pp. 390-405.

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

[128] E. Gassiat, *Déconvolution Aveugle*, Ph.D. Thesis, Université de Paris Sud, Jan. 1988.

[129] G.B. Giannakis, J.M. Mendel, "Identification of Nonminimum Phase Systems Using Higher Order Statistics", *IEEE Trans. Acoustics, Speech, Signal Processing*, Vol. 37, No. 3, March 1989, pp. 360-377.

[130] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, Maryland, 1991.

[131] G.H. Golub, W. Kahan, "Calculating the Singular Values and Pseudo-Inverse of a Matrix", *SIAM J. Numer. Anal.*, Vol. 2, 1965, pp. 205-224.

[132] G.H. Golub, C. Reinsch, "Singular Value Decomposition and Least-Squares Solution", *Numer. Math.*, Vol. 14, 1970, pp. 403-420.

[133] W. Greub, *Multilinear Algebra*, Springer-Verlag, N.Y., 1978.

[134] V.S. Grigorascu, *Tenseurs Structurés, Produits d'Ordre Supérieur et Cumulants*, Ph.D. Thesis, Université de Paris-Sud, Sept. 1996.

[135] M. Haardt, K. Hüper, J. Moore, J. Nossek, "Simultaneous Schur Decomposition of Several Matrices to Achieve Automatic Pairing in Multidimensional Harmonic Retrieval Problems", *Signal Processing VIII: Theories and Applications, Proc. of EUSIPCO-96*, Trieste, Italy, Sept. 10-13, Vol. 1, pp. 531-534.

[136] G.G. Hall, *Matrices and Tensors. Int. Encyclopedia of Physical Chemistry and Chemical Physics, Topic 1: Mathematical Techniques*, Vol. 4, Pergamon Press Oxford, 1963.

[137] P.C. Hansen, *Regularization Tools. A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems.* Tech. Report UNIC-92-03, Danish Computing Center for Research and Education, Technical University of Denmark, June, 1992.

[138] R.A. Harshman, M.E. Lundy, "The PARAFAC Model for Three-Way Factor Analysis and Multidimensional Scaling," in: H.G. Law, C.W. Snyder, J.A. Hattie, R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis*, Praeger, N.Y., 1984, pp. 122-215.

[139] R.A. Harshman, "Foundations of the PARAFAC Procedure: Model and Conditions for an "Explanatory" Multi-Mode Factor Analysis", *UCLA Working Papers in Phonetics*, Vol. 16, No. 1, 1970.

[140] R.A. Harshman, M.E. Lundy, "PARAFAC: Parallel Factor Analysis", *Comp. Stat. Data Anal.*, Vol. 18, No. 39, 1994, pp. 39-72.

[141] R. Harshman, P. Ladefoged, L. Goldstein, "Factor Analysis of Tongue Shapes", *J. Acoust. Soc. Am.*, Vol. 62, No. 3, Sept. 1977, pp. 693-707.

[142] J. Hastad, "Tensor Rank is NP-Complete", *Journal of Algorithms*, Vol. 11, 1990, pp. 644-654.

[143] J. Hérault, C. Jutten, B. Ans, "Détection de Grandeurs Primitives dans un Message Composite par une Architecture Neuromimétique en Apprentissage Non Supervisé", *Proc. 10th GRETSI Colloquium*, Nice, France, May 20-24, 1985, pp. 1017-1022.

[144] R. Hermann, *Linear and Tensor Algebra*, Mathematical Scientific Press Bookline, 1973.

[145] F.A. Hinchey, *Vectors and Tensors for Engineers and Scientists*, Wiley, New Delhi, 1976.

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

[147] P.J. Huber, *Robust Statistics*, Wiley, 1981.

[148] H. Jeffreys, B.S. Jeffreys, *Methods of Mathematical Physics*, Cambridge University Press, 1956.

[149] D.H. Johnson, D.E. Dudgeon, *Array Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[150] I.T. Jolliffe, *Principal Component Analysis*, Springer, N.Y., 1986.

[151] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, J. Caelen, "Blind Separation of Sources: An Algorithm for Separation of Convolutive Mixtures", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 275-278.

[152] C. Jutten, J.-F. Cardoso, "Separation of Sources: Really Blind?", *Proc. NOLTA'95*, Las Vegas, U.S.A., Dec. 10-14, 1995, Vol. 1, pp. 79-84.



[153] C. Jutten, J. Herault, "Blind Separation of Sources, Part I: an Adaptive Algorithm based on Neuromimetic Architecture", *Signal Processing*, No. 24, 1991, pp. 1-10.

[154] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators*, Springer-Verlag, N.Y., 1982.

[155] D.C. Kay, *Theory and Problems of Tensor Calculus*, McGraw-Hill, 1988.

[156] L.E. Kay, M. Clore, A. Bax, A. Gronenbaum, "Four-Dimensional Heteronuclear Triple-Resonance NMR Spectroscopy of Interleukin-1$\beta$ in Solution", *Science*, 249, 1990, pp. 411-414.

[157] M.G. Kendall, A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, Griffin, London, 1977.

[158] H. Kiers, W. Krijnen, "An Efficient Algorithm for PARAFAC of Three-Way Data with Large Numbers of Observational Units", *Psychometrika*, Vol. 56, No. 147, 1991.

[159] E.G. Kogbetliantz, "Solution of Linear Systems by Diagonalization of the Coefficients Matrix", *Quart. Appl. Math.*, No. 13, 1955, pp. 123-132.

[160] H. Krim, M. Viberg, "Two Decades of Array Signal Processing Research", *IEEE Signal Proc. Mag.*, July 1996, pp. 67-94.

[161] P.M. Kroonenberg, J. de Leeuw, "Principal Component Analysis of Three-Mode Data by means of Alternating Least Squares Algorithms", *Psychometrika*, Vol. 45, 1980, pp. 69-97.

[162] P.M. Kroonenberg, "Three-Mode Principal Component Analysis: Illustrated with an Example from Attachment Theory", in: H.G. Law, C.W. Snyder, J.A. Hattie, R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis*, Praeger, N.Y., 1984, pp. 64-103.

[163] P.M. Kroonenberg, *Three-Mode Principal Component Analysis*, DSWO Press, Leiden, 1983.

[164] J.B. Kruskal, "More Factors than Subjects, Tests and Treatments: an Indeterminacy Theorem for Canonical Decomposition and Individual Differences Scaling", *Psychometrika*, Vol. 41, No. 3, Sept. 1976, pp. 281-293.

[165] J.B. Kruskal, "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics", *Lin. Alg. Appl.*, No. 18, 1977, pp. 95-138.

[166] J.B. Kruskal, R.A. Harshman, M.E. Lundy, "How 3-MFA Data Can Cause Degenerate PARAFAC Solutions, among Other Relationships", in: R. Coppi, S. Bolasco (Eds.), *Multiway Data Analysis*, North-Holland, 1989, pp. 115-122.

[167] J.B. Kruskal, "Rank, Decomposition, and Uniqueness for 3-Way and $N$-Way Arrays", in: R. Coppi, S. Bolasco (Eds.), *Multiway Data Analysis*, North-Holland, 1989, pp. 7-18.

[168] S.Y. Kung, "A New Identification and Model Reduction Algorithm with Singular value Decomposition", *Proc. 12th Asilomar Conf. on Circuits and Systems*, Pacific Grove, CA, U.S.A., Nov. 1978, pp. 705-714.

[169] J.L. Lacoume, M.Gaeta, P.O. Amblard, "From Order 2 to HOS: New Tools and Applications", *Signal Processing VI: Theories and Applications, Proc. of EUSIPCO-92*, Brussels, Aug. 24-27, 1992, Vol.1, pp. 91-98.

[170] J.L. Lacoume, M.Gaeta, "Complex Random Variables: A Tensorial Approach", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 81-84.

[171] J.L. Lacoume, "Approche Tensorielle des Variables Aléatoires", Institut National Polytechnique de Grenoble, Centre d'Etude des Phénomènes Aléatoires et Géophysiques, Tech. Report No. 29/91.

[172] J.L. Lacoume, P. Ruiz, "Separation of Independent Sources from Correlated Inputs", *IEEE Trans. Signal Processing*, Vol. 40, No. 12, Dec. 1992, pp. 3074-3078.

[173] J.L. Lacoume, P. Comon, P.O. Amblard, *Statistiques d'Ordre Elevé en Traitement du Signal*, Masson (in preparation).

[174] J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, Elsevier, 1992.

[175] P. Ladefoged, R. Harshman, L. Goldstein, L. Rice, "Generating Vocal Tract Shapes from Formant Frequencies", *J. Acoust. Soc. Am.*, Vol. 64, No. 4, 1978, pp. 1027-1035.

[176] B. Laheld, J.-F. Cardoso, "Adaptive Source Separation without Prewhitening", *Proc. EUSIPCO-94*, Edinburgh, Scotland, U.K., Sept. 13-16, 1994, Vol. 1, pp. 183-186.

[177] S. Lang, *Algebra*, Addison Wesley, Reading, 1984.

[178] Gilles Legrand, *Algèbre Linéaire et Multilinéaire et Géometrie Différentielle*, Masson, Paris, 1971.

[179] E.L. Lehmann, *Testing Statistical Hypotheses*, Wiley, N.Y., 1959.



[180] D. Leibovici, R. Sabatier, "A Singular Value Decomposition of a k-Ways Array for a Principal Components Analysis of Multi-Way Data, the PTA-kmodes", submitted to *Lin. Alg. Appl.*.

[181] V.P. Leonov, A. Shiryaev, "On the Technique of Computing Semi-Invariants", *Theory of Probability Applications*, No. 4, 1959, pp. 319-329.

[182] S.E. Leurgans, R.T. Ross, R.B. Abel, "A Decomposition for Three-Way Arrays", *SIAM J. Matrix Anal. Appl.*, Vol. 14, No. 4, Oct. 1993, pp. 1064-1083.

[183] D. Lovelock, H. Rund, *Tensors, Differential Forms, and Variational Principles*, Dover Publications Inc., N.Y., 1989.

[184] M. Marcus, *Finite Dimensional Multilinear Algebra*, Dekker, N.Y., 1975.

[185] G.J. Marseille, R. de Beer, A.F. Mehlkopf, D. van Ormondt, "Optimization of Sample Positions in Magnetic Resonance Spectroscopy", *Proc. ProRISC IEEE Workshop on Circuits, Systems and Signal Processing*, Houthalen, Belgium, March 24-25, 1993, pp. 233-238.

[186] *MATLAB User's Guide. High-Performance Numeric Computation and Visualization Software*, The MathWorks Inc., Aug. 1992.

[187] S. Mayrargue, "A Blind Spatio-Temporal Equalizer for a Radio-Mobile Channel Using the Constant-Modulus Algorithm (CMA)", *Proc. ICASSP-94*, Vol. 4, 1994, pp. 317-320.

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

[189] J. McWhirter, "Algorithmic Engineering in Adaptive Signal Processing", *IEE Proceedings-F*, Vol. 139, No. 3, June 1992, pp. 226-232.

[190] J. McWhirter, I. Proudler, "Algorithmic Engineering: a Worked Example", *Signal Processing VI: Theories and Applications, Proc. of EUSIPCO-92*, Brussels, Aug. 24-27, 1992, Vol.1, pp. 5-12.

[191] J. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *Proc. IEEE*, Vol. 79, No. 3, March 1991, pp. 278-305.

[192] J. Mendel, X. Zhao, C. Ngo, "A New Formula to Calculate Cumulants and Some Applications of It", in: J.L. Lacoume (Ed.), *Higher Order Statistics. Proc. Int. Signal Processing Workshop on Higher Order Statistics*, Chamrousse, France, July 10-12, 1991, pp. 95-98.

[193] W. Merx, M.S. Yoon, J. Han, "The Role of Local Disparity in Conduction and Local Recovery Time on Ventricular Vulnerability to Fibrillation", *Am. Heart J.*, Vol. 94, No. 5, 1977, pp. 603-610.

[194] M. Moonen, *Jacobi-type Updating Algorithms for Signal Processing, Systems Identification and Control*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., 1990.

[195] M. Moonen, B. De Moor (Eds.), *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1995.

[196] J.J. More, "The Levenberg-Marquardt Algorithm: Implementation and Theory", in: G.A. Watson (Ed.), *Numerical Analysis, Lecture Notes in Mathematics*, Vol. 630, Springer-Verlag, 1977, pp. 105-116.

[197] A.P. Morgan, A.J. Sommese, L.T. Watson, "Finding All Isolated Solutions to Polynomial Systems Using HOMPACK", *ACM Trans. Mathematical Software*, Vol. 15, No. 2, June 1989, pp. 93-122.

[198] R.E. Mortensen, *Random Signals and Systems*, Wiley, 1987.

[199] E. Moulines, P. Duhamel, J.-F. Cardoso, S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters", *Proc. ICASSP-94*, Vol. 4, 1994, pp. 573-576.

[200] M. Mouly, M.-B. Pautet, *The GSM System for Mobile Communications*, 1982.

[201] J. Nagel, M. Schaldach, "Processing the Abdominal Fetal ECG using a new method", in: P. Rolfe (Ed.), *Fetal and Neonatal Physiological Measurements*, Pitman Medical, Londen, U.K., 1980.

[202] A.K. Nandi (Ed.), *Journal of the Franklin Institute*, Special Issue on *HOS in Signal Processing*, Vol. 333B, No. 3, May 1996.

[203] D.G. Nathcott, *Multilinear Algebra*, Cambridge University Press, Cambridge, 1984.

[204] C.L. Nikias, "Higher Order Spectra in Signal Processing", *Signal Processing V: Theories and Applications, Proc. EUSIPCO-90*, 1990, pp. 35-41.

[205] C.L. Nikias, J.M. Mendel, "Signal Processing with Higher-Order Spectra", *IEEE Signal Proc. Mag.*, July 1993, pp. 10-37.

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

[207] D. Nix, G. Papcun, J. Hogden, I. Zlokamik, "Two Cross-Linguistics Factors Underlying Tongue Shapes for Vowels", *J. Acoust. Soc. Am.*, Vol. 99, No. 6, 1996, pp. 3707-3717.

[208] T. Oostendorp, *Modelling the Fetal ECG*, Ph.D. Thesis, K.U.Nijmegen, The Netherlands, 1989.



[209] D. Otte, P. Van de Ponseele, J. Leuridan, "Operational Deflection Shapes in Multisource Environments", *Proc. 8th Int. Modal Analysis Conf.*, Kissimmee, FL, U.S.A., Jan. 29 - Feb. 1, 1990, pp. 213-220.

[210] D. Otte, "The use of SVD for the study of multivariate noise and vibration problems", in: B. De Moor, M. Moonen (Eds.), *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1995, pp. 357-366.

[211] D. Otte, "Development and Evaluation of Singular Value Analysis Methodologies for Studying Multivariate Noise and Vibration Problems", Ph.D. Thesis, K.U.Leuven, Mech. Dept., May 1994.

[212] D. Otte, J. Leuridan, M. Hermanski, H. Ostholt, "The Use of Principal Component Analysis and Operating Deflection Shapes for Noise Problem Identification on a Combine Harvester", *Proc. 15th Int. Seminar on Modal Analysis*, Leuven, Belgium, Sept. 19-21, 1990, pp. 1365-1377.

[213] P. Paatero, "A Weighted Non-Negative Least Squares Algorithm for Three-Way "PARAFAC" Factor Analysis", to appear in: *Chemom. Intell. Lab. Syst.*, Special Issue *2nd Internet Conf. in Chemometrics (INCINC'96).*

[214] D. Pal, *Balanced Stochastic Realization and Model Reduction*, Master's Thesis, Washington State University, E.E. Dept., 1982.

[215] J.P. Phelan, "Tests of Fetal Well-Being Using the Fetal Heart Rate", in: J.A.D. Spencer (Ed.), *Fetal Monitoring: Physiology and Techniques of Antenatal and Intrapartum Assessment*, Castle House Publ. Ltd., Turnbridge Wells, 1989.

[216] R. Plonsey, *Bioelectric Phenomena*, McGraw-Hill, N.Y., 1969.

[217] J.G. Proakis, *Digital Communications*, McGraw-Hill, 1989.

[218] I. Proudler, J. McWhirter, "Algorithmic Engineering in Adaptive Signal Processing II - Worked Examples", *IEE Proceedings VIS*, Vol. 141, No. 1, Feb. 1994, pp. 19-26.

[219] D.V.B. Rao, S.Y. Kung, "A State Space Approach for the 2-D Harmonic Retrieval Problem", *Proc. ICASSP-84*, pp. 4.10.1-3.

[220] W.S. Rayens, B.C. Mitchell, "Two-Factor Degeneracies and a Stabilization of PARAFAC", to appear in: *Chemom. Intell. Lab. Syst.*, Special Issue *2nd Internet Conf. in Chemometrics (INCINC'96).*

[221] P.A. Regalia, S.K. Mitra, "Kronecker Products, Unitary Matrices and Signal Processing Applications", *SIAM Rev.*, Vol. 31, No. 4, Dec. 1989, pp. 586-613.

[222] R. Roy, *ESPRIT: Estimation of Signal Parameters via Rotational Invariance Techniques*, Ph.D. Thesis, Stanford University, Aug. 1987.

[223] A. Rutherford, *Vectors, Tensors, and the Basic Equations of Fluid Mechanics*, Dover Publications Inc., N.Y., 1989.

[224] M.K. Sain, "The Growing Algebraic Presence in Systems Engineering: An Introduction", *Proc. IEEE*, Vol. 64, No. 1, Jan. 1976, pp. 96-111.

[225] E. Sanchez, B.R. Kowalski, "Tensorial Calibration: II. Second-Order Calibration", *Journal of Chemometrics*, Vol. 2, pp. 265-280, 1988.

[226] E. Sanchez, B.R. Kowalski, "Tensorial Resolution: a Direct Trilinear Decomposition", *Journal of Chemometrics*, Vol. 4, pp. 29-45, 1990.

[227] R. Sands, F. Young, "Component Models for Three-Way Data: an Alternating Least Squares Algorithm with Optimal Scaling Features", *Psychometrika*, Vol. 45, 1980, pp. 39-67.

[228] B.S. Schiffrin, D. Clement, "Routine Antepartum Fetal Heart Rate Monitoring", in: J.A.D. Spencer (Ed.), *Fetal Monitoring: Physiology and Techniques of Antenatal and Intrapartum Assessment*, Castle House Publ. Ltd., Turnbridge Wells, 1989.

[229] R.O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Trans. Antennas and Prop.*, Vol. 34, No. 3, March 1986, pp. 276-280.

[230] R.O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, Ph.D. Thesis, Stanford University, E.E. Dept., Nov. 1981.

[231] M. Schmutz, "Optimal and Suboptimal Separable Expansions for 3D-Signal Processing", *Pattern Recognition Letters*, No. 8, Nov. 1988, Elsevier Science Publishers B.V., North-Holland, pp. 217-220.

[232] T. Schreiber, M. Richter, D.T. Kaplan, "Fetal ECG Extraction with Nonlinear State Space Projections", submitted to: *IEEE Trans. Biomed. Eng.*.

[233] L. Schwartz, *Les Tenseurs*, Herman, Paris, 1975.

[234] S. Seidel et al., "Path Loss, Scattering and Multipath Cellular and Microcellular Radiotelephone", *IEEE Trans. VT*, Vol. 40, Nov. 1991, pp. 721-730.

[235] R.J. Serfling, *Approximations Theorems of Mathematical Statistics*, Wiley, 1980.



[236] I.R. Shafarevich, *Basic Algebraic Geometry*, Grundlehren der Mathematischen Wissenschaften 213, Springer-Verlag, N.Y., 1974.

[237] S. Shamsunder, G.B. Giannakis, "Modeling of Non-Gaussian Array Data using Cumulants: DOA Estimation of More Sources with Less Sensors", *Signal Processing*, No. 30, Elsevier, 1993, pp. 279-297.

[238] S. Shamsunder, *Array and Multichannel Signal Processing using Higher Order and Cyclic Statistics*, Ph.D. Thesis, University of Virginia, School of Engineering and Applied Science, Jan. 1994.

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

[240] K.C. Sharman, "Adaptive Algorithms for Estimating the Complete Covariance Eigenstructure", *Proc. ICASSP-86*, Tokyo, 1986, pp. 1401-1404.

[241] M.A. Shcherbakov, "Fast Estimation of Wiener Kernels of Nonlinear Systems in the Frequency Domain", *Proc. IEEE Signal Processing workshop on HOS*, July 21-23, 1997, Banff, Alberta, Canada, pp. 117-121.

[242] W.F. Sheppard, *From Determinant to Tensor*, Chelsea, N.Y., 1962.

[243] H. Shono et al., "Fetal Heart Rate Recorder for Long-Duration Use in Active Full Term Pregnant Women", *Obstet. Gynecol.*, Vol. 83, No. 2, Feb. 1994, pp. 301-305.

[244] D. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction", *Proc. ICASSP-94*, Vol. 4, pp. 585-588.

[245] A. Souloumiac, J.-F. Cardoso, "Performances en Séparation de Sources", *Proc. GRETSI*, Juan les Pins, France, 1993, pp. 321-324.

[246] A. Souloumiac, J.-F. Cardoso, *Jacobi Angles for Simultaneous Diagonalization*, SIAM J. Matr. Anal. Appl., Vol. 17, No. 1, Jan. 1996, pp. 161-164.

[247] E. Sourouchyari, "Blind Separation of Sources, Part III: Stability Analysis", *Signal Processing*, No. 24, 1991, pp. 21-29.

[248] M. Spivak, *Differential Geometry*, Publish or Perish, Berkeley, 1970.

[249] G.W. Stewart, "Error and Perturbation Bounds for Subspaces Associated with Certain Eigenvalue Problems", *SIAM Rev.*, Vol. 15, No. 4, Oct. 1973, pp. 727-764.

[250] G.W. Stewart, "On the Early History of the Singular Value Decomposition", *SIAM Rev.*, Vol. 35, No. 4, Dec. 1993, pp. 551-566.

[251] S. Swales, M. Beach, D. Edwards, J. McGeehan, "The Performance Enhancement of Multibeam Adaptive Base-Station Antennas for Cellular Land Mobile Radio Systems", *IEEE Trans. VT*, Vol. 39, No. 1, Feb. 1990, pp. 56-67.

[252] A. Swami, G. Giannakis, "Editorial. Higher-Order Statistics", *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 53, Nos. 2-3, Sept. 1996, pp. 89-91.

[253] S. Talwar, M. Viberg, A. Paulraj, "Blind Estimation of Multiple Co-Channel Digital Signals Arriving at an Antenna Array", *Proc. 27th Asilomar Conf. Signals, Syst. Comp.*, IEEE, Vol. 1, 1993, pp. 349-353.

[254] J. ten Berghe, J. de Leeuw, P.M. Kroonenberg, "Some Additional Results on Principal Components Analysis of Three-Mode Data by means of Alternating Least Squares Algorithms", *Psychometrika*, No. 52, 1987, pp. 183-191.

[255] T.N. Thiele, "The Theory of Observations", *Ann. Math. Statist.*, No. 2, 1931, pp. 165-308.

[256] N. Thirion, *Séparation d'Ondes en Prospection Sismique*, Ph.D. Thesis, CEPHAG, Grenoble, France, Sept. 1995.

[257] L. Tong, R. Liu, V. Soon, Y.-F. Huang, "Indeterminacy and Identifiability of Blind Identification", *IEEE Trans. Circuits and Systems*, Vol. 38, No. 5, May 1991, pp. 499-509.

[258] L. Tong, G. Xu, T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: a Time Domain Approach", *IEEE Trans. Informat. Th.*, Vol. 40, March 1994, pp. 340-349.

[259] L.R. Tucker, "The Extension of Factor Analysis to Three-Dimensional Matrices", in: H. Gulliksen, N. Frederiksen (Eds.), *Contributions to mathematical psychology*, Holt, Rinehart & Winston, N.Y., 1964, pp. 109-127.

[260] L.R. Tucker, "Some Mathematical Notes on Three-Mode Factor Analysis", *Psychometrika*, No. 31, 1966, pp. 279-311.

[261] R. Vaccaro, D. Tufts, G.F. Boudreaux-Bartels, "Advances in Principal Component Analysis", in: E. Deprettere (Ed.), *SVD and Signal Processing. Algorithms, Applications and Architectures*, Elsevier Science Publishers B.V., North-Holland, 1988, pp. 115-148.

[262] R. Vaccaro (Ed.), *SVD and Signal Processing, II. Algorithms, Applications and Architectures*, Elsevier, Amsterdam, 1991.



[263] J. Vanderschoot, D. Callaerts, W. Sansen, J. Vandewalle, G. Vantrappen, J. Janssens, "Two Methods for Optimal MECG Elimination and FECG Detection from Skin Electrode Signals", *IEEE Trans. Biomed. Eng.*, Vol. 34, No. 3, March 1987, pp. 233-243.

[264] J. Vanderschoot, G. Vantrappen, J. Janssens, J. Vandewalle, W. Sansen, "A Reliable Method for Fetal ECG Extraction from Abdominal Recordings", in: F.H. Roger et al. (Eds.), *Medical Informatics Europe 84*, Lecture Notes in Medical Informatics, Vol. 24, Springer Verlag, Berlin,1984, pp. 249-254.

[265] A.-J. van der Veen, A. Paulraj, "A Constant Modulus Factorization Technique for Smart Antenna Applications in Mobile Communications", *Proc. SPIE'94*, Vol. 2296, pp. 230-241.

[266] A.-J. van der Veen, A. Paulraj, "An Analytical Constant Modulus Algorithm", *IEEE Trans. Signal Processing*, Vol. 44, No. 5, May 1996, pp. 1136-1155.

[267] A.-J. van der Veen, S. Talwar, A. Paulraj, "Blind Estimation of Multiple Digital Signals Transmitted over FIR Channels", *IEEE Signal Processing Letters*, Vol. 2, May 1995, pp. 99-102.

[268] J. Vandewalle, D. Callaerts, "Singular Value Decomposition: a Powerful Concept and Tool in Signal Processing", *Mathematics in Signal Processing II*, 1990, pp. 539-559.

[269] J. Vandewalle, B. De Moor, "On the Use of the Singular Value Decomposition in Identification and Signal Processing", *NATO ASI Series: Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*, Vol. F70, 1991, pp. 321-360.

[270] H.P. van Geijn, F.J.A. Copray (Eds.), *A Critical Appraisal of Fetal Surveillance*, Section IV - Fetal Heart Rate Monitoring, Elsevier, 1994.

[271] P. Van Overschee, B. De Moor, "Subspace Algorithms for the Stochastic Identification Problem", *Automatica*, Vol. 29, No. 3, 1993, pp. 649-660.

[272] P. Van Overschee, *Subspace Identification. Theory - Implementation - Application*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., Feb. 1995.

[273] F. Vanpoucke, *Algorithms and Architectures for Adaptive Array Signal Processing*, Ph.D. Thesis, K.U.Leuven, E.E. Dept., Feb. 1995.

[274] B.D. Van Veen, K.M. Buckley, "Beamforming: a Versatile Approach to Spatial Filtering", *IEEE ASSP Magazine*, April 1988, pp. 4-24.

[275] P. Vary, L. Kittel, "Technical Trends in Digital Mobile Radio", in: *Proc. ProRISC IEEE Workshop on Circuits, Systems and Signal Processing*, Houthalen, Belgium, March 24-25, 1993, pp. 23-34.

[276] J. Verschelde, R. Cools, "An Ada Workbench for Solving Polynomial Systems by Homotopy Continuation Methods", *Ada-Belgium Newsletter*, Vol. 2, No. 1, Feb. 1993, pp. 23-40.

[277] D.L. Wallace, "Asymptotic Approximations to Distributions", *Annals Math. Stat.*, Vol. 29, 1958, pp. 635-654.

[278] B. Widrow, S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[279] T.J. Willmore, *Total Curvature in Riemannian Geometry*, Ellis Horwood, Chichester, 1982.

[280] Q.-S. Zheng, A.J.M. Spencer, "Tensors which Characterize Anisotropies", *Int. Journal of Engineering Science*, Vol. 31, No. 5, May 1993, pp. 679-695.

Lieven De Lathauwer was born on November 10, 1969 in Aalst, Belgium. He received the degree of electrotechnical-mechanical engineering, specialization control engineering, from the Katholieke Universiteit Leuven, Belgium in July 1992. Since then he is working as a researcher at the Electrical Engineering Department E.S.A.T. of the K.U.Leuven, first on an I.W.O.N.L. grant (Belgian Institute for the Advancement of Scientific Research in Industry and Agriculture), from October 1994 on an I.W.T. grant (Flemish Institute for Scientific and Technological Research in Industry), and from October 1996 on a K.U.Leuven Ph.D. grant.