

Tensor Networks for Dimensionality Reduction and Large-Scale Optimization Part 1 Low-Rank Tensor Decompositions

Andrzej Cichocki

Namgil Lee

Ivan Oseledets

Anh-Huy Phan

Qibin Zhao

Danilo P. Mandic

now

the essence of knowledge

Boston — Delft

Foundations and Trends[®] in Machine Learning

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

A. Cichocki *et al.*. *Tensor Networks for Dimensionality Reduction and Large-Scale Optimization*

Part 1 Low-Rank Tensor Decompositions. Foundations and Trends[®] in Machine Learning, vol. 9, no. 4-5, pp. 249–429, 2016.

This Foundations and Trends[®] issue was typeset in L^AT_EX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-223-5
© 2017 A. Cichocki *et al.*

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends[®] in Machine Learning
Volume 9, Issue 4-5, 2016
Editorial Board

Editor-in-Chief

Michael Jordan
University of California, Berkeley
United States

Editors

Peter Bartlett <i>UC Berkeley</i>	Geoffrey Hinton <i>University of Toronto</i>	Andrew Moore <i>CMU</i>
Yoshua Bengio <i>University of Montreal</i>	Aapo Hyvarinen <i>HIIT, Finland</i>	John Platt <i>Microsoft Research</i>
Avrim Blum <i>CMU</i>	Leslie Pack Kaelbling <i>MIT</i>	Luc de Raedt <i>University of Freiburg</i>
Craig Boutilier <i>University of Toronto</i>	Michael Kearns <i>UPenn</i>	Christian Robert <i>U Paris-Dauphine</i>
Stephen Boyd <i>Stanford University</i>	Daphne Koller <i>Stanford University</i>	Sunita Sarawagi <i>IIT Bombay</i>
Carla Brodley <i>Tufts University</i>	John Lafferty <i>University of Chicago</i>	Robert Schapire <i>Princeton University</i>
Inderjit Dhillon <i>UT Austin</i>	Michael Littman <i>Brown University</i>	Bernhard Schoelkopf <i>MPI Tübingen</i>
Jerome Friedman <i>Stanford University</i>	Gabor Lugosi <i>Pompeu Fabra University</i>	Richard Sutton <i>University of Alberta</i>
Kenji Fukumizu <i>ISM, Japan</i>	David Madigan <i>Columbia University</i>	Larry Wasserman <i>CMU</i>
Zoubin Ghahramani <i>University of Cambridge</i>	Pascal Massart <i>University of Paris-Sud</i>	Bin Yu <i>UC Berkeley</i>
David Heckerman <i>Microsoft Research</i>	Andrew McCallum <i>UMass Amherst</i>	
Tom Heskes <i>Radboud University</i>	Marina Meila <i>University of Washington</i>	

Editorial Scope

Topics

Foundations and Trends[®] in Machine Learning publishes survey and tutorial articles on the theory, algorithms and applications of machine learning, including the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive, and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends[®] in Machine Learning, 2016, Volume 9, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends[®] in Machine Learning
Vol. 9, No. 4-5 (2016) 249–429
© 2017 A. Cichocki *et al.*
DOI: 10.1561/22000000059



Tensor Networks for Dimensionality Reduction and Large-Scale Optimization

Part 1 Low-Rank Tensor Decompositions

Andrzej Cichocki

RIKEN Brain Science Institute (BSI), Japan and
Skolkovo Institute of Science and Technology (SKOLTECH)
a.cichocki@riken.jp

Namgil Lee

RIKEN BSI, namgil.lee@riken.jp

Ivan Oseledets

Skolkovo Institute of Science and Technology (SKOLTECH) and
Institute of Numerical Mathematics of Russian Academy of Sciences
i.oseledets@skolkovotech.ru

Anh-Huy Phan

RIKEN BSI, phan@brain.riken.jp

Qibin Zhao

RIKEN BSI, qbzha@brain.riken.jp

Danilo P. Mandic

Department of Electrical and Electronic Engineering
Imperial College London
d.mandic@imperial.ac.uk

Contents

1	Introduction and Motivation	2
1.1	Challenges in Big Data Processing	3
1.2	Tensor Notations and Graphical Representations	4
1.3	Curse of Dimensionality and Generalized Separation of Variables for Multivariate Functions	12
1.4	Advantages of Multiway Analysis via Tensor Networks	20
1.5	Scope and Objectives	21
2	Tensor Operations and Tensor Network Diagrams	24
2.1	Basic Multilinear Operations	24
2.2	Graphical Representation of Fundamental Tensor Networks	44
2.3	Generalized Tensor Network Formats	62
3	Constrained Tensor Decompositions: From Two-way to Multiway Component Analysis	66
3.1	Constrained Low-Rank Matrix Factorizations	66
3.2	The CP Format	69
3.3	The Tucker Tensor Format	75
3.4	Higher Order SVD (HOSVD) for Large-Scale Problems	84
3.5	Tensor Sketching Using Tucker Model	94
3.6	Multiway Component Analysis (MWCA)	103

3.7	Nonlinear Tensor Decompositions – Infinite Tucker	112
4	Tensor Train Decompositions: Graphical Interpretations and Algorithms	115
4.1	Tensor Train Decomposition – Matrix Product State	115
4.2	Matrix TT Decomposition – Matrix Product Operator	121
4.3	Links Between CP, BTD Formats and TT/TC Formats	126
4.4	Quantized Tensor Train (QTT) – Blessing of Dimensionality	129
4.5	Basic Operations in TT Formats	135
4.6	Algorithms for TT Decompositions	145
5	Discussion and Conclusions	159
	Acknowledgements	161
	References	162

Abstract

Modern applications in engineering and data science are increasingly based on multidimensional data of exceedingly high volume, variety, and structural richness. However, standard machine learning algorithms typically scale exponentially with data volume and complexity of cross-modal couplings - the so called curse of dimensionality - which is prohibitive to the analysis of large-scale, multi-modal and multi-relational datasets. Given that such data are often efficiently represented as multiway arrays or tensors, it is therefore timely and valuable for the multidisciplinary machine learning and data analytic communities to review low-rank tensor decompositions and tensor networks as emerging tools for dimensionality reduction and large scale optimization problems. Our particular emphasis is on elucidating that, by virtue of the underlying low-rank approximations, tensor networks have the ability to alleviate the curse of dimensionality in a number of applied areas. In Part 1 of this monograph we provide innovative solutions to low-rank tensor network decompositions and easy to interpret graphical representations of the mathematical operations on tensor networks. Such a conceptual insight allows for seamless migration of ideas from the flat-view matrices to tensor network operations and vice versa, and provides a platform for further developments, practical applications, and non-Euclidean extensions. It also permits the introduction of various tensor network operations without an explicit notion of mathematical expressions, which may be beneficial for many research communities that do not directly rely on multilinear algebra. Our focus is on the Tucker and tensor train (TT) decompositions and their extensions, and on demonstrating the ability of tensor networks to provide linearly or even super-linearly (e.g., logarithmically) scalable solutions, as illustrated in detail in Part 2 of this monograph.

1

Introduction and Motivation

This monograph aims to present a coherent account of ideas and methodologies related to tensor decompositions (TDs) and tensor networks models (TNs). Tensor decompositions (TDs) decompose complex data tensors of exceedingly high dimensionality into their factor (component) tensors and matrices, while tensor networks (TNs) decompose higher-order tensors into sparsely interconnected small-scale factor matrices and/or low-order core tensors. These low-order core tensors are called “components”, “blocks”, “factors” or simply “cores”. In this way, large-scale data can be approximately represented in highly compressed and distributed formats.

In this monograph, the TDs and TNs are treated in a unified way, by considering TDs as simple tensor networks or sub-networks; the terms “tensor decompositions” and “tensor networks” will therefore be used interchangeably. Tensor networks can be thought of as special graph structures which break down high-order tensors into a set of sparsely interconnected low-order core tensors, thus allowing for both enhanced interpretation and computational advantages. Such an approach is valuable in many application contexts which require the computation of eigenvalues and the corresponding eigenvectors of extremely high-dimensional linear or nonlinear operators. These operators typically describe the coupling between many degrees of freedom within real-world physical systems; such degrees of freedom are often only weakly coupled. Indeed, quantum physics provides evidence that couplings between multiple data channels usually do not exist among all

the degrees of freedom but mostly locally, whereby “relevant” information, of relatively low-dimensionality, is embedded into very large-dimensional measurements (Verstraete *et al.*, 2008; Schollwöck, 2013; Orús, 2014; Murg *et al.*, 2015).

Tensor networks offer a theoretical and computational framework for the analysis of computationally prohibitive large volumes of data, by “dissecting” such data into the “relevant” and “irrelevant” information, both of lower dimensionality. In this way, tensor network representations often allow for super-compression of datasets as large as 10^{50} entries, down to the affordable levels of 10^7 or even less entries (Osleedets and Tyrtysnikov, 2009; Dolgov and Khoromskij, 2013; Kazeev *et al.*, 2013a, 2014; Kressner *et al.*, 2014a; Vervliet *et al.*, 2014; Dolgov and Khoromskij, 2015; Liao *et al.*, 2015; Bolten *et al.*, 2016).

With the emergence of the big data paradigm, it is therefore both timely and important to provide the multidisciplinary machine learning and data analytic communities with a comprehensive overview of tensor networks, together with an example-rich guidance on their application in several generic optimization problems for huge-scale structured data. Our aim is also to unify the terminology, notation, and algorithms for tensor decompositions and tensor networks which are being developed not only in machine learning, signal processing, numerical analysis and scientific computing, but also in quantum physics/chemistry for the representation of, e.g., quantum many-body systems.

1.1 Challenges in Big Data Processing

The volume and structural complexity of modern datasets are becoming exceedingly high, to the extent which renders standard analysis methods and algorithms inadequate. Apart from the huge Volume, the other features which characterize big data include Veracity, Variety and Velocity (see Figures 1.1(a) and (b)). Each of the “V features” represents a research challenge in its own right. For example, high volume implies the need for algorithms that are scalable; high Velocity requires the processing of big data streams in near real-time; high Veracity calls for robust and predictive algorithms for noisy, incomplete

and/or inconsistent data; high Variety demands the fusion of different data types, e.g., continuous, discrete, binary, time series, images, video, text, probabilistic or multi-view. Some applications give rise to additional “V challenges”, such as Visualization, Variability and Value. The Value feature is particularly interesting and refers to the extraction of high quality and consistent information, from which meaningful and interpretable results can be obtained.

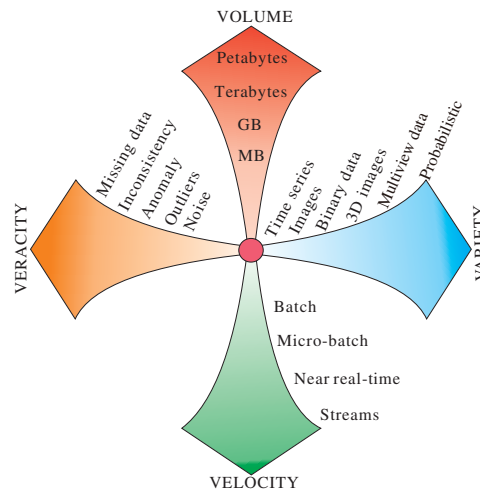
Owing to the increasingly affordable recording devices, extreme-scale volumes and variety of data are becoming ubiquitous across the science and engineering disciplines. In the case of multimedia (speech, video), remote sensing and medical/biological data, the analysis also requires a paradigm shift in order to efficiently process massive datasets within tolerable time (velocity). Such massive datasets may have billions of entries and are typically represented in the form of huge block matrices and/or tensors. This has spurred a renewed interest in the development of matrix/tensor algorithms that are suitable for very large-scale datasets. We show that tensor networks provide a natural sparse and distributed representation for big data, and address both established and emerging methodologies for tensor-based representations and optimization. Our particular focus is on low-rank tensor network representations, which allow for huge data tensors to be approximated (compressed) by interconnected low-order core tensors.

1.2 Tensor Notations and Graphical Representations

Tensors are multi-dimensional generalizations of matrices. A matrix (2nd-order tensor) has two modes, rows and columns, while an N th-order tensor has N modes (see Figures 1.2–1.7); for example, a 3rd-order tensor (with three-modes) looks like a cube (see Figure 1.2). Subtensors are formed when a subset of tensor indices is fixed. Of particular interest are *fibers* which are vectors obtained by fixing every tensor index but one, and *matrix slices* which are two-dimensional sections (matrices) of a tensor, obtained by fixing all the tensor indices but two. It should be noted that block matrices can also be represented by tensors, as illustrated in Figure 1.3 for 4th-order tensors.

1.2. Tensor Notations and Graphical Representations

(a)



(b)



Figure 1.1: A framework for extremely large-scale data analysis. (a) The 4V challenges for big data. (b) A unified framework for the 4V challenges and the potential applications based on tensor decomposition approaches.

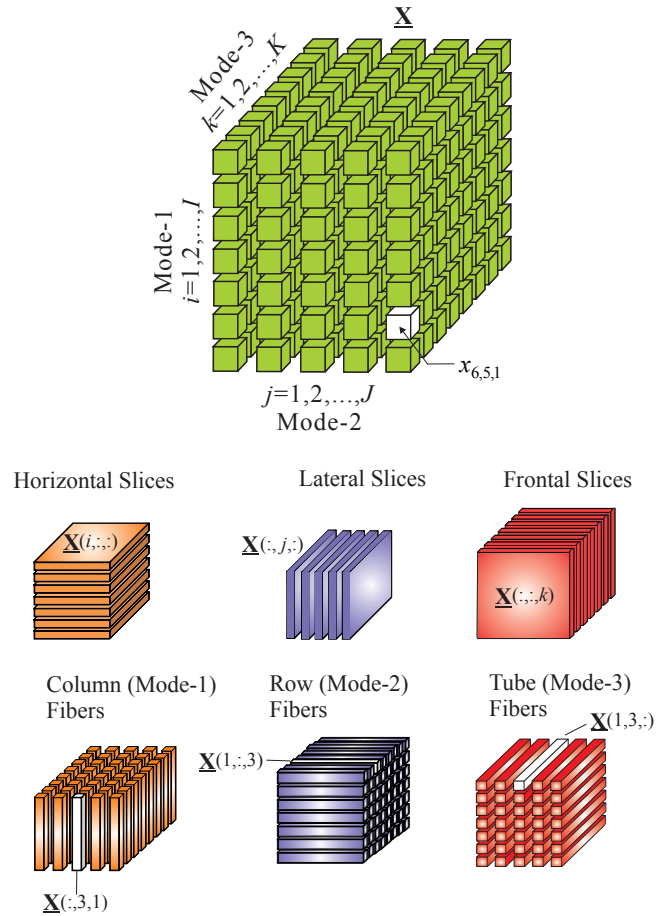


Figure 1.2: A 3rd-order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$, with entries $x_{i,j,k} = \underline{\mathbf{X}}(i, j, k)$, and its subtensors: slices (middle) and fibers (bottom). All fibers are treated as column vectors.

We adopt the notation whereby tensors (for $N \geq 3$) are denoted by bold underlined capital letters, e.g., $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. For simplicity, we assume that all tensors are real-valued, but it is, of course, possible to define tensors as complex-valued or over arbitrary fields. Matrices are denoted by boldface capital letters, e.g., $\mathbf{X} \in \mathbb{R}^{I \times J}$, and vectors (1st-order tensors) by boldface lower case letters, e.g., $\mathbf{x} \in \mathbb{R}^J$. For example, the columns of the matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R] \in \mathbb{R}^{I \times R}$ are

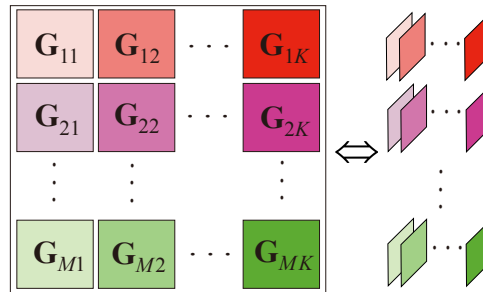


Figure 1.3: A block matrix and its representation as a 4th-order tensor, created by reshaping (or a projection) of blocks in the rows into lateral slices of 3rd-order tensors.

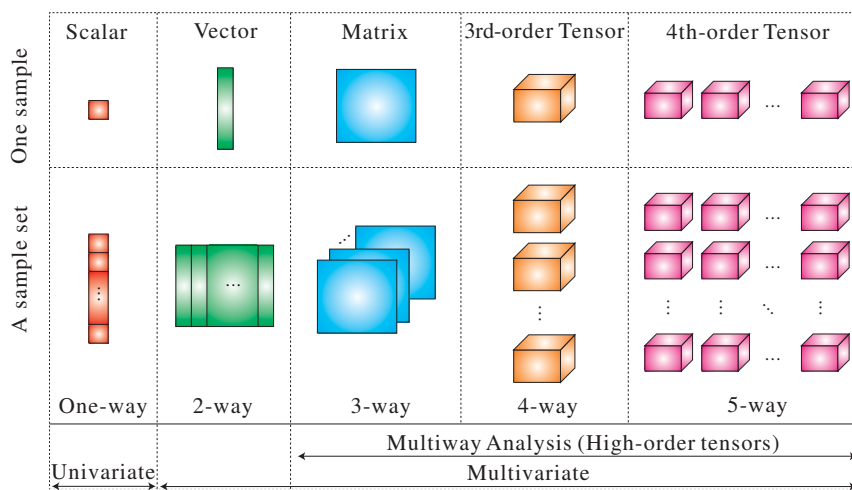
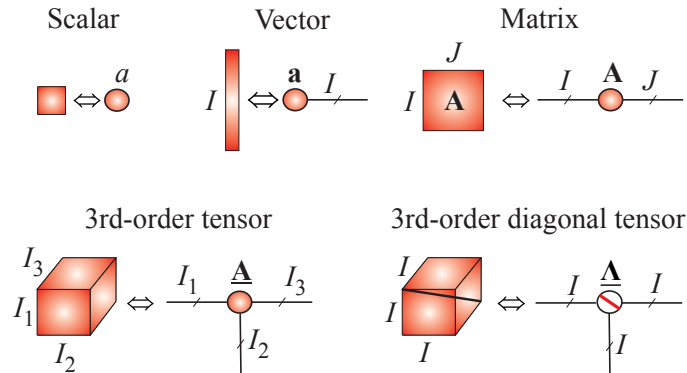


Figure 1.4: Graphical representation of multiway array (tensor) data of increasing structural complexity and “Volume” (see (Olivieri, 2008) for more detail).

the vectors denoted by $\mathbf{a}_r \in \mathbb{R}^I$, while the elements of a matrix (scalars) are denoted by lowercase letters, e.g., $a_{ir} = \mathbf{A}(i, r)$ (see Table 1.1).

A specific entry of an N th-order tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $x_{i_1, i_2, \dots, i_N} = \mathbf{X}(i_1, i_2, \dots, i_N) \in \mathbb{R}$. The order of a tensor is the number of its “modes”, “ways” or “dimensions”, which can include space, time, frequency, trials, classes, and dictionaries. The term “size” stands for the number of values that an index can take in a particular

(a)



(b)

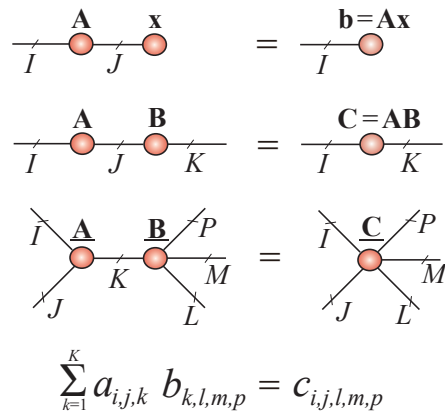


Figure 1.5: Graphical representation of tensor manipulations. (a) Basic building blocks for tensor network diagrams. (b) Tensor network diagrams for matrix-vector multiplication (top), matrix by matrix multiplication (middle) and contraction of two tensors (bottom). The order of reading of indices is anti-clockwise, from the left position.

mode. For example, the tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is of order N and size I_n in all modes- n ($n = 1, 2, \dots, N$). Lower-case letters e.g, i, j are used for the subscripts in running indices and capital letters I, J denote the upper bound of an index, i.e., $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$. For a positive integer n , the shorthand notation $\langle n \rangle$ denotes the set of indices $\{1, 2, \dots, n\}$.

Table 1.1: Basic matrix/tensor notation and symbols.

$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$	N th-order tensor of size $I_1 \times I_2 \times \cdots \times I_N$
$x_{i_1, i_2, \dots, i_N} = \underline{\mathbf{X}}(i_1, i_2, \dots, i_N)$	(i_1, i_2, \dots, i_N) th entry of $\underline{\mathbf{X}}$
$x, \mathbf{x}, \mathbf{X}$	scalar, vector and matrix
$\underline{\mathbf{G}}, \underline{\mathbf{S}}, \underline{\mathbf{G}}^{(n)}, \underline{\mathbf{X}}^{(n)}$	core tensors
$\underline{\Lambda} \in \mathbb{R}^{R \times R \times \cdots \times R}$	N th-order diagonal core tensor with nonzero entries λ_r on the main diagonal
$\mathbf{A}^T, \mathbf{A}^{-1}, \mathbf{A}^\dagger$	transpose, inverse and Moore–Penrose pseudo-inverse of a matrix \mathbf{A}
$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R] \in \mathbb{R}^{I \times R}$	matrix with R column vectors $\mathbf{a}_r \in \mathbb{R}^I$, with entries a_{ir}
$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{A}^{(n)}, \mathbf{B}^{(n)}, \mathbf{U}^{(n)}$	component (factor) matrices
$\underline{\mathbf{X}}_{(n)} \in \mathbb{R}^{I_n \times I_1 \cdots I_{n-1} I_{n+1} \cdots I_N}$	mode- n matricization of $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$
$\underline{\mathbf{X}}_{\langle n \rangle} \in \mathbb{R}^{I_1 I_2 \cdots I_n \times I_{n+1} \cdots I_N}$	mode- $(1, \dots, n)$ matricization of $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$
$\underline{\mathbf{X}}(:, i_2, i_3, \dots, i_N) \in \mathbb{R}^{I_1}$	mode-1 fiber of a tensor $\underline{\mathbf{X}}$ obtained by fixing all indices but one (a vector)
$\underline{\mathbf{X}}(:, :, i_3, \dots, i_N) \in \mathbb{R}^{I_1 \times I_2}$	slice (matrix) of a tensor $\underline{\mathbf{X}}$ obtained by fixing all indices but two
$\underline{\mathbf{X}}(:, :, :, i_4, \dots, i_N)$	subtensor of $\underline{\mathbf{X}}$, obtained by fixing several indices
$R, (R_1, \dots, R_N)$	tensor rank R and multilinear rank
\circ, \odot, \otimes	outer, Khatri–Rao, Kronecker products
\otimes_L, \boxtimes	Left Kronecker, strong Kronecker products
$\mathbf{x} = \text{vec}(\underline{\mathbf{X}})$	vectorization of $\underline{\mathbf{X}}$
$\text{tr}(\bullet)$	trace of a square matrix
$\text{diag}(\bullet)$	diagonal matrix

Table 1.2: Terminology used for tensor networks across the machine learning/scientific computing and quantum physics/chemistry communities.

Machine Learning	Quantum Physics
N th-order tensor	rank- N tensor
high/low-order tensor	tensor of high/low dimension
ranks of TNs	bond dimensions of TNs
unfolding, matricization	grouping of indices
tensorization	splitting of indices
core	site
variables	open (physical) indices
ALS Algorithm	one-site DMRG or DMRG1
MALS Algorithm	two-site DMRG or DMRG2
column vector $\mathbf{x} \in \mathbb{R}^{I \times 1}$	ket $ \Psi\rangle$
row vector $\mathbf{x}^T \in \mathbb{R}^{1 \times I}$	bra $\langle\Psi $
inner product $\langle\mathbf{x}, \mathbf{x}\rangle = \mathbf{x}^T \mathbf{x}$	$\langle\Psi \Psi\rangle$
Tensor Train (TT)	Matrix Product State (MPS) (with Open Boundary Conditions (OBC))
Tensor Chain (TC)	MPS with Periodic Boundary Conditions (PBC)
Matrix TT	Matrix Product Operators (with OBC)
Hierarchical Tucker (HT)	Tree Tensor Network State (TTNS) with rank-3 tensors

Notations and terminology used for tensors and tensor networks differ across the scientific communities (see Table 1.2); to this end we employ a unifying notation particularly suitable for machine learning and signal processing research, which is summarized in Table 1.1.

Even with the above notation conventions, a precise description of tensors and tensor operations is often tedious and cumbersome, given the multitude of indices involved. To this end, in this monograph, we grossly simplify the description of tensors and their mathematical operations through diagrammatic representations borrowed from physics and quantum chemistry (see (Orús, 2014) and references therein). In this way, tensors are represented graphically by nodes of any geometrical shapes (e.g., circles, squares, dots), while each outgoing line (“edge”, “leg”, “arm”) from a node represents the indices of a specific mode (see Figure 1.5(a)). In our adopted notation, each scalar (zero-order tensor), vector (first-order tensor), matrix (2nd-order tensor), 3rd-order tensor or higher-order tensor is represented by a circle (or rectangular), while the order of a tensor is determined by the number of lines (edges) connected to it. According to this notation, an N th-order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ is represented by a circle (or any shape) with N branches each of size I_n , $n = 1, 2, \dots, N$ (see Section 2). An interconnection between two circles designates a contraction of tensors, which is a summation of products over a common index (see Figure 1.5(b) and Section 2).

Block tensors, where each entry (e.g., of a matrix or a vector) is an individual subtensor, can be represented in a similar graphical form, as illustrated in Figure 1.6. Hierarchical (multilevel block) matrices are also naturally represented by tensors and vice versa, as illustrated in Figure 1.7 for 4th-, 5th- and 6th-order tensors. All mathematical operations on tensors can be therefore equally performed on block matrices.

In this monograph, we make extensive use of tensor network diagrams as an intuitive and visual way to efficiently represent tensor decompositions. Such graphical notations are of great help in studying and implementing sophisticated tensor operations. We highlight the significant advantages of such diagrammatic notations in the description of tensor manipulations, and show that most tensor operations can

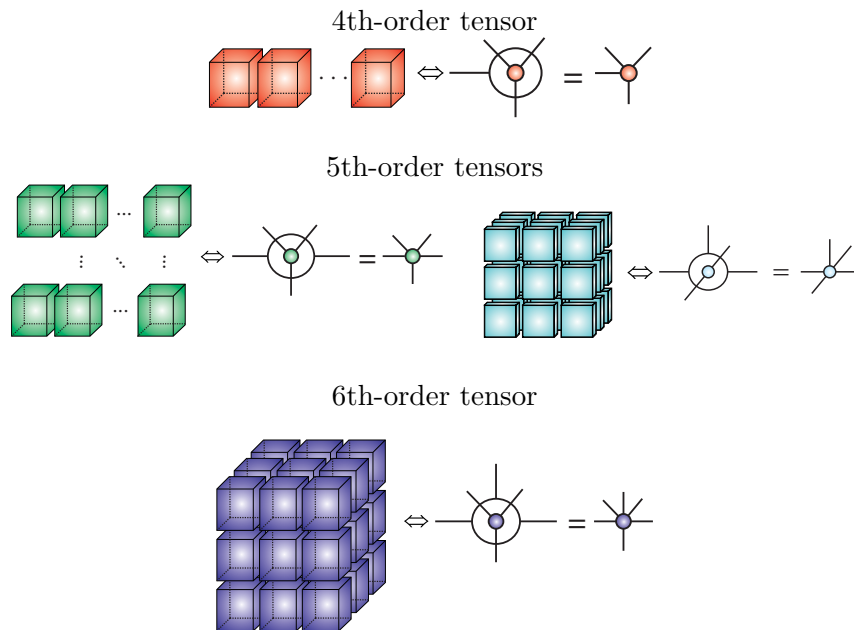


Figure 1.6: Graphical representations and symbols for higher-order block tensors. Each block represents either a 3rd-order tensor or a 2nd-order tensor. The outer circle indicates a global structure of the block tensor (e.g. a vector, a matrix, a 3rd-order block tensor), while the inner circle reflects the structure of each element within the block tensor. For example, in the top diagram a vector of 3rd order tensors is represented by an outer circle with one edge (a vector) which surrounds an inner circle with three edges (a 3rd order tensor), so that the whole structure designates a 4th-order tensor.

be visualized through changes in the architecture of a tensor network diagram.

1.3 Curse of Dimensionality and Generalized Separation of Variables for Multivariate Functions

1.3.1 Curse of Dimensionality

The term *curse of dimensionality* was coined by [Bellman \(1961\)](#) to indicate that the number of samples needed to estimate an arbitrary function with a given level of accuracy grows exponentially with the

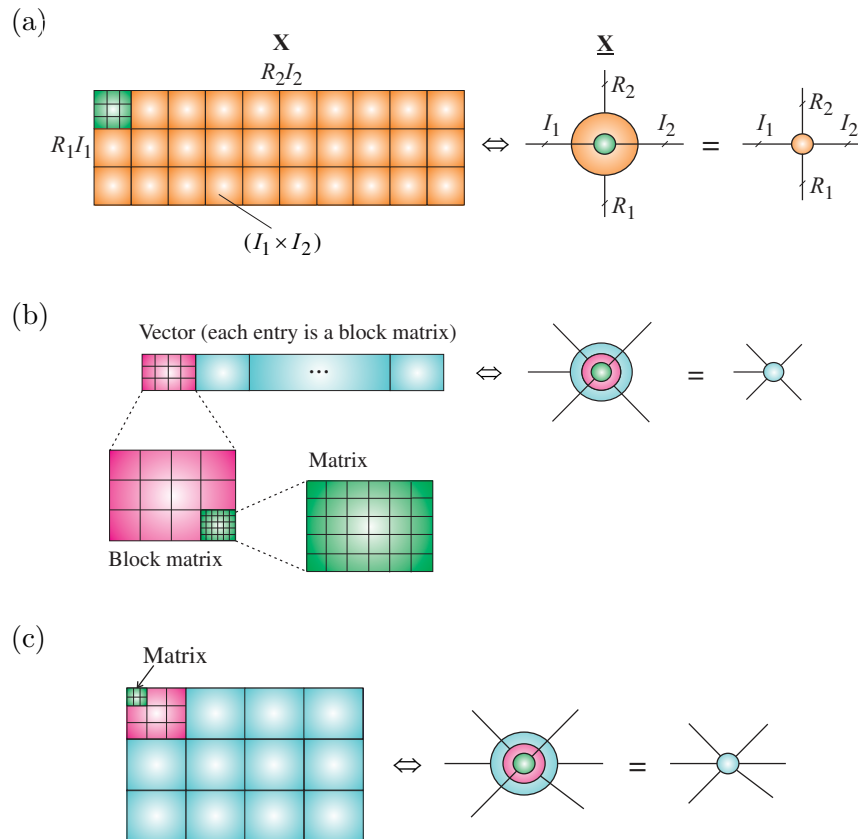


Figure 1.7: Hierarchical matrix structures and their symbolic representation as tensors. (a) A 4th-order tensor representation for a block matrix $\mathbf{X} \in \mathbb{R}^{R_1 I_1 \times R_2 I_2}$ (a matrix of matrices), which comprises block matrices $\mathbf{X}_{r_1, r_2} \in \mathbb{R}^{I_1 \times I_2}$. (b) A 5th-order tensor. (c) A 6th-order tensor.

number of variables, that is, with the dimensionality of the function. In a general context of machine learning and the underlying optimization problems, the “curse of dimensionality” may also refer to an exponentially increasing number of parameters required to describe the data/system or an extremely large number of degrees of freedom. The term “curse of dimensionality”, in the context of tensors, refers to the

phenomenon whereby the number of elements, I^N , of an N th-order tensor of size $(I \times I \times \dots \times I)$ grows exponentially with the tensor order, N . Tensor volume can therefore easily become prohibitively big for multi-way arrays for which the number of dimensions (“ways” or “modes”) is very high, thus requiring enormous computational and memory resources to process such data. The understanding and handling of the inherent dependencies among the excessive degrees of freedom create both difficult to solve problems and fascinating new opportunities, but comes at a price of reduced accuracy, owing to the necessity to involve various approximations.

We show that the curse of dimensionality can be alleviated or even fully dealt with through tensor network representations; these naturally cater for the excessive volume, veracity and variety of data (see Figure 1.1) and are supported by efficient tensor decomposition algorithms which involve relatively simple mathematical operations. Another desirable aspect of tensor networks is their relatively small-scale and low-order *core tensors*, which act as “building blocks” of tensor networks. These core tensors are relatively easy to handle and visualize, and enable super-compression of the raw, incomplete, and noisy huge-scale datasets. This also suggests a solution to a more general quest for new technologies for processing of exceedingly large datasets within affordable computation times.

To address the curse of dimensionality, this work mostly focuses on approximative low-rank representations of tensors, the so-called low-rank tensor approximations (LRTA) or low-rank tensor network decompositions.

1.3.2 Separation of Variables and Tensor Formats

A tensor is said to be in a *full format* when it is represented as an original (raw) multidimensional array (Klus and Schütte, 2015), however, distributed storage and processing of high-order tensors in their full format is infeasible due to the curse of dimensionality. The *sparse format* is a variant of the full tensor format which stores only the nonzero entries of a tensor, and is used extensively in software tools such as the

Tensor Toolbox (Bader and Kolda, 2015) and in the sparse grid approach (Garcke *et al.*, 2001; Bungartz and Griebel, 2004; Hackbusch, 2012).

As already mentioned, the problem of huge dimensionality can be alleviated through various distributed and compressed tensor network formats, achieved by low-rank tensor network approximations. The underpinning idea is that by employing tensor networks formats, both computational costs and storage requirements may be dramatically reduced through distributed storage and computing resources. It is important to note that, except for very special data structures, a tensor cannot be compressed without incurring some compression error, since a low-rank tensor representation is only an approximation of the original tensor.

The concept of compression of multidimensional large-scale data by tensor network decompositions can be intuitively explained as follows. Consider the approximation of an N -variate function $f(\mathbf{x}) = f(x_1, x_2, \dots, x_N)$ by a finite sum of products of individual functions, each depending on only one or a very few variables (Bebendorf, 2011; Dolgov, 2014; Cho *et al.*, 2016; Trefethen, 2017). In the simplest scenario, the function $f(\mathbf{x})$ can be (approximately) represented in the following separable form

$$f(x_1, x_2, \dots, x_N) \cong f^{(1)}(x_1)f^{(2)}(x_2) \cdots f^{(N)}(x_N). \quad (1.1)$$

In practice, when an N -variate function $f(\mathbf{x})$ is discretized into an N th-order array, or a tensor, the approximation in (1.1) then corresponds to the representation by rank-1 tensors, also called elementary tensors (see Section 2). Observe that with I_n , $n = 1, 2, \dots, N$ denoting the size of each mode and $I = \max_n \{I_n\}$, the memory requirement to store such a full tensor is $\prod_{n=1}^N I_n \leq I^N$, which grows exponentially with N . On the other hand, the separable representation in (1.1) is completely defined by its factors, $f^{(n)}(x_n)$, ($n = 1, 2, \dots, N$), and requires only $\sum_{n=1}^N I_n \ll I^N$ storage units. If x_1, x_2, \dots, x_N are statistically independent random variables, their joint probability density function is equal to the product of marginal probabilities, $f(\mathbf{x}) = f^{(1)}(x_1)f^{(2)}(x_2) \cdots f^{(N)}(x_N)$, in an exact analogy to outer products of elementary tensors. Unfortunately, the form of separability in (1.1) is rather rare in practice.

The concept of tensor networks rests upon generalized (full or partial) separability of the variables of a high dimensional function. This can be achieved in different tensor formats, including:

- The Canonical Polyadic (CP) format (see Section 3.2), where

$$f(x_1, x_2, \dots, x_N) \cong \sum_{r=1}^R f_r^{(1)}(x_1) f_r^{(2)}(x_2) \cdots f_r^{(N)}(x_N), \quad (1.2)$$

in an exact analogy to (1.1). In a discretized form, the above CP format can be written as an N th-order tensor

$$\underline{\mathbf{F}} \cong \sum_{r=1}^R \mathbf{f}_r^{(1)} \circ \mathbf{f}_r^{(2)} \circ \cdots \circ \mathbf{f}_r^{(N)} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, \quad (1.3)$$

where $\mathbf{f}_r^{(n)} \in \mathbb{R}^{I_n}$ denotes a discretized version of the univariate function $f_r^{(n)}(x_n)$, symbol \circ denotes the outer product, and R is the tensor rank.

- The Tucker format, given by

$$f(x_1, \dots, x_N) \cong \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} g_{r_1, \dots, r_N} f_{r_1}^{(1)}(x_1) \cdots f_{r_N}^{(N)}(x_N), \quad (1.4)$$

and its distributed tensor network variants (see Section 3.3),

- The Tensor Train (TT) format (see Section 4.1), in the form

$$f(x_1, x_2, \dots, x_N) \cong \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_{N-1}=1}^{R_{N-1}} f_{r_1}^{(1)}(x_1) f_{r_1 r_2}^{(2)}(x_2) \cdots f_{r_{N-2} r_{N-1}}^{(N-2)}(x_{N-1}) f_{r_{N-1}}^{(N)}(x_N), \quad (1.5)$$

with the equivalent compact matrix representation

$$f(x_1, x_2, \dots, x_N) \cong \mathbf{F}^{(1)}(x_1) \mathbf{F}^{(2)}(x_2) \cdots \mathbf{F}^{(N)}(x_N), \quad (1.6)$$

where $\mathbf{F}^{(n)}(x_n) \in \mathbb{R}^{R_{n-1} \times R_n}$, with $R_0 = R_N = 1$.

- The Hierarchical Tucker (HT) format (also known as the Hierarchical Tensor format) can be expressed via a hierarchy of nested

separations in the following way. Consider nested nonempty disjoint subsets u, v , and $t = u \cup v \subset \{1, 2, \dots, N\}$, then for some $1 \leq N_0 < N$, with $u_0 = \{1, \dots, N_0\}$ and $v_0 = \{N_0 + 1, \dots, N\}$, the HT format can be expressed as

$$f(x_1, \dots, x_N) \cong \sum_{r_{u_0}=1}^{R_{u_0}} \sum_{r_{v_0}=1}^{R_{v_0}} g_{r_{u_0}, r_{v_0}}^{(12\dots N)} f_{r_{u_0}}^{(u_0)}(\mathbf{x}_{u_0}) f_{r_{v_0}}^{(v_0)}(\mathbf{x}_{v_0}),$$

$$f_{r_t}^{(t)}(\mathbf{x}_t) \cong \sum_{r_u=1}^{R_u} \sum_{r_v=1}^{R_v} g_{r_u, r_v, r_t}^{(t)} f_{r_u}^{(u)}(\mathbf{x}_u) f_{r_v}^{(v)}(\mathbf{x}_v),$$

where $\mathbf{x}_t = \{x_i : i \in t\}$. See Section 2.2.1 for more detail.

Example. In a particular case for $N=4$, the HT format can be expressed by

$$f(x_1, x_2, x_3, x_4) \cong \sum_{r_{12}=1}^{R_{12}} \sum_{r_{34}=1}^{R_{34}} g_{r_{12}, r_{34}}^{(1234)} f_{r_{12}}^{(12)}(x_1, x_2) f_{r_{34}}^{(34)}(x_3, x_4),$$

$$f_{r_{12}}^{(12)}(x_1, x_2) \cong \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} g_{r_1, r_2, r_{12}}^{(12)} f_{r_1}^{(1)}(x_1) f_{r_2}^{(2)}(x_2),$$

$$f_{r_{34}}^{(34)}(x_3, x_4) \cong \sum_{r_3=1}^{R_3} \sum_{r_4=1}^{R_4} g_{r_3, r_4, r_{34}}^{(34)} f_{r_3}^{(3)}(x_3) f_{r_4}^{(4)}(x_4).$$

The Tree Tensor Network States (TTNS) format, which is an extension of the HT format, can be obtained by generalizing the two subsets, u, v , into a larger number of disjoint subsets u_1, \dots, u_m , $m \geq 2$. In other words, the TTNS can be obtained by more flexible separations of variables through products of larger numbers of functions at each hierarchical level (see Section 2.2.1 for graphical illustrations and more detail).

All the above approximations adopt the form of “sum-of-products” of single-dimensional functions, a procedure which plays a key role in all tensor factorizations and decompositions.

Indeed, in many applications based on multivariate functions, very good approximations are obtained with a surprisingly small number of factors; this number corresponds to the tensor rank, R , or tensor

network ranks, $\{R_1, R_2, \dots, R_N\}$ (if the representations are exact and minimal). However, for some specific cases this approach may fail to obtain sufficiently good low-rank TN approximations. The concept of generalized separability has already been explored in numerical methods for high-dimensional density function equations (Liao *et al.*, 2015; Trefethen, 2017; Cho *et al.*, 2016) and within a variety of huge-scale optimization problems (see Part 2 of this monograph).

To illustrate how tensor decompositions address excessive volumes of data, if all computations are performed on a CP tensor format in (1.3) and not on the raw N th-order data tensor itself, then instead of the original, *exponentially growing*, data dimensionality of I^N , the number of parameters in a CP representation reduces to NIR , which *scales linearly* in the tensor order N and size I (see Table 4.4). For example, the discretization of a 5-variate function over 100 sample points on each axis would yield the difficulty to manage $100^5 = 10,000,000,000$ sample points, while a rank-2 CP representation would require only $5 \times 2 \times 100 = 1000$ sample points.

Although the CP format in (1.2) effectively bypasses the curse of dimensionality, the CP approximation may involve numerical problems for very high-order tensors, which in addition to the intrinsic uncloseness of the CP format (i.e., difficulty to arrive at a canonical format), the corresponding algorithms for CP decompositions are often ill-posed (de Silva and Lim, 2008). As a remedy, greedy approaches may be considered which, for enhanced stability, perform consecutive rank-1 corrections (Lim and Comon, 2010). On the other hand, many efficient and stable algorithms exist for the more flexible Tucker format in (1.4), however, this format is not practical for tensor orders $N > 5$ because the number of entries of both the original data tensor and the core tensor (expressed in (1.4) by elements g_{r_1, r_2, \dots, r_N}) scales exponentially in the tensor order N (curse of dimensionality).

In contrast to CP decomposition algorithms, TT tensor network formats in (1.5) exhibit both very good numerical properties and the ability to control the error of approximation, so that a desired accuracy of approximation is obtained relatively easily. The main advantage of the TT format over the CP decomposition is the ability to provide stable

quasi-optimal rank reduction, achieved through, for example, truncated singular value decompositions (tSVD) or adaptive cross-approximation (Oseledets and Tyrtyshnikov, 2010; Bebendorf, 2011; Khoromskij and Veit, 2016). This makes the TT format one of the most stable and simple approaches to separate latent variables in a sophisticated way, while the associated TT decomposition algorithms provide full control over low-rank TN approximations¹. In this monograph, we therefore make extensive use of the TT format for low-rank TN approximations and employ the TT toolbox software for efficient implementations (Oseledets *et al.*, 2012). The TT format will also serve as a basic prototype for high-order tensor representations, while we also consider the Hierarchical Tucker (HT) and the Tree Tensor Network States (TTNS) formats (having more general tree-like structures) whenever advantageous in applications.

Furthermore, we address in depth the concept of tensorization of structured vectors and matrices to convert a wide class of huge-scale optimization problems into much smaller-scale interconnected optimization sub-problems which can be solved by existing optimization methods (see Part 2 of this monograph).

The tensor network optimization framework is therefore performed through the two main steps:

- Tensorization of data vectors and matrices into a high-order tensor, followed by a distributed approximate representation of a cost function in a specific low-rank tensor network format.
- Execution of all computations and analysis in tensor network formats (i.e., using only core tensors) that scale linearly, or even sub-linearly (quantized tensor networks), in the tensor order N . This yields both the reduced computational complexity and distributed memory requirements.

¹Although similar approaches have been known in quantum physics for a long time, their rigorous mathematical analysis is still a work in progress (see (Oseledets, 2011; Orús, 2014) and references therein).

1.4 Advantages of Multiway Analysis via Tensor Networks

In this monograph, we focus on two main challenges in huge-scale data analysis which are addressed by tensor networks: (i) an approximate representation of a specific cost (objective) function by a tensor network while maintaining the desired accuracy of approximation, and (ii) the extraction of physically meaningful latent variables from data in a sufficiently accurate and computationally affordable way. The benefits of multiway (tensor) analysis methods for large-scale datasets then include:

- Ability to perform all mathematical operations in tractable tensor network formats;
- Simultaneous and flexible distributed representations of both the structurally rich data and complex optimization tasks;
- Efficient compressed formats of large multidimensional data achieved via tensorization and low-rank tensor decompositions into low-order factor matrices and/or core tensors;
- Ability to operate with noisy and missing data by virtue of numerical stability and robustness to noise of low-rank tensor/matrix approximation algorithms;
- A flexible framework which naturally incorporates various diversities and constraints, thus seamlessly extending the standard, flat view, Component Analysis (2-way CA) methods to multiway component analysis;
- Possibility to analyze linked (coupled) blocks of large-scale matrices and tensors in order to separate common/correlated from independent/uncorrelated components in the observed raw data;
- Graphical representations of tensor networks allow us to express mathematical operations on tensors (e.g., tensor contractions and reshaping) in a simple and intuitive way, and without the explicit use of complex mathematical expressions.

In that sense, this monograph both reviews current research in this area and complements optimisation methods, such as the Alternating Direction Method of Multipliers (ADMM) (Boyd *et al.*, 2011).

Tensor decompositions (TDs) have been already adopted in widely diverse disciplines, including psychometrics, chemometrics, biometric, quantum physics/information, quantum chemistry, signal and image processing, machine learning, and brain science (Smilde *et al.*, 2004; Tao *et al.*, 2007; Kroonenberg, 2008; Kolda and Bader, 2009; Hackbusch, 2012; Favier and de Almeida, 2014; Cichocki *et al.*, 2009, 2015b). This is largely due to their advantages in the analysis of data that exhibit not only large volume but also very high variety (see Figure 1.1), as in the case in bio- and neuroinformatics and in computational neuroscience, where various forms of data collection include sparse tabular structures and graphs or hyper-graphs.

Moreover, tensor networks have the ability to efficiently parameterize, through structured compact representations, very general high-dimensional spaces which arise in modern applications (Kressner *et al.*, 2014b; Cichocki, 2014; Zhang *et al.*, 2015; Corona *et al.*, 2015; Litsarev and Oseledets, 2016; Khoromskij and Veit, 2016; Benner *et al.*, 2016). Tensor networks also naturally account for intrinsic multidimensional and distributed patterns present in data, and thus provide the opportunity to develop very sophisticated models for capturing multiple interactions and couplings in data – these are more physically insightful and interpretable than standard pair-wise interactions.

1.5 Scope and Objectives

Review and tutorial papers (Kolda and Bader, 2009; Lu *et al.*, 2011; Grasedyck *et al.*, 2013; Cichocki *et al.*, 2015b; de Almeida *et al.*, 2015; Sidiropoulos *et al.*, 2016; Papalexakis *et al.*, 2016; Bachmayr *et al.*, 2016) and books (Smilde *et al.*, 2004; Kroonenberg, 2008; Cichocki *et al.*, 2009; Hackbusch, 2012) dealing with TDs and TNs already exist, however, they typically focus on standard models, with no explicit links to very large-scale data processing topics or connections to a wide class of optimization problems. The aim of this monograph is therefore to

extend beyond the standard Tucker and CP tensor decompositions, and to demonstrate the perspective of TNs in extremely large-scale data analytics, together with their role as a mathematical backbone in the discovery of hidden structures in prohibitively large-scale data. Indeed, we show that TN models provide a framework for the analysis of linked (coupled) blocks of tensors with millions and even billions of non-zero entries.

We also demonstrate that TNs provide natural extensions of 2-way (matrix) Component Analysis (2-way CA) methods to multi-way component analysis (MWCA), which deals with the extraction of desired components from multidimensional and multimodal data. This paradigm shift requires new models and associated algorithms capable of identifying core relations among the different tensor modes, while guaranteeing linear/sub-linear scaling with the size of datasets².

Furthermore, we review tensor decompositions and the associated algorithms for very large-scale linear/multilinear dimensionality reduction problems. The related optimization problems often involve structured matrices and vectors with over a billion entries (see (Grasedyck *et al.*, 2013; Dolgov, 2014; Garreis and Ulbrich, 2016) and references therein). In particular, we focus on Symmetric Eigenvalue Decomposition (EVD/PCA) and Generalized Eigenvalue Decomposition (GEVD) (Dolgov *et al.*, 2014; Kressner *et al.*, 2014a; Kressner and Uschmajew, 2016), SVD (Lee and Cichocki, 2015), solutions of overdetermined and undetermined systems of linear algebraic equations (Oseledets and Dolgov, 2012; Dolgov and Savostyanov, 2014), the Moore–Penrose pseudo-inverse of structured matrices (Lee and Cichocki, 2016b), and Lasso problems (Lee and Cichocki, 2016a). Tensor networks for extremely large-scale multi-block (multi-view) data are also discussed, especially TN models for orthogonal Canonical Correlation Analysis (CCA) and related Partial Least Squares (PLS) problems. For convenience, all these problems are reformulated as constrained optimization problems

²Usually, we assume that huge-scale problems operate on at least 10^7 parameters.

which are then, by virtue of low-rank tensor networks reduced to manageable lower-scale optimization sub-problems. The enhanced tractability and scalability is achieved through tensor network contractions and other tensor network transformations.

The methods and approaches discussed in this work can be considered a both an alternative and complementary to other emerging methods for huge-scale optimization problems like random coordinate descent (RCD) scheme (Nesterov, 2012; Richtárik and Takáč, 2016), sub-gradient methods (Nesterov, 2014), alternating direction method of multipliers (ADMM) (Boyd *et al.*, 2011), and proximal gradient descent methods (Parikh and Boyd, 2014) (see also (Cevher *et al.*, 2014; Hong *et al.*, 2016) and references therein).

This monograph systematically introduces TN models and the associated algorithms for TNs/TDs and illustrates many potential applications of TDs/TNS. The dimensionality reduction and optimization frameworks (see Part 2 of this monograph) are considered in detail, and we also illustrate the use of TNs in other challenging problems for huge-scale datasets which can be solved using the tensor network approach, including anomaly detection, tensor completion, compressed sensing, clustering, and classification.

References

- E. Acar and B. Yener. Unsupervised multiway data analysis: A literature survey. *IEEE Transactions on Knowledge and Data Engineering*, 21:6–20, 2009.
- I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Physical Review Letters*, 59(7):799, 1987.
- A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15:2773–2832, 2014.
- D. Anderson, S. Du, M. Mahoney, C. Melgaard, K. Wu, and M. Gu. Spectral gap error bounds for improving CUR matrix decomposition and the Nyström method. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 19–27, 2015.
- W. Austin, G. Ballard, and T. G. Kolda. Parallel tensor compression for large-scale scientific data. *arXiv preprint arXiv:1510.06689*, 2015.
- F. R. Bach and M. I. Jordan. Kernel independent component analysis. *The Journal of Machine Learning Research*, 3:1–48, 2003.
- M. Bachmayr, R. Schneider, and A. Uschmajew. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Foundations of Computational Mathematics*, pages 1–50, 2016.
- B. W. Bader and T. G. Kolda. MATLAB tensor toolbox version 2.6, February 2015. URL <http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/>.

- J. Ballani and L. Grasedyck. Tree adaptive approximation in the hierarchical tensor format. *SIAM Journal on Scientific Computing*, 36(4):A1415–A1431, 2014.
- J. Ballani, L. Grasedyck, and M. Kluge. A review on adaptive low-rank approximation techniques in the hierarchical tensor format. In *Extraction of Quantifiable Information from Complex Systems*, pages 195–210. Springer, 2014.
- G. Ballard, A. R. Benson, A. Druinsky, B. Lipshitz, and O. Schwartz. Improving the numerical stability of fast matrix multiplication algorithms. *arXiv preprint arXiv:1507.00687*, 2015a.
- G. Ballard, A. Druinsky, N. Knight, and O. Schwartz. Brief announcement: Hypergraph partitioning for parallel sparse matrix-matrix multiplication. In *Proceedings of the 27th ACM on Symposium on Parallelism in Algorithms and Architectures*, pages 86–88. ACM, 2015b.
- G. Barcza, Ö. Legeza, K. H. Marti, and M. Reiher. Quantum information analysis of electronic states at different molecular structures. In *Physical Review A*, 83(1):012508, 2011.
- K. Batselier and N. Wong. A constructive arbitrary-degree Kronecker product decomposition of tensors. *arXiv preprint arXiv:1507.08805*, 2015.
- K. Batselier, H. Liu, and N. Wong. A constructive algorithm for decomposing a tensor into a finite sum of orthonormal rank-1 terms. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1315–1337, 2015.
- M. Bebendorf. Adaptive cross-approximation of multivariate functions. *Constructive Approximation*, 34(2):149–179, 2011.
- M. Bebendorf, C. Kuske, and R. Venn. Wideband nested cross approximation for Helmholtz problems. *Numerische Mathematik*, 130(1):1–34, 2015.
- R. E. Bellman. *Adaptive Control Processes*. Princeton University Press, Princeton, NJ, 1961.
- P. Benner, V. Khoromskaia, and B. N. Khoromskij. A reduced basis approach for calculation of the Bethe–Salpeter excitation energies by using low-rank tensor factorisations. *Molecular Physics*, 114(7-8):1148–1161, 2016.
- A. R. Benson, J. D. Lee, B. Rajwa, and D. F. Gleich. Scalable methods for nonnegative matrix factorizations of near-separable tall-and-skinny matrices. In *Proceedings of Neural Information Processing Systems (NIPS)*, pages 945–953, 2014. URL <http://arxiv.org/abs/1402.6964>.
- D. Bini. Tensor and border rank of certain classes of matrices and the fast evaluation of determinant inverse matrix and eigenvalues. *Calcolo*, 22(1):209–228, 1985.

- M. Bolten, K. Kahl, and S. Sokolović. Multigrid methods for tensor structured Markov chains with low rank approximation. *SIAM Journal on Scientific Computing*, 38(2):A649–A667, 2016. URL <http://adsabs.harvard.edu/abs/2016arXiv160506246B>.
- S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- A. Bruckstein, D. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.
- H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.
- C. Caiafa and A. Cichocki. Generalizing the column-row matrix decomposition to multi-way arrays. *Linear Algebra and its Applications*, 433(3):557–573, 2010.
- C. Caiafa and A. Cichocki. Computing sparse representations of multidimensional signals using Kronecker bases. *Neural Computation*, 25(1):186–220, 2013.
- C. Caiafa and A. Cichocki. Stable, robust, and super-fast reconstruction of tensors using multi-way projections. *IEEE Transactions on Signal Processing*, 63(3):780–793, 2015.
- J. D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3):283–319, 1970.
- V. Cevher, S. Becker, and M. Schmidt. Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics. *IEEE Signal Processing Magazine*, 31(5):32–43, 2014.
- G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavský, and A. Yeredor. Joint matrix decompositions and blind source separation: A survey of methods, identification, and applications. *IEEE Signal Processing Magazine*, 31(3):34–43, 2014.
- V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.
- T.-L. Chen, D. D. Chang, S.-Y. Huang, H. Chen, C. Lin, and W. Wang. Integrating multiple random sketches for singular value decomposition. *arXiv e-prints*, 2016.

- H. Cho, D. Venturi, and G. E. Karniadakis. Numerical methods for high-dimensional probability density function equations. *Journal of Computational Physics*, 305:817–837, 2016.
- J. H. Choi and S. Vishwanathan. DFacTo: Distributed factorization of tensors. In *Advances in Neural Information Processing Systems*, pages 1296–1304, 2014.
- W. Chu and Z. Ghahramani. Probabilistic models for incomplete multi-dimensional arrays. In *JMLR Workshop and Conference Proceedings Volume 5: AISTATS 2009*, volume 5, pages 89–96. Microtome Publishing (paper) Journal of Machine Learning Research, 2009.
- A. Cichocki. Tensor decompositions: A new concept in brain data analysis? *arXiv preprint arXiv:1305.0395*, 2013a.
- A. Cichocki. Era of big data processing: A new approach via tensor networks and tensor decompositions (invited). In *Proceedings of the International Workshop on Smart Info-Media Systems in Asia (SISA2013)*, September 2013b. URL <http://arxiv.org/abs/1403.2048>.
- A. Cichocki. Tensor networks for big data analytics and large-scale optimization problems. *arXiv preprint arXiv:1407.3124*, 2014.
- A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, Ltd, 2003.
- A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, Chichester, 2009.
- A. Cichocki, S. Cruces, and S. Amari. Log-determinant divergences revisited: Alpha-beta and gamma log-det divergences. *Entropy*, 17(5):2988–3034, 2015a.
- A. Cichocki, D. Mandic, A.-H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163, 2015b.
- N. Cohen and A. Shashua. Convolutional rectifier networks as generalized tensor decompositions. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 955–963, 2016.
- N. Cohen, O. Sharir, and A. Shashua. On the expressive power of deep learning: A tensor analysis. In *29th Annual Conference on Learning Theory*, pages 698–728, 2016.
- P. Comon. Tensors: a brief introduction. *IEEE Signal Processing Magazine*, 31(3):44–53, 2014.

- P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 2010.
- P. G. Constantine and D. F. Gleich. Tall and skinny QR factorizations in MapReduce architectures. In *Proceedings of the Second International Workshop on MapReduce and its Applications*, pages 43–50. ACM, 2011.
- P. G. Constantine, D. F. Gleich, Y. Hou, and J. Templeton. Model reduction with MapReduce-enabled tall and skinny singular value decomposition. *SIAM Journal on Scientific Computing*, 36(5):S166–S191, 2014.
- E. Corona, A. Rahimian, and D. Zorin. A Tensor-Train accelerated solver for integral equations in complex geometries. *arXiv preprint arXiv:1511.06029*, November 2015.
- C. Crainiceanu, B. Caffo, S. Luo, V. Zipunnikov, and N. Punjabi. Population value decomposition, a framework for the analysis of image populations. *Journal of the American Statistical Association*, 106(495):775–790, 2011. URL <http://pubs.amstat.org/doi/abs/10.1198/jasa.2011.ap10089>.
- A. Critch and J. Morton. Algebraic geometry of matrix product states. *Symmetry, Integrability and Geometry: Methods and Applications (SIGMA)*, 10:095, 2014. .
- A. J. Critch. *Algebraic Geometry of Hidden Markov and Related Models*. PhD thesis, University of California, Berkeley, 2013.
- A. L. F. de Almeida, G. Favier, J. C. M. Mota, and J. P. C. L. da Costa. Overview of tensor decompositions with applications to communications. In R. F. Coelho, V. H. Nascimento, R. L. de Queiroz, J. M. T. Romano, and C. C. Cavalcante, editors, *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*, chapter 12, pages 325–355. CRC Press, 2015.
- F. De la Torre. A least-squares framework for component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6):1041–1055, 2012.
- L. De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 28:642–666, 2006.
- L. De Lathauwer. Decompositions of a higher-order tensor in block terms — Part I and II. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1022–1066, 2008. URL <http://publi-etis.ensea.fr/2008/De08e>. Special Issue on Tensor Decompositions and Applications.

- L. De Lathauwer. Blind separation of exponential polynomials and the decomposition of a tensor in rank- $(L_r, L_r, 1)$ terms. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1451–1474, 2011.
- L. De Lathauwer and D. Nion. Decompositions of a higher-order tensor in block terms – Part III: Alternating least squares algorithms. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1067–1083, 2008.
- L. De Lathauwer, B. De Moor, and J. Vandewalle. A Multilinear singular value decomposition. *SIAM Journal on Matrix Analysis Applications*, 21:1253–1278, 2000a.
- L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors. *SIAM Journal of Matrix Analysis and Applications*, 21(4):1324–1342, 2000b.
- W. de Launey and J. Seberry. The strong Kronecker product. *Journal of Combinatorial Theory, Series A*, 66(2):192–213, 1994. URL <http://dblp.uni-trier.de/db/journals/jct/jcta66.html#LauneyS94>.
- V. de Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30:1084–1127, 2008.
- A. Desai, M. Ghashami, and J. M. Phillips. Improved practical matrix sketching with guarantees. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1678–1690, 2016.
- I. S. Dhillon. Fast Newton-type methods for nonnegative matrix and tensor approximation. The NSF Workshop, Future Directions in Tensor-Based Computation and Modeling, 2009.
- E. Di Napoli, D. Fabregat-Traver, G. Quintana-Ortí, and P. Bientinesi. Towards an efficient use of the BLAS library for multilinear tensor contractions. *Applied Mathematics and Computation*, 235:454–468, 2014.
- S. V. Dolgov. *Tensor Product Methods in Numerical Simulation of High-dimensional Dynamical Problems*. PhD thesis, Faculty of Mathematics and Informatics, University Leipzig, Germany, 2014.
- S. V. Dolgov and B. N. Khoromskij. Two-level QTT-Tucker format for optimized tensor calculus. *SIAM Journal on Matrix Analysis and Applications*, 34(2):593–623, 2013.
- S. V. Dolgov and B. N. Khoromskij. Simultaneous state-time approximation of the chemical master equation using tensor product formats. *Numerical Linear Algebra with Applications*, 22(2):197–219, 2015.

- S. V. Dolgov and D. V. Savostyanov. Alternating minimal energy methods for linear systems in higher dimensions. *SIAM Journal on Scientific Computing*, 36(5):A2248–A2271, 2014.
- S. V. Dolgov, B. N. Khoromskij, I. V. Oseledets, and D. V. Savostyanov. Computation of extreme eigenvalues in higher dimensions using block tensor train format. *Computer Physics Communications*, 185(4):1207–1216, 2014.
- P. Drineas and M. W. Mahoney. A randomized algorithm for a tensor-based generalization of the singular value decomposition. *Linear Algebra and its Applications*, 420(2):553–571, 2007.
- G. Ehlers, J. Sólyom, Ö. Legeza, and R.M. Noack. Entanglement structure of the Hubbard model in momentum space. *Physical Review B*, 92(23):235116, 2015.
- M. Espig, M. Schuster, A. Killaitis, N. Waldren, P. Wähnert, S. Handschuh, and H. Auer. TensorCalculus library, 2012. URL <http://gitorious.org/tensorcalculus>.
- F. Esposito, T. Scarabino, A. Hyvärinen, J. Himberg, E. Formisano, S. Comani, G. Tedeschi, R. Goebel, E. Seifritz, and F. Di Salle. Independent component analysis of fMRI group studies by self-organizing clustering. *NeuroImage*, 25(1):193–205, 2005.
- G. Evenbly and G. Vidal. Algorithms for entanglement renormalization. *Physical Review B*, 79(14):144108, 2009.
- G. Evenbly and S. R. White. Entanglement renormalization and wavelets. *Physical Review Letters*, 116(14):140403, 2016.
- H. Fanaee-T and J. Gama. Tensor-based anomaly detection: An interdisciplinary survey. *Knowledge-Based Systems*, 2016.
- G. Favier and A. de Almeida. Overview of constrained PARAFAC models. *EURASIP Journal on Advances in Signal Processing*, 2014(1):1–25, 2014.
- J. Garcke, M. Griebel, and M. Thess. Data mining with sparse grids. *Computing*, 67(3):225–253, 2001.
- S. Garreis and M. Ulbrich. Constrained optimization with low-rank tensors and applications to parametric problems with PDEs. *SIAM Journal on Scientific Computing* (accepted), 2016.
- M. Ghashami, E. Liberty, and J. M. Phillips. Efficient frequent directions algorithm for sparse matrices. *arXiv preprint arXiv:1602.00412*, 2016.

- V. Giovannetti, S. Montangero, and R. Fazio. Quantum multiscale entanglement renormalization ansatz channels. *Physical Review Letters*, 101(18):180503, 2008.
- S. A. Goreinov, E. E. Tyrtshnikov, and N. L. Zamarashkin. A theory of pseudo-skeleton approximations. *Linear Algebra and its Applications*, 261:1–21, 1997a.
- S. A. Goreinov, N. L. Zamarashkin, and E. E. Tyrtshnikov. Pseudo-skeleton approximations by matrices of maximum volume. *Mathematical Notes*, 62(4):515–519, 1997b.
- L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2029–2054, 2010.
- L. Grasedyck, D. Kessner, and C. Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36:53–78, 2013.
- A. R. Groves, C. F. Beckmann, S. M. Smith, and M. W. Woolrich. Linked independent component analysis for multimodal data fusion. *NeuroImage*, 54(1):2198–21217, 2011.
- Z.-C. Gu, M. Levin, B. Swingle, and X.-G. Wen. Tensor-product representations for string-net condensed states. *Physical Review B*, 79(8):085118, 2009.
- M. Haardt, F. Roemer, and G. Del Galdo. Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. *IEEE Transactions on Signal Processing*, 56:3198–3213, July 2008.
- W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*, volume 42 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2012.
- W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *Journal of Fourier Analysis and Applications*, 15(5):706–722, 2009.
- N. Halko, P. Martinsson, and J. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.
- S. Handschuh. *Numerical Methods in Tensor Networks*. PhD thesis, Faculty of Mathematics and Informatics, University Leipzig, Germany, Leipzig, Germany, 2015.
- R. A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

- F. L. Hitchcock. Multiple invariants and generalized rank of a p -way matrix or tensor. *Journal of Mathematics and Physics*, 7:39–79, 1927.
- S. Holtz, T. Rohwedder, and R. Schneider. The alternating linear scheme for tensor optimization in the tensor train format. *SIAM Journal on Scientific Computing*, 34(2), 2012. URL <http://dx.doi.org/10.1137/100818893>.
- M. Hong, M. Razaviyayn, Z. Q. Luo, and J. S. Pang. A unified algorithmic framework for block-structured optimization involving big data with applications in machine learning and signal processing. *IEEE Signal Processing Magazine*, 33(1):57–77, 2016.
- H. Huang, C. Ding, D. Luo, and T. Li. Simultaneous tensor subspace selection and clustering: The equivalence of high order SVD and K-means clustering. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 327–335. ACM, 2008.
- R. Hübener, V. Nebendahl, and W. Dür. Concatenated tensor network states. *New Journal of Physics*, 12(2):025004, 2010.
- C. Hubig, I. P. McCulloch, U. Schollwöck, and F. A. Wolf. Strictly single-site DMRG algorithm with subspace expansion. *Physical Review B*, 91(15):155115, 2015.
- T. Huckle, K. Waldherr, and T. Schulte-Herbrüggen. Computations in quantum tensor networks. *Linear Algebra and its Applications*, 438(2):750–781, 2013.
- A. Hyvärinen. Independent component analysis: Recent advances. *Philosophical Transactions of the Royal Society A*, 371(1984):20110534, 2013.
- I. Jeon, E. E. Papalexakis, C. Faloutsos, L. Sael, and U. Kang. Mining billion-scale tensors: Algorithms and discoveries. *The VLDB Journal*, pages 1–26, 2016.
- B. Jiang, F. Yang, and S. Zhang. Tensor and its Tucker core: The invariance relationships. *arXiv e-prints arXiv:1601.01469*, January 2016.
- U. Kang, E. E. Papalexakis, A. Harpale, and C. Faloutsos. GigaTensor: Scaling tensor analysis up by 100 times – algorithms and discoveries. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*, pages 316–324, August 2012.
- Y.-J. Kao, Y.-D. Hsieh, and P. Chen. Uni10: An open-source library for tensor network algorithms. In *Journal of Physics: Conference Series*, volume 640, page 012040. IOP Publishing, 2015.
- L. Karlsson, D. Kressner, and A. Uschmajew. Parallel algorithms for tensor completion in the CP format. *Parallel Computing*, 57:222–234, 2016.

- J.-P. Kauppi, J. Hahne, K. R. Müller, and A. Hyvärinen. Three-way analysis of spectrospatial electromyography data: Classification and interpretation. *PloS One*, 10(6):e0127231, 2015.
- V. A. Kazeev and B. N. Khoromskij. Low-rank explicit QTT representation of the Laplace operator and its inverse. *SIAM Journal on Matrix Analysis and Applications*, 33(3):742–758, 2012.
- V. A. Kazeev, B. N. Khoromskij, and E. E. Tyrtshnikov. Multilevel Toeplitz matrices generated by tensor-structured vectors and convolution with logarithmic complexity. *SIAM Journal on Scientific Computing*, 35(3):A1511–A1536, 2013a.
- V. A. Kazeev, O. Reichmann, and C. Schwab. Low-rank tensor structure of linear diffusion operators in the TT and QTT formats. *Linear Algebra and its Applications*, 438(11):4204–4221, 2013b.
- V. A. Kazeev, M. Khammash, M. Nip, and C. Schwab. Direct solution of the chemical master equation using quantized tensor trains. *PLoS Computational Biology*, 10(3):e1003359, 2014.
- B. N. Khoromskij. $O(d \log N)$ -quantics approximation of N - d tensors in high-dimensional numerical modeling. *Constructive Approximation*, 34(2):257–280, 2011a.
- B. N. Khoromskij. Tensors-structured numerical methods in scientific computing: Survey on recent advances. *Chemometrics and Intelligent Laboratory Systems*, 110(1):1–19, 2011b. URL <http://www.mis.mpg.de/de/publications/preprints/2010/prepr2010-21.html>.
- B. N. Khoromskij and A. Veit. Efficient computation of highly oscillatory integrals by using QTT tensor approximation. *Computational Methods in Applied Mathematics*, 16(1):145–159, 2016.
- H.-J. Kim, E. Ollila, V. Koivunen, and H. V. Poor. Robust iteratively reweighted Lasso for sparse tensor factorizations. In *IEEE Workshop on Statistical Signal Processing (SSP)*, pages 420–423, 2014.
- S. Klus and C. Schütte. Towards tensor-based methods for the numerical approximation of the Perron-Frobenius and Koopman operator. *arXiv e-prints arXiv:1512.06527*, December 2015.
- T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
- D. Kressner and C. Tobler. Algorithm 941: HTucker—A MATLAB toolbox for tensors in hierarchical Tucker format. *ACM Transactions on Mathematical Software*, 40(3):22, 2014.

- D. Kressner and A. Uschmajew. On low-rank approximability of solutions to high-dimensional operator equations and eigenvalue problems. *Linear Algebra and its Applications*, 493:556–572, 2016.
- D. Kressner, M. Steinlechner, and A. Uschmajew. Low-rank tensor methods with subspace correction for symmetric eigenvalue problems. *SIAM Journal on Scientific Computing*, 36(5):A2346–A2368, 2014a.
- D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by Riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014b.
- P. M. Kroonenberg. *Applied Multiway Data Analysis*. John Wiley & Sons Ltd, New York, 2008.
- J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95–138, 1977.
- V. Kuleshov, A. T. Chaganty, and P. Liang. Tensor factorization via matrix factorization. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, pages 507–516, 2015.
- N. Lee and A. Cichocki. Estimating a few extreme singular values and vectors for large-scale matrices in Tensor Train format. *SIAM Journal on Matrix Analysis and Applications*, 36(3):994–1014, 2015.
- N. Lee and A. Cichocki. Tensor train decompositions for higher order regression with LASSO penalties. In *Workshop on Tensor Decompositions and Applications (TDA2016)*, 2016a.
- N. Lee and A. Cichocki. Regularized computation of approximate pseudoinverse of large matrices using low-rank tensor train decompositions. *SIAM Journal on Matrix Analysis and Applications*, 37(2):598–623, 2016b. URL <http://adsabs.harvard.edu/abs/2015arXiv150601959L>.
- N. Lee and A. Cichocki. Fundamental tensor operations for large-scale data analysis using tensor network formats. *Multidimensional Systems and Signal Processing* (accepted), 2016c.
- J. Li, C. Battaglino, I. Perros, J. Sun, and R. Vuduc. An input-adaptive and in-place approach to dense tensor-times-matrix multiply. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, page 76. ACM, 2015.
- M. Li and V. Monga. Robust video hashing via multilinear subspace projections. *IEEE Transactions on Image Processing*, 21(10):4397–4409, 2012.

- S. Liao, T. Vejchodský, and R. Erban. Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks. *Journal of the Royal Society Interface*, 12(108):20150233, 2015.
- A. P. Liavas and N. D. Sidiropoulos. Parallel algorithms for constrained tensor factorization via alternating direction method of multipliers. *IEEE Transactions on Signal Processing*, 63(20):5450–5463, 2015.
- L. H. Lim and P. Comon. Multiarray signal processing: Tensor decomposition meets compressed sensing. *Comptes Rendus Mecanique*, 338(6):311–320, 2010.
- M. S. Litsarev and I. V. Oseledets. A low-rank approach to the computation of path integrals. *Journal of Computational Physics*, 305:557–574, 2016.
- H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. A survey of multilinear subspace learning for tensor data. *Pattern Recognition*, 44(7):1540–1551, 2011.
- M. Lubasch, J. I. Cirac, and M.-C. Banuls. Unifying projected entangled pair state contractions. *New Journal of Physics*, 16(3):033014, 2014. URL <http://stacks.iop.org/1367-2630/16/i=3/a=033014>.
- C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and tensor-train tensors. *SIAM Journal on Matrix Analysis and Applications*, 34(2):470–494, 2013.
- M. W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, 3(2):123–224, 2011.
- M. W. Mahoney and P. Drineas. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106:697–702, 2009.
- M. W. Mahoney, M. Maggioni, and P. Drineas. Tensor-CUR decompositions for tensor-based data. *SIAM Journal on Matrix Analysis and Applications*, 30(3):957–987, 2008.
- H. Matsueda. Analytic optimization of a MERA network and its relevance to quantum integrability and wavelet. *arXiv preprint arXiv:1608.02205*, 2016.
- A. Y. Mikhalev and I. V. Oseledets. Iterative representing set selection for nested cross-approximation. *Numerical Linear Algebra with Applications*, 2015.
- L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11:50–59, 1960.

- J. Morton. Tensor networks in algebraic geometry and statistics. *Lecture at Networking Tensor Networks, Centro de Ciencias de Benasque Pedro Pascual, Benasque, Spain*, 2012.
- M. Mørup. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Review: Data Mining and Knowledge Discovery*, 1(1):24–40, 2011.
- V. Murg, F. Verstraete, R. Schneider, P. R. Nagy, and O. Legeza. Tree tensor network state with variable tensor order: An efficient multireference method for strongly correlated systems. *Journal of Chemical Theory and Computation*, 11(3):1027–1036, 2015.
- N. Nakatani and G. K. L. Chan. Efficient tree tensor network states (TTNS) for quantum chemistry: Generalizations of the density matrix renormalization group algorithm. *The Journal of Chemical Physics*, 2013. URL <http://arxiv.org/pdf/1302.2298.pdf>.
- Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.
- Y. Nesterov. Subgradient methods for huge-scale optimization problems. *Mathematical Programming*, 146(1-2):275–297, 2014.
- N. H. Nguyen, P. Drineas, and T. D. Tran. Tensor sparsification via a bound on the spectral norm of random tensors. *Information and Inference*, page iav004, 2015.
- M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.
- A. Novikov and R. A. Rodomanov. Putting MRFs on a tensor train. In *Proceedings of the International Conference on Machine Learning (ICML '14)*, 2014.
- A. C. Olivieri. Analytical advantages of multivariate data processing. One, two, three, infinity? *Analytical Chemistry*, 80(15):5713–5720, 2008.
- R. Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014.
- I. V. Oseledets. Approximation of $2^d \times 2^d$ matrices using tensor decomposition. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2130–2145, 2010.
- I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

- I. V. Oseledets and S. V. Dolgov. Solution of linear systems and matrix inversion in the TT-format. *SIAM Journal on Scientific Computing*, 34(5): A2718–A2739, 2012.
- I. V. Oseledets and E. E. Tyrtyshnikov. Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM Journal on Scientific Computing*, 31(5):3744–3759, 2009.
- I. V. Oseledets and E. E. Tyrtyshnikov. TT cross-approximation for multidimensional arrays. *Linear Algebra and its Applications*, 432(1):70–88, 2010.
- I. V. Oseledets, S. V. Dolgov, V. A. Kazeev, D. Savostyanov, O. Lebedeva, P. Zhlobich, T. Mach, and L. Song. TT-Toolbox, 2012. URL <https://github.com/oseledets/TT-Toolbox>.
- E. E. Papalexakis, N. Sidiropoulos, and R. Bro. From K-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors. *IEEE Transactions on Signal Processing*, 61(2):493–506, 2013.
- E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(2):16, 2016.
- N. Parikh and S. P. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.
- D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Information & Computation*, 7(5):401–430, July 2007. URL <http://dl.acm.org/citation.cfm?id=2011832.2011833>.
- R. Pfeifer, G. Evenbly, S. Singh, and G. Vidal. NCON: A tensor network contractor for MATLAB. *arXiv preprint arXiv:1402.0939*, 2014.
- N. Pham and R. Pagh. Fast and scalable polynomial kernels via explicit feature maps. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 239–247. ACM, 2013.
- A.-H. Phan and A. Cichocki. Extended HALS algorithm for nonnegative Tucker decomposition and its applications for multiway analysis and classification. *Neurocomputing*, 74(11):1956–1969, 2011.
- A.-H. Phan, P. Tichavský, and A. Cichocki. Fast alternating ls algorithms for high order candecomp/parafac tensor factorizations. *IEEE Transactions on Signal Processing*, 61(19):4834–4846, 2013a.

- A.-H. Phan, P. Tichavský, and A. Cichocki. Tensor deflation for CANDECOMP/PARAFAC – Part I: Alternating subspace update algorithm. *IEEE Transactions on Signal Processing*, 63(22):5924–5938, 2015a.
- A.-H. Phan, A. Cichocki, A. Uschmajew, P. Tichavský, G. Luta, and D. Mandic. Tensor networks for latent variable analysis. Part I: Algorithms for tensor train decomposition. *ArXiv e-prints*, 2016.
- A.-H. Phan and A. Cichocki. Tensor decompositions for feature extraction and classification of high dimensional datasets. *Nonlinear Theory and its Applications, IEICE*, 1(1):37–68, 2010.
- A.-H. Phan, A. Cichocki, P. Tichavský, D. Mandic, and K. Matsuoka. On revealing replicating structures in multiway data: A novel tensor decomposition approach. In *Proceedings of the 10th International Conference LVA/ICA, Tel Aviv, March 12–15*, pages 297–305. Springer, 2012.
- A.-H. Phan, A. Cichocki, P. Tichavský, R. Zdunek, and S. R. Lehky. From basis components to complex structural patterns. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26–31, 2013*, pages 3228–3232, 2013b. URL <http://dx.doi.org/10.1109/ICASSP.2013.6638254>.
- A.-H. Phan, P. Tichavský, and A. Cichocki. Low complexity damped Gauss-Newton algorithms for CANDECOMP/PARAFAC. *SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 34(1):126–147, 2013c. URL <http://arxiv.org/pdf/1205.2584.pdf>.
- A.-H. Phan, P. Tichavský, and A. Cichocki. Low rank tensor deconvolution. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP*, pages 2169–2173, April 2015b. URL <http://dx.doi.org/10.1109/ICASSP.2015.7178355>.
- S. Ragnarsson. *Structured Tensor Computations: Blocking Symmetries and Kronecker Factorization*. PhD dissertation, Cornell University, Department of Applied Mathematics, 2012.
- M. V. Rakhuba and I. V. Oseledets. Fast multidimensional convolution in low-rank tensor formats via cross-approximation. *SIAM Journal on Scientific Computing*, 37(2):A565–A582, 2015.
- P. Richtárik and M. Takáč. Parallel coordinate descent methods for big data optimization. *Mathematical Programming*, 156:433–484, 2016.
- J. Salmi, A. Richter, and V. Koivunen. Sequential unfolding SVD for tensors with applications in array signal processing. *IEEE Transactions on Signal Processing*, 57:4719–4733, 2009.

- U. Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011.
- U. Schollwöck. Matrix product state algorithms: DMRG, TEBD and relatives. In *Strongly Correlated Systems*, pages 67–98. Springer, 2013.
- N. Schuch, I. Cirac, and D. Pérez-García. PEPS as ground states: Degeneracy and topology. *Annals of Physics*, 325(10):2153–2192, 2010.
- N. Sidiropoulos, R. Bro, and G. Giannakis. Parallel factor analysis in sensor array processing. *IEEE Transactions on Signal Processing*, 48(8):2377–2388, 2000.
- N. D. Sidiropoulos. Generalizing Caratheodory’s uniqueness of harmonic parameterization to N dimensions. *IEEE Transactions on Information Theory*, 47(4):1687–1690, 2001.
- N. D. Sidiropoulos. Low-rank decomposition of multi-way arrays: A signal processing perspective. In *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2004)*, July 2004. URL <http://www.sandia.gov/~tgkolda/tdw2004/Nikos04.pdf>.
- N. D. Sidiropoulos and R. Bro. On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics*, 14(3):229–239, 2000.
- N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. Tensor decomposition for signal processing and machine learning. *arXiv e-prints arXiv:1607.01668*, 2016.
- A. Smilde, R. Bro, and P. Geladi. *Multi-way Analysis: Applications in the Chemical Sciences*. John Wiley & Sons Ltd, New York, 2004.
- S. M. Smith, A. Hyvärinen, G. Varoquaux, K. L. Miller, and C. F. Beckmann. Group-PCA for very large fMRI datasets. *NeuroImage*, 101:738–749, 2014.
- L. Sorber, M. Van Barel, and L. De Lathauwer. Optimization-based algorithms for tensor decompositions: Canonical Polyadic Decomposition, decomposition in rank- $(L_r, L_r, 1)$ terms and a new generalization. *SIAM Journal on Optimization*, 23(2), 2013.
- L. Sorber, I. Domanov, M. Van Barel, and L. De Lathauwer. Exact line and plane search for tensor optimization. *Computational Optimization and Applications*, 63(1):121–142, 2016.
- M. Sørensen and L. De Lathauwer. Blind signal separation via tensor decomposition with Vandermonde factor. Part I: Canonical polyadic decomposition. *IEEE Transactions on Signal Processing*, 61(22):5507–5519, 2013. URL <http://dx.doi.org/10.1109/TSP.2013.2276416>.

- M. Sørensen, L. De Lathauwer, P. Comon, S. Icart, and L. Deneire. Canonical Polyadic Decomposition with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1190–1213, 2012.
- M. Steinlechner. Riemannian optimization for high-dimensional tensor completion. Technical report, Technical report MATHICSE 5.2015, EPF Lausanne, Switzerland, 2015.
- M. M. Steinlechner. *Riemannian Optimization for Solving High-Dimensional Problems with Low-Rank Tensor Structure*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2016.
- E. M. Stoudenmire and S. R. White. Minimally entangled typical thermal state algorithms. *New Journal of Physics*, 12(5):055026, 2010.
- J. Sun, D. Tao, and C. Faloutsos. Beyond streams and graphs: Dynamic tensor analysis. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pages 374–383. ACM, 2006.
- S. K. Suter, M. Makhynia, and R. Pajarola. TAMRESH – tensor approximation multiresolution hierarchy for interactive volume visualization. *Computer Graphics Forum*, 32(3):151–160, 2013.
- Y. Tang, R. Salakhutdinov, and G. Hinton. Tensor analyzers. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013), Atlanta, USA, 2013*.
- D. Tao, X. Li, X. Wu, and S. Maybank. General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1700–1715, 2007.
- P. Tichavský and A. Yeredor. Fast approximate joint diagonalization incorporating weight matrices. *IEEE Transactions on Signal Processing*, 47(3):878–891, 2009.
- M. K. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, pages 567–574, 2009.
- C. Tobler. *Low-rank tensor methods for linear systems and eigenvalue problems*. PhD thesis, ETH Zürich, 2012.
- L. N. Trefethen. Cubature, approximation, and isotropy in the hypercube. *SIAM Review*, Forthcoming, 2017. URL <https://people.maths.ox.ac.uk/trefethen/hypercube.pdf>.
- V. Tresp, C. Esteban, Y. Yang, S. Baier, and D. Krompaß. Learning with memory embeddings. *arXiv preprint arXiv:1511.07972*, 2015.

- J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher. Randomized single-view algorithms for low-rank matrix approximation. *arXiv e-prints*, 2016.
- L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.
- L. R. Tucker. The extension of factor analysis to three-dimensional matrices. In H. Gulliksen and N. Frederiksen, editors, *Contributions to Mathematical Psychology*, pages 110–127. Holt, Rinehart and Winston, New York, 1964.
- A. Uschmajew and B. Vandereycken. The geometry of algorithms using hierarchical tensors. *Linear Algebra and its Applications*, 439:133–166, 2013.
- N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen. A new truncation strategy for the higher-order singular value decomposition. *SIAM Journal on Scientific Computing*, 34(2):A1027–A1052, 2012.
- M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 2350, pages 447–460, Copenhagen, Denmark, May 2002.
- F. Verstraete, V. Murg, and I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008.
- N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer. Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis. *IEEE Signal Processing Magazine*, 31(5):71–79, 2014.
- G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91(14):147902, 2003.
- S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman. Robust iterative fitting of multilinear models. *IEEE Transactions on Signal Processing*, 53(8):2678–2689, 2005.
- S. Wahls, V. Koivunen, H. V. Poor, and M. Verhaegen. Learning multidimensional Fourier series with tensor trains. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 394–398. IEEE, 2014.
- D. Wang, H. Shen, and Y. Truong. Efficient dimension reduction for high-dimensional matrix-valued data. *Neurocomputing*, 190:25–34, 2016.
- H. Wang and M. Thoss. Multilayer formulation of the multiconfiguration time-dependent Hartree theory. *Journal of Chemical Physics*, 119(3):1289–1299, 2003.

- H. Wang, Q. Wu, L. Shi, Y. Yu, and N. Ahuja. Out-of-core tensor approximation of multi-dimensional matrices of visual data. *ACM Transactions on Graphics*, 24(3):527–535, 2005.
- S. Wang and Z. Zhang. Improving CUR matrix decomposition and the Nystrom approximation via adaptive sampling. *The Journal of Machine Learning Research*, 14(1):2729–2769, 2013.
- Y. Wang, H.-Y. Tung, A. Smola, and A. Anandkumar. Fast and guaranteed tensor decomposition via sketching. In *Advances in Neural Information Processing Systems*, pages 991–999, 2015.
- S. R. White. Density-matrix algorithms for quantum renormalization groups. *Physical Review B*, 48(14):10345, 1993.
- Z. Xu, F. Yan, and Y. Qi. Infinite Tucker decomposition: Nonparametric Bayesian models for multiway data analysis. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, ICML '12, pages 1023–1030. Omnipress, July 2012.
- Y. Yang and T. Hospedales. Deep multi-task representation learning: A tensor factorisation approach. *arXiv preprint arXiv:1605.06391*, 2016. URL <http://adsabs.harvard.edu/abs/2016arXiv160506391Y>.
- T. Yokota, Q. Zhao, and A. Cichocki. Smooth PARAFAC decomposition for tensor completion. *IEEE Transactions on Signal Processing*, 64(20):5423–5436, 2016.
- T. Yokota, N. Lee, and A. Cichocki. Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria. *IEEE Transactions on Signal Processing*, accepted, 2017.
- Z. Zhang, X. Yang, I. V. Oseledets, G. E. Karniadakis, and L. Daniel. Enabling high-dimensional hierarchical uncertainty quantification by ANOVA and tensor-train decomposition. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(1):63–76, 2015.
- H. H. Zhao, Z. Y. Xie, Q. N. Chen, Z. C. Wei, J. W. Cai, and T. Xiang. Renormalization of tensor-network states. *Physical Review B*, 81(17):174411, 2010.
- Q. Zhao, C. Caiafa, D. P. Mandic, Z. C. Chao, Y. Nagasaka, N. Fujii, L. Zhang, and A. Cichocki. Higher order partial least squares (HOPLS): A generalized multilinear regression method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1660–1673, 2013a.

- Q. Zhao, G. Zhou, T. Adali, L. Zhang, and A. Cichocki. Kernelization of tensor-based models for multiway data analysis: Processing of multidimensional structured data. *IEEE Signal Processing Magazine*, 30(4):137–148, 2013b.
- Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki. Tensor ring decomposition. *CoRR*, abs/1606.05535, 2016. URL <http://arxiv.org/abs/1606.05535>.
- S. Zhe, Y. Qi, Y. Park, Z. Xu, I. Molloy, and S. Chari. DinTucker: Scaling up Gaussian process models on large multidimensional arrays. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016. URL <http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11959/11888>.
- G. Zhou and A. Cichocki. Fast and unique Tucker decompositions via multiway blind source separation. *Bulletin of Polish Academy of Science*, 60(3):389–407, 2012a.
- G. Zhou and A. Cichocki. Canonical Polyadic Decomposition based on a single mode blind source separation. *IEEE Signal Processing Letters*, 19(8):523–526, 2012b.
- G. Zhou, A. Cichocki, and S. Xie. Fast nonnegative matrix/tensor factorization based on low-rank approximation. *IEEE Transactions on Signal Processing*, 60(6):2928–2940, June 2012.
- G. Zhou, A. Cichocki, Q. Zhao, and S. Xie. Efficient nonnegative Tucker decompositions: Algorithms and uniqueness. *IEEE Transactions on Image Processing*, 24(12):4990–5003, 2015.
- G. Zhou, A. Cichocki, Y. Zhang, and D. P. Mandic. Group component analysis for multiblock data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, (in print), 2016a.
- G. Zhou, Q. Zhao, Y. Zhang, T. Adali, S. Xie, and A. Cichocki. Linked component analysis from matrices to high-order tensors: Applications to biomedical data. *Proceedings of the IEEE*, 104(2):310–331, 2016b.