Maolin Che
Yimin Wei

Theory and Computation of Complex Tensors and its Applications

Springer

Theory and Computation of Complex Tensors and its Applications

Maolin Che • Yimin Wei

Theory and Computation of Complex Tensors and its Applications

Springer

Maolin Che
School of Economics Mathematics
Southwestern University of Finance
and Economics
Chengdu, Sichuan, China

Yimin Wei
School of Mathematical Sciences
Fudan University
Shanghai, China

ISBN 978-981-15-2058-7 ISBN 978-981-15-2059-4 (eBook)
https://doi.org/10.1007/978-981-15-2059-4

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This book can be divided into five categories based on the main purposes: (1) the development of tensor spectral theory; (2) the study of tensor complementarity problems using structured tensors; (3) the development and enrichment of the theory of nonnegative tensors; (4) the presentation of new numerical algorithms for solving the real tensor rank-one approximation and computing the US- or U-eigenpairs of complex tensors; (5) the study of randomized algorithms for the computation of the approximate Tucker and Tensor Train decompositions.

There are eight chapters in this book. In Chap. 1, we give some examples to illustrate that tensors can be abstracted from some real and mathematical objects. Some basic operations and definitions of tensors, for example, tensor-matrix multiplication, the Frobenius norm and inner product of tensors, rank-one and symmetric tensors, are introduced. We also provide a summary of the relevant background for the tensor spectral theory, the Perron–Frobenius theorem of nonnegative tensors, and the tensor rank-one approximation problem.

In Chap. 2, we generalize the pseudo-spectral theory of matrices to tensors. We obtain the fundamental properties of the tensor ϵ-pseudo-spectrum, leading to alternative definitions of the tensor ϵ-pseudo-spectrum. We also consider the stability of homogeneous dynamical systems. Similarly, we derive the fundamental properties for the ϵ-pseudo-spectrum of tensor polynomial eigenvalue problems. Furthermore, we discuss the implications of the ϵ-pseudo-spectrum on computing the backward errors of an approximate eigenpair of a tensor polynomial and the distance from a regular tensor polynomial to its nearest irregular tensor polynomial.

In Chap. 3, we analyze the perturbation of tensor eigenvalue problems. We consider the first-order perturbation results for the algebraically simple Z- and H-eigenvalues of tensors and H-eigenvalues of tensor polynomials with relative Frobenius normwise or componentwise perturbations. Based on the perturbation for the algebraically simple Z-eigenvalue of a symmetric real tensor and mode-symmetric embedding, we obtain the perturbation of the algebraically simple singular value of a real tensor. Specifically, we focus on the perturbation for the smallest eigenvalue of an irreducible and symmetric nonsingular $\mathcal{M}$-tensor for relative componentwise perturbations.

In Chap. 4, we first analyze the first-order necessary conditions for the solution of the tensor complementarity problem. From the properties of copositive tensors, we prove that the problem with copositive tensors has a nonempty and compact solution set. We also consider a special case via structured tensors.

In Chap. 5, we introduce the sign nonsingular tensors and derive the relationship between the combinatorial determinant and the permanent of nonnegative tensors. We generalize the results from doubly stochastic matrices to totally plane stochastic tensors and obtain a probabilistic algorithm for locating a positive diagonal in a nonnegative tensor. We obtain a normalization algorithm to convert some nonnegative tensors to plane stochastic tensors. We obtain a lower bound for the minimum of the axial N-index assignment problem using the set of plane stochastic tensors.

Chapter 6 deals with the local optimal rank-one approximation of a real tensor via neural networks. We prove that the solution of the neural network is locally asymptotically stable in the sense of Lyapunov stability theory. We define the tensor restricted singular pairs and present several numerical algorithms for computing them. Similarly, we use the neural networks for the computation of the local optimal generalized H-eigenpairs of symmetric-definite tensor pairs.

Chapter 7 presents the iterative algorithms (QRCST or QRCT) for computing the US- and U-eigenpairs of complex tensors. Specifically, we derive a higher order power type method for computing a US- or a U-eigenpair, similar to the higher-order power method for computing the best rank-one approximation of a real tensor.

In Chap. 8, we design adaptive randomized algorithms for computing the approximate tensor decompositions. For a low multilinear rank approximation of a real tensor with unknown multilinear rank, we analyze its probabilistic error bound under certain assumptions. Finally, we also consider the tensor train approximations of the tensors. Based on the bounds of the singular values of sub-Gaussian matrices with independent columns or independent rows, we analyze these randomized algorithms. Several illustrated numerical examples are provided.

Chapters 2 and 3 are based closely on [1–3]. The main content in Chap. 4 comes from [4]. Chapters 5, 6, and 7 are adopted from [5–7]. Chapter 8 is from [8].

In this book, the computations are carried out in MATLAB Version 2013a and the MATLAB Tensor Toolbox [9] on a laptop with an Intel Core i5-4200M CPU (2.50 GHz) and a 8.00 GB RAM. All floating point numbers in each example have four digits after the decimal point. For $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, we assume that "all i_n" and "all n" mean $i_n = 1, 2, \ldots, I_n$ and $n = 1, 2, \ldots, N$, respectively; for $\mathcal{A} \in CT_{N,I}$, we assume that "all i_n" means $i_n = 1, 2, \ldots, I$ for all n. We assume that "all l" means "$l = 0, 1, \ldots, L$".

We would like to thank Prof. Andrzej Cichocki for computing the tensor rank-one approximation via neural network models, Prof. Guoyin Li for the tensor pseudo-spectral theory, and Prof. Changjiang Bu for the study of plane stochastic tensors.

We also thank Prof. Liqun Qi of the Hong Kong Polytechnic University, who leads us to the research of tensor spectral theory and always encourages us to explore the research. We would like to thank Prof. Eric King-wah Chu and Prof. Sanzheng Qiao, who read this book carefully and provide feedback during the writing process.

M. Che is supported by the Fundamental Research Funds for the Central Universities under grant JBK1804026 and the National Natural Science Foundation of China under grant 11901471. Y. Wei is supported by the National Natural Science Foundation of China under grant 11771099, the Innovation Program of Shanghai Municipal Education Commission, Key Laboratory of Mathematics for Nonlinear Sciences of Fudan University, and Shanghai Key Laboratory of Contemporary Applied Mathematics.

Chengdu, China — Maolin Che
Shanghai, China — Yimin Wei
July, 2018

References

1. M. Che, Y. Wei, An inequality for the Perron pair of an irreducible and symmetric nonnegative tensor with application. J. Oper. Res. Soc. China **5**(1), 65–82 (2017)
2. M. Che, G. Li, L. Qi, Y. Wei, Pseudo-spectra theory of tensors and tensor polynomial eigenvalue problems. Linear Algebra Appl. **533**(15), 536–572 (2017)
3. M. Che, L. Qi, Y. Wei, Perturbation bounds of tensor eigenvalue and singular value problems with even order. Linear Multilinear Algebra **64**(4), 622–652 (2016)
4. M. Che, L. Qi, Y. Wei, Positive-definite tensors to nonlinear complementarity problems. J. Optim. Theory Appl. **168**(2), 475–487 (2016)
5. M. Che, C. Bu, L. Qi, Y. Wei, Nonnegative tensors revisited: plane stochastic tensors. Linear Multilinear Algebra **67**(7), 1364–1391 (2019)
6. M. Che, A. Cichocki, Y. Wei, Neural networks for computing best rank-one approximations of tensors and its applications. Neurocomputing **267**(3), 114–133 (2017)
7. M. Che, L. Qi, Y. Wei, Iterative algorithms for computing US- and U-eigenpairs of complex tensors. J. Comput. Appl. Math. **317**, 547–564 (2017)
8. M. Che, Y. Wei, Randomized algorithms for the approximations of Tucker and the tensor train decompositions. Adv. Comput. Math. **45**(1), 395–428 (2019)
9. B. Bader, T. Kolda, *Matlab Tensor Toolbox Version 2.6* (2015). http://www.sandia.gov/~tgkolda/TensorToolbox/

Contents

Chapter 1
Introduction

An increasing number of applications in signal processing, data analysis and higher-order statistics, as well as independent component analysis [1–5] involve the manipulation of data whose elements are addressed by more than two indices. In the literature, these higher-order extensions of vectors (first-order) and matrices (second-order) are called *higher-order tensors*, *multi-dimensional matrices*, or *multiway arrays*.

Tensor problems have wide applications in chemometrics, signal processing and high order statistics [4]. For the theory and applications of tensors, we refer to Comon et al. [6], Kolda and Bader [7], Cichocki et al. [8], Yang and Yang [9], Qi and Luo [10], Wei and Ding [11], and Qi et al. [12].

We use $\mathbb{C}$ or $\mathbb{R}$ to denote the complex or real field. A tensor is an Nth-order array of numbers denoted by script notation $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ with entries given by $a_{i_1 i_2 \ldots i_N} \in \mathbb{C}$ for all i_n and n. When all the I_n are the same, i.e., $I_n = I$ for all n, $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is called an Nth-order I-dimensional complex tensor. We use $RT_{N,I}$ (or $CT_{N,I}$) to denote the set of all Nth-order I-dimensional real (or complex) tensors.

1.1 Examples for Tensors

Tensors can be abstracted from some real and mathematical objects.

Example 1.1.1 (Homogeneous Polynomials) Given $\mathbf{c} \in \mathbb{R}^I$, the first-degree homogeneous polynomial with respect to $\mathbf{x} \in \mathbb{R}^I$ can be represented as $\mathbf{c}^\top \mathbf{x} = c_1 x_1 + c_2 x_2 + \cdots + c_I x_I$. Given $\mathbf{A} \in \mathbb{R}^{I \times J}$, the second-degree homogeneous polynomial with respect to $\mathbf{x} \in \mathbb{R}^I$ and $\mathbf{y} \in \mathbb{R}^J$ can be represented as $\mathbf{x}^\top \mathbf{A} \mathbf{y} = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} x_i y_j$.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_1

In general, the Nth-degree homogeneous polynomial, with respect to $\mathbf{x}_1 \in \mathbb{R}^{I_1}$, $\mathbf{x}_2 \in \mathbb{R}^{I_2}, \ldots, \mathbf{x}_N \in \mathbb{R}^{I_N}$, can be denoted by

$$p_N(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \ldots i_N} x_{1,i_1} x_{2,i_2} \ldots x_{N,i_N},$$

where $a_{i_1 i_2 \ldots i_N}$ is the coefficient of the term $x_{1,i_1} x_{2,i_2} \ldots x_{N,i_N}$ in $p_N(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ and x_{n,i_n} is the i_nth entry of $\mathbf{x}_n$ for all i_n and n. If we set a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ such that its $(i_1, i_2, \ldots, i_N)$-entry is $a_{i_1 i_2 \ldots i_N}$, then the polynomial can be expressed as

$$p_N(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top.$$

The tensor-vector multiplication will be introduced in the next section.

Example 1.1.2 (The Discretization of Multivariate Functions) Suppose that

$$f(x, y, z) = \frac{1}{x + y + z}, \quad g(x, y, z) = \cos(x + y + z),$$

$$h(x, y, z) = \exp((-0.01 + 4\pi\iota)(x + y + z - 2)) + \exp((-0.02 + 4.2\pi\iota)(x + y + z - 2)),$$

where $1 < x, y, z < 2$ and $\iota = \sqrt{-1}$.

Let $\{x_1, x_2, \ldots, x_I\}$ be any monotonically increasing sequence in the open interval $(1, 2)$. When the values of x, y and z are chosen from $\{x_1, x_2, \ldots, x_I\}$, we define three tensors $\mathcal{A}, \mathcal{B} \in RT_{3,I}$ and $C \in CT_{3,I}$ such that

$$a_{ijk} := f(x_i, x_j, x_k), \quad b_{ijk} = g(x_i, x_j, x_k), \quad c_{ijk} = h(x_i, x_j, x_k).$$

From the definitions of f, g and h, it is clear that $\mathcal{A} \in RT_{3,I}$ is symmetric and positive, $\mathcal{B} \in RT_{3,I}$ is symmetric but not nonnegative and $C \in CT_{3,I}$ is complex symmetric.

Example 1.1.3 (The Associated Tensors of Uniform Hypergraphs) Analogous to spectral graph theory [13], adjacency tensors and Laplacian tensors have been introduced in spectral hypergraph theory. The notations related to the hypergraph can be referred to [14].

A hypergraph $\mathbb{H}$ is a pair $(\mathbb{V}, \mathbb{E})$, where $\mathbb{E} \subseteq \mathfrak{P}(\mathbb{V})$. The elements of $\mathbb{V} = \mathbb{V}(\mathbb{H})$ are referred to as vertices and the elements of $\mathbb{E} = \mathbb{E}(\mathbb{H})$ are called hyperedges. A hypergraph $\mathbb{H}$ is said to be N-uniform for an integer $N \geq 2$, if, for all $\mathbf{e} \in \mathbb{E}(\mathbb{H})$, the cardinality of $\mathbf{e}$ is N. Such an N-uniform hypergraph is also called an N-graph.

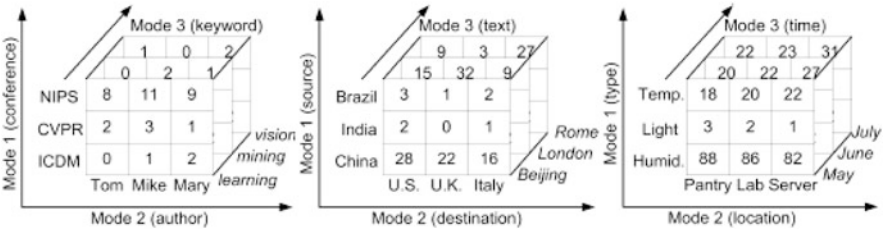

Fig. 1.1 Examples of the third-order tensor data

For a given N-uniform hypergraph $\mathbb{H} = (\mathbb{V}, \mathbb{E})$, the adjacency tensor $\mathcal{A} \in RT_{N,I}$, with respect to $\mathbb{H}$, is defined as [15–18]:

$$a_{i_1 i_2 \ldots i_N} = \frac{1}{(N-1)!} \begin{cases} 1, & \text{if } \{i_1, i_2, \ldots, i_N\} \in \mathbb{E}; \\ 0, & \text{otherwise.} \end{cases}$$

Obviously, $a_{i_1 i_2 \ldots i_N} = 0$ if at least two indices are the same. Note that $\mathcal{A}$ is symmetric and nonnegative. The degree of $i \in \mathbb{V}$ is defined as $d(i) = |\{e_{\mathbf{p}} : i \in \mathbf{p} \in \mathbb{V}\}|$. We assume that every vertex has at least one edge. Thus, $d(i) > 0$ for $i \in \mathbb{V}$. The degree tensor $\mathcal{D}(\mathbb{H})$ is an Nth-order I-dimensional diagonal tensor, with its main diagonal elements as $d(i)$. The Laplacian tensor $\mathcal{L}$ of $\mathbb{H}$ is defined by $\mathcal{D}(\mathbb{H}) - \mathcal{A}$. The signless Laplacian tensor $\mathcal{L}^+$ of $\mathbb{H}$ is defined by $\mathcal{D}(\mathbb{H}) + \mathcal{A}$ [19–22]. Note that $\mathcal{L}$ is symmetric and nonnegative and $\mathcal{L}^+$ is symmetric.

Example 1.1.4 (Description of Complex Social Networks [23]) Third-order tensors are three-dimensional (3D), with some examples shown in Fig. 1.1. The left part of Fig. 1.1 illustrates social network analysis data organized in three modes of conference, author and keyword. The web graph mining data organized in three modes of source, destination and text is demonstrated in the middle part of Fig. 1.1. Lastly, the environmental sensor monitoring data organized in the three modes of type, location and time is demonstrated in the right part of Fig. 1.1.

Generally speaking, there are two kinds of tensors: a data structure, which admits different dimensions according to the complexity of the data; or an operator, where it possesses different meanings in different situations. All tensors mentioned in this book can be viewed as a data structure.

1.2 Basics of Tensors

Throughout this book, I, J, and N are reserved to denote the index upper bounds, unless stated otherwise. We use small letters $x, u, v, \ldots$ for scalars, small bold letters $\mathbf{x}, \mathbf{u}, \mathbf{v}, \ldots$ for vectors, bold capital letters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ for matrices,

and calligraphic letters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$ for higher-order tensors. This notation is consistently used for lower-order parts of a given structure. For example, the entry with row index i and column index j in a matrix $\mathbf{A}$, i.e., $(\mathbf{A})_{ij}$, is written by a_{ij} (also $(\mathbf{x})_i = x_i$ and $(\mathcal{A})_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N}$).

We use $\cdot^\top$, $\bar{\cdot}$, $|\cdot|$ and $\cdot^*$ to denote the transpose, complex conjugate, element-wise modulus and complex conjugated transpose, respectively. We use $\|\cdot\|_2$ and $\|\cdot\|_F$ to denote the 2-norm and the Frobenius norm, respectively. We use $\Re(\mathbf{z})$ and $\Im(\mathbf{z})$ to denote the real and imaginary parts of a vector $\mathbf{z} \in \mathbb{C}^I$, respectively. The argument of $z \in \mathbb{C}$ is denoted by $\arg(z) \in (-\pi, \pi]$. Also, $\mathbf{0}_I \in \mathbb{C}^I$ is the zero vector. We use $\mathbf{I}_I \in \mathbb{C}^{I \times I}$ and $\mathbf{0}_{I \times J} \in \mathbb{C}^{I \times J}$ to denote the identity and the zero matrices, respectively. Lastly, $\iota = \sqrt{-1}$ and $\mathbb{S}_I$ are the imaginary unit and the symmetric group on the set $\{1, 2, \ldots, I\}$, respectively.

We use parentheses to denote the concatenation of two or more vectors, e.g., $(\mathbf{a}, \mathbf{b})$ is equivalent to $(\mathbf{a}^\top, \mathbf{b}^\top)^\top$, where $\mathbf{a} \in \mathbb{C}^I$ and $\mathbf{b} \in \mathbb{C}^J$ are two column vectors. Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$, $\mathbf{x} > \mathbf{y}$ and $\mathbf{x} \geq \mathbf{y}$ mean $x_i > y_i$ and $x_i \geq y_i$, respectively, for all i. Similarly, we can also define $\mathbf{x} < \mathbf{y}$ and $\mathbf{x} \leq \mathbf{y}$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$. Finally, we introduce the following notations:

$$\mathbb{R}_{+}^{I_1 \times I_2 \times \cdots \times I_N} := \{\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} : a_{i_1 i_2 \ldots i_N} \geq 0, i_n = 1, 2, \ldots, I, n = 1, 2, \ldots, N\};$$

$$\mathbb{R}_{++}^{I_1 \times I_2 \times \cdots \times I_N} := \{\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} : a_{i_1 i_2 \ldots i_N} > 0, i_n = 1, 2, \ldots, I, n = 1, 2, \ldots, N\}.$$

$\mathbb{R}_{+}^{I_1 \times I_2 \times \cdots \times I_N}$ is called the set of all nonnegative tensors and $\mathbb{R}_{++}^{I_1 \times I_2 \times \cdots \times I_N}$ is called the set of all positive tensors in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. In particular, when $N = 1$, we have $\mathbb{R}_{+}^{I} := \{\mathbf{x} \in \mathbb{R}^I : x_i \geq 0, i = 1, 2, \ldots, I\}$ and $\mathbb{R}_{++}^{I} := \{\mathbf{x} \in \mathbb{R}^I : x_i > 0, i = 1, 2, \ldots, I\}$. For a given $\mathbf{x} \in \mathbb{C}^I$, $\mathrm{diag}(\mathbf{x})$ denotes the diagonal matrix whose main diagonal entries are the entries of $\mathbf{x}$.

1.2.1 Basic Operations

The mode-n product [7] of a complex tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ by a matrix $\mathbf{B} \in \mathbb{C}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{B}$, is a tensor $\mathcal{C} \in \mathbb{C}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$, with entries

$$c_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n = 1}^{I_n} a_{i_1 i_2 \ldots i_N} b_{j i_n},$$

for all i_n and n.

In particular, the mode-n multiplication of a complex tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ by a vector $\mathbf{x} \in \mathbb{C}^{I_n}$ is denoted by $\mathcal{A} \bar{\times}_n \mathbf{x}$. If we set $\mathcal{C} = \mathcal{A} \times_n \mathbf{x}^\top \in$

$\mathbb{C}^{I_1\times\cdots\times I_{n-1}\times I_{n+1}\times\cdots\times I_N}$, then we have element-wise [7],

$$c_{i_1\ldots i_{n-1}i_{n+1}\ldots i_N} = \sum_{i_n=1}^{I_n} a_{i_1\ldots i_{n-1}i_n i_{n+1}\ldots i_N} x_{i_n}.$$

For any given tensor $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$ and the matrices $\mathbf{F} \in \mathbb{C}^{J_n\times I_n}$ and $\mathbf{G} \in \mathbb{C}^{J_m\times I_m}$, one has [7]

$$\begin{cases} (\mathcal{A}\times_n \mathbf{F})\times_m \mathbf{G} = (\mathcal{A}\times_m \mathbf{G})\times_n \mathbf{F} = \mathcal{A}\times_n \mathbf{F}\times_m \mathbf{G}; \\ (\mathcal{A}\times_n \mathbf{F})\times_n \mathbf{G} = \mathcal{A}\times_n (\mathbf{G}\cdot\mathbf{F}), \quad \text{with} \quad J_n = I_m, \end{cases}$$

with different integers m and n, where '$\cdot$' represents the multiplication of two matrices.

For a given $\mathbf{x} \in \mathbb{C}^I$ and $\mathcal{A} \in CT_{N,I}$, we introduce the following two notations [24]:

$$\mathcal{A}\mathbf{x}^{N-1} := \mathcal{A}\times_2 \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top, \ \mathcal{A}\mathbf{x}^N := \mathcal{A}\times_1 \mathbf{x}^\top \times_2 \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top.$$

Scalar products and the Frobenius norm of a tensor are extensions of the well-known definitions, from matrices to tensors of arbitrary order [7, 25]. Suppose that $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$, the scalar product $\langle\mathcal{A}, \mathcal{B}\rangle$ is defined as [25]

$$\langle\mathcal{A}, \mathcal{B}\rangle = \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} b_{i_1 i_2\ldots i_N}\overline{a_{i_1 i_2\ldots i_N}},$$

and the Frobenius norm of a tensor $\mathcal{A}$ is given by $\|\mathcal{A}\|_F = \sqrt{\langle\mathcal{A}, \mathcal{A}\rangle}$.

1.2.2 Structured Tensors

We recommend [10] for a thorough survey of structured tensors. For given N vectors $\mathbf{x}_n \in \mathbb{C}^{I_n}$ with all n, if the entries of $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$ can be represented as $a_{i_1 i_2\ldots i_N} = x_{1,i_1}x_{2,i_2}\ldots x_{N,i_N}$, where x_{n,i_n} is the i_nth element of $\mathbf{x}_n$, then $\mathcal{A}$ is a complex rank-one tensor [26, 27] given by

$$\mathcal{A} = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_N,$$

where "$\otimes$" denotes the outer (tensor) product. If $\mathbf{x}_n \in \mathbb{R}^{I_n}$ with all n, then $\mathcal{A}$ is a real rank-one tensor.

For any $\mathcal{A} \in CT_{N,I}$, $\mathcal{A}$ is *complex symmetric* [28], if $a_{i_1 i_2\ldots i_N}$ is invariant by any permutation $\pi \in \mathbb{S}_I$, that is, $a_{i_1 i_2\ldots i_N} = a_{i_{\pi(1)} i_{\pi(2)}\ldots i_{\pi(N)}}$ for all i_n and n. For any

$\mathcal{A} \in RT_{N,I}$, $\mathcal{A}$ is *real symmetric* [24, 29], if $a_{i_1 i_2 \ldots i_N}$ is invariant by any permutation π, that is, $a_{i_1 i_2 \ldots i_N} = a_{i_{\pi(1)} i_{\pi(2)} \ldots i_{\pi(N)}}$, for all i_n and n. For a given vector $\mathbf{x} \in \mathbb{C}^I$, if the entries of $\mathcal{A} \in CT_{N,I}$ can be represented as

$$a_{i_1 i_2 \ldots i_N} = x_{i_1} x_{i_2} \ldots x_{i_N},$$

where x_{i_n} is the i_nth element of $\mathbf{x}$, then $\mathcal{A}$ is a complex symmetric rank-one tensor given by

$$\mathcal{A} = \underbrace{\mathbf{x} \otimes \mathbf{x} \otimes \cdots \otimes \mathbf{x}}_{N} := \mathbf{x}^{\otimes N}.$$

If $\mathbf{x} \in \mathbb{R}^I$, then $\mathbf{x}^{\otimes N}$ is a real symmetric rank-one tensor.

For any given symmetric tensor $\mathcal{A} \in RT_{N,I}$ with an even N, if $\mathcal{A}\mathbf{x}^N > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^I$, then $\mathcal{A}$ is positive definite [24]; if $\mathcal{A}\mathbf{x}^N \geq 0$ for all $\mathbf{x} \in \mathbb{R}^I$, then $\mathcal{A}$ is positive semi-definite [24]; if $\mathcal{A}\mathbf{x}^N \geq 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^I_+$, then $\mathcal{A}$ is strictly copositive [30]; if $\mathcal{A}\mathbf{x}^N \geq 0$ for all $\mathbf{x} \in \mathbb{R}^I_+$, then $\mathcal{A}$ is copositive [30].

When we consider positive definite tensors and copositive tensors, symmetry is unnecessary. According to [31], we can relate a tensor $\mathcal{A} \in RT_{N,I}$ to a symmetric tensor $\mathcal{B} \in RT_{N,I}$ as follows. For any $\mathcal{A} \in RT_{N,I}$, there is a unique symmetric tensor $\mathcal{B} \in RT_{N,I}$ such that $\mathcal{A}\mathbf{x}^N = \mathcal{B}\mathbf{x}^N$ for all $\mathbf{x} \in \mathbb{R}^I$.

We call an index set $\{i_1, i_2, \ldots, i_N\}$ a permutation of another index set $\{k_1, k_2, \ldots, k_N\}$ if the former is a rearrangement of the latter, denoting this operation by π, that is $\pi(k_1, k_2, \ldots, k_N) = \{i_1, i_2, \ldots, i_N\}$. Denote the set of all distinct permutations of an index set $\{k_1, k_2, \ldots, k_N\}$ by $\Sigma(k_1, k_2, \ldots, k_N)$.

Note that $|\Sigma(k_1, k_2, \ldots, k_N)|$, the cardinality of $\Sigma(k_1, k_2, \ldots, k_N)$, is variant for different index sets. Then the entries of $\mathcal{B}$ are given as

$$b_{j_1 j_2 \ldots j_N} = \frac{\sum_{\pi \in \Sigma(k_1, k_2, \ldots, k_N)} a_{\pi(k_1, k_2, \ldots, k_N)}}{|\Sigma(k_1, k_2, \ldots, k_N)|}$$

for all i_n, j_n and n. Here, we call $\mathcal{B}$ a symmetrization of $\mathcal{A}$.

Suppose that $\mathcal{A} \in CT_{N,I}$, the I-tuple $\{a_{1\pi_2(1)\ldots\pi_N(1)}, a_{2\pi_2(2)\ldots\pi_N(2)}, \ldots, a_{I\pi_2(I)\ldots\pi_N(I)}\}$ is a *diagonal* [32] of the tensor $\mathcal{A}$ associated with $\pi_n \in \mathbb{S}_I$ and $n = 2, 3, \ldots, N$. In particular, $\{a_{11\ldots1}, a_{22\ldots2}, \ldots, a_{II\ldots I}\}$ is the main diagonal of the tensor $\mathcal{A}$. A diagonal is *positive*, if its elements are positive.

The product $\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}$ is the *diagonal product* [32] of the tensor $\mathcal{A}$ associated with $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$. Meanwhile, the sum $\sum_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}$ is the *diagonal sum* of $\mathcal{A}$ associated with $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$.

A tensor $\mathcal{A} \in RT_{N,I}$ is *nonnegative* [33], if the elements are nonnegative, and we denote the set of all nonnegative tensors by $NT_{N,I}$; a tensor $\mathcal{D} \in RT_{N,I}$ is *diagonal* [24], if the entries not in the main diagonal are zero. In particular, if the entries on the main diagonal of any diagonal tensor $\mathcal{A} \in CT_{N,I}$ are 1, then $\mathcal{A}$ is the unit tensor or the identity tensor, denoted by $\mathcal{I}$.

Analogous to the reducible matrices [34, Chapter 2], $\mathcal{A} \in RT_{N,I}$ is *reducible* [33], if there exists a nonempty proper index subset $\mathbb{I} \subset \{1, 2, \ldots, I\}$ such that

$$a_{i_1 \ldots i_N} = 0, \text{ for all } i_1 \in \mathbb{I} \text{ and } i_2, \ldots, i_N \notin \mathbb{I}.$$

Otherwise, $\mathcal{A}$ is *irreducible*. Similarly, we can define the irreducibility of any $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$.

Qi [24] introduces the principal symmetric sub-tensors of any symmetric tensor $\mathcal{A} \in RT_{N,I}$ and proves that if $\mathcal{A}$ is positive definite, then all of its principal symmetric sub-tensors are also positive definite. Now, we introduce the definition of a sub-tensor of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$.

Suppose that $L_n \leq I_n$ are positive integers for all n. We use $\mathbb{Q}_{L_n, I_n}$ to denote the set of $C_{I_n}^{L_n}$ increasing sequences $\omega = (\omega_1, \omega_2, \ldots, \omega_{L_n})$ such that $1 \leq \omega_1 < \omega_2 < \ldots < \omega_{L_n} \leq I_n$, where $C_{I_n}^{L_n} = \frac{I_n!}{L_n!(I_n - L_n)!}$. If $\alpha_n \in \mathbb{Q}_{P_n, I_n}$, where $P_n \leq I_n$ is a positive integer with $n = 1, 2, \ldots, N$, then $\mathcal{A}[\alpha_1|\alpha_2|\ldots|\alpha_N] \in \mathbb{C}^{P_1 \times P_2 \times \cdots \times P_N}$ is a sub-tensor of $\mathcal{A}$, whose $(i_1, i_2, \ldots, i_N)$-entry is $a_{\alpha_{1,i_1}, \alpha_{2,i_2}, \ldots, \alpha_{N,i_N}}$.

The $(i_1, \ldots, i_{m-1}, i_{m+1}, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N)$th mode-$(m, n)$ slice of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as a matrix in $\mathbb{C}^{I_m \times I_n}$, denoted by $\mathcal{A}_{m,n}^{(i)}$, equaling $\mathcal{A} \times_1 \mathbf{e}_{1,i_1}^\top \cdots \times_{m-1} \mathbf{e}_{m-1,i_{m-1}}^\top \times_{m+1} \mathbf{e}_{m+1,i_{m+1}}^\top \cdots \times_{n-1} \mathbf{e}_{n-1,i_{n-1}}^\top \times_{n+1} \mathbf{e}_{n+1,i_{n+1}}^\top \cdots \times_N \mathbf{e}_{N,i_N}^\top$, where $\mathbf{e}_{n,i_n}$ is the i_nth column of $\mathbf{I}_{I_n} \in \mathbb{R}^{I_n \times I_n}$ for all i_n and $m < n$. Note that for a given i, all the $(i, i, \ldots, i)$th mode-(m, n) slices, or the ith mode-(m, n) slices, of any complex symmetric tensor $\mathcal{A} \in CT_{N,I}$ are the same complex symmetric matrix with all $m < n$.

1.3 Basic Results

1.3.1 Tensor Spectral Theory

The eigenvalue problem of tensors can be regarded as the generalizations of the eigenvalue problem of matrices (i.e., matrix standard eigenvalue problems, matrix generalized eigenvalue problems and matrix polynomial eigenvalue problems).The eigenvalue problem of tensors are widely used in polynomial optimization [35], spectral hypergraph theory [16, 36], higher-order Markov chain [37], image science [38] and other fields. Very recently, the eigenvalue problem of tensors, positive semi-definite tensors and copositive tensors have been used to study some physical problems, such as the quantum spin state, the quantum field theory and liquid crystals [39–41].

In 2005, Qi [24] defines two kinds of eigenvalues and investigates relative results similar to the matrix eigenvalues. Independently, Lim [29] proposes another definition of eigenvalues, eigenvectors, singular values, and singular vectors for

tensors based on a constrained variational approach, in the flavor of the Rayleigh quotient for symmetric matrix eigenvalues [42, Chapter 8].

Definition 1.3.1 ([24]) Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric. If there exist a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ and $\lambda \in \mathbb{C}$ such that $\mathcal{A}\mathbf{x}^{N-1} = \lambda \mathbf{x}^{[N-1]}$, where $\mathbf{x}^{[N-1]} = (x_1^{N-1}, x_2^{N-1}, \ldots, x_I^{N-1})^\top$, then $(\lambda; \mathbf{x})$ is called an eigenpair of $\mathcal{A}$. The spectrum $\Lambda(\mathcal{A})$, and the spectral radius $\rho(\mathcal{A})$ of $\mathcal{A}$ are defined as

$$\Lambda(\mathcal{A}) = \{\lambda : \lambda \text{ is an eigenvalue of } \mathcal{A}\}, \quad \rho(\mathcal{A}) = \max\{|\lambda| : \lambda \in \Lambda(\mathcal{A})\}.$$

Moreover, if $\mathbf{x} \in \mathbb{R}^I$ and $\lambda \in \mathbb{R}$, then $(\lambda; \mathbf{x})$ is called an H-eigenpair of $\mathcal{A}$.

Definition 1.3.2 Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric. If there exist a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ and $\lambda \in \mathbb{C}$ such that $\mathcal{A}\mathbf{x}^{N-1} = \lambda \mathbf{x}$ and $\mathbf{x}^*\mathbf{x} = 1$, then $(\lambda; \mathbf{x})$ is called an E-eigenpair of $\mathcal{A}$ [43, Definition 5.1.1].

Moreover, if $\mathbf{x} \in \mathbb{R}^I$ is unit and $\lambda \in \mathbb{R}$, then $(\lambda; \mathbf{x})$ is called a Z-eigenpair of $\mathcal{A}$.

Note that the E-eigenpair of any real symmetric tensor, defined by Definition 1.3.2, is different from the E-eigenpair in [24].

In 2008, Qi et al. [44] introduce the D-eigenvalues for a diffusion kurtosis tensor and indicate that the largest, the smallest and the average D-eigenvalues correspond with the largest, the smallest and the average apparent kurtosis coefficients of a water molecule in the space, respectively. The strong ellipticity condition plays an important role in nonlinear elasticity and in materials. In 2009, Qi et al. [45] define the M-eigenvalues for an elasticity tensor and prove that the strong ellipticity condition holds if and only if the smallest M-eigenvalue is positive. Hu et al. [35] investigate properties of the determinants of tensors, and their applications. It is pointed out in [33, 46–48] that the generalized eigenvalue framework unifies several definitions of eigenvalues of tensors, such as eigenvalues and H-, E-, Z- and D-eigenvalues. Ding and Wei [11, 49] focus on the properties and perturbations of the spectra of regular tensor pairs and extend several classical results from matrices or matrix pairs to tensor pairs.

Kolda and Mayo [43] derive a shifted symmetric higher-order power method (SS-HOPM) for computing the Z-eigenpairs of real symmetric tensors and indicate that SS-HOPM can be viewed as a generalization of the power iteration method for matrices or the symmetric higher-order power method. Kolda and Mayo [48] present the adaptive power method for solving the tensor generalized eigenvalue problem associated with symmetric positive tensor pairs, which is an extension of SS-HOPM for finding the Z-eigenpairs. Cui et al. [47] propose a new approach for computing all real eigenvalues (that is, Z- and H-eigenvalues) of real symmetric tensors sequentially, from the largest to the smallest. Chen et al. [50] derive an upper bound for the number of equivalence classes of generalized tensor eigenpairs using mixed volumes. Based on this bound and the structures of tensor eigenvalue problems, they propose two homotopy algorithms for the tensor eigenvalue problem. Using the state-of-the-art L-BFGS approach, Chang et al. [51] develop a first-

order optimization algorithm for computing the H- and Z-eigenvalues of large scale sparse real symmetric tensors. Batselier and Wong [52] derive the QR algorithm for computing the Z-eigenpairs of a real symmetric tensor, based on the symmetric QR algorithm for the real symmetric matrix eigenvalue problem.

Chang et al. [33] generalize the Perron-Frobenius theorem for nonnegative matrices to the class of nonnegative tensors. We state the Perron-Frobenius theorem for nonnegative tensors as follows:

Theorem 1.3.1 *If $\mathcal{A} \in RT_{N,I}$ is nonnegative, then there exist $\lambda_0 \geq 0$ and a nonzero $\mathbf{x}_0 \in \mathbb{R}^I_+$ such that*

$$\mathcal{A}\mathbf{x}_0^{N-1} = \lambda_0 \mathbf{x}_0^{[N-1]}. \tag{1.3.1}$$

Theorem 1.3.2 *If $\mathcal{A} \in RT_{N,I}$ is irreducible nonnegative, then $(\lambda_0; \mathbf{x}_0)$ in (1.3.1) satisfies:* (1) *λ_0 is an* H*-eigenvalue;* (2) *all components of $\mathbf{x}_0$ are positive;* (3) *if λ is an eigenvalue with nonnegative eigenvector, then $\lambda = \lambda_0$; moreover, the nonnegative eigenvector is unique up to a multiplicative constant;* (4) *if λ is an eigenvalue of $\mathcal{A}$, then $|\lambda| \leq \lambda_0$.*

More similar results can be found in [53, 54]. Ng et al. [37] derive an iterative method (denoted by NQZ) for computing the spectral radius of an irreducible nonnegative tensor. Zhang and Qi [55] establish an explicit linear convergence rate of the NQZ method for nonnegative tensors under certain conditions. Liu et al. [56] propose an inverse iterative method for computing the Perron pair of an irreducible nonnegative third-order tensor and prove that this method converges quadratically and is positivity preserving in the sense that the vectors approximating the Perron vector are strictly positive in each iteration. By combining the idea of Newton's method with the idea of the Noda iteration, Liu et al. [57] present a Newton-Noda iteration (NNI) for computing the Perron pair of a weakly irreducible nonnegative tensor. A survey on eigenvalues of nonnegative tensors can be found in [58]. Li and Ng [59, 60] extend the well-known column sum bound of the spectral radius for nonnegative matrices to the tensor case, and also derive an upper bound of the spectral radius for a nonnegative tensor via the largest eigenvalue of a symmetric tensor. Chen et al. [61] introduce three new classes of symmetric nonnegative tensors and discuss their properties and applications in the context of polynomial and tensor optimization.

1.3.2 Real Tensor Rank-One Approximations

The rank-one approximation of a real tensor is a special case of tensor low-rank approximations. The common tensor low-rank approximations consist of approximated canonical polyadic (CP) decompositions, approximated Tucker decompositions and approximated tensor train (TT) decompositions. We recommend

[7, 62, 63] and their references for thorough surveys of these three types of tensor decompositions. Note that when the order of tensors is two, then these three tensor decompositions reduce to the singular value decompositions (SVD). However, when the order of the tensors is larger than 2, these three kinds of tensor low-rank approximations have essential differences. The low CP-rank approximation is generally ill-posed [64] for the case of CP-rank larger than 1, contrary to the low Tucker-rank approximation.

The problem of the best rank-one approximation of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is to find a real scalar $\sigma \in \mathbb{R}$ and N unit vectors $\mathbf{x}_n \in \mathbb{R}^{I_n}$ with all n to minimize

$$\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} (a_{i_1 i_2 \ldots i_N} - \sigma \cdot (x_{1,i_1} x_{2,i_2} \ldots x_{N,i_N}))^2,$$

where x_{n,i_n} is the i_nth element of $\mathbf{x}_n \in \mathbb{R}^{I_n}$ for all i_n and n, and $\sigma \in \mathbb{R}$. Note that the best tensor rank-1 approximation is in fact always well-posed. The Kurash-Kuhn-Tucker (KKT) conditions for the rank-one approximation of real tensors are given below.

Definition 1.3.3 ([29]) Suppose that $\mathbf{x}_n \in \mathbb{R}^{I_n}$ satisfies $\|\mathbf{x}_n\|_2 = 1$ with all n and $\sigma \in \mathbb{R}$. For a given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, if $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ solves the following system of nonlinear equations $F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)_{-n} = \sigma \mathbf{x}_n$ with $F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)_{-n} = \mathcal{A} \times_1 \mathbf{x}_1^\top \cdots \times_{n-1} \mathbf{x}_{n-1}^\top \times_{n+1} \mathbf{x}_{n+1}^\top \cdots \times_N \mathbf{x}_N^\top$, then σ and the unit vectors $\mathbf{x}_n$ are a singular value of $\mathcal{A}$ and the mode-n singular vector associated to σ, respectively.

Another analogue of the Perron-Frobenius theorem is proved for nonnegative normalized singular pairs in [64], $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, defined in Definition 1.3.3, is called the normalized singular pair of $\mathcal{A}$. For all n, when the entries of $\mathbf{u}_n$ are nonnegative, $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is called the nonnegative normalized singular pair of $\mathcal{A}$. The singular value and singular value inclusion sets for tensors are investigated in [65].

There are several numerical methods for computing a tensor rank-one approximation, such as the alternating least squares (ALS) or higher-order power method (HOPM) [26, 27], the truncated higher-order singular value decomposition, optimization methods based on the product of several Grassmannian manifolds [66–68], semi-definite relaxation methods [69], and sequential rank-one approximation and projection [70]. Recently, Jiang and Kong [71] study the uniqueness of the best rank-one approximation of a tensor under the Frobenius norm. Espig and Khachatryan [72] analyze the convergence of the alternating least squares algorithm. For applications of best rank-one and low multilinear rank approximations, we refer to Cichocki et al. [2], Yang et al. [73], Konakli and Sudret [74], Shah et al. [75], da Silva et al. [76] and the references therein. Applications in machine learning can be found in [77–79].

Many scholars have researched the computation of the symmetric rank-one approximations of real symmetric tensors; see Friedland [80], Kofidis and Regalia

[81], Qi [82], Qi et al. [83], Hu et al. [84], Ni and Wang [85], Wang and Qi [86], and Jiang et al. [87] and the references therein.

For a given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_1 \times I_2 \times I_2}$, if the entries of $\mathcal{A}$ satisfy the symmetric property: $a_{i_1 i_2 i_3 i_4} = a_{i_2 i_1 i_3 i_4} = a_{i_1 i_2 i_4 i_3}$ with $i_1, i_2 = 1, 2, \ldots, I_1$ and $i_3, i_4 = 1, 2, \ldots, I_2$, then we call $\mathcal{A}$ partially symmetric.

The fourth-order partially symmetrity have received much attention [45, 88–93]. Zhang et al. [94] prove that the best rank-one approximation of a symmetric tensor is its best symmetric rank-one approximation. Similarly, we can prove that the best rank-one approximation of a fourth order partially symmetric tensor is its best partially symmetric rank-one approximation. The rank-one approximation problem of a partially symmetric tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_1 \times I_2 \times I_2}$ is to find a real scalar $\sigma \in \mathbb{R}$ and two unit vectors $\mathbf{x}_n \in \mathbb{R}^{I_n}$ with $n = 1, 2$ to minimize

$$\sum_{i_1, i_2 = 1}^{I_1} \sum_{i_3, i_4 = 1}^{I_2} (a_{i_1 i_2 i_3 i_4} - \sigma \cdot (x_{1,i_1} x_{1,i_2} x_{2,i_3} x_{2,i_4}))^2. \tag{1.3.2}$$

The minimization problem (1.3.2) is equivalent to finding two unit vectors $\mathbf{x}_n \in \mathbb{R}^{I_n}$ ($\|\mathbf{x}_n\|_2 = 1; n = 1, 2$) to maximize

$$\max |\mathcal{A}\mathbf{x}_1^2\mathbf{x}_2^2|, \tag{1.3.3}$$

with $\mathcal{A}\mathbf{x}_1^2\mathbf{x}_2^2 := \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_1^\top \times_3 \mathbf{x}_2^\top \times_4 \mathbf{x}_2^\top$. The biquadratic optimization problems arise from the strong ellipticity condition problem in solid mechanics [95–98] and the entanglement problem in quantum physics [88, 99]. Before considering the KKT conditions for the minimization problem (1.3.2), we introduce two notations:

$$\mathcal{A}\mathbf{x}_1^2\mathbf{x}_2 := \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_1^\top \times_4 \mathbf{x}_2^\top, \ \mathcal{A}\mathbf{x}_1\mathbf{x}_2^2 := \mathcal{A} \times_2 \mathbf{x}_1^\top \times_3 \mathbf{x}_2^\top \times_4 \mathbf{x}_2^\top.$$

For any maximizer $(\mathbf{x}_1, \mathbf{x}_2)$ of (1.3.3), by the optimality theory [100], there exist $\lambda, \sigma \in \mathbb{R}$ such that

$$\mathcal{A}\mathbf{x}_1^2\mathbf{x}_2 = \sigma\mathbf{x}_2, \ \mathcal{A}\mathbf{x}_1\mathbf{x}_2^2 = \lambda\mathbf{x}_1, \ \|\mathbf{x}_1\|_2 = \|\mathbf{x}_2\|_2 = 1. \tag{1.3.4}$$

The optimal conditions can further be simplified with (1.3.4) and $\lambda = \sigma$. If σ, $\mathbf{x}_1$ and $\mathbf{x}_2$ are real solutions of (1.3.4), then σ is said to be an M-eigenvalue of $\mathcal{A}$, and $\mathbf{x}_1$ and $\mathbf{x}_2$ are said to be the first and the second (left and right in [45]) M-eigenvector of $\mathcal{A}$, associated with σ, respectively.

There are several numerical methods for solving the minimization problem (1.3.2), such as ALS or HOPM [92] and semi-definite programming (SDP) relaxations [93].

1.3.3 Complex Tensor Rank-One Approximations

Entanglement has been identified as a resource central to quantum information processing and we are motivated to investigate its quantification in the bipartite and multipartite pure states [101]. A useful tool for quantifying the amount of entanglement of a state is given by the so-called *entanglement measures*. The geometric measure of entanglement is one of most natural entanglement measures for pure states in bipartite and multipartite systems [102, 103]. This measure is the injective tensor norm [104], which appears in the theory of operator algebra [105]. It also has applications in many-body physics [106, 107], entanglement witnesses [108, 109] and the study of quantum channel capacities [110–112].

The best rank-one approximation of complex tensors is the core problem for computing the geometric measure for pure states. For $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, we need to find a real scalar $\sigma \in \mathbb{R}$ and N unitary vectors $\mathbf{x}_n \in \mathbb{C}^{I_n}$ with all n to minimize

$$\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \left| a_{i_1 i_2 \ldots i_N} - \sigma \cdot (x_{1,i_1} x_{2,i_2} \ldots x_{N,i_N}) \right|^2,$$

where x_{n,i_n} is the i_nth element of $\mathbf{x}_n \in \mathbb{C}^{I_n}$ for all i_n and n, and $\sigma \in \mathbb{R}$. By means of the Wirtinger calculus of complex functions [113–116], the corresponding KKT conditions are given below.

Definition 1.3.4 For a given $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, let $\mathbf{v}_n \in \mathbb{C}^{I_n}$ be nonzero vectors with $\|\mathbf{v}_n\|_2 = 1$ $(n = 1, 2, \ldots, N)$ and let $\sigma \in \mathbb{R}$. If $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ solves the following system of nonlinear equations

$$F(\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_N)_{-n} = \sigma \mathbf{v}_n, \quad \overline{F}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)_{-n} = \sigma \overline{\mathbf{v}}_n,$$

where

$$\begin{aligned} F(\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_N)_{-n} &= \mathcal{A} \times_1 \overline{\mathbf{v}}_1^\top \cdots \times_{n-1} \overline{\mathbf{v}}_{n-1}^\top \times_{n+1} \overline{\mathbf{v}}_{n+1}^\top \cdots \times_N \overline{\mathbf{v}}_N^\top \\ &= \mathcal{A} \times_1 \mathbf{v}_1^* \cdots \times_{n-1} \mathbf{v}_{n-1}^* \times_{n+1} \mathbf{v}_{n+1}^* \cdots \times_N \mathbf{v}_N^*, \\ \overline{F}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)_{-n} &= \overline{\mathcal{A}} \times_1 \mathbf{v}_1^\top \cdots \times_{n-1} \mathbf{v}_{n-1}^\top \times_{n+1} \mathbf{v}_{n+1}^\top \cdots \times_N \mathbf{v}_N^\top, \end{aligned}$$

then $\mathbf{v}_n$ and σ are called the mode-n unitary eigenvector (the mode-n U-eigenvector) and unitary eigenvalue (U-eigenvalue) of $\mathcal{A}$, respectively. We call $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ a U-eigenpair of $\mathcal{A}$.

In particular, if $\mathcal{A}$ is complex symmetric, then all $\mathbf{v}_n$ are identical, denoted by $\mathbf{v}$, and $(\sigma; \mathbf{v})$ is a US-eigenpair of $\mathcal{A}$. Hu et al. [117] consider how to use the spectral theory of nonnegative tensors for computing the geometric measure of entanglement in multipartite pure states. Ni et al. [28] define the concept of the U-eigenvalue of a complex tensor, the US-eigenvalue of a complex symmetric tensor and the

best complex rank-one approximation. They also derive an upper bound on the number of distinct US-eigenvalues of a complex symmetric tensor. Based on the theory of the spherical optimization problem with complex variables, Ni and Bai [118] design an iterative algorithm for computing the US-eigenpairs of complex symmetric tensors. Che et al. [119] present the complex-valued neural networks for solving the quantum eigenvalue problem for multipartite pure states. Wang et al. [120] derive the partial orthogonal rank-one decomposition of complex symmetric tensors based on the Takagi factorization.

References

1. J. Cardoso, High-order contrasts for independent component analysis. Neural Comput. **11**(1), 157–192 (1999)
2. A. Cichocki, N. Lee, I. Oseledets, A. Phan, Q. Zhao, D. Mandic, Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. Found. Trends Mach. Learn. **9**(4-5), 249–429 (2016)
3. P. Comon, Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
4. P. Comon, Tensor decompositions: state of the art and applications, in *Mathematics in Signal Processing, V (Coventry, 2000)*. The Institute of Mathematics and its Applications Conference Series, vol. 71 (Oxford University, Oxford, 2002), pp. 1–24
5. C. Nikias, J. Mendel, Signal processing with higher-order spectra. IEEE Signal Process. Mag. **10**(3), 10–37 (1993)
6. P. Comon, G. Golub, L. Lim, B. Mourrain, Symmetric tensors and symmetric tensor rank. SIAM J. Matrix Anal. Appl. **30**(3), 1254–1279 (2008)
7. T. Kolda, B. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
8. A. Cichocki, R. Zdunek, A. Phan, S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation* (Wiley, New York, 2009)
9. Y. Yang, Q. Yang, *A Study on Eigenvalues of Higher-Order Tensors and Related Polynomial Optimization Problems* (Science, Beijing, 2015)
10. L. Qi, Z. Luo, *Tensor Analysis: Spectral Theory and Special Tensors* (Society of Industrial and Applied Mathematics, Philadelphia, 2017)
11. Y. Wei, W. Ding, *Theory and Computation of Tensors: Multi-Dimensional Arrays* (Academic, Amsterdam, 2016)
12. L. Qi, H. Chen, Y. Chen, *Tensor Eigenvalues and their Applications.*, vol. 39 (Springer: Singapore, 2018)
13. F. Chung, *Spectral Graph Theory* (American Mathematical Society, New York, 1997)
14. C. Berge, Graphs and Hypergraphs, in *North-Holland Mathematical Library*, vol. 45 (North-Holland, Amsterdam, 1976)
15. S. Bulo, M. Pelillo, New bounds on the clique number of graphs based on spectral hypergraph theory, in *Learning and Intelligent Optimization* (Springer, New York, 2009), pp. 45–58
16. J. Cooper, A. Dutle, Spectra of uniform hypergraphs. Linear Algebra Appl. **436**(9), 3268–3292 (2012)
17. K. Pearson, T. Zhang, On spectral hypergraph theory of the adjacency tensor. Graphs and Combinatorics **30**(5), 1233–1248 (2014)
18. J. Xie, A. Chang, On the Z-eigenvalues of the adjacency tensors for uniform hypergraphs. Linear Algebra Appl. **439**(8), 2195–2204 (2013)

19. S. Hu, L. Qi, J. Shao, Cored hypergraphs, power hypergraphs and their Laplacian H-eigenvalues. Linear Algebra Appl. **439**(10), 2980–2998 (2013)
20. S. Hu, L. Qi, J. Xie, The largest Laplacian and signless Laplacian H-eigenvalues of a uniform hypergraph. Linear Algebra Appl. **469**, 1–27 (2015)
21. V. Nikiforov, Analytic methods for uniform hypergraphs. Linear Algebra Appl. **457**, 455–535 (2014)
22. L. Qi, J. Shao, Q. Wang, Regular uniform hypergraphs, s-cycles, s-paths and their largest Laplacian H-eigenvalues. Linear Algebra Appl. **443**, 215–227 (2014)
23. H. Lu, K. Plataniotis, A. Venetsanopoulos, *Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data* (CRC, Boca Raton, 2013)
24. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
25. L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000)
26. L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(r_1, r_2, ..., r_n) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000)
27. T. Zhang, G. Golub, Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. **23**(2), 534–550 (2001)
28. G. Ni, L. Qi, M. Bai, Geometric measure of entanglement and U-eigenvalues of tensors. SIAM J. Matrix Anal. Appl. **35**(1), 73–87 (2014)
29. L. Lim, Singular values and eigenvalues of tensors: a variational approach, in *IEEE CAMSAP 2005: First International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (IEEE, Piscataway, 2005), pp. 129–132
30. L. Qi, Symmetric nonnegative tensors and copositive tensors. Linear Algebra Appl. **439**(1), 228–238 (2013)
31. Z. Chen, L. Qi, Circulant tensors with applications to spectral hypergraph theory and stochastic process. J. Ind. Manag. Optim. **12**(4), 1227–1247 (2013)
32. S. Dow, P. Gibson, Permanents of d-dimensional matrices. Linear Algebra Appl. **90**, 133–145 (1987)
33. K. Chang, K. Pearson, T. Zhang, Perron-Frobenius theorem for nonnegative tensors. Commun. Math. Sci. **6**(2), 507–520 (2008)
34. A. Berman, R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Society for Industrial and Applied Mathematics, Philadelphia, 1994)
35. S. Hu, G. Li, L. Qi, Y. Song, Finding the maximum eigenvalue of essentially nonnegative symmetric tensors via sum of squares programming. J. Optim. Theory Appl. **158**(3), 713–738 (2013)
36. G. Li, L. Qi, G. Yu, The Z-eigenvalues of a symmetric tensor and its application to spectral hypergraph theory. Numer. Linear Algebra Appl. **20**(6), 1001–1029 (2013)
37. M. Ng, L. Qi, G. Zhou, Finding the largest eigenvalue of a nonnegative tensor. SIAM J. Matrix Anal. Appl. **31**(3), 1090–1099 (2009)
38. L. Qi, G. Yu, E. Wu, Higher order positive semi-definite diffusion tensor imaging. SIAM J. Imag. Sci. **3**(3), 416–433 (2010)
39. F. Bohnetwaldraff, D. Braun, O. Giraud, Tensor eigenvalues and entanglement of symmetric states. Phys. Rev. A **94**(4), 042324 (2016)
40. G. Gaeta, E. Virga, Octupolar order in three dimensions. Eur. Phys. J. E **39**(11), 113 (2016)
41. K. Kannike, Vacuum stability of a general scalar potential of a few fields. Eur. Phys. J. C **76**(6), 324 (2016)
42. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)
43. T. Kolda, J. Mayo, Shifted power method for computing tensor eigenpairs. SIAM J. Matrix Anal. Appl. **32**(4), 1095–1124 (2011)
44. L. Qi, Y. Wang, E. Wu, D-eigenvalues of diffusion kurtosis tensors. J. Comput. Appl. Math. **221**(1), 150–157 (2008)

45. L. Qi, H. Dai, D. Han, Conditions for strong ellipticity and M-eigenvalues. Front. Math. China **4**(2), 349–364 (2009)
46. K. Chang, K. Pearson, T. Zhang, On eigenvalue problems of real symmetric tensors. J. Math. Anal. Appl. **350**(1), 416–422 (2009)
47. C. Cui, Y. Dai, J. Nie, All real eigenvalues of symmetric tensors. SIAM J. Matrix Anal. Appl. **35**(4), 1582–1601 (2014)
48. T. Kolda, J. Mayo, An adaptive shifted power method for computing generalized tensor eigenpairs. SIAM J. Matrix Anal. Appl. **35**(4), 1563–1581 (2014)
49. W. Ding, Y. Wei, Generalized tensor eigenvalue problems. SIAM J. Matrix Anal. Appl. **36**(3), 1073–1099 (2015)
50. L. Chen, L. Han, L. Zhou, Computing tensor eigenvalues via homotopy methods. SIAM J. Matrix Anal. Appl. **37**(1), 290–319 (2016)
51. J. Chang, Y. Chen, L. Qi, Computing eigenvalues of large scale sparse tensors arising from a hypergraph. SIAM J. Sci. Comput. **38**(6), A3618–A3643 (2016)
52. K. Batselier, N. Wong, *A QR Algorithm for Symmetric Tensors* (2014). ArXiv preprint:1411.1926v1
53. Q. Yang, Y. Yang, Further results for Perron-Frobenius theorem for nonnegative tensors II. SIAM J. Matrix Anal. Appl. **32**(4), 1236–1250 (2011)
54. Y. Yang, Q. Yang, Further results for Perron-Frobenius theorem for nonnegative tensors. SIAM J. Matrix Anal. Appl. **31**(5), 2517–2530 (2010)
55. L. Zhang, L. Qi, Linear convergence of an algorithm for computing the largest eigenvalue of a nonnegative tensor. Numer. Linear Algebra Appl. **19**(5), 830–841 (2012)
56. C. Liu, C. Guo, W. Lin, A positivity preserving inverse iteration for finding the perron pair of an irreducible nonnegative third order tensor. SIAM J. Matrix Anal. Appl. **37**(3), 911–932 (2016)
57. C. Liu, C. Guo, W. Lin, Newton-Noda iteration for finding the Perron pair of a weakly irreducible nonnegative tensor. Numer. Math. **137**(1), 63–90 (2017)
58. K. Chang, L. Qi, T. Zhang, A survey on the spectral theory of nonnegative tensors. Numer. Linear Algebra Appl. **20**(6), 891–912 (2013)
59. W. Li, M. Ng, The perturbation bound for the spectral radius of a nonnegative tensor. Adv. Numer. Anal., Article ID 109525, 10 (2014)
60. W. Li, M. Ng, Some bounds for the spectral radius of nonnegative tensors. Numer. Math. **130**(2), 315–335 (2015)
61. B. Chen, S. He, Z. Li, S. Zhang, On new classes of nonnegative symmetric tensors. SIAM J. Optim. **27**, 292–318 (2017)
62. A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, H. Phan, Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Process. Mag. **32**(2), 145–163 (2015)
63. L. Grasedyck, D. Kressner, C. Tobler, A literature survey of low-rank tensor approximation techniques. GAMM-Mitt. **36**(1), 53–78 (2013)
64. Y. Qi, P. Comon, L. Lim, Uniqueness of nonnegative tensor approximations. IEEE Trans. Inf. Theory **62**(4), 2170–2183 (2016)
65. Y. Miao, C. Li, Y. Wei, Z-singular value and Z-singular value inclusion sets for tensors. Jpn. J. Ind. Appl. Math. **36**, 1055–1087 (2019)
66. L. Eldén, B. Savas, A Newton-Grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor. SIAM J. Matrix Anal. Appl. **31**(2), 248–271 (2009)
67. M. Ishteva, P. Absil, S. Van Huffel, L. De Lathauwer, Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. SIAM J. Matrix Anal. Appl. **32**(1), 115–135 (2011)
68. B. Savas, L. Lim, Quasi-Newton methods on Grassmannians and multilinear approximations of tensors. SIAM J. Sci. Comput. **32**(6), 3352–3393 (2010)
69. J. Nie, L. Wang, Semidefinite relaxations for best rank-1 tensor approximations. SIAM J. Matrix Anal. Appl. **35**(3), 1155–1179 (2014)

70. A. da Silva, P. Comon, A. de Almeida, A finite algorithm to compute rank-1 tensor approximations. Signal Processing Lett. **23**, 959–963 (2016)
71. Y. Jiang, K. Xu, On the uniqueness and perturbation to the best rank-one approximation of a tensor. SIAM J. Matrix Anal. Appl. **36**(2), 775–792 (2015)
72. M. Espig, A. Khachatryan, *Convergence of Alternating Least Squares Optimisation for Rank-One Approximation to High Order Tensors* (2015). ArXiv preprint:1503.05431
73. Y. Yang, Y. Feng, X. Huang, J. Suykens, Rank-1 tensor properties with applications to a class of tensor optimization poblems. SIAM J. Optim. **26**(1), 171–196 (2016)
74. K. Konakli, B. Sudret, Uncertainty quantification in high-dimensional spaces with low-rank tensor approximations, in *1st International Conference on Uncertainty Quantification in Computational Sciences and Engineering* (2015)
75. P. Shah, N. Rao, G. Tang, *Optimal Low-Rank Tensor Recovery from Separable Measurements: Four Contractions Suffice* (2015). ArXiv preprint: 1505.04085
76. A. da Silva, P. Comon, A. de Almeida, *Rank-1 Tensor Approximation Methods and Application to Deflation* (2015). ArXiv preprint:1508.05273
77. A. Novikov, M. Trofimov, I. Oseledets, *Exponential Machines* (2016). ArXiv preprint:1605.03795
78. E. Stoudenmire, D. Schwab, Supervised learning with quantum-inspired tensor networks (2016). ArXiv preprint:1605.05775
79. E. Stoudenmire, D. Schwab, Supervised learning with tensor networks, in *Advances in Neural Information Processing Systems*, vol. 29 (Curran Associates Inc., New York, 2016), pp. 4799–4807
80. S. Friedland, Best rank one approximation of real symmetric tensors can be chosen symmetric. Front. Math. China **8**(1), 19–40 (2013)
81. E. Kofidis, P. Regalia, On the best rank-1 approximation of higher-order supersymmetric tensors. SIAM J. Matrix Anal. Appl. **23**(3), 863–884 (2002)
82. L. Qi, The best rank-one approximation ratio of a tensor space. SIAM J. Matrix Anal. Appl. **32**(2), 430–442 (2011)
83. L. Qi, F. Wang, Y. Wang, Z-eigenvalue methods for a global polynomial optimization problem. Math. Program. **118**(Serires A), 301–316 (2009)
84. S. Hu, Z. Huang, L. Qi, Finding the extreme Z-eigenvalues of tensors via a sequential semidefinite programming method. Numer. Linear Algebra Appl. **20**(6), 972–984 (2013)
85. G. Ni, Y. Wang, On the best rank-1 approximation to higher-order symmetric tensors. Math. Comput. Model. **46**(9), 1345–1352 (2007)
86. Y. Wang, L. Qi, On the successive supersymmetric rank-1 decomposition of higher-order supersymmetric tensors. Numer. Linear Algebra Appl. **14**(6), 503–519 (2007)
87. B. Jiang, S. Ma, S. Zhang, Tensor principal component analysis via convex optimization. Math. Program. **150**(2), 423–457 (2015)
88. G. Dahl, J. Leinaas, J. Myrheim, E. Ovrum, A tensor product matrix approximation problem in quantum physics. Linear Algebra Appl. **420**(2-3), 711–725 (2007)
89. D. Han, H. Dai, L. Qi, Conditions for strong ellipticity of anisotropic elastic materials. J. Elast. **97**(1), 1–13 (2009)
90. C. Ling, J. Nie, L. Qi, Y. Ye, Biquadratic optimization over unit spheres and semidefinite programming relaxations. SIAM J. Optim. **20**(3), 1286–1310 (2009)
91. A. So, Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems. Math. Program. **129**(2), 357–382 (2011)
92. Y. Wang, L. Qi, X. Zhang, A practical method for computing the largest M-eigenvalue of a fourth-order partially symmetric tensor. Numer. Linear Algebra Appl. **16**(7), 589–601 (2009)
93. X. Zhang, C. Ling, L. Qi, Semidefnite relaxation bounds for bi-quadratic optimization problems with quadratic constraints. J. Global Optim. **49**(2), 293–311 (2011)
94. X. Zhang, C. Ling, L. Qi, The best rank-1 approximation of a symmetric tensor and related spherical optimization problems. SIAM J. Matrix Anal. Appl. **33**(3), 806–821 (2012)
95. J. Knowles, E. Sternberg, On the ellipticity of the equations of nonlinear elastostatics for a special material. J. Elast. **5**(3-4), 341–361 (1975)

96. J. Knowles, E. Sternberg, On the failure of ellipticity of the equations for finite elastostatic plane strain. Arch. Ration. Mech. Anal. **63**(63), 321–336 (1976)
97. P. Rosakis, Ellipticity and deformations with discontinuous gradients in finite elastostatics. Arch. Ration. Mech. Anal. **109**(109), 1–37 (1990)
98. Y. Wang, M. Aron, A reformulation of the strong ellipticity conditions for unconstrained hyperelastic media. J. Elast. **44**(1), 89–96 (1996)
99. N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **48**(10), 696–702 (1935)
100. J. Nocedal, S. Wright, *Numerical Optimization* (Springer, New York, 1999)
101. M. Nielsen, I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, Cambridge, 2000)
102. A. Shimony, Degree of entanglementa. Ann. N. Y. Acad. Sci. **755**(1), 675–679 (1995)
103. T. Wei, P. Goldbart, Geometric measure of entanglement and applications to bipartite and multipartite quantum states. Phys. Rev. A **68**(4), 042307 (2003)
104. M. Hayashi, D. Markham, M. Murao, M. Owari, S. Virmani, The geometric measure of entanglement for a symmetric pure state with non-negative amplitudes. J. Math. Phys. **50**(12), 122104 (2009)
105. A. Defant, K. Floret, *Tensor Norms and Operator Ideals*, in *North-Holland Mathematics Studies*, vol. 176 (North-Holland Publishing Co., Amsterdam, 1993)
106. D. Markham, J. Anders, V. Vedral, M. Murao, A. Miyake, Survival of entanglement in thermal states. Europhys. Lett. **81**(4), 40006 (2008)
107. Y. Nakata, D. Markham, M. Murao, Thermal robustness of multipartite entanglement of the one-dimensional spin-1/2 XY model. Phys. Rev. A **79**(4), 042313 (2009)
108. O. Gühne, G. Tóth, Entanglement detection. Phys. Rep. **474**(1), 1–75 (2009)
109. M. Hayashi, D. Markham, M. Murao, M. Owari, S. Virmani, Entanglement of multiparty-stabilizer, symmetric, and antisymmetric states. Phys. Rev. A **77**(1), 012104 (2008)
110. D. Buhr, M. Carrington, T. Fugleberg, R. Kobes, G. Kunstatter, D. McGillis, C. Pugh, D. Ryckman, Geometrical entanglement of highly symmetric multipartite states and the Schmidt decomposition. J. Phys. A Math. Theor. **44**(36), 365305 (2011)
111. C. González-Guillén, Multipartite maximally entangled states in symmetric scenarios. Phys. Rev. A **86**(2), 022304 (2012)
112. R. Werner, A. Holevo, Counterexample to an additivity conjecture for output purity of quantum channels. J. Math. Phys. **43**(9), 4353–4357 (2002)
113. D. Brandwood, A complex gradient operator and its application in adaptive array theory. IEE Proc. F: Commun. Radar and Signal Process. **130**(1), 11–16 (1983)
114. K. Kreutz-Delgado, *The Complex Gradient Operator and the CR-Calculus* (2009). ArXiv preprint:0906.4835
115. R. Remmert, *Theory of Complex Functions* (Springer, New York, 1991)
116. A. Van Den Bos, Complex gradient and Hessian, in *IEEE Proceedings Vision, Image and Signal Processing*, vol. 141 (1994), pp. 380–383
117. S. Hu, L. Qi, G. Zhang, Computing the geometric measure of entanglement of multipartite pure states by means of non-negative tensors. Phys. Rev. A **93**(1), 012304 (2016)
118. G. Ni, M. Bai, Spherical optimization with complex variables for computing US-eigenpairs. Comput. Optim. Appl. **65**(3), 1–22 (2016)
119. M. Che, L. Qi, Y. Wei, G. Zhang, Geometric measures of entanglement in multipartite pure states via complex-valued neural networks. Neurocomputing **313**, 25–38 (2018)
120. X. Wang, M. Che, Y. Wei, Partial orthogonal rank-one decomposition of complex symmetric tensors based on the Takagi factorization. J. Comput. Appl. Math. **332**, 56–71 (2018)

Chapter 2
The Pseudo-Spectrum Theory

Pseudo-spectra, developed and popularized mainly by Trefethen and his co-authors [1–5], are an important tool for assessing the global sensitivity of matrix eigenvalues to perturbations in the matrix. Most research has emphasised on pseudospectra of standard and generalized eigenvalue problems. Pseudospectra for square matrix polynomials are defined and characterized by Tisseur and Higham [6], Higham and Tisseur [7], and Lancaster and Psarrakos [8], for the generalization of the ϵ-pseudo-spectrum to matrix polynomials.

To our best knowledge, an investigation of ϵ-pseudo-spectra for tensors and tensor polynomials seems to be little in the literature. The main purpose of this chapter is to bridge this gap by introducing ϵ-pseudo-spectra for tensors and tensor polynomial eigenvalue problems, and develop some fundamental properties.

2.1 Preliminaries

We recall the definition of eigenvalues and associated eigenvectors of $\mathcal{A} \in CT_{N,I}$. This definition is presented in [9] when $\mathcal{A} \in RT_{N,I}$ is symmetric.

Definition 2.1.1 Let $\mathcal{A} \in CT_{N,I}$. If there exist a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ and $\lambda \in \mathbb{C}$ such that

$$\mathcal{A}\mathbf{x}^{N-1} = \lambda \mathbf{x}^{[N-1]},$$

then λ and $\mathbf{x}$ are called an eigenvalue and the associated eigenvector of $\mathcal{A}$, respectively.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_2

We call $(\lambda; \mathbf{x})$ as an eigenpair of $\mathcal{A}$. The spectrum $\Lambda(\mathcal{A})$ of $\mathcal{A}$ is defined as

$$\Lambda(\mathcal{A}) = \{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } \mathcal{A}\}.$$

Furthermore, if $\mathcal{A} \in RT_{N,I}$, $\mathbf{x} \in \mathbb{R}^I$ and $\lambda \in \mathbb{R}$, then $(\lambda; \mathbf{x})$ is called an H-eigenpair of $\mathcal{A}$.

We now recall the definition of determinant of $\mathcal{A} \in CT_{N,I}$ introduced in Hu et al. [10]. If $\mathcal{A} \in RT_{N,I}$ is symmetric, then the following definition reduces to the definition of determinant in [9].

Definition 2.1.2 ([10, Definition 1.2]) The determinant of Nth-order I-dimensional complex tensors denoted by DET, is defined as the irreducible polynomial with variables $v_{i_1 \ldots i_N}$ for all i_n and n such that it is the resultant of the polynomial system

$$\mathcal{V}\mathbf{x}^{N-1} = \mathbf{0}_I,$$

where $\mathcal{V} \in CT_{N,I}$.

The value of the determinant for a given tensor $\mathcal{A} \in CT_{N,I}$ is denoted by $\det(\mathcal{A})$, and is defined as the evaluation of DET at the point $v_{i_1 \ldots i_N} = a_{i_1 \ldots i_N}$ for all i_n and n. For any $\mathcal{A} \in CT_{N,I}$, if $\det(\mathcal{A}) \neq 0$, then $\mathcal{A}$ is said to be nonsingular.

An interesting property of the determinant is that, for $\mathcal{A} \in CT_{N,I}$, $\det(\mathcal{A}) = 0$ if and only if $\mathcal{A}\mathbf{x}^{N-1} = \mathbf{0}_I$ has a nonzero solution in $\mathbb{C}^I$ [10, Theorem 3.1]. We know that $\lambda \in \Lambda(\mathcal{A})$ if and only if $\det(\mathcal{A} - \lambda\mathcal{I}) = 0$. It is clear that $\det(\mathcal{A} - \lambda\mathcal{I})$ is a scalar polynomial with complex coefficients. This together with the fundamental theorem of algebra gives us

$$\Lambda(\mathcal{A}) \neq \emptyset. \tag{2.1.1}$$

Another important property of the determinant of any symmetric tensor $\mathcal{A} \in RT_{N,I}$ has been established by Qi [9, Theorem 1 (b)], which is stated as $\det(\mathcal{A}) = \prod_{\lambda_i \in \Lambda(\mathcal{A})} \lambda_i$. The result also holds for the case of complex-valued tensors [10, Corollary 6.5].

For $(L+1)$ given tensors $\mathcal{A}_l \in CT_{N,I}$ with all l, we define a tensor polynomial $\mathcal{P}$ corresponding to the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ as

$$\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^L\mathcal{A}_L, \tag{2.1.2}$$

for all $z \in \mathbb{C}$.

Definition 2.1.3 ([11, Definition 2.5]) For a tensor polynomial $\mathcal{P}$ as in (2.1.2), if there exist a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ and $\lambda \in \mathbb{C}$ such that $\mathcal{P}(\lambda)\mathbf{x}^{N-1} = \mathbf{0}_I$, then λ and $\mathbf{x}$ are called an eigenvalue and the associated eigenvector of $\mathcal{P}$, respectively. We call $(\lambda; \mathbf{x})$ as an eigenpair of $\mathcal{P}$. The set of eigenvalues of the tensor polynomial $\mathcal{P}$ is defined by $\Lambda(\mathcal{P}) = \{\lambda : \det(\mathcal{P}(\lambda)) = 0\} = \{\lambda : \lambda \text{ is an eigenvalue of } \mathcal{P}\}$.

Furthermore, if $\mathbf{x} \in \mathbb{R}^I$, $\lambda \in \mathbb{R}$ and $\mathcal{A}_l \in RT_{N,I}$ with all l, then $(\lambda; \mathbf{x})$ is called an H-eigenpair of $\mathcal{P}$. We next introduce the definition of a regular tensor polynomial corresponding to any given tensor tuple.

Definition 2.1.4 ([11, Definition 2.6]) For a tensor polynomial $\mathcal{P}$ as in (2.1.2), if for all $\lambda \in \mathbb{C}$, $\det(\mathcal{P}(\lambda)) = 0$ holds, then we call $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ as singular. Otherwise the tensor tuple is said to be regular.

Furthermore, we say the tensor polynomial $\mathcal{P}$ is singular (resp. regular) if the corresponding tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ is singular (resp. regular).

According to Definition 2.1.4, if there exists $\hat{\lambda} \in \mathbb{C}$ such that $\det(\mathcal{P}(\hat{\lambda})) \neq 0$, then $\mathcal{P}$ is a regular tensor polynomial. In this chapter, we only consider the tensor polynomial eigenvalue problem where its associated $\mathcal{P}$ is regular.

We can choose another tensor tuple $\{\tilde{\mathcal{A}}_0, \tilde{\mathcal{A}}_1, \ldots, \tilde{\mathcal{A}}_L\}$ such that $\tilde{\mathcal{A}}_L = \sum_{l=0}^{L} \hat{\lambda}^l \mathcal{A}_l$ and there is a one-to-one map between $\Lambda(\mathcal{P})$ and $\Lambda(\tilde{\mathcal{P}})$, where $\det(\tilde{\mathcal{A}}_L) \neq 0$ and $\tilde{\mathcal{P}}(z) = \tilde{\mathcal{A}}_0 + z\tilde{\mathcal{A}}_1 + \cdots + z^L\tilde{\mathcal{A}}_L$ with all $z \in \mathbb{C}$. Hence, we can assume that $\mathcal{A}_L$ is nonsingular, that is, $\det(\mathcal{A}_L) \neq 0$. As shown in [11], when $\mathcal{A}_L$ is nonsingular, $\Lambda(\mathcal{P})$ is a finite subset of $\mathbb{C}$.

It is well known that the matrix eigenvalue and the generalized eigenvalue problem are special cases of the matrix polynomial eigenvalue problem [12]. For the tensor polynomial eigenvalue problem, we have the following conclusions.

Remark 2.1.1 For $(L + 1)$ given tensors $\mathcal{A}_l \in CT_{N,I}$ with all l, suppose that $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^L\mathcal{A}_L$ with $z \in \mathbb{C}$. The following three special cases of Definition 2.1.3 have been discussed in the literature.

(a) When $L = 1$, $\mathcal{A}_1 = -\mathcal{I}$ and $\mathcal{A}_0 \in RT_{N,I}$ is symmetric, Definition 2.1.3 reduces to the eigenvalue problem with real symmetric tensors proposed in [9].
(b) When $L = 1$ and $\mathcal{A}_l \in RT_{N,I}$ with $l = 0, 1$, Definition 2.1.3 has been considered by Chang et al. [13] where $(\lambda; \mathbf{x})$ is called a generalized eigenpair of the tensor pair $\{\mathcal{A}_0, \mathcal{A}_1\}$. For a related work, see [14]. Furthermore, Ding and Wei [15] present several properties of the generalized eigenpairs of the tensor pair $\{\mathcal{A}_0, \mathcal{A}_1\}$ with $\mathcal{A}_l \in CT_{N,I}$ $(l = 0, 1)$.
(c) When $L = 2$, $(\lambda; \mathbf{x})$, satisfying Definition 2.1.3, is called an eigenpair of the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2\}$.

2.2 Pseudo-Spectrum of a Complex Tensor

We first define the ϵ-pseudo-spectrum of $\mathcal{A} \in CT_{N,I}$ in the Frobenius norm and characterize its fundamental properties. We reveal the relationship between the ϵ-pseudo-spectrum and the spectrum (i.e., the set of all eigenvalues). Finally, we define the ϵ-pseudo-spectral abscissa and ϵ-pseudo-spectral radius with respect to the Frobenius norm.

2.2.1 Definition and Properties

The unstructured perturbations bounds in the Frobenius and spectral norms lead to the same matrix ϵ-pseudo-spectrum. Lim [16] provides some definitions for the ϵ-pseudo-spectrum of a tensor by its smallest singular value [17]. However, we define the ϵ-pseudo-spectrum of a tensor in term of unstructured perturbations bounds in the Frobenius norm. These two definitions of the ϵ-pseudo-spectrum of a tensor are, in general, independent (see Remark 2.2.3).

Definition 2.2.1 (ϵ-Pseudo-Spectrum for Tensors) Let $\epsilon \geq 0$. Then the ϵ-pseudo-spectrum of $\mathcal{A} \in CT_{N,I}$ is defined as

$$\Lambda_\epsilon(\mathcal{A}) = \Big\{\lambda \in \mathbb{C} : (\mathcal{A}+\mathcal{E})\mathbf{x}^{N-1} = \lambda\,\mathbf{x}^{[N-1]} \text{ for some } \mathcal{E} \in CT_{N,I}$$
$$\text{with } \|\mathcal{E}\|_F \leq \epsilon \text{ and nonzero vectors } \mathbf{x} \in \mathbb{C}^I\Big\}.$$

When $\epsilon = 0$, we have $\Lambda_0(\mathcal{A}) = \Lambda(\mathcal{A})$. If $\mathcal{A}$ is a matrix, then the definition of the ϵ-pseudo-spectrum $\Lambda_\epsilon(\mathcal{A})$ reduces to the ϵ-pseudo-spectrum of $\mathcal{A}$ with respect to unstructured perturbations bounds in the Frobenius norm [2]. We now present some fundamental properties of the ϵ-pseudo-spectrum $\Lambda_\epsilon(\mathcal{A})$.

Proposition 2.2.1 *Let $\mathcal{A} \in CT_{N,I}$ and $\epsilon \geq 0$. Then the following results hold.*

(a) *For all $0 \leq \epsilon_1 \leq \epsilon_2$, we have $\Lambda_{\epsilon_1}(\mathcal{A}) \subseteq \Lambda_{\epsilon_2}(\mathcal{A})$.*

(b) *If there exists a positive integer $P \in \{1, 2, \ldots, I\}$ such that $\mathcal{A}[P+1 : I|\underbrace{1:P|\ldots|1:P}_{N-1}] \in \mathbb{C}^{(I-P)\times P\times\cdots\times P}$ and $\mathcal{A}[1:P|P+1:I|\ldots|P+1:I] \in \mathbb{C}^{P\times(I-P)\times\cdots\times(I-P)}$ are zero tensors, then,*

$$\Lambda_\epsilon(\mathcal{A}_1) \cup \Lambda_\epsilon(\mathcal{A}_2) \subseteq \Lambda_\epsilon(\mathcal{A}),$$

where $\mathcal{A}_1 = \mathcal{A}[1:P|1:P|\ldots|1:P] \in CT_{N,P}$ and $\mathcal{A}_2 = \mathcal{A}[P+1:I|P+1:I|\ldots|P+1:I] \in CT_{N,I-P}$.

(c) *If $\alpha, \beta \in \mathbb{C}$, then $\Lambda_{\epsilon|\beta|}(\alpha\mathcal{I} + \beta\mathcal{A}) = \{\alpha\} + \beta\Lambda_\epsilon(\mathcal{A})$.*

(d) **(Equivalent definition via rank-one perturbation)** *A complex number λ belongs to the ϵ-pseudo-spectrum of $\mathcal{A}$ if and only if*

$$(\mathcal{A}+\mathcal{E})\mathbf{x}^{N-1} = \lambda\mathbf{x}^{[N-1]}, \tag{2.2.1}$$

for some rank-one tensors $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$ and nonzero vectors $\mathbf{x} \in \mathbb{C}^I$.

(e) *For the spectrum and ϵ-pseudo-spectrum, we have $\cap_{\epsilon>0}\Lambda_\epsilon(\mathcal{A}) = \Lambda(\mathcal{A})$.*

(f) *Let $\mathcal{A}_i = \mathcal{A}[1:i|1:i|\ldots|1:i] \in CT_{N,i}$ with all i. Then we have*

$$\Lambda_\epsilon(\mathcal{A}_1) \subseteq \Lambda_\epsilon(\mathcal{A}_2) \subseteq \cdots \subseteq \Lambda_\epsilon(\mathcal{A}_I) = \Lambda_\epsilon(\mathcal{A}).$$

Proof Parts (a), (e) and (f) are immediate from the definition of the ϵ-pseudo-spectrum.

[Proof of (b)]. If $\lambda \in \Lambda_\epsilon(\mathcal{A}_1)$, then there exist a tensor $\mathcal{E}_1 \in CT_{N,P}$ with $\|\mathcal{E}_1\|_F \leq \epsilon$ and a nonzero vector $\mathbf{x} \in \mathbb{C}^P$ such that

$$(\mathcal{A}_1 + \mathcal{E}_1)\mathbf{x}^{N-1} = \lambda \mathbf{x}^{[N-1]}. \tag{2.2.2}$$

Let $\mathbf{y} = (\mathbf{x}, \mathbf{0}_{I-P})$ and $\mathcal{E} \in CT_{N,I}$ be a tensor such that $\mathcal{E}[1:P|1:P|\ldots|1:P] = \mathcal{E}_1$ and it takes the zero value otherwise. Then, we have $\|\mathcal{E}\|_F = \|\mathcal{E}_1\|_F \leq \epsilon$. Note that the entries of $\mathcal{A}(P+1:I|1:P|\ldots,1:P) \in \mathbb{R}^{(I-P)\times P\times\cdots\times P}$ are zeros. From (2.2.2), one has $(\mathcal{A}+\mathcal{E})\mathbf{y}^{N-1} = \lambda \mathbf{y}^{[N-1]}$. Then $\lambda \in \Lambda_\epsilon(\mathcal{A})$, which implies that $\Lambda_\epsilon(\mathcal{A}_1) \subseteq \Lambda_\epsilon(\mathcal{A})$. Similarly, we can prove $\Lambda_\epsilon(\mathcal{A}_2) \subseteq \Lambda_\epsilon(\mathcal{A})$. Thus part (b) follows.

[Proof of (c)]. If $\beta = 0$, then part (d) has been established in Corollary 3 of [9]. We only need to consider the case when $\beta \neq 0$. Let $\lambda \in \Lambda_{\epsilon|\beta|}(\alpha\mathcal{I} + \beta\mathcal{A})$, there exist $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq |\beta|\epsilon$, and a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that

$$(\alpha\mathcal{I} + \beta\mathcal{A} + \mathcal{E})\mathbf{x}^{N-1} = \lambda \mathbf{x}^{[N-1]}.$$

We prove that

$$\left(\mathcal{A} + \frac{\mathcal{E}}{\beta}\right)\mathbf{x}^{N-1} = \frac{\lambda - \alpha}{\beta}\mathbf{x}^{[N-1]}.$$

It is obvious that $\|\mathcal{E}/\beta\|_F \leq \epsilon$ and there exists $\tilde{\lambda} \in \Lambda_\epsilon(\mathcal{A})$ such that $\tilde{\lambda} = (\lambda - \alpha)/\beta$. Thus $\Lambda_{\epsilon|\beta|}(\alpha\mathcal{I} + \beta\mathcal{A}) \subseteq \{\alpha\} + \beta\Lambda_\epsilon(\mathcal{A})$.

To prove the reverse inclusion, let $\lambda \in \Lambda_\epsilon(\mathcal{A})$ and $\tilde{\lambda} = \beta\lambda + \alpha$. Then there exist a tensor $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$, and a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that $(\mathcal{A}+\mathcal{E})\mathbf{x}^{N-1} = \lambda\mathbf{x}^{[N-1]}$. It implies that

$$(\alpha\mathcal{I} + \beta\mathcal{A} + \beta\mathcal{E})\mathbf{x}^{N-1} = (\beta\lambda + \alpha)\mathbf{x}^{[N-1]} = \tilde{\lambda}\mathbf{x}^{[N-1]}.$$

$\tilde{\lambda} \in \Lambda_{\epsilon|\beta|}(\alpha\mathcal{I} + \beta\mathcal{A})$, and the result follows.

[Proof of (d)] Let λ be an element in the ϵ-pseudo-spectrum $\Lambda_\epsilon(\mathcal{A})$. There exist $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$ and a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that

$$(\mathcal{A} + \mathcal{E})\mathbf{x}^{N-1} = \lambda\mathbf{x}^{[N-1]}.$$

Let $\mathbf{v} = \mathbf{x}/\|\mathbf{x}\|_2$. Then $\mathbf{v}$ is an unit vector whose image $\mathbf{y} := (\mathcal{A}-\lambda\mathcal{I})\mathbf{v}^{N-1}$ satisfies $\mathbf{y} = \mathcal{E}\,\mathbf{v}^{N-1}$. Define $\tilde{\mathcal{E}} = \mathbf{y} \otimes \overline{\mathbf{v}}^{\otimes N-1} \in CT_{N,I}$. We have $\tilde{\mathcal{E}}\mathbf{v}^{N-1} = \mathbf{y}$ and

$$(\mathcal{A} - \tilde{\mathcal{E}} - \lambda\mathcal{I})\mathbf{v}^{N-1} = (\mathcal{A} - \lambda\mathcal{I})\mathbf{v}^{N-1} - \tilde{\mathcal{E}}\mathbf{v}^{N-1} = \mathcal{E}\mathbf{v}^{N-1} - \mathbf{y} = \mathbf{0}_I,$$

which implies that λ is also an eigenvalue of $\mathcal{A} - \tilde{\mathcal{E}}$. The Frobenius norm of $\tilde{\mathcal{E}}$ is at most ε, because

$$\|\tilde{\mathcal{E}}\|_F^2 = \sum_{i_1,i_2,\ldots,i_N=1}^{I} |(\mathcal{E}\,\mathbf{v}^{N-1})_{i_1}|^2\,|\overline{v}_{i_2}|^2 \ldots |\overline{v}_{i_N}|^2 \le \|\mathcal{E}\|_F^2 \|\mathbf{v}\|_2^{4(N-1)} \le \epsilon^2.$$

Then (2.2.1) holds with a rank-one tensor $\tilde{\mathcal{E}}$. The converse is straightforward. □

Next, we present a Gerschgorin type result which informs that the ϵ-pseudo-spectrum of $\mathcal{A}$ lie in the union of I disks in $\mathbb{C}$. These I disks have the diagonal elements of the tensor as their centers, with the sums of the moduli of the off-diagonal elements and a constant multiple of ϵ as their radii. The result is developed in [18, Theorem 12ϵ] for the matrix case.

Theorem 2.2.1 (Gerschgorin Theorem for ϵ-Pseudo-Spectrum) *Let $\mathcal{A} \in CT_{N,I}$ and $\epsilon \ge 0$. Then*

$$\Lambda_\epsilon(\mathcal{A}) \subseteq \bigcup_{1\le i\le I} \left\{ \lambda \in \mathbb{C} : |\lambda - a_{ii\ldots i}| \le \sum_{\substack{1\le i,i_2,\ldots,i_N\le I \\ (i_2,\ldots,i_N)\neq(i,\ldots,i)}} |a_{ii_2\ldots i_N}| + I^{\frac{N-1}{2}}\epsilon \right\}.$$

Proof Let $\lambda \in \Lambda_\epsilon(\mathcal{A})$. There exist $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \le \epsilon$ and a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that

$$(\mathcal{A} + \mathcal{E})\mathbf{x}^{N-1} = \lambda \mathbf{x}^{[N-1]}.$$

Let $i \in \{1, 2, \ldots, I\}$ be such that $|x_i| = \max\{|x_j| : j = 1, 2, \ldots, I\}$. It implies that

$$(a_{ii\ldots i} + e_{ii\ldots i} - \lambda)x_i^{N-1} = \sum_{\substack{1\le i,i_2,\ldots,i_N\le I \\ (i_2,\ldots,i_N)\neq(i,\ldots,i)}} \left(a_{ii_2\ldots i_N} + e_{ii_2\ldots i_N}\right) x_{i_2}\ldots x_{i_N}\,.$$

From the choice of i and $\mathbf{x} \neq \mathbf{0}_I$, one has $x_i \neq 0$ and $\frac{|x_j|}{|x_i|} \le 1$ for all j. Thus

$$\begin{aligned} |a_{ii\ldots i} + e_{ii\ldots i} - \lambda| &\le \sum_{\substack{1\le i,i_2,\ldots,i_N\le I \\ (i_2,\ldots,i_N)\neq(i,\ldots,i)}} |a_{ii_2\ldots i_N} + e_{ii_2\ldots i_N}| \left|\frac{x_{i_2}}{x_i}\right| \ldots \left|\frac{x_{i_N}}{x_i}\right| \\ &\le \sum_{1\le i,i_2,\ldots,i_N\le I\,(i_2,\ldots,i_N)\neq(i,\ldots,i)} |a_{ii_2\ldots i_N} + e_{ii_2\ldots i_N}|. \end{aligned}$$

We then have

$$|\lambda - a_{ii\ldots i}| \leq \sum_{\substack{1\leq i,i_2,\ldots,i_N\leq I\\(i_2,\ldots,i_N)\neq(i,\ldots,i)}} |a_{ii_2\ldots i_N}| + \sum_{1\leq i_2,\ldots,i_N\leq I} |e_{ii_2\ldots i_N}|.$$

It follows from the Hölder inequality that

$$\sum_{1\leq i_2,\ldots,i_N\leq I} |e_{ii_2\ldots i_N}| \leq \sqrt{I^{N-1}\sum_{1\leq i_2,\ldots,i_N\leq I} |e_{ii_2\ldots i_N}|^2} \leq I^{\frac{N-1}{2}}\|\mathcal{E}\|_F.$$

Thus

$$|\lambda - a_{ii\ldots i}| \leq \sum_{\substack{1\leq i,i_2,\ldots,i_N\leq I\\(i_2,\ldots,i_N)\neq(i,\ldots,i)}} |a_{ii_2\ldots i_N}| + I^{\frac{N-1}{2}}\epsilon,$$

and the result follows. □

Remark 2.2.1 In the matrix case, Gerschgorin bounds on pseudo-spectra can be improved (see [19] for a recent development). It would be natural to ask how the techniques can be used to improve the previous theorem [20].

2.2.2 Computational Interpretation of ϵ-Pseudo-Spectrum

In the computation of the eigenvalues of $\mathcal{A} \in CT_{N,I}$, we terminate the algorithm after obtaining a number λ with $\left\|(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}^{N-1}\right\|_2 \leq \epsilon$ for some small tolerance ϵ and a unit vector $\mathbf{v} \in \mathbb{C}^I$. The next theorem proves that one obtains an element of the ϵ-pseudo-spectrum of $\mathcal{A}$.

Theorem 2.2.2 *Let $\epsilon \geq 0$. A complex number λ belongs to the ϵ-pseudo-spectrum of $\mathcal{A} \in CT_{N,I}$ if and only if there exists a unit vector $\mathbf{v}$ such that $\left\|(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}^{N-1}\right\|_2 \leq \epsilon$.*

Proof ($\Rightarrow$) Let λ be an element of the ϵ-pseudo-spectrum of $\mathcal{A}$. We proceed by contradiction and assume that for all unit vectors $\mathbf{v} \in \mathbb{C}^I$, we have

$$\left\|(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}^{N-1}\right\|_2 > \epsilon.$$

Then for any $\mathcal{A}+\mathcal{E} \in CT_{N,I}$ which is a small perturbation of $\mathcal{A}$ (with $\mathcal{E}$ having the Frobenius norm at most ϵ) and for any unit vector $\mathbf{v} \in \mathbb{C}^I$, one has

$$\left\|(\mathcal{A}+\mathcal{E} - \lambda\mathcal{I})\mathbf{v}^{N-1}\right\|_2 \geq \left\|(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}^{N-1}\right\|_2 - \left\|\mathcal{E}\mathbf{v}^{N-1}\right\|_2 > \epsilon - \epsilon = 0$$

by the triangle inequality, which implies that λ cannot be an eigenvalue of $\mathcal{A}+\mathcal{E}$. This contradicts the assumption that λ is an element of the ϵ-pseudo-spectrum of $\mathcal{A}$.

($\Leftarrow$) Note that there exists a unit vector $\mathbf{v}\in\mathbb{C}^I$ such that $\left\|(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}^{N-1}\right\|_2\leq\epsilon$. Let $\mathbf{y}:=(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}^{N-1}$ and $\mathcal{E}=\mathbf{y}\otimes\overline{\mathbf{v}}^{\otimes N-1}$. Then we have $\|\mathbf{y}\|_2\leq\epsilon$, $\mathcal{E}\mathbf{v}^{N-1}=\mathbf{y}$ and

$$(\mathcal{A}-\mathcal{E}-\lambda\mathcal{I})\mathbf{v}^{N-1}=(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}^{N-1}-\mathcal{E}\mathbf{v}^{N-1}=\mathbf{y}-\mathbf{y}=\mathbf{0}_I.$$

$\|\mathcal{E}\|_F\leq\|\mathbf{y}\|_2\|\mathbf{v}\|_2^{N-1}\leq\epsilon$. Then λ is an element of the ϵ-pseudo-spectrum of $\mathcal{A}$. □

According to Theorem 2.2.2, the ϵ-pseudo-spectrum of $\mathcal{A}\in CT_{N,I}$ can be also represented as

$$\Lambda_\epsilon(\mathcal{A})=\left\{\lambda\in\mathbb{C}:\min_{\|\mathbf{x}\|_2=1}\left\|(\mathcal{A}-\lambda\mathcal{I})\mathbf{x}^{N-1}\right\|_2\leq\epsilon\right\}. \tag{2.2.3}$$

As a corollary of Theorem 2.2.2, we know that $\Lambda_\epsilon(\mathcal{A})$ always contains the $(\rho_{N,I}^{-1}\varepsilon)$-neighborhood of the spectrum $\Lambda(\mathcal{A})$ for some constant $\rho_{N,I}$ depending on the order of the tensor $\mathcal{A}$ and the dimension of the underlying space.

Corollary 2.2.1 *Let $\epsilon\geq 0$ and $\rho_{N,I}=\min_{\mathbf{x}\in\mathbb{C}^I}\left\{\left\|\mathbf{x}^{[N-1]}\right\|_2:\|\mathbf{x}\|_2=1\right\}>0$. Then*

$$\Lambda(\mathcal{A})+\Delta_{\rho_{N,I}^{-1}\epsilon}\subseteq\Lambda_\epsilon(\mathcal{A}), \tag{2.2.4}$$

where $\Delta_{\rho_{N,I}^{-1}\epsilon}=\{z\in\mathbb{C}:|z|\leq\rho_{N,I}^{-1}\epsilon\}$.

Remark 2.2.2 Proposition 2.2.2 shows that there exist tensors $\mathcal{A}\in CT_{N,I}$ such that $\Lambda(\mathcal{A})+\Delta_{\rho_{N,I}^{-1}\epsilon}=\Lambda_\epsilon(\mathcal{A})$ for all $\epsilon\geq 0$.

We present an example illustrating the sharpness of (2.2.4) in Corollary 2.2.1.

Example 2.2.1 Let

$$\mathbf{M}=\begin{pmatrix}0&1\\0&0\end{pmatrix}\in\mathbb{C}^{2\times 2}.$$

Let N be an even number. Define $\mathcal{A}\in CT_{N,2}$ as

$$a_{i_1,\ldots,i_N}=m_{i_1i_2}m_{i_3i_4},\ldots,m_{i_{N-1}i_N},$$

where m_{ij} denotes the (i,j)-element of $\mathbf{M}$. For all $\mathbf{x}=(x_1,x_2)^\top\in\mathbb{C}^2$, it can be verified that

$$\mathcal{A}\mathbf{x}^{N-1}=\left(\mathbf{x}^\top\mathbf{M}\mathbf{x}\right)^{\frac{N}{2}-1}(\mathbf{M}\mathbf{x})=(x_1x_2)^{\frac{N}{2}-1}(x_2,0)^\top.$$

Thus the only eigenvalue of $\mathcal{A}$ is zero, and hence $\Lambda(\mathcal{A})=\{0\}$.

For $\epsilon > 0$, let

$$\mathbf{M}^{\epsilon} = \begin{pmatrix} 0 & 1 \\ \epsilon & 0 \end{pmatrix},$$

and $\mathcal{A}^{\epsilon} \in CT_{N,2}$ be defined by $a^{\epsilon}_{i_1,\ldots,i_N} = m^{\epsilon}_{i_1 i_2} m_{i_3 i_4} \cdots m_{i_{N-1} i_N}$. Direct verification proves that

$$\left\| \mathcal{A}^{\epsilon} - \mathcal{A} \right\|_F \leq \epsilon,$$

and for all $\mathbf{x} = (x_1, x_2)^{\top} \in \mathbb{C}^2$, we have

$$\mathcal{A}^{\epsilon}\mathbf{x}^{N-1} = \left(\mathbf{x}^{\top}\mathbf{M}\mathbf{x}\right)^{\frac{N}{2}-1} (\mathbf{M}^{\epsilon}\mathbf{x}) = (x_1 x_2)^{\frac{N}{2}-1}(x_2, \epsilon x_1)^{\top}.$$

Let $\mathbf{u} = (1, \sqrt[N]{\epsilon})^{\top}$. Then $\mathcal{A}^{\epsilon}\mathbf{u}^{N-1} = (\sqrt{\epsilon}, \epsilon^{\frac{3N-2}{2N}})^{\top}$, and

$$\mathcal{A}^{\epsilon}\mathbf{u}^{N-1} = \sqrt{\epsilon}\mathbf{u}^{[N-1]}.$$

It illustrates that $\Lambda_{\epsilon}(\mathcal{A})$ contains an element with modulus with $\sqrt{\epsilon}$, which is larger than $\rho^{-1}_{N,2}\epsilon$ when ϵ is small. Thus (2.2.4) is strict in this example.

In the following theorem, we develop a lower bound of the distance between any scalar $z_0 \in \mathbb{C}$ and the ϵ-pseudo-spectrum of $\mathcal{A} \in CT_{N,I}$.

Theorem 2.2.3 *Let $\epsilon \geq 0$, $z_0 \in \mathbb{C}$ and $\mathcal{A} \in CT_{N,I}$. Then*

$$\operatorname{dist}(z_0, \Lambda_{\epsilon}(\mathcal{A})) \geq \rho^{-1}_{N,I} \left(\min_{\|\mathbf{x}\|_2 = 1} \left\| (\mathcal{A} - z_0\mathcal{I})\mathbf{x}^{N-1} \right\|_2 - \epsilon \right),$$

where $\rho_{N,I} = \min\limits_{\mathbf{x} \in \mathbb{C}^I} \left\{ \left\| \mathbf{x}^{[N-1]} \right\|_2 : \|\mathbf{x}\|_2 = 1 \right\} > 0.$

Proof Let λ be an element of the ϵ-pseudo-spectrum of $\mathcal{A}$. Then there exist a tensor $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$ and a nonzero vector $\mathbf{x}_0 \in \mathbb{C}^I$ such that

$$(\mathcal{A} + \mathcal{E})\mathbf{x}_0^{N-1} = \lambda \mathbf{x}_0^{[N-1]}.$$

For any scalar $z_0 \in \mathbb{C}$, we obtain

$$(\mathcal{A} - z_0\mathcal{I})\mathbf{x}_0^{N-1} + \mathcal{E}\mathbf{x}_0^{N-1} = (\lambda - z_0)\mathbf{x}_0^{[N-1]}.$$

Thus

$$\left\| (\mathcal{A} - z_0\mathcal{I})\mathbf{x}_0^{N-1} + \mathcal{E}\mathbf{x}_0^{N-1} \right\|_2 = \left\| (\lambda - z_0)\mathbf{x}_0^{[N-1]} \right\|_2.$$

By the triangle inequality and the definition of $\rho_{N,I}$, we get

$$|\lambda - z_0|\rho_{N,I} \geq \min_{\|\mathbf{x}\|_2=1} \left\{ \left\| (\mathcal{A} - z_0\mathcal{I})\mathbf{x}^{N-1} \right\|_2 - \|\mathcal{E}\|_F \, \|\mathbf{x}\|_2^{N-1} \right\}.$$

Since $\|\mathcal{E}\|_F \leq \epsilon$, we have

$$\text{dist}(z_0, \Lambda_\epsilon(\mathcal{A})) \geq \rho_{N,I}^{-1} \left(\min_{\|\mathbf{x}\|_2=1} \left\| (\mathcal{A} - z_0\mathcal{I})\mathbf{x}^{N-1} \right\|_2 - \epsilon \right).$$

Then it is completely proved. □

Furthermore, we consider the real ϵ-pseudo-spectrum of $\mathcal{A} \in RT_{N,I}$, defined as

$$\Lambda_\epsilon^{\mathrm{H}}(\mathcal{A}) = \Big\{ \lambda \in \mathbb{R} : (\mathcal{A} + \mathcal{E})\mathbf{x}^{N-1} = \lambda \, \mathbf{x}^{[N-1]} \text{ for some } \mathcal{E} \in RT_{N,I}$$
$$\text{with } \|\mathcal{E}\|_F \leq \epsilon \text{ and nonzero vectors } \mathbf{x} \in \mathbb{R}^I \Big\}.$$

If $\epsilon = 0$, then the real 0-pseudo-spectrum consists of all H-eigenvalues of $\mathcal{A}$. When N is even, we reveal a relationship between the real ϵ-pseudo-spectrum and the real numerical range of $\mathcal{A}$, where the latter is given by Ding and Wei [15],

$$W(\mathcal{A}) = \left\{ \mathcal{A}\mathbf{v}^N : \|\mathbf{v}\|_N = 1, \mathbf{v} \in \mathbb{R}^I \right\}$$

where $\|\mathbf{v}\|_N^N = |v_1|^N + |v_2|^N + \cdots + |v_I|^N$ for all vectors $\mathbf{v} \in \mathbb{R}^I$.

Theorem 2.2.4 *Let N be even, $\epsilon \geq 0$ and $\mathcal{A} \in RT_{N,I}$. Denote $\delta_{N,I} = \max_{\mathbf{v}\in\mathbb{R}^I}\{\|\mathbf{v}\|_2^N : \|\mathbf{v}\|_N = 1\} > 0$. Then*

$$\Lambda_\epsilon^{\mathrm{H}}(\mathcal{A}) \subseteq W(\mathcal{A}) + \Delta_{\delta_{N,I}\epsilon}.$$

Proof Let $\lambda \in \Lambda_\epsilon^{\mathrm{H}}(\mathcal{A})$. There exist a tensor $\mathcal{E} \in RT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$ and a nonzero vector $\mathbf{x} \in \mathbb{R}^I$ such that

$$(\mathcal{A} + \mathcal{E})\mathbf{x}^{N-1} = \lambda \mathbf{x}^{[N-1]}.$$

Let $\mathbf{v} = \frac{\mathbf{x}}{\|\mathbf{x}\|_N}$, we have

$$(\mathcal{A} + \mathcal{E})\mathbf{v}^{N-1} = \lambda \mathbf{v}^{[N-1]}.$$

Multiplying $\mathbf{v}^\top$, we have $(\mathcal{A} + \mathcal{E})\mathbf{v}^N = \lambda \|\mathbf{v}\|_N^N = \lambda$, which implies that

$$\lambda = (\mathcal{A} + \mathcal{E})\mathbf{v}^N = \mathcal{A}\mathbf{v}^N + \mathcal{E}\mathbf{v}^N.$$

Obviously, the first term is an element of the real numerical range $W(\mathcal{A})$ while the second satisfies

$$\left|\mathcal{E}\mathbf{v}^N\right| \le \|\mathcal{E}\|_F \|\mathbf{v}\|_2^N \le \epsilon\delta_{N,I}.$$

The the proof is completed. □

Next, we present explicit expressions for $\rho_{N,I}$ and $\delta_{N,I}$. For any unit vector $\mathbf{x} \in \mathbb{C}^I$, if we set $x_i = r_i e^{\iota\theta_i}$ with $r_i \in \mathbb{R}$ and $\theta_i \in (-\pi, \pi]$, then $\mathbf{r} = (r_1, r_2, \ldots, r_I)^\top \in \mathbb{R}^I$ is a unit vector and

$$\rho_{N,I} = \min_{\mathbf{x}\in\mathbb{C}^I} \left\{ \left\| \mathbf{x}^{[N-1]} \right\|_2 : \|\mathbf{x}\|_2 = 1 \right\} = \min_{\mathbf{x}\in\mathbb{R}^I} \left\{ \left\| \mathbf{x}^{[N-1]} \right\|_2 : \|\mathbf{x}\|_2 = 1 \right\}.$$

When $N = 2$, it is clear that $\rho_{N,I} = \delta_{N,I} \equiv 1$. We only need to consider the case of $N \ge 3$. We consider the following constrained optimization problem

$$\text{(P)} \ \min \left\| \mathbf{x}^{[N-1]} \right\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_2 = 1; \ \mathbf{x} \in \mathbb{R}^I.$$

The Lagrangian function for the optimization problem is

$$L(\mathbf{x}, \mu) = \left\| \mathbf{x}^{[N-1]} \right\|_2^2 - \mu(\|\mathbf{x}\|_2^2 - 1).$$

Any critical point of the Lagrangian function satisfies

$$\begin{cases} 2(N-1)x_i^{2(N-1)-1} - 2\mu x_i = 0, \quad i = 1, 2, \ldots, I, \\ x_1^2 + x_2^2 + \cdots + x_N^2 = 1. \end{cases}$$

It implies that $\mu = (N-1)(x_1^{2(N-1)} + x_2^{2(N-1)} + \cdots + x_I^{2(N-1)})$, either $x_i = 0$ or $\mu = (N-1)x_i^{2N-4}$. Let $\mathbb{I} = \{1 \le i \le I : x_i = 0\}$ and T be the cardinality of $\mathbb{I}$. Then we have $0 \le T \le I - 1$ and

$$x_i = \left(\frac{\mu}{N-1} \right)^{1/(2N-4)} \quad \text{for all } i \notin \mathbb{I}$$

with $\mu = (N-1)\left(\frac{1}{I-T}\right)^{N-2} = (N-1)(I-T)^{2-N}$, and

$$\left\| \mathbf{x}^{[N-1]} \right\|_2^2 = (I-T)^{1-N} I.$$

Table 2.1 The values of $\rho_{N,I}$ and $\delta_{N,I}$ for integer pairs (N, I) with $N \in \{2, 3, 4, 5, 6\}$ and $I \in \{5, 6, 7, 8, 9, 10\}$

	$\rho_{N,I}$					$\delta_{N,I}$				
I	$N := 2$	3	4	5	6	$N := 2$	3	4	5	6
5	1.00	0.45	0.20	0.09	0.04	1.00	2.24	5.00	11.18	25.00
6	1.00	0.41	0.17	0.07	0.03	1.00	2.45	6.00	14.70	36.00
7	1.00	0.38	0.14	0.05	0.02	1.00	2.65	7.00	18.52	49.00
8	1.00	0.35	0.13	0.04	0.02	1.00	2.83	8.00	22.63	64.00
9	1.00	0.33	0.11	0.04	0.01	1.00	3.00	9.00	27.00	81.00
10	1.00	0.32	0.10	0.03	0.01	1.00	3.16	10.00	31.62	100.00

Note with $N \geq 3$ that

$$I^{1-N} \leq (I-1)^{1-N} \leq \cdots \leq 1^{1-N} = 1.$$

The optimal solution of the optimization problem (P) are attained when $T = 0$ and $x_1 = \cdots = x_I = \sqrt{1/I}$. The value of $\rho_{N,I} = I^{(2-N)/2} \in (0, 1]$. Similar, we can show that the value of $\delta_{N,I}$ is $I^{N/2-1} \geq 1$. For some given pairs (N, I), the values of $\rho_{N,I}$ and $\delta_{N,I}$ are shown in Table 2.1.

We next illustrate the ϵ-pseudo-spectrum of diagonal tensors.

Proposition 2.2.2 *If $\mathcal{D} \in CT_{N,I}$ is a diagonal tensor with all main diagonal entries $d_i \in \mathbb{C}$ for all i, then the ϵ-pseudo-spectrum of $\mathcal{D}$ is $\{d_1, d_2, \ldots, d_I\} + \Delta_{\rho_{N,I}^{-1}\epsilon}$.*

Proof By Corollary 2.2.1, it is evident that $\{d_1, d_2, \ldots, d_I\} + \Delta_{\rho_{N,I}^{-1}\epsilon} \subseteq \Lambda_\epsilon(\mathcal{D})$ holds. We will show that $\Lambda_\epsilon(\mathcal{D}) \subseteq \{d_1, d_2, \ldots, d_I\} + \Delta_{\rho_{N,I}^{-1}\epsilon}$ holds for any diagonal tensor $\mathcal{D} \in CT_{N,I}$.

Let $\lambda \in \Lambda_\epsilon(\mathcal{D})$. If $\lambda \in \{d_1, d_2, \ldots, d_I\}$, then the result follows trivially. We assume that $\lambda \notin \{d_1, d_2, \ldots, d_I\}$. From (2.2.3), we consider the optimization problem,

$$\min \left\| (\mathcal{D} - \lambda \mathcal{I})\mathbf{x}^{N-1} \right\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_2 = 1;\ \mathbf{x} \in \mathbb{C}^I.$$

According to the explicit expression of $\rho_{N,I}$, the above optimization problem is equivalent to

$$\min \left\| (\mathcal{D} - \lambda \mathcal{I})\mathbf{x}^{N-1} \right\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_2 = 1;\ \mathbf{x} \in \mathbb{R}^I.$$

The corresponding Lagrangian function is

$$L(\mathbf{x}, \mu) = |d_1 - \lambda|^2 x_1^{2(N-1)} + |d_2 - \lambda|^2 x_2^{2(N-1)} + \cdots + |d_I - \lambda|^2 x_I^{2(N-1)} - \mu(\|\mathbf{x}\|_2^2 - 1).$$

Differentiating $L(\mathbf{x}, \mu)$ with respect to $\mathbf{x}$ and μ, we have

$$\begin{cases} 2(N-1)|d_i - \lambda|^2 x_i^{2(N-1)-1} - 2\mu x_i = 0, \quad i = 1, 2, \ldots, I, \\ x_1^2 + x_2^2 + \cdots + x_N^2 = 1. \end{cases}$$

These lead to

$$\mu = (N-1)(|d_1 - \lambda|^2 x_1^{2(N-1)} + |d_2 - \lambda|^2 x_2^{2(N-1)} + \cdots + |d_I - \lambda|^2 x_I^{2(M-1)}),$$

and either $x_i = 0$ or $\mu = (N-1)|d_i - \lambda|^2 x_i^{2N-4}$. Let $\mathbb{I} = \{1 \le i \le I : x_i = 0\}$ and T be the cardinality of $\mathbb{I}$. We have $0 \le T \le I - 1$ and

$$x_i = \left(\frac{\mu}{N-1} \frac{1}{|d_i - \lambda|^2} \right)^{1/(2N-4)}, \quad \text{for all } i \notin \mathbb{I}.$$

We prove that

$$\mu = (N-1) \left(\sum_{i \notin \mathbb{I}} \left(\frac{1}{|d_i - \lambda|^2} \right)^{1/(N-2)} \right)^{2-N}.$$

Let $k \notin \mathbb{I}$ be such that $|\lambda - d_k| = \min\{|d_i - \lambda| : i \notin \mathbb{I}\}$. Then

$$\mu \ge (N-1)(I-T)^{2-N} |d_k - \lambda|^2.$$

Since

$$\begin{aligned} \frac{\mu}{N-1} &= |d_1 - \lambda|^2 x_1^{2(N-1)} + |d_2 - \lambda|^2 x_2^{2(N-1)} + \cdots + |d_I - \lambda|^2 x_I^{2(N-1)} \\ &= \left\| (\mathcal{D} - \lambda \mathcal{I}) \mathbf{x}^{N-1} \right\|_2^2. \end{aligned}$$

Then

$$\min_{\|\mathbf{x}\|_2 = 1} \left\| (\mathcal{D} - \lambda \mathcal{I}) \mathbf{x}^{N-1} \right\|_2 \ge \min_{0 \le T \le I-1} \left\{ (N-T)^{\frac{2-N}{2}} |d_k - \lambda| \right\} = \rho_{N,I} |d_k - \lambda|.$$

Thus $\lambda \in \{d_1, d_2, \ldots, d_I\} + \Delta_{\rho_{N,I}^{-1} \epsilon}$ and this completes the proof. □

Remark 2.2.3 (Comparison with Other Definitions) Lim [16] defines the ϵ-pseudo-spectrum of a cubical tensor as

$$\sigma_\epsilon^\Sigma(\mathcal{A}) = \{\lambda \in \mathbb{C} : \sigma_{\min}(\mathcal{A} - \lambda \mathcal{I}) < \epsilon\},$$

where $\sigma_{\min}$ denotes the smallest singular value of $\mathcal{A} \in CT_{N,I}$.

Without loss of generality, $\sigma \in \mathbb{R}_+$ is a singular value of $\mathcal{A} \in CT_{3,I}$ if there exist three unitary vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^I$ such that

$$\begin{cases} \mathcal{A} \times_2 \mathbf{y}^\top \times_3 \mathbf{z}^\top = \sigma \mathbf{x}, \\ \mathcal{A} \times_1 \mathbf{x}^\top \times_3 \mathbf{z}^\top = \sigma \mathbf{y}, \\ \mathcal{A} \times_1 \mathbf{x}^\top \times_2 \mathbf{y}^\top = \sigma \mathbf{z}, \\ \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \|\mathbf{z}\|_2 = 1. \end{cases}$$

We notice that Lim's definition and Definition 2.2.1 are different. To see this, consider $\mathcal{A} = \mathcal{I} \in RT_{3,2}$ where $\mathcal{I}$ is the identity tensor in $RT_{3,2}$ and $\epsilon > 0$. The preceding proposition shows that $1 - \sqrt{2}\epsilon \in \sigma_\epsilon(\mathcal{A})$. However, $1 - \sqrt{2}\epsilon \notin \sigma_\epsilon^\Sigma(\mathcal{A})$ because $\sigma_{\min}(\sqrt{2}\epsilon\mathcal{I}) = \sqrt{2}\epsilon > \epsilon$.

2.2.3 Pseudo-Spectral Abscissa and Radius

We can define the ϵ-pseudo-spectral abscissa and ϵ-pseudo-spectral radius of a tensor in the Frobenius norm as the matrix cases.

Definition 2.2.2 (ϵ-Pseudo-Spectral Abscissa and Radius) Let $\mathcal{A} \in CT_{N,I}$ and $\epsilon \geq 0$. We define its ϵ-pseudo-spectral abscissa as

$$\alpha_\epsilon(\mathcal{A}) = \max_{\lambda \in \Lambda_\epsilon(\mathcal{A})} \left\{ \Re(z) : z^{N-1} = \lambda \right\}.$$

The maximum in the definition of the ϵ-pseudo-spectral abscissa is taken over the set consisting of all the $(N-1)$th roots of λ.

If $\epsilon = 0$, then we refer the 0-pseudo-spectral abscissa of $\mathcal{A}$ as the spectral abscissa of $\mathcal{A}$, and denote it as $\alpha(\mathcal{A})$.

We define the ϵ-pseudo-spectral radius by

$$\rho_\epsilon(\mathcal{A}) = \max_{\lambda \in \Lambda_\epsilon(\mathcal{A})} \left\{ |z| : z^{N-1} = \lambda \right\}.$$

If $N = 2$, then the definition of the ϵ-pseudo-spectral abscissa/radius of a tensor reduces to that for a matrix. The above definition of the ϵ-pseudo-spectral abscissa has a close connection to the existence of an asymptotically stable solution of a homogeneous dynamical system under small perturbations. Consider the following homogeneous dynamical system (differential inclusion)

$$\mathbf{u}'(t) \in \left(\mathcal{A}\mathbf{u}(t)^{N-1} \right)^{[\frac{1}{N-1}]} \text{ with } \mathbf{u}(t) \in \mathbb{C}^I,\ t \in [0, +\infty), \tag{2.2.5}$$

and its perturbed version

$$\mathbf{u}'(t) \in \left((\mathcal{A}+\mathcal{E})\mathbf{u}(t)^{N-1}\right)^{[\frac{1}{N-1}]} \text{ with } \mathbf{u}(t) \in \mathbb{C}^I,\ t \in [0,+\infty), \tag{2.2.6}$$

where for a given $\epsilon > 0$, $\mathcal{E} \in CT_{N,I}$ satisfies $\|\mathcal{E}\|_F \leq \epsilon$. $\mathbf{a}^{[\frac{1}{N-1}]} = \{(z_1,\ldots,z_I)^\top \in \mathbb{C}^I : z_1^{N-1} = a_1,\ldots,z_I^{N-1} = a_I\}$ for all $\mathbf{a} = (a_1,\ldots,a_I)^\top \in \mathbb{C}^I$.

The next theorem indicates that an asymptotically stable solution exists for the above homogeneous dynamical system and its perturbed version, if the ϵ-pseudo-spectral abscissa of $\mathcal{A}$ is negative. In the case $\epsilon = 0$, the stability of homogeneous dynamical system is discussed in [21] in terms of the tensor logarithm norm (an upper bound of the spectral abscissa).

Theorem 2.2.5 *Let $\epsilon \geq 0$ and $\mathcal{A} \in CT_{N,I}$. If $\alpha_\epsilon(\mathcal{A}) < 0$, then for each tensor $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$, the perturbed homogeneous dynamical system* (2.2.5) *has an asymptotically stable solution $\bar{\mathbf{u}}(t) \in \mathbb{C}^I$, in the sense that $\lim_{t\to+\infty} |\bar{\mathbf{u}}(t)| = \mathbf{0}_I$.*

Proof Let $\mathcal{E} \in CT_{N,I}$ with $\|\mathcal{E}\|_F \leq \epsilon$ and $\lambda \in \Lambda(\mathcal{A}+\mathcal{E})$ (which is possible by (2.1.1)). Then there exists a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that

$$(\mathcal{A}+\mathcal{E})\mathbf{x}^{N-1} = \lambda\,\mathbf{x}^{[N-1]}. \tag{2.2.7}$$

In particular, $\lambda \in \Lambda_\epsilon(\mathcal{A})$. Moreover, as $\alpha_\epsilon(\mathcal{A}) < 0$, there exists $z \in \mathbb{C}$ such that $z^{N-1} = \lambda$ and $\Re(z) < 0$. Considering (2.2.5), let $\bar{\mathbf{u}}(t) = \gamma e^{tz}\mathbf{x}$ with $\gamma \in \mathbb{R}$. Then we have $\bar{\mathbf{u}}'(t) = \gamma z e^{tz}\mathbf{x}$ and

$$\begin{aligned}(\mathcal{A}+\mathcal{E})\big(\bar{\mathbf{u}}(t)\big)^{N-1} &= \big(\gamma^{N-1}(e^{tz})^{N-1}\big)(\mathcal{A}+\mathcal{E})\mathbf{x}^{N-1}\\ &= \big(\gamma^{N-1}(e^{tz})^{N-1}\big)\lambda\mathbf{x}^{[N-1]}\\ &= \big(\gamma^{N-1}(e^{tz})^{N-1}\big)z^{N-1}\mathbf{x}^{[N-1]}.\end{aligned}$$

It shows that $\bar{\mathbf{u}}'(t)$ satisfies (2.2.5). Moreover, from $\Re(z) < 0$, we have

$$\lim_{t\to+\infty} |\bar{\mathbf{u}}(t)| = \lim_{t\to+\infty} |\gamma| e^{\Re(z)t}|\mathbf{x}| = \mathbf{0}_I.$$

Thus $\bar{\mathbf{u}}(t)$ is an asymptotically stable solution for the perturbed homogeneous dynamical system (2.2.5). □

If $\alpha(\mathcal{A}) < 0$, then the unperturbed homogeneous dynamical system (2.2.6) has an asymptotically stable solution $\bar{\mathbf{u}}(t) \in \mathbb{C}^I$.

Now we illustrate how the Gerschgorin result for the ϵ-pseudo-spectra of tensors and Theorem 2.2.5 can be used to determine the existence of an asymptotically stable solution for a homogeneous dynamical system.

Example 2.2.2 Denote a permutation of (i_1, i_2, i_3, i_4) by $\pi(i_1, i_2, i_3, i_4)$. Consider $\mathcal{A} \in CT_{4,2}$ defined by

$$a_{1111} = -8 + \iota,\ a_{2222} = -6,\ a_{1122} = a_{1212} = a_{1221} = a_{2112} = a_{2121} = a_{2211} = \frac{1}{6},$$

and, for any π

$$a_{i_1 i_2 i_3 i_4} = 0 \text{ for all } (i_1, i_2, i_3, i_4) \notin \{(1,1,1,1), (2,2,2,2), \pi(1,1,2,2)\}.$$

Let $\epsilon = 1$. For each $\mathcal{E} \in CT_{4,2}$ with $\|\mathcal{E}\|_F \le 1$, suppose that $\mathbf{u}(t)$ satisfies (2.2.5) with $N = 4$ and $I = 2$. We apply Theorem 2.2.1 that

$$\Lambda_1(\mathcal{A}) \subseteq \{\lambda \in \mathbb{C} : |\lambda - (-8 + \iota)| \le \frac{1}{2} + 2\sqrt{2}\} \cup \{\lambda \in \mathbb{C} : |\lambda + 6| \le \frac{1}{2} + 2\sqrt{2}\}.$$

Direct computation verifies that, for any $\lambda = |\lambda| e^{\iota\theta} \in \Lambda_1(\mathcal{A})$, one has $|\lambda| \ge 1$ and $\theta \in [\underline{\theta}, \overline{\theta}]$ for some $\underline{\theta}$ and $\overline{\theta}$ such that $\frac{\pi}{2} < \underline{\theta} \le \overline{\theta} < \frac{3\pi}{2}$. Observe that one of the cubic roots of λ is $|\lambda|^{\frac{1}{3}} e^{\iota \frac{\theta+2\pi}{3}}$ where $\frac{\theta+2\pi}{3} \in [\frac{\underline{\theta}+2\pi}{3}, \frac{\overline{\theta}+2\pi}{3}]$ and

$$\frac{5\pi}{6} < \frac{\underline{\theta} + 2\pi}{3} \le \frac{\overline{\theta} + 2\pi}{3} < \frac{7\pi}{6},$$

which implies that $\alpha_1(\mathcal{A}) < 0$ and the existence of asymptotically stable solution follows from Theorem 2.2.5.

In many potential applications of pseudo-spectra, the perturbation of the tensor often has some specific structure. For example, if the tensor is real, then it is also reasonable that the perturbation is real. This motivates us to define the structured ϵ-pseudo-spectra and ϵ-pseudo-spectral abscissa.

Definition 2.2.3 (Structured ϵ-Pseudo-Spectrum and ϵ-Pseudo-Spectral Abscissa) Let $\mathbb{S}$ be a subspace of $CT_{N,I}$. Suppose that $\epsilon \ge 0$ and $\mathcal{A} \in \mathbb{S}$. We define the structured ϵ-pseudo-spectrum and ϵ-pseudo-spectral abscissa respectively as

$$\Lambda_\epsilon^{\mathbb{S}}(\mathcal{A}) = \Big\{\lambda \in \mathbb{C} : (\mathcal{A} + \mathcal{E})\mathbf{x}^{N-1} = \lambda\, \mathbf{x}^{[N-1]} \text{ for some } \mathcal{E} \in \mathbb{S}$$
$$\text{with } \|\mathcal{E}\|_F \le \epsilon \text{ and nonzero vectors } \mathbf{x} \in \mathbb{C}^I\Big\},$$

and

$$\alpha_\epsilon^{\mathbb{S}}(\mathcal{A}) = \max_{\lambda \in \Lambda_\epsilon^{\mathbb{S}}(\mathcal{A})} \Big\{\Re(z) : z^{N-1} = \lambda\Big\}.$$

If $\epsilon = 0$, then the structured 0-pseudo-spectral abscissa of $\mathcal{A}$ equals to the structured spectral abscissa of A, denoted by $\alpha^{\mathbb{S}}(\mathcal{A})$.

It is obvious that, for all $\mathcal{A} \in \mathbb{S}$, $\Lambda_\epsilon^{\mathbb{S}}(\mathcal{A}) \subseteq \Lambda_\epsilon(\mathcal{A})$ and $\alpha_\epsilon^{\mathbb{S}}(\mathcal{A}) \leq \alpha_\epsilon(\mathcal{A})$. Moreover, using a similar argument, one can obtain the following result concerning the stability of the homogeneous dynamical system with structured perturbations. For simplicity, we omit the proof.

Theorem 2.2.6 *Let $\mathbb{S}$ be a subspace of $CT_{N,I}$, $\epsilon \geq 0$ and $\mathcal{A} \in \mathbb{S}$. Suppose that $\alpha_\epsilon^{\mathbb{S}}(\mathcal{A}) < 0$. Then for each tensor $\mathcal{E} \in \mathbb{S}$ with $\|\mathcal{E}\|_F \leq \epsilon$, the perturbed homogeneous dynamical system (2.2.5) has an asymptotically stable solution $\bar{\mathbf{u}}(t) \in \mathbb{C}^I$, in the sense that $\lim_{t\to+\infty} |\bar{\mathbf{u}}(t)| = \mathbf{0}_I$.*

If $\alpha^{\mathbb{S}}(\mathcal{A}) < 0$, then the unperturbed homogeneous dynamical system (2.2.5) has an asymptotically stable solution $\bar{\mathbf{u}}(t) \in \mathbb{C}^I$.

2.3 Pseudo-Spectrum for Tensor Polynomials

In this section, we present the definition of the ϵ-pseudo-spectrum of a tensor polynomial associated with a given tensor tuple and derive its basic properties. We then consider the backward error of the approximate eigenpairs of tensor polynomials in the Frobenius norm. For a given regular tensor polynomial $\mathcal{P}$, we measure its distance from the nearest singular tensor polynomial.

2.3.1 Definitions and Properties

For $\mathcal{A}_l, \Delta\mathcal{A}_l \in CT_{N,I}$ with all l and $z \in \mathbb{C}$, we define a tensor polynomial $\mathcal{P}_\Delta$, associated with the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$, as follows:

$$\mathcal{P}_\Delta(z) := \mathcal{P}_{\Delta\mathcal{A}_0,\ldots,\Delta\mathcal{A}_L}(z) = \mathcal{P}(z) + \big(\Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + z^2\Delta\mathcal{A}_2 + \cdots + z^L\Delta\mathcal{A}_L\big). \tag{2.3.1}$$

Then, for a given nonnegative weight vector $\mathbf{w} \in \mathbb{R}_+^{L+1}$ and any $z \in \mathbb{C}$, we introduce the set

$$B(\mathcal{P}, \mathbf{w}, \epsilon) = \{\mathcal{P} + \Delta\mathcal{P} : \Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + z^2\Delta\mathcal{A}_2 + \cdots + z^L\Delta\mathcal{A}_L, \quad \text{where } \|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon \text{ for all } l\} \tag{2.3.2}$$

and the scalar polynomial

$$p_{\mathbf{w}}(z) = w_0 + w_1 z + \cdots + w_L z^L. \tag{2.3.3}$$

Next, we introduce the definition of the ϵ-pseudo-spectrum of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ in the Frobenius norm.

Definition 2.3.1 (ϵ-Pseudo-Spectrum of Tensor Polynomials) For a tensor polynomial $\mathcal{P}$ as in (2.3.1), the ϵ-pseudo-spectrum of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ is defined as

$$\begin{aligned}\Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) &= \{\lambda \in \mathbb{C} : \det(\mathcal{P}_\Delta(\lambda)) = 0 \text{ for some } \mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)\} \\ &= \{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } \mathcal{P}_\Delta \text{ for some } \mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)\},\end{aligned}$$

with $\epsilon \geq 0$.

If $\lambda \in \Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$, then λ is called an ϵ-pseudo eigenvalue of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$. If $\epsilon = 0$, then the 0-pseudo-spectrum equals the spectrum of $\mathcal{P}$, denoted by $\Lambda(\mathcal{P})$.

In the special case of $L = 1$, if we choose $\mathcal{A}_1 = -\mathcal{I}$, $w_0 = 1$ and $w_1 = 0$, then Definition 2.3.1 reduces to the definition of the ϵ-pseudo-spectrum of the tensor $\mathcal{A}_0$ (see Definition 2.2.1). As in Proposition 2.2.1, similar basic properties of the ϵ-pseudo-spectrum of a tensor polynomial can also be derived. Below, we focus on the equivalent definitions of the ϵ-pseudo-spectrum of a tensor polynomial in terms of rank-one tensors.

Proposition 2.3.1 *For a tensor polynomial $\mathcal{P}$ as in (2.3.1), Then with $\epsilon \geq 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$, a complex number λ belongs to $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ if and only if there exists a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that*

$$\mathcal{P}_\Delta(\lambda)\mathbf{x}^{N-1} = \mathbf{0}_I, \tag{2.3.4}$$

for some $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$, where $\mathcal{P}_\Delta(z) = \mathcal{P}(z) + \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \cdots + z^L\Delta\mathcal{A}_L$ and $\Delta\mathcal{A}_l$ are rank-one tensors with $\|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon$ and all l.

Proof Let λ be an element in the ϵ-pseudo-spectrum $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$. Then there exist $\Delta\mathcal{A}_l \in CT_{N,I}$ with $\|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon$ and a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ such that

$$(\mathcal{P}(\lambda) + \Delta\mathcal{P}(\lambda))\mathbf{x}^{N-1} = \mathbf{0}_I,$$

where

$$\Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \cdots + z^L\Delta\mathcal{A}_L. \tag{2.3.5}$$

Let $\mathbf{v} = \mathbf{x}/\|\mathbf{x}\|_2$, then $\mathbf{v}$ is a unitary vector. For each l, define

$$\mathbf{z}_l = \Delta\mathcal{A}_l\mathbf{v}^{N-1}, \quad \widetilde{\Delta\mathcal{A}_l} = \mathbf{z}_l \otimes \overline{\mathbf{v}}^{\otimes N-1},$$

it is easy to see $\widetilde{\Delta\mathcal{A}_l}\mathbf{v}^{N-1} = \mathbf{z}_l$. Then we have

$$(\mathcal{P}(\lambda) + \widetilde{\Delta\mathcal{P}(\lambda)})\mathbf{v}^{N-1} = \big(\mathcal{P}(\lambda) + \Delta\mathcal{P}(\lambda)\big)\mathbf{v}^{N-1} = \mathbf{0}_I,$$

where $\widetilde{\Delta\mathcal{P}(z)} = \widetilde{\Delta\mathcal{A}_0} + z\widetilde{\Delta\mathcal{A}_1} + \cdots + z^L\widetilde{\Delta\mathcal{A}_L}$.

Thus λ is an eigenvalue of $\mathcal{P} + \widetilde{\Delta\mathcal{P}}$. Since the Frobenius norm of each $\widetilde{\Delta\mathcal{A}_l}$ is at most $w_l\,\epsilon$, for each l,

$$\left\|\widetilde{\Delta\mathcal{A}_l}\right\|_F^2 = \sum_{i_1,i_2,\ldots i_N=1}^{I} |(\Delta\mathcal{A}_l\,\mathbf{v}^{N-1})_{i_1}|^2\,|\overline{v}_{i_2}|^2 \ldots |\overline{v}_{i_N}|^2 \le \|\Delta\mathcal{A}_l\|_F^2 \|\mathbf{v}\|_2^{4(N-1)} = w_l^2\epsilon^2.$$

Then (2.3.4) holds with these chosen rank-one tensors $\widetilde{\Delta\mathcal{A}_l}$.

The result follows as the reverse implication is immediate by the definition. □

In order to deduce another equivalent characterization of the ϵ-pseudo-spectrum $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ with respect to a nonnegative weighting vector $\mathbf{w} \in \mathbb{R}^{L+1}$, we cite the following lemma without the proof.

Lemma 2.3.1 *For given unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^I$, there exists a tensor $\mathcal{B} \in CT_{N,I}$ with $\|\mathcal{B}\|_F = 1$ such that $\mathcal{B}\mathbf{x}^{N-1} = \mathbf{y}$.*

In the special case where $N = 2$, this theorem has been established in [7, Theorem 2.1] or [6, Lemma 2.1].

Theorem 2.3.1 *For a tensor polynomial $\mathcal{P}$ as in (2.3.1), Then with $\epsilon \ge 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$, the ϵ-pseudo-spectrum with respect to $\mathbf{w}$ equals*

$$\Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) = \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 \le p_{\mathbf{w}}(|\lambda|)\epsilon\right\}, \tag{2.3.6}$$

with $p_{\mathbf{w}}(|\lambda|) = w_0 + w_1|\lambda| + \cdots + w_L|\lambda|^L$.

Proof Let $\mathscr{S}$ denote the set in (2.3.6). If $\lambda \in \Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$, then there exist a nonzero vector $\mathbf{x} \in \mathbb{C}^I$ and a tensor polynomial $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ such that $\mathcal{P}_\Delta(\lambda)\mathbf{x}^{N-1} = \mathbf{0}_I$. As $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$, there exists $\Delta\mathcal{A}_l \in CT_{N,I}$ with all l such that

$$\mathcal{P}_\Delta = \mathcal{P} + \Delta\mathcal{P},$$

where $\Delta\mathcal{P}(z)$ is defined in (2.3.5). We have

$$\mathcal{P}(\lambda)\mathbf{x}^{N-1} = -\Delta\mathcal{P}(\lambda)\mathbf{x}^{N-1}.$$

It implies that

$$\min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 = \min_{\|\mathbf{x}\|_2=1} \left\|\Delta\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 \le \|\Delta\mathcal{P}(\lambda)\|_F \le p_{\mathbf{w}}(|\lambda|)\epsilon,$$

where the last inequality holds from $p_{\mathbf{w}}$ in (2.3.1). Thus $\lambda \in \mathscr{S}$ and we have

$$\Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) \subseteq \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 \le p_{\mathbf{w}}(|\lambda|)\epsilon\right\}.$$

To prove the converse, we suppose that $\lambda \in \mathscr{S}$ and we want to prove that $\lambda \in \Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$. Let $\mathbf{u} \in \mathbb{C}^I$ with $\|\mathbf{u}\|_2 = 1$ be such that

$$\min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 := \left\|\mathcal{P}(\lambda)\mathbf{u}^{N-1}\right\|_2.$$

Let $\mathbf{y} = \mathcal{P}(\lambda)\mathbf{u}^{N-1}$. If $\mathbf{y} = \mathbf{0}_I$, then we obtain that $\lambda \in \Lambda(\mathcal{P})$ and $\mathbf{u}$ is an eigenvector corresponding to λ, and the result follows.

Next, we assume that $\mathbf{y}$ is a nonzero vector, then there exists $\mathcal{H} \in CT_{N,I}$ with $\|\mathcal{H}\|_F = 1$ such that $\mathcal{H}\mathbf{u}^{N-1} = -\mathbf{y}/\|\mathbf{y}\|_2$ (see Lemma 2.3.1). Let $\mathcal{E} = \|\mathbf{y}\|_2\mathcal{H}$. Then $(\mathcal{P}(\lambda) + \mathcal{E})\mathbf{u}^{N-1} = \mathbf{y} + \|\mathbf{y}\|_2\mathcal{H}\mathbf{u}^{N-1} = \mathbf{0}_I$ and

$$\|\mathcal{E}\|_F = \|\mathbf{y}\|_2 = \left\|\mathcal{P}(\lambda)\mathbf{u}^{N-1}\right\|_2 = \min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 \le p_{\mathbf{w}}(|\lambda|)\epsilon.$$

For all l, define $\mathcal{A}_l \in CT_{N,I}$ as

$$\Delta\mathcal{A}_l = \operatorname{sign}(\lambda^l) w_l p_{\mathbf{w}}(|\lambda|)^{-1}\mathcal{E},$$

where the sign of $z \in \mathbb{C}$ is defined as

$$\operatorname{sign}(z) = \begin{cases} \overline{z}/|z|, & z \neq 0, \\ 0, & z = 0. \end{cases} \tag{2.3.7}$$

By constructions, we see that $\|\Delta\mathcal{A}_l\|_F \le w_l\epsilon$. Let $\Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \cdots + z^L\Delta\mathcal{A}_L$ with $z \in \mathbb{C}$. Then

$$\Delta\mathcal{P}(\lambda) = \sum_{l=0}^{L} \lambda^l \Delta\mathcal{A}_l = \left(\sum_{l=0}^{L} |\lambda^l|\right) p_{\mathbf{w}}(|\lambda|)^{-1}\mathcal{E} = \mathcal{E},$$

Thus $\big(\mathcal{P}(\lambda) + \Delta\mathcal{P}(\lambda)\big)\mathbf{u}^{N-1} = \big(\mathcal{P}(\lambda) + \mathcal{E}\big)\mathbf{u}^{M-1} = \mathbf{0}_I$. As $\mathbf{u} \neq \mathbf{0}_I$, we see that $\lambda \in \Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$. □

Remark 2.3.1 For the case of $L = 1$, with $\mathcal{A}_1 = -\mathcal{I}$ and $(w_0, w_1) = (1, 0)$, (2.3.6) reduces to the equivalent characterization for the ϵ-pseudo-spectrum of a tensor as in (2.2.3).

If all the coefficients of the matrix polynomial with respect to any nonnegative weight are real matrices, then its ϵ-pseudo-spectrum is symmetric with respect to the real axis. The symmetry holds for tensor polynomials. In the special case where $N = 2$, which has been established in [8, Proposition 2.1].

Theorem 2.3.2 *For a tensor polynomial $\mathcal{P}$ as in (2.3.1), then with $\epsilon \ge 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$, $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ is symmetric with respect to the real axis.*

Proof Let $\lambda \in \Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$. It suffices to show that its conjugate $\overline{\lambda}$ also belongs to $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$. The tensor polynomial $(\mathcal{A}_0 + \Delta\mathcal{A}_0) + \lambda(\mathcal{A}_1 + \Delta\mathcal{A}_1) + \cdots + \lambda^L(\mathcal{A}_L + \Delta\mathcal{A}_L) \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ satisfies

$$\det\Big((\mathcal{A}_0 + \Delta\mathcal{A}_0) + \lambda(\mathcal{A}_1 + \Delta\mathcal{A}_1)\cdots + \lambda^L(\mathcal{A}_L + \Delta\mathcal{A}_L)\Big) = 0,$$

where $\Delta\mathcal{A}_l \in CT_{N,I}$ for all l. Since $\mathcal{A}_l \in RT_{N,I}$, then we have

$$\begin{aligned}&\det\Big(\overline{(\mathcal{A}_0 + \Delta\mathcal{A}_0) + \lambda(\mathcal{A}_1 + \Delta\mathcal{A}_1)\cdots + \lambda^L(\mathcal{A}_L + \Delta\mathcal{A}_L)}\Big)\\ &= \det\Big((\mathcal{A}_0 + \overline{\Delta\mathcal{A}_0}) + \overline{\lambda}(\mathcal{A}_1 + \overline{\Delta\mathcal{A}_1})\cdots + \overline{\lambda}^L(\mathcal{A}_L + \overline{\Delta\mathcal{A}_L})\Big) = 0.\end{aligned}$$

Hence, $\overline{\lambda} \in \Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$. □

Next, we apply a simple property of the scalar polynomial $z \mapsto \det(\mathcal{P}(z))$.

Proposition 2.3.2 *For a tensor polynomial $\mathcal{P}$ as in (2.3.1), if $\mathcal{A}_L$ is nonsingular, then the degree of the scalar polynomial* $\det(\mathcal{P}(z))$ *is $IL(N-1)^{I-1}$ and its leading coefficient is* $\det(\mathcal{A}_L)$.

In particular, there are $IL(N-1)^{I-1}$ eigenvalues, counting multiplicity, of the tensor polynomial $\mathcal{P}$.

Proof According to [10, Proposition 2.4], we see that $\det(\mathcal{A}) \in \mathbb{C}[\mathcal{A}]$ gives rise to an irreducible and homogeneous polynomial of degree $IL(N-1)^{I-1}$, where $\mathcal{A} \in CT_{N,I}$ and $\mathbb{C}[\mathcal{A}]$ is the polynomial ring consisting of polynomials in indeterminate variables $\{a_{i_1 i_2 \ldots i_N}\}$ with coefficients in $\mathbb{C}$.

Thus for $z \in \mathbb{C}$ and a tensor polynomial $\mathcal{P}$, $\det(\mathcal{P}(z))$ can be treated as an irreducible and homogeneous polynomial in $\mathbb{C}[\mathcal{P}(z)]$ of degree $IL(N-1)^{I-1}$, where $\mathbb{C}[\mathcal{P}(z)]$ is the polynomial ring consisting of polynomials in indeterminate variables $\{\mathcal{P}(z)_{i_1 i_2 \ldots i_N}\}$ with coefficients in $\mathbb{C}$. Each entry $\mathcal{P}(z)_{i_1 i_2 \ldots i_N}$ is a scalar polynomial of z with degree L. It shows that $\det(\mathcal{P}(z))$ is of degree $IL(N-1)^{I-1}$.

In the following, we consider to check the leading coefficient of $\det(\mathcal{P}(z))$. Note that for $\alpha \in \mathbb{C}$ and $\mathcal{A} \in CT_{N,I}$, $\det(\alpha\mathcal{A}) = \alpha^{IL(N-1)^{I-1}} \det(\mathcal{A})$ [10, Corollary 2.6]. It follows that

$$\lim_{|z|\to\infty} \frac{\det(\mathcal{P}(z))}{z^{IL(N-1)^{I-1}}} = \lim_{|z|\to\infty} \det\left(\frac{\mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^l\mathcal{A}_l}{z^L}\right) = \det(\mathcal{A}_L) \neq 0,$$

where the last relation follows from the assumption that $\mathcal{A}_L$ is invertible. Then the leading coefficient of $z \mapsto \det(\mathcal{P}(z))$ is $\det(\mathcal{A}_L)$.

Finally, the result follows from the fundamental theorem of algebra. □

For a given tensor polynomial $\mathcal{P}$, if there is a perturbation $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ with identically zero determinant, then $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ coincides with the whole complex plane and a priori, the ϵ-pseudo-spectrum with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ may be unbounded. If the leading coefficient $\mathcal{A}_L$ is invertible, then the tensor polynomial $\mathcal{P}$ has exactly

$IL(N-1)^{I-1}$ eigenvalues, counting multiplicities, according to Proposition 2.3.2. For sufficiently small ϵ, the ϵ-pseudo-spectrum with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ must be bounded and consist of no more than $IL(N-1)^{I-1}$ connected components.

These possible scenarios motivate us to investigate when $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ will be bounded. Next, we present a necessary and sufficient condition for $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ to be bounded, for any $\epsilon \geq 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$. In the special case of $N = 2$, which has been developed in [8, Theorem 2.2].

Theorem 2.3.3 (Boundedness of Pseudo-Spectra for Tensor Polynomials) *For a tensor polynomial $\mathcal{P}$ as in (2.3.1), then with $\epsilon \geq 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$, the ϵ-pseudo-spectrum $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ with respect to $\mathbf{w}$ is bounded if and only if the $(w_L\,\epsilon)$-pseudo-spectrum of $\mathcal{A}_L$ does not contain the origin.*

Proof Assume that $0 \notin \Lambda_{w_L\,\epsilon}(\mathcal{A}_L)$. Then $\det(\mathcal{A}_L + \Delta\mathcal{A}_L) \neq 0$ for any $\Delta\mathcal{A}_L \in CT_{N,I}$ with $\|\Delta\mathcal{A}_L\|_F \leq w_L\epsilon$. We have

$$\zeta_\epsilon = \min\{|\det(\mathcal{A}_L + \Delta\mathcal{A}_L)| : \|\Delta\mathcal{A}_L\|_F \leq w_L\epsilon\} > 0.$$

Since $B(\mathcal{P}, \mathbf{w}, \epsilon)$ is compact, there exists a positive constant $M_\epsilon > 0$ such that any perturbation $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ has the form

$$\mathcal{P}_\Delta(z) = (\mathcal{A}_0 + \Delta\mathcal{A}_0) + z(\mathcal{A}_1 + \Delta\mathcal{A}_1)\cdots + z^L(\mathcal{A}_L + \Delta\mathcal{A}_L)$$

with $\|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon$ and all l. For any $z \in \mathbb{C}$ with $|z| > M_\epsilon$, we have

$$\begin{aligned}\left|\det(\mathcal{P}_\Delta(z)) - \det(\mathcal{A}_L + \Delta\mathcal{A}_L)z^{IL(N-1)^{I-1}}\right| &< \zeta_\epsilon \left|z^{IL(N-1)^{I-1}}\right| \\ &\leq \left|\det(\mathcal{A}_L + \Delta\mathcal{A}_L)z^{IL(N-1)^{I-1}}\right|.\end{aligned}$$

The first strict inequality follows because $\det(\mathcal{P}_\Delta(z))$ is a scalar polynomial with degree $IL(N-1)^{I-1}$ and the leading coefficient $\det(\mathcal{A}_L + \Delta\mathcal{A}_L)$; the second inequality holds by the definition of ζ_ϵ. It implies that $\det(\mathcal{P}_\Delta(z)) \neq 0$ for all $z \in \mathbb{C}$ with $|z| > M_\epsilon$. That is to say,

$$\{z : \det(\mathcal{P}_\Delta(z)) = 0\} \subseteq \{z : |z| \leq M_\epsilon\} \text{ for all } \mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon),$$

and $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ is bounded.

To prove the converse, assume that $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ is bounded but there exists a perturbed tensor polynomial $\mathcal{P}_{\widehat{\Delta}} \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ of the form

$$\mathcal{P}_{\widehat{\Delta}}(z) = (\mathcal{A}_0 + \widehat{\Delta\mathcal{A}_0}) + z(\mathcal{A}_1 + \widehat{\Delta\mathcal{A}_1})\cdots + z^L(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_L}),$$

for any $z \in \mathbb{C}$, where $\|\widehat{\Delta\mathcal{A}_l}\|_F \leq w_l\epsilon$ and $\det(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_L}) = 0$. If $w_L\,\epsilon = 0$, then $\det(\mathcal{A}_L) = 0$ which contradicts our assumption that $\mathcal{A}_L$ is nonsingular. Thus, we can assume that $w_L\,\epsilon > 0$. Moreover, we can also suppose that at least one of the

coefficients of the scalar polynomial $\det(\mathcal{P}_{\widehat{\Delta}}(z))$ is nonzero; otherwise $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) = \mathbb{C}$, a contradiction. Let β_τ be a nonzero coefficient of z^τ $(\tau = 0, 1, \ldots, IL(N-1)^{I-1} - 1)$ in the scalar polynomial $\det(\mathcal{P}_{\widehat{\Delta}}(z))$. The set $\{C \in CT_{N,I} : \|C\|_F \leq w_L\,\epsilon, \det(\mathcal{A}_L + C) \neq 0\}$ is dense in $\{C \in CT_{N,I} : \|C\|_F \leq w_L\,\epsilon\}$. There exists a sequence of tensors $\{\widehat{\Delta\mathcal{A}_{L,k}}\}_{k\in\mathbb{N}} \subset CT_{N,I}$ such that $\lim_{k\to\infty} \widehat{\Delta\mathcal{A}_{L,k}} = \widehat{\Delta\mathcal{A}_L}$ and

$$\det(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_{L,k}}) \neq 0 \text{ and } \|\widehat{\Delta\mathcal{A}_{L,k}}\|_F \leq w_L\,\epsilon, \quad (k = 1, 2, \ldots).$$

Define

$$\widehat{\mathcal{P}}_k(z) = (\mathcal{A}_0 + \widehat{\Delta\mathcal{A}_0}) + z(\mathcal{A}_1 + \widehat{\Delta\mathcal{A}_1}) + \cdots + z^L(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_{L,k}})$$

and

$$h_k(z) = \det(\widehat{\mathcal{P}}_k(z)).$$

Obviously, $\widehat{\mathcal{P}}_k \in B(\mathcal{P}, \mathbf{w}, \epsilon)$. In particular, any root of h_k lies in $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$. Denote the coefficient of z^τ in h_k by $\beta_{\tau,k}$. By the continuity, we know that $\beta_{\tau,k} \to \beta_\tau \neq 0$ as $k \to \infty$. Moreover, since $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ is bounded and the leading coefficient of h_k is $\det(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_{L,k}})$, the $(IL(N-1)^{I-1} - \tau)$-th elementary symmetric function of the roots of h_k equals to $\pm\beta_{\tau,k}/\det(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_{L,k}})$ and is bounded for all k. Note that $\beta_{\tau,k} \to \beta_\tau \neq 0$ and $\det(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_{L,k}}) \to \det(\mathcal{A}_L + \widehat{\Delta\mathcal{A}_L}) = 0$. This is a contradiction. □

If the ϵ-pseudo-spectrum of a tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ is bounded, then the following theorem gives the maximum possible number of connected components of the pseudo-spectrum of $\mathcal{P}$. It also builds a bridge between the spectra of $\mathcal{P}$ and $\mathcal{P}_\Delta$. In the special case $N = 2$, which has been established in [8, Theorem 2.3].

Theorem 2.3.4 *For a tensor polynomial $\mathcal{P}$ as in (2.3.1), if $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ is bounded, then it has no more than $IL(N-1)^{I-1}$ connected components, and any $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ has at least one eigenvalue in each one of these components, with $\epsilon \geq 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$.*

Furthermore, $\mathcal{P}$ and $\mathcal{P}_\Delta$ have the same number of eigenvalues, counting multiplicities, in each connected component of $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$.

Proof Assume that $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ is bounded. It follows from Theorem 2.3.3 that any perturbed tensor polynomial $\mathcal{P}_\Delta \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ of $\mathcal{P}$, of the form

$$\mathcal{P}_\Delta(z) = (\mathcal{A}_0 + \Delta\mathcal{A}_0) + z(\mathcal{A}_1 + \Delta\mathcal{A}_1) \cdots + z^L(\mathcal{A}_L + \Delta\mathcal{A}_L),$$

for any $z \in \mathbb{C}$, must satisfy $\det(\mathcal{A}_L + \Delta\mathcal{A}_L) \neq 0$. That is the leading coefficient of the polynomial $\det(\mathcal{P}_\Delta(z))$ is nonzero. As a consequence, for the tensor polynomial $\mathcal{P}_\Delta$, there are at most $LI(N-1)^{LI-1}$ eigenvalues, counting multiplicities and the

same for every member of the family of tensor polynomials

$$\mathcal{P}_{\Delta,t}(z) = (1-t)\mathcal{P}(z) + t\mathcal{P}_{\Delta}(z), \quad t \in [0,1],$$

for any $z \in \mathbb{C}$. Moreover, $\mathcal{P}_{\Delta,t}$ belongs to $B(\mathcal{P}, \mathbf{w}, \epsilon)$ and all of its eigenvalues lie in $\Lambda_\epsilon(\mathcal{P})$.

The coefficients of the scalar polynomial $\det(\mathcal{P}_{\Delta,t}(z))$ are continuous functions of $t \in [0,1]$. By continuity of the zeros, the eigenvalues of $\mathcal{P}_{\Delta,t}$ form a continuous path from the eigenvalues of $\mathcal{P} = \mathcal{P}_{\Delta,0}$ to that of $\mathcal{P}_\Delta = \mathcal{P}_{\Delta,1}$. Thus, if $\mathcal{P}$ has k eigenvalues (counting multiplicities) in a connected component $\mathcal{G}$ of $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ and its $IL(N-1)^{I-1} - k$ remaining eigenvalues are isolated in $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})\backslash\mathcal{G}$, as for the eigenvalues of every $\mathcal{P}_{\Delta,t}$, $t \in [0,1]$. Consequently, $\mathcal{P}_\Delta$ has exactly k eigenvalues in $\mathcal{G}$, counting multiplicities. Thus $\mathcal{P}$ and $\mathcal{P}_\Delta$ have the same number of eigenvalues, counting multiplicities, in each connected component of $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$.

Finally, each bounded connected component of $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ contains at least one eigenvalue of the tensor polynomial $\mathcal{P}$. By the above discussion, it contains at least one eigenvalue of the perturbed tensor $\mathcal{P}_\Delta$. Hence, $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ cannot have more than $IL(N-1)^{I-1}$ connected components. □

Since the origin lies in $\Lambda_{w_L\epsilon}(\mathcal{A}_L)$ if and only if $\min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{A}_L\mathbf{x}^{N-1}\right\|_2 \leq w_L\epsilon$ (see (2.2.3)), we also have the following corollary.

Corollary 2.3.1 *For any $\epsilon \geq 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$, if $\min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{A}_L\mathbf{x}^{L-1}\right\|_2 > w_L\epsilon$, then $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$ consists of no more than $IL(N-1)^{I-1}$ bounded connected components, each containing one or more eigenvalues of the tensor polynomial $\mathcal{P}$.*

We also obtain the following corollary for the ϵ-pseudo-spectrum of a tensor $\mathcal{A} \in CT_{N,I}$.

Corollary 2.3.2 *Let $\epsilon \geq 0$. For $\mathcal{A} \in CT_{N,I}$, its ϵ-pseudo-spectrum is bounded and has no more than $I(N-1)^{I-1}$ connected components. Moreover, if $\mathcal{A} \in RT_{N,I}$, then its ϵ-pseudo-spectrum is symmetric with respect to the real axis.*

2.3.2 Backward Errors

For a tensor polynomial $\mathcal{P}$ as in (2.3.1), a definition of the normwise backward error of an approximate eigenpair $(\tilde{\lambda}; \tilde{\mathbf{x}})$ of $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ is

$$\begin{aligned}\eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}}) := \min \Big\{ \epsilon : \Big(\mathcal{P}(\tilde{\lambda}) + \Delta\mathcal{P}(\tilde{\lambda})\Big)\tilde{\mathbf{x}}^{N-1} = \mathbf{0}_I, \\ \text{with } \Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \ldots + z^L\Delta\mathcal{A}_L \\ \text{for all } z \in \mathbb{C} \text{ and } \|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon,\ l = 0, 1, \ldots, L\Big\},\end{aligned}$$

and the backward error of the approximate eigenvalue $\tilde{\lambda}$ is characterized by

$$\eta_{\mathbf{w}}(\tilde{\lambda}) := \min_{\tilde{\mathbf{x}} \neq \mathbf{0}_I} \eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}}) = \min_{\|\tilde{\mathbf{x}}\|_2 = 1} \eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}}).$$

Our first result presents an explicit expression for $\eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}})$ and makes precise the inituition that if the residual $\mathbf{r} = \mathcal{P}(\tilde{\lambda})\tilde{\mathbf{x}}^{N-1}$ is small, then we have an accurate approximate eigenpair.

Comparing Definition 2.3.1 with the normwise backward error of an approximate eigenpair $(\tilde{\lambda}; \tilde{\mathbf{x}})$ of $\mathcal{P}$, it is obvious that the ϵ-pseudo-spectrum can be expressed in term of the normwise backward error of $\tilde{\lambda}$ as $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) = \{\tilde{\lambda} \in \mathbb{C} : \eta_{\mathbf{w}}(\tilde{\lambda}) \leq \epsilon\}$. In the matrix case where $N = 2$, this theorem has been established in [6, Lemma 2.2].

Theorem 2.3.5 *Let $(\tilde{\lambda}; \tilde{\mathbf{x}})$ be an approximate eigenpair of $\mathcal{P}$. Then, the normwise backward error $\eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}})$ is given by*

$$\eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}}) = \frac{\|\mathbf{r}\|_2}{p_{\mathbf{w}}(|\tilde{\lambda}|)\|\tilde{\mathbf{x}}\|_2^{N-1}}, \tag{2.3.8}$$

where $p_{\mathbf{w}}$ is given in (2.3.3) and $\mathbf{r} = \mathcal{P}(\tilde{\lambda})\tilde{\mathbf{x}}^{N-1}$.

Proof For any $\epsilon \geq 0$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$, according to the definition of the normwise backward error of an approximate eigenpair $(\tilde{\lambda}; \tilde{\mathbf{x}})$ of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w}$, there exist some tensor polynomials $\Delta\mathcal{P} \in B(\mathcal{P}, \mathbf{w}, \epsilon)$ such that $\left(\mathcal{P}(\tilde{\lambda}) + \Delta\mathcal{P}(\tilde{\lambda})\right)\tilde{\mathbf{x}}^{N-1} = \mathbf{0}_I$. For any $z \in \mathbb{C}$, let $\Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 \cdots + z^L\Delta\mathcal{A}_L$ with $\|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon$ and all l. Then we have

$$\mathbf{r} = \mathcal{P}(\tilde{\lambda})\tilde{\mathbf{x}}^{N-1} = -\Delta\mathcal{P}(\tilde{\lambda})\tilde{\mathbf{x}}^{N-1} = -\left(\Delta\mathcal{A}_0\tilde{\mathbf{x}}^{N-1} + \tilde{\lambda}\Delta\mathcal{A}_1\tilde{\mathbf{x}}^{N-1} \cdots + \tilde{\lambda}^L\Delta\mathcal{A}_L\tilde{\mathbf{x}}^{N-1}\right).$$

It means that

$$\|\mathbf{r}\|_2 \leq \left(\sum_{l=0}^{L} w_l|\tilde{\lambda}|^l\right)\|\tilde{\mathbf{x}}\|_2^{N-1}\epsilon = p_{\mathbf{w}}(|\tilde{\lambda}|)\|\tilde{\mathbf{x}}\|_2^{N-1}\epsilon.$$

We see that the right-hand side of (2.3.8) is a lower bound for $\eta_{\mathbf{w}}(\tilde{\lambda}, \tilde{\mathbf{x}})$. This lower bound is attained by choosing

$$\Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 \cdots + z^L\Delta\mathcal{A}_L,$$

with $z \in \mathbb{C}$, where, for all l,

$$\Delta\mathcal{A}_l = -\frac{w_l\epsilon}{p_{\mathbf{w}}(|\tilde{\lambda}|)}\text{sign}(\tilde{\lambda}^l)\mathbf{r} \otimes \overline{\tilde{\mathbf{x}}}^{\otimes N-1}/\|\tilde{\mathbf{x}}\|_2^{2(N-1)}.$$

Here, $\text{sign}(\tilde{\lambda}^l)$ is the sign function defined in (2.3.7) and $\mathbf{r} = \mathcal{P}(\tilde{\lambda})\tilde{\mathbf{x}}^{N-1} = -\Delta\mathcal{P}(\tilde{\lambda})\tilde{\mathbf{x}}^{N-1}$. □

2.3.3 Nearest Irregular Tensor Polynomials

Regularity is a nice property possessed by tensor polynomial eigenvalue problems in applications. The distance from a regular tensor polynomial to the nearest irregular one is therefore of great interest. The notion of the ϵ-pseudo-spectrum of a polynomial tensor sheds light on this problem.

For a tensor polynomial $\mathcal{P}$ as in (2.3.1), suppose that $\mathcal{P}(z)$ is regular. The distance from $\mathcal{P}$ to its nearest irregular tensor polynomial with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ can be defined as

$$\begin{aligned}\delta_{\mathbf{w}}(\mathcal{P}) = \min\{\epsilon : &\det(\mathcal{P}(z) + \Delta\mathcal{P}(z)) = 0 \\ &\text{with } \Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \ldots + z^L\Delta\mathcal{A}_L \\ &\text{for all } z \in \mathbb{C} \text{ and } \|\Delta\mathcal{A}_l\|_F \le w_l\epsilon,\ l = 0, 1, \ldots, L\}.\end{aligned} \tag{2.3.9}$$

From Definition 2.3.1 of $\Lambda_{\mathbf{w},\epsilon}(\mathcal{P})$, it is evident that

$$\delta_{\mathbf{w}}(\mathcal{P}) \ge \min\{\epsilon : \Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) = \mathbb{C}\}.$$

In general, the inequality can be strict. As an immediate consequence of Theorem 2.3.1, a lower bound of $\delta_{\mathbf{w}}(P)$ is given as follows,

$$\delta_{\mathbf{w}}(\mathcal{P}) \ge \max_{\lambda\in\mathbb{C}} \min_{\|\mathbf{x}\|_2=1} \frac{\left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2}{p_{\mathbf{w}}(|\lambda|)}.$$

2.4 Further Discussions

In this section, we illustrate our theory via an example, conclude our main results and list some problems for future research.

2.4.1 An Example

According to the ϵ-pseudo-spectrum of any given tensor polynomial with respect to the weights $\mathbf{w} \in \mathbb{R}_+^{L+1}$, we have

$$\Lambda_{\mathbf{w},\epsilon}(\mathcal{P}) = \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{x}\|_2=1} \left\|\mathcal{P}(\lambda)\mathbf{x}^{N-1}\right\|_2 \le p_{\mathbf{w}}(|\lambda|)\epsilon\right\}.$$

Let $\mathcal{P}(z) = \mathcal{A} - z\mathcal{B}$ and $\mathbf{w} = (1, 0)^\top$. Then the ϵ-pseudo-spectrum of the tensor generalized eigenvalue problem with the pair $\{\mathcal{A}, \mathcal{B}\}$ can be expressed as

$$\Lambda_\epsilon(\mathcal{A}, \mathcal{B}) = \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{x}\|_2=1} \left\|(\mathcal{A} - \lambda\mathcal{B})\mathbf{x}^{N-1}\right\|_2 \le \epsilon\right\}.$$

Given $\mathcal{A}, \mathcal{B} \in CT_{N,I}$, the pair $\{\mathcal{A}, \mathcal{B}\}$ is a diagonalizable tensor pair (see [15, Section 2.3]), if there exist two nonsingular matrices $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{C}^{I\times I}$ such that

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{Q}_1 \times_2 \mathbf{Q}_2 \cdots \times_N \mathbf{Q}_2, \quad \mathcal{T} = \mathcal{B} \times_1 \mathbf{Q}_1 \times_2 \mathbf{Q}_2 \cdots \times_N \mathbf{Q}_2$$

are diagonal. Furthermore, if $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are also unitary, then the pair $\{\mathcal{A}, \mathcal{B}\}$ is called a unitarily diagonalizable tensor pair. Let the main diagonal entries of $\mathcal{S}$ and $\mathcal{T}$ be $\{s_1, s_2, \ldots, s_I\}$ and $\{t_1, t_2, \ldots, t_I\}$, respectively.

For similarity, we suppose that all t_i are nonzero with all i. For the unitary diagonalizable tensor pair $\{\mathcal{A}, \mathcal{B}\}$ and $\mathbf{y} = \mathbf{Q}_2^{-1}\mathbf{x}$, one has $\mathcal{S}\mathbf{y}^{N-1} = \big(\mathcal{A} \times_1 \mathbf{Q}_1 \times_2 \mathbf{Q}_2 \cdots \times_N \mathbf{Q}_2\big)(\mathbf{Q}_2^{-1}\mathbf{x})^{N-1} = \mathbf{Q}_1(\mathcal{A}\mathbf{x}^{N-1})$ and $\mathcal{T}\mathbf{y}^{N-1} = \mathbf{Q}_1(\mathcal{B}\mathbf{x}^{N-1})$. It follows that the corresponding ϵ-pseudo-spectrum can be represented as

$$\begin{aligned}
\Lambda_\epsilon(\mathcal{A}, \mathcal{B}) &= \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{x}\|_2=1} \left\|(\mathcal{A} - \lambda\mathcal{B})\mathbf{x}^{N-1}\right\|_2 \le \epsilon\right\} \\
&= \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{y}\|_2=1} \left\|(\mathcal{S} - \lambda\mathcal{T})\mathbf{y}^{N-1}\right\|_2 \le \epsilon\right\}, \text{ with } \mathbf{y} = \mathbf{Q}_2^{-1}\mathbf{x} \\
&= \left\{\lambda \in \mathbb{C} : \min_{\|\mathbf{y}\|_2=1} \left\|(\mathcal{D} - \lambda\mathcal{I})\mathbf{y}^{N-1}\right\|_2 \le \epsilon\right\},
\end{aligned}$$

where $\mathcal{D}$ is a diagonal tensor with main diagonal entries $d_i = s_i/t_i$ with all i. According to Proposition 2.2.2, we have

$$\Lambda_\epsilon(\mathcal{A}, \mathcal{B}) = \{d_1, d_2, \ldots, d_I\} + \Delta_{\rho_{N,I}^{-1}\epsilon}.$$

Next, we illustrate the ϵ-pseudo-spectrum of tensor generalized eigenvalue problem with a unitary diagonalizable tensor pair.

Let $(N, I) = (3, 2)$, $(s_1, s_2) = (1 + \imath, -1 - \imath)$ and $(t_1, t_2) = (1, 1)$. For different $\epsilon > 0$, the generalized eigenvalues of the tensor pair $\{\mathcal{A} + \Delta\mathcal{A}, \mathcal{B}\}$ are displayed in Fig. 2.1, where $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{C}^{2\times 2}$ are two randomized unitary matrices and $\Delta\mathcal{A} \in CT_{3,2}$ satisfies $\|\Delta\mathcal{A}\|_F \le \epsilon$.

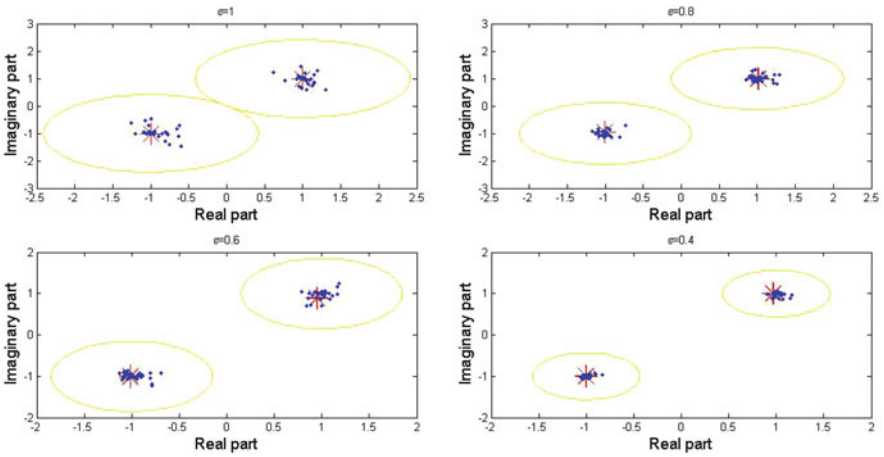

Fig. 2.1 Illustration for the ϵ-pseudo-spectrum of tensor generalized eigenvalue problem with the unitary diagonalizable tensor pair $\{\mathcal{A}, \mathcal{B}\}$. For a given ϵ, the number of samples is 20. Red stars represent the generalized eigenvalues of the tensor pair $\{\mathcal{A}, \mathcal{B}\}$, blue dots for the pseudo generalized eigenvalues and yellow cycles for the boundary of $\Lambda_\epsilon(\mathcal{A}, \mathcal{B})$

2.4.2 Conclusion and Remarks

Although we extend the pseudo-spectrum theory from matrices to tensors, there are a few important distinctions between matrix and tensor pseudo-spectra.

1. In the literature of matrix pseudo-spectra, there are various equivalent definitions [2, 3]. In this chapter, we only generalize one of their natural extensions by the Frobenius norm. It is not clear whether other forms of equivalent definitions can be extended to the tensor case.
2. Many statements of the tensor pseudo-spectra involve the scalar $\rho_{N,I}$ (first mentioned in Corollary 2.2.1) where N is the degree of the associated tensor and I is the dimension of the underlying space. In the matrix case, the scalar $\rho_{N,I} = 1$ is independent of I.
3. The computation and display of matrix pseudospectra have been well-studied, see for example [3, 22–25] and the references therein. In particular, the boundary structure of the matrix pseudo-spectra is reasonably well-understood, paving the way in their computation and visualization. There is a lack in computational tools for locating the boundaries of tensor pseudo-spectra. One is limited to computing points in the tensor pseudo-spectra via randomized techniques. How to efficiently compute and visualize the tensor pseudo-spectra is still an open problem.
4. We present the pseudospectrum for generalized tensor eigenvalues, and a set to locate this pseudospectrum [26].

Malyshev and Sadkane [27] investigate a component-wise ϵ-pseudo-spectrum of a matrix $\mathbf{A} \in \mathbb{R}^{I\times I}$ and discuss how to compute it. Rump [28] characterizes the structured (complex and real) ϵ-pseudo-spectrum for a number of structures and points that there is no significant difference to the usual, unstructured ϵ-pseudo-spectrum. Tisseur and Higham [6] discuss a structured ϵ-pseudo-spectrum for matrix polynomial eigenvalue problems.

Let $\mathbf{I} \in CT_{N,I}$ be the entire one tensor, that is, $\mathbf{I}_{i_1 i_2 \ldots i_N} = 1$ for all i_n and n.

Definition 2.4.1 Let $\epsilon \geq 0$. The component-wise ϵ-pseudo-spectrum of $\mathcal{A} \in CT_{N,I}$ is defined as

$$\Lambda_\epsilon^c(\mathcal{A}) = \Big\{\lambda \in \mathbb{C} : (\mathcal{A}+\mathcal{E})\mathbf{x}^{N-1} = \lambda\,\mathbf{x}^{[N-1]} \text{ for some } \mathcal{E} \in CT_{N,I}$$
$$\text{with } |\mathcal{E}| \leq \epsilon\mathbf{I} \text{ and nonzero vectors } \mathbf{x} \in \mathbb{C}^I\Big\},$$

where $|\mathcal{A}|$ means the modulus of $\mathcal{A}$, that is, all entries of $|\mathcal{A}|$ are $|a_{i_1 i_2 \ldots i_N}|$ for all i_n and n. Here, $|\mathcal{E}| \leq \epsilon\mathbf{I}$ means the modulus of each entries of $\mathcal{E}$ is less or equal to ϵ.

If $\epsilon = 0$, then the component-wise 0-pseudo-spectrum equals to $\Lambda(\mathcal{A})$.

In order to define the component-wise ϵ-pseudo-spectrum of $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$. For a tensor polynomial $\mathcal{P}$ as in (2.3.1), we set

$$B_c(\mathcal{P}, \mathbf{w}, \epsilon) = \Big\{\mathcal{P} + \Delta\mathcal{P} : \Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \ldots + z^L\Delta\mathcal{A}_L$$
$$\text{for all } z \in \mathbb{C} \text{ and } |\Delta\mathcal{A}_l| \leq w_l\epsilon\,\mathbf{I},\ l = 0, 1, \ldots, L\big\}.$$

Definition 2.4.2 For a tensor polynomial $\mathcal{P}$ as in (2.3.1) and $\epsilon > 0$, the component-wise ϵ-pseudo-spectrum of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ is defined as

$$\Lambda_{\mathbf{w},\epsilon}^c(\mathcal{P}) = \{\lambda \in \mathbb{C} : \det(\mathcal{P}_\Delta(\lambda)) = 0 \text{ and } \mathcal{P}_\Delta \in B_c(\mathcal{P}, \mathbf{w}, \epsilon)\}$$
$$= \{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } \mathcal{P}_\Delta \in B_c(\mathcal{P}, \mathbf{w}, \epsilon)\}.$$

We introduce the structured ϵ-pseudo-spectrum of any $\mathcal{A} \in CT_{N,I}$ in the Frobenius norm (see Definition 2.2.3). We now define the structured ϵ-pseudo-spectrum of any tensor polynomial with respect to nonnegative weights in the Frobenius norm.

Let $\mathbb{S}$ be a subspace of $CT_{N,I}$. For a tensor polynomial $\mathcal{P}$ as in (2.3.1), we define

$$B^{\mathbb{S}}(\mathcal{P}, \mathbf{w}, \epsilon) = \Big\{\mathcal{P} + \Delta\mathcal{P} : \Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \ldots + z^L\Delta\mathcal{A}_L$$
$$\text{for all } z \in \mathbb{C} \text{ and } \|\Delta\mathcal{A}_l\|_F \leq w_l\epsilon,\ l = 0, 1, \ldots, L\big\},$$

and

$$B_c^{\mathbb{S}}(\mathcal{P}, \mathbf{w}, \epsilon) = \Big\{ \mathcal{P} + \Delta\mathcal{P} : \Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \ldots + z^L \Delta\mathcal{A}_L \\ \text{for all } z \in \mathbb{C} \text{ and } |\Delta\mathcal{A}_l| \leq w_l \epsilon \mathbf{I},\ l = 0, 1, \ldots, L \Big\},$$

with $\Delta\mathcal{A}_l \in \mathbb{S}$ and $\mathbf{w} \in \mathbb{R}_+^{L+1}$.

Definition 2.4.3 For a tensor polynomial $\mathcal{P}$ as in (2.3.1) and $\epsilon \geq 0$, the structured ϵ-pseudo-spectrum of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ is defined as

$$\Lambda_{\mathbf{w},\epsilon}^{\mathbb{S}}(\mathcal{P}) = \{\lambda \in \mathbb{C} : \det(\mathcal{P}_\Delta(\lambda)) = 0 \text{ and } \mathcal{P}_\Delta \in B^{\mathbb{S}}(\mathcal{P}, \mathbf{w}, \epsilon)\} \\ = \{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } \mathcal{P}_\Delta \in B^{\mathbb{S}}(\mathcal{P}, \mathbf{w}, \epsilon)\}.$$

Definition 2.4.4 For a tensor polynomial $\mathcal{P}$ as in (2.3.1) and $\epsilon \geq 0$, the structured component-wise ϵ-pseudo-spectrum of the tensor polynomial $\mathcal{P}$ with respect to $\mathbf{w} \in \mathbb{R}_+^{L+1}$ is defined as

$$\Lambda_{\mathbf{w},\epsilon}^{c,\mathbb{S}}(\mathcal{P}) = \{\lambda \in \mathbb{C} : \det(\mathcal{P}_\Delta(\lambda)) = 0 \text{ and } \mathcal{P}_\Delta \in B_c^{\mathbb{S}}(\mathcal{P}, \mathbf{w}, \epsilon)\} \\ = \{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } \mathcal{P}_\Delta \in B_c^{\mathbb{S}}(\mathcal{P}, \mathbf{w}, \epsilon)\}.$$

References

1. L. Trefethen, Pseudospectra of matrices, in *Numerical Analysis 1991 (Dundee, 1991)*. Pitman Research Notes in Mathematics Series, vol. 260 (Longman Science and Technology, Harlow, 1992), pp. 234–266
2. L. Trefethen, Pseudospectra of linear operators. SIAM Rev. **39**(3), 383–406 (1997)
3. L. Trefethen, Computation of pseudospectra, in *Acta Numerica, 1999*. Acta Numerica, vol. 8 (Cambridge University, Cambridge, 1999), pp. 247–295
4. L. Trefethen, M. Embree, *Spectra and Pseudospectra. The Behaviour of Nonnormal Matrices and Operators* (Princeton University, Princeton, 2006)
5. T. Wright, L. Trefethen, Pseudospectra of rectangular matrices. IMA J. Numer. Anal. **32**, 501–519 (2002)
6. F. Tisseur, N. Higham, Structured pseudospectra for polynomial eigenvalue problems, with applications. SIAM J. Matrix Anal. Appl. **23**(1), 187–208 (2001)
7. N. Higham, F. Tisseur, More on pseudospectra for polynomial eigenvalue problems and applications in control theory. Linear Algebra Appl. **351/352**, 435–453 (2002)
8. P. Lancaster, P. Psarrakos, On the pseudospectra of matrix polynomials. SIAM J. Matrix Anal. Appl. **27**(1), 115–129 (2005)
9. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
10. S. Hu, Z. Huang, C. Ling, L. Qi, On determinants and eigenvalue theory of tensors. J. Symb. Comput. **50**, 508–531 (2013)
11. M. Che, L. Qi, Y. Wei, Perturbation bounds of tensor eigenvalue and singular value problems with even order. Linear Multilinear Algebra **64**(4), 622–652 (2016)
12. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)

13. K. Chang, K. Pearson, T. Zhang, Perron-Frobenius theorem for nonnegative tensors. Commun. Math. Sci. **6**(2), 507–520 (2008)
14. T. Zhang, Existence of real eigenvalues of real tensors. Nonlinear Anal. Theory Methods Appl. **74**(8), 2862–2868 (2011)
15. W. Ding, Y. Wei, Generalized tensor eigenvalue problems. SIAM J. Matrix Anal. Appl. **36**(3), 1073–1099 (2015)
16. L. Lim, *Spectrum and Pseudospectrum of a Tensor* (*Talk*) (University of California, Berkeley, 2008)
17. L. Lim, Singular values and eigenvalues of tensors: a variational approach, in *IEEE CAMSAP 2005: First International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (IEEE, Piscataway, 2005), pp. 129–132
18. M. Embree, L. Trefethen, Generalizing eigenvalue theorems to pseudospectra theorems. SIAM J. Sci. Comput. **23**(2), 583–590 (2001)
19. V. Kostić, L. Cvetković, D. Cvetković, Pseudospectra localizations and their applications. Numer. Linear Algebra Appl. **23**, 356–372 (2016)
20. J. He, C. Li, Y. Wei, Pseudospectra localization sets of tensors with applications. J. Comput. Appl. Math. **369**(e112580) (2020)
21. W. Ding, Z. Hou, Y. Wei, Tensor logarithmic norm and its applications. Numer. Linear Algebra Appl. **23**(6), 989–1006 (2016)
22. T. Braconnier, N. Higham, Computing the field of values and pseudospectra using the Lanczos method with continuation. BIT Numer. Math. **36**(3), 422–440 (1996)
23. S. Lui, Computation of pseudospectra by continuation. SIAM J. Sci. Comput. **18**(2), 565–573 (1997)
24. T. Wright, *Eigtool* (2002). http://www.comlab.ox.ac.uk/pseudospectra/eigtool/
25. T. Wright, L. Trefethen, Large-scale computation of pseudospectra using ARPACK and eigs. SIAM J. Sci. Comput. **23**(2), 591–605 (2001)
26. C. Li, Q. Liu, Y. Wei, Pseudospectra localizations for generalized tensor eigenvalues to seek more positive definite tensors. Comput. Appl. Math. **38**, Article: 183 (2019)
27. A. Malyshev, M. Sadkane, Componentwise pseudospectrum of a matrix. Linear Algebra Appl. **378**, 283–288 (2004)
28. S. Rump, Eigenvalues, pseudospectrum and structured perturbations. Linear Algebra Appl. **413**(2-3), 567–593 (2006)

Chapter 3
Perturbation Theory

3.1 Preliminaries

Backward errors and condition numbers are important in modern numerical linear algebra. If a numerical algorithm is backward stable, then the product of the condition number and the backward error is the first order asymptotic error bound of the solution. We recommend the monographs Stewart and Sun [1], Golub and Van Loan [2], Wilkinson [3], and Kågström and Ruhe [4] for the background of the matrix eigenvalue problem. These consist of the matrix standard eigenvalue problem, the matrix generalized eigenvalue problem and the matrix polynomial eigenvalue problem. The perturbation theory of eigenvalue problems for matrices attracts extensive attention of many researchers, and the interested readers are referred to [1, 3, 5–7].

Similarly, the eigenvalue problems for tensors in this chapter consist of the tensor standard eigenvalue problem (i.e., the eigenvalue problem and the E-eigenvalue problem), the tensor generalized eigenvalue problem and the tensor polynomial eigenvalue problem. In Sect. 1.3.1, we introduce the definition of the real symmetric tensor standard eigenvalue problem, i.e., (H-) eigenpairs, and E- or Z-eigenpairs and the numerical algorithms for the computation of H- and Z-eigenpairs. A systematic study of the perturbation theory of the eigenvalue problem for tensors appears to be lacking in the literature. The main purpose is to bridge this gap by considering the first-order perturbation bound of the eigenvalue problem for tensors with relative normwise or componentwise perturbations.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_3

3.1.1 Definitions

In 2005, the definition of symmetric tensors are introduced by Qi [8] and Lim [9]. Now we introduce the definition of mode-symmetric tensors, as a generalization of symmetric tensors. When $N = 2$, both symmetric and mode-symmetric tensors are reduced to symmetric matrices.

Definition 3.1.1 For all i_n and n, if the entries of $\mathcal{A} \in RT_{N,I}$ satisfy the following formulae

$$a_{i_1 i_2 \ldots i_N} = a_{i_2 i_3 \ldots i_N i_1} = \cdots = a_{i_N i_1 \ldots i_{N-1}},$$

then $\mathcal{A}$ is called mode-symmetric.

Suppose that $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^I$ have unit norms and are orthogonal to each other. If $\mathcal{A} \in RT_{3,I}$ is given by

$$\mathcal{A} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} + \mathbf{b} \otimes \mathbf{c} \otimes \mathbf{a} + \mathbf{c} \otimes \mathbf{a} \otimes \mathbf{b},$$

then $\mathcal{A}$ is mode-symmetric [10]. Sometimes, we call $\mathcal{A}$ a cyclically symmetric tensor [11].

For a given n, we define the mode-n E-eigenpairs of $\mathcal{A} \in RT_{N,I}$, which is a generalization of the E-eigenpairs of the real symmetric tensors. In the following definition, if $\mathbf{x}^*\mathbf{x}$ is replaced by $\mathbf{x}^\top\mathbf{x}$, this definition extends that of Qi [8].

Definition 3.1.2 Assume that $\mathcal{A} \in RT_{N,I}$. For a given n, if there exist a unitary vector $\mathbf{x}_n \in \mathbb{C}^I$ and $\lambda_n \in \mathbb{C}$ such that

$$\mathcal{A} \times_1 \mathbf{x}_n^\top \cdots \times_{n-1} \mathbf{x}_n^\top \times_{n+1} \mathbf{x}_n^\top \cdots \times_N \mathbf{x}_n^\top = \lambda_n \mathbf{x}_n, \tag{3.1.1}$$

then $(\lambda_n; \mathbf{x}_n)$ is called a mode-n E-eigenpair of $\mathcal{A}$.

If $\mathbf{x}_n \in \mathbb{R}^I$ and $\lambda_n \in \mathbb{R}$, then $(\lambda_n; \mathbf{x}_n)$ is called a mode-n Z-eigenpair of $\mathcal{A}$. Moreover, the mode-n E-and Z-spectra of $\mathcal{A}$ are defined, respectively, as

$$E_n(\mathcal{A}) = \{\lambda \in \mathbb{C} : \lambda \text{ is a mode-}n \text{ E-eigenvalue of } \mathcal{A}\},$$
$$Z_n(\mathcal{A}) = \{\lambda \in \mathbb{C} : \lambda \text{ is a mode-}n \text{ Z-eigenvalue of } \mathcal{A}\}.$$

For a given n, we also define the mode-n eigenpairs of $\mathcal{A} \in RT_{N,I}$, which are the generalization of the eigenpairs of the real symmetric tensors [8].

Definition 3.1.3 Suppose that $\mathcal{A} \in RT_{N,I}$. For a given n, if there exist a nonzero vector $\mathbf{x}_n \in \mathbb{C}^I$ and $\lambda_n \in \mathbb{C}$ such that

$$\mathcal{A} \times_1 \mathbf{x}_n^\top \cdots \times_{n-1} \mathbf{x}_n^\top \times_{n+1} \mathbf{x}_n^\top \cdots \times_N \mathbf{x}_n^\top = \lambda_n \mathbf{x}_n^{[N-1]}, \tag{3.1.2}$$

then $(\lambda_n; \mathbf{x}_n)$ is called a mode-n eigenpair of $\mathcal{A}$.

If $\mathbf{x}_n \in \mathbb{R}^I$ and $\lambda_n \in \mathbb{R}$, then $(\lambda_n; \mathbf{x}_n)$ is called a mode-n H-eigenpair of $\mathcal{A}$. Moreover, the mode-n spectrum $\Lambda_n(\mathcal{A})$ of $\mathcal{A}$ is defined as

$$\Lambda_n(\mathcal{A}) = \{\lambda \in \mathbb{C} : \lambda \text{ is a mode-}n \text{ eigenvalue of } \mathcal{A}\}.$$

The mode-n spectral radius $\rho_n(\mathcal{A})$ is defined as $\max\{|\lambda| : \lambda \in \Lambda_n(\mathcal{A})\}$. The spectral radius of $\mathcal{A}$ is denoted by $\rho(\mathcal{A}) = \max\{\rho_1(\mathcal{A}), \rho_2(\mathcal{A}), \ldots, \rho_N(\mathcal{A})\}$.

For all n, the mode-n eigenvectors of $\mathcal{A} \in RT_{N,I}$ can be viewed as the generalization of the left and right eigenvectors of $\mathbf{A} \in \mathbb{R}^{I\times I}$. Some properties of the mode-n eigenpairs of $\mathcal{A}$ are presented in the following:

(a) Given symmetric or mode-symmetric $\mathcal{A} \in RT_{N,I}$ and $\mathbf{x} \in \mathbb{C}^I$, we have the following equalities

$$\begin{aligned}\mathcal{A} \times_2 \mathbf{x}^\top \times_3 \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top &= \cdots = \mathcal{A} \times_1 \mathbf{x}^\top \cdots \times_{n-1} \mathbf{x}^\top \times_{n+1} \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top \\ &= \cdots = \mathcal{A} \times_1 \mathbf{x}^\top \times_2 \mathbf{x}^\top \cdots \times_{N-1} \mathbf{x}^\top.\end{aligned}$$

(b) Generally speaking, for all n, $\Lambda_n(\mathcal{A})$ are different sets. For any $m \neq n$, $\rho_m(\mathcal{A}) \neq \rho_n(\mathcal{A})$. Similar results for the mode-n E- and Z-spectra of $\mathcal{A} \in RT_{N,I}$ hold.

(c) Let $\mathcal{A}$ be symmetric or mode-symmetric. If $(\lambda; \mathbf{x})$ is a mode-n eigenpair of $\mathcal{A}$, then for all m and n, $(\lambda; \mathbf{x})$ is a mode-m eigenpair of $\mathcal{A}$. Furthermore, $\Lambda_n(\mathcal{A})$ are the same set with all n, denoted by $\Lambda(\mathcal{A})$.

We illustrate why (a) and (c) hold. Without loss of generality, let $N = 4$. We have

$$\begin{cases}(\mathcal{A} \times_2 \mathbf{x}^\top \times_3 \mathbf{x}^\top \times_4 \mathbf{x}^\top)_i = \sum\limits_{jkl=1}^{I} a_{ijkl} x_j x_k x_l, \\ (\mathcal{A} \times_1 \mathbf{x}^\top \times_3 \mathbf{x}^\top \times_4 \mathbf{x}^\top)_j = \sum\limits_{ikl=1}^{I} a_{ijkl} x_i x_k x_l = \sum\limits_{kli=1}^{I} a_{jkli} x_k x_l x_i, \\ (\mathcal{A} \times_1 \mathbf{x}^\top \times_2 \mathbf{x}^\top \times_4 \mathbf{x}^\top)_k = \sum\limits_{ijl=1}^{I} a_{ijkl} x_i x_k x_l = \sum\limits_{lij=1}^{I} a_{klij} x_k x_i x_j, \\ (\mathcal{A} \times_1 \mathbf{x}^\top \times_2 \mathbf{x}^\top \times_3 \mathbf{x}^\top)_l = \sum\limits_{ijk=1}^{I} a_{ijkl} x_i x_j x_k = \sum\limits_{ijk=1}^{I} a_{lijk} x_i x_j x_k.\end{cases}$$

According to Definitions 3.1.1 and 3.1.3, we have

$$a_{ijkl} = a_{jkli} = a_{klij} = a_{lijk},$$

Table 3.1 All approximated mode-n spectral radii of the positive tensors $\mathcal{A}$, $\mathcal{B}$ and C

	The mode-1 spectral radius	The mode-2 spectral radius	The mode-3 spectral radius
$\mathcal{A}$	4.7284	4.7340	4.7539
$\mathcal{B}$	4.7686	4.7686	4.7686
C	3.6167	3.6167	3.6167

for all i, j, k and l, then (a) and (c) hold for the case of symmetric or mode-symmetric tensors. We now illustrate (b) via the following example.

Example 3.1.1 ([12]) Consider the positive tensor $\mathcal{A} \in \mathbb{R}^{3\times3\times3}$ with

$$\begin{cases} a_{111} = 0.4333,\ a_{121} = 0.4278,\ a_{131} = 0.4140,\ a_{211} = 0.8154,\ a_{221} = 0.0199, \\ a_{231} = 0.5598,\ a_{311} = 0.0643,\ a_{321} = 0.3815,\ a_{331} = 0.8834,\ a_{112} = 0.4866, \\ a_{122} = 0.8087,\ a_{132} = 0.2073,\ a_{212} = 0.7641,\ a_{222} = 0.9924,\ a_{232} = 0.8752, \\ a_{312} = 0.6708,\ a_{322} = 0.8296,\ a_{332} = 0.1325,\ a_{113} = 0.3871,\ a_{123} = 0.0769, \\ a_{133} = 0.3151,\ a_{213} = 0.1355,\ a_{223} = 0.7727,\ a_{233} = 0.4089,\ a_{313} = 0.9715, \\ a_{323} = 0.7726,\ a_{333} = 0.5526. \end{cases}$$

According to Definition 3.1.1, we define the mode-symmetric tensor $\mathcal{B} \in RT_{3,3}$ in the following strategy: choose two tensors $\mathcal{P}, \mathcal{Q} \in RT_{3,3}$ such that

$$p_{ijk} = a_{jki}, \quad q_{ijk} = a_{kij}, \quad b_{ijk} = \frac{a_{ijk} + p_{ijk} + q_{ijk}}{3},$$

with $i, j, k = 1, 2, 3$. Consider the symmetric tensor $C \in RT_{3,3}$ with

$$\begin{cases} c_{111} = 0.0517, & c_{112} = 0.3579, & c_{113} = 0.5298, & c_{122} = 0.7544, & c_{123} = 0.2156, \\ c_{133} = 0.3612, & c_{222} = 0.3943, & c_{223} = 0.0146, & c_{233} = 0.6718, & c_{333} = 0.9723. \end{cases}$$

Since the entries of $\mathcal{A}$, $\mathcal{B}$ and C are positive, for $n = 1, 2, 3$, we can compute all approximated mode-n spectral radii of these three tensors via the NQZ method, which are shown in Table 3.1.

For a given tensor $\mathcal{A} \in RT_{N,I}$, if $(\lambda; \mathbf{x})$ is the mode-1 eigenpair of $\mathcal{A}$, then (3.1.2) in Definition 3.1.3 can be simplified as $\mathcal{A}\mathbf{x}^{N-1} = \lambda\mathbf{x}^{[N-1]}$. If $\mathcal{A}$ is symmetric, then Qi [8] deduces some properties of the mode-1 eigenpairs.

Qi [8] defines the determinant of a symmetric tensor $\mathcal{A} \in RT_{N,I}$. The following definition extends Hu et al. [13] and we name it as the mode-n determinant of $\mathcal{A} \in RT_{N,I}$ with a given n.

Definition 3.1.4 ([13, Definition 1.2]) For a given n, the determinant of Nth-order I-dimensional real tensors denoted by DET_n, is defined as the irreducible polynomial with variables $v_{i_1\ldots i_N} \in \mathbb{R}$ for all i_n and n, such that it is the resultant of

the polynomial system

$$\mathcal{A} \times_1 \mathbf{x}^\top \cdots \times_{n-1} \mathbf{x}^\top \times_{n+1} \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top = \mathbf{0}_I,$$

where $\mathcal{V} \in RT_{N,I}$.

Moreover, the value of the mode-n determinant of a given tensor $\mathcal{A} \in RT_{N,I}$ is denoted by $\det_n(\mathcal{A})$, and is defined as the evaluation of DET_n at the point $\{v_{i_1 \ldots i_N} = a_{i_1 \ldots i_N} : i_n = 1, 2, \ldots, I; n = 1, 2, \ldots, N\}$. For a given $\mathcal{A} \in RT_{N,I}$, if $\det_n(\mathcal{A}) \neq 0$, then $\mathcal{A}$ is said to be mode-n nonsingular.

When $n = 1$, Hu et al. [13, Corollary 6.5] derive that

$$\det_1(\mathcal{A}) = \prod_{\lambda_i \in \Lambda_1(\mathcal{A})} \lambda_i.$$

According to Definitions 3.1.3 and 3.1.4, we can derive the more general result for all n:

$$\det_n(\mathcal{A}) = \prod_{\lambda_i \in \Lambda_n(\mathcal{A})} \lambda_i.$$

Chang et al. [14] consider the tensor generalized eigenvalue problem, in which the tensors belong to $RT_{N,I}$. Note that the tensor generalized eigenvalue problem is a special case of the tensor polynomial eigenvalue problem. We introduce the definition of a tensor polynomial, determined by some tensors in $RT_{N,I}$.

For $\mathcal{A}_l \in RT_{N,I}$ with all l, the tensor polynomial $\mathcal{P}$, determined by the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_2, \ldots, \mathcal{A}_L\}$, is defined as

$$\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^L \mathcal{A}_L, \tag{3.1.3}$$

for all $z \in \mathbb{C}$.

Definition 3.1.5 Suppose that $\mathcal{P}$ is a tensor polynomial defined in (3.1.3). For a given n, if there exist a nonzero vector $\mathbf{x}_n \in \mathbb{C}^I$ and $\lambda_n \in \mathbb{C}$ such that

$$\mathcal{P}(\lambda_n) \times_1 \mathbf{x}_n^\top \cdots \times_{n-1} \mathbf{x}_n^\top \times_{n+1} \mathbf{x}_n^\top \cdots \times_N \mathbf{x}_n^\top = \mathbf{0}_I$$

then $(\lambda_n; \mathbf{x}_n)$ is called a mode-n eigenpair of $\mathcal{P}$.

If $\mathbf{x}_n \in \mathbb{R}^I$ and $\lambda_n \in \mathbb{R}$, then $(\lambda_n; \mathbf{x}_n)$ is called a mode-n H-eigenpair of $\mathcal{P}$. Meanwhile, the mode-n spectrum of $\mathcal{P}$ is denoted by

$$\begin{aligned}\Lambda_n(\mathcal{P}) &= \{\lambda \in \mathbb{C} : \det_n(\mathcal{P}(\lambda)) = 0\} \\ &= \{\lambda \in \mathbb{C} : \lambda \text{ is a mode-}n \text{ eigenvalue of } \mathcal{P}\}.\end{aligned}$$

Remark 3.1.1 In general, for all n, $\Lambda_n(\mathcal{P})$ are different. If all tensors $\mathcal{A}_l \in RT_{N,I}$ are symmetric or mode-symmetric, then $\Lambda_n(\mathcal{P})$ are identical, denoted by $\Lambda(\mathcal{P})$. Here, $(\lambda_n; \mathbf{x}_n)$ is called as an eigenpair of $\mathcal{P}$.

Ding and Wei [15] introduce the regular tensor pair $\{\mathcal{A}, \mathcal{B}\}$, as a generalization of the regular matrix pair [1], where $\mathcal{A}, \mathcal{B} \in RT_{N,I}$. In general, according to Definition 3.1.4 and the definition of $\mathcal{P}$, we give the definition of the mode-n regular tensor tuple $\{\mathcal{A}_0, \mathcal{A}_2, \ldots, \mathcal{A}_L\}$, where $\mathcal{A}_l \in RT_{N,I}$ with all l and a given n.

Definition 3.1.6 Suppose that $\mathcal{P}$ is a tensor polynomial defined in (3.1.3). For a given n, the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ is mode-n singular if $\det_n(\mathcal{P}(\lambda)) = 0$ for all $\lambda \in \mathbb{C}$. Otherwise the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ is said to be mode-n regular.

Furthermore, we say that the tensor polynomial $\mathcal{P}$ is mode-n singular (or mode-n regular) if the corresponding tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ is mode-n singular (or mode-n regular).

We suppose that the tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ is mode-n regular with a given n, where $\mathcal{A}_l \in RT_{N,I}$ with all l. Then according to Definition 3.1.6, there exists a $\hat{\lambda} \in \mathbb{C}$ such that $\det_n(\mathcal{P}(\hat{\lambda})) \neq 0$. Hence we can select another tensor tuple $\left\{\tilde{\mathcal{A}}_0, \tilde{\mathcal{A}}_1, \ldots, \tilde{\mathcal{A}}_L\right\}$ such that $\tilde{\mathcal{A}}_L = \sum_{l=0}^{L} \tilde{\lambda}^l \mathcal{A}_l$ and there is a one-to-one map between $\Lambda_n(\mathcal{P})$ and $\Lambda_n(\tilde{\mathcal{P}})$, where $\det_n(\tilde{\mathcal{A}}_L) \neq 0$ and $\tilde{\mathcal{P}}(\lambda) = \tilde{\mathcal{A}}_0 + \lambda\tilde{\mathcal{A}}_1 + \cdots + +\lambda^L \tilde{\mathcal{A}}_L$.

Thus, for the mode-n regular tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$, we can assume, without loss of generality, that $\mathcal{A}_L$ is nonsingular. For all n, all the mode-n spectra of $\mathcal{P}$, determined by $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$, are finite subsets of $\mathbb{C}$.

For a given tensor tuple $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$, where $\mathcal{A}_l \in RT_{N,I}$ with all l, we can also define another tensor polynomial $\mathcal{P}$

$$\mathcal{P}(\alpha, \beta) = \alpha^L \mathcal{A}_L + \alpha^{L-1}\beta\mathcal{A}_{L-1} + \cdots + \alpha\beta^{L-1}\mathcal{A}_1 + \beta^L \mathcal{A}_0,$$

for all $\alpha, \beta \in \mathbb{C}$.

It is obvious that $\mathcal{P}(\cdot, \cdot)$ is a homogeneous polynomial on α and β. The relationship between $\mathcal{P}(\lambda)$ and $\mathcal{P}(\alpha, \beta)$ is listed below. If $\beta \neq 0$, then $\mathcal{P}(\alpha, \beta) = \beta^L \mathcal{P}(\alpha/\beta)$; and if $\alpha \neq 0$, then $\mathcal{P}(\alpha, \beta) = \alpha^L \hat{\mathcal{P}}(\beta/\alpha)$, where

$$\hat{\mathcal{P}}\left(\frac{\beta}{\alpha}\right) = \mathcal{A}_L + \frac{\beta}{\alpha}\mathcal{A}_{L-1} + \cdots + \left(\frac{\beta}{\alpha}\right)^{L-1} \mathcal{A}_1 + \left(\frac{\beta}{\alpha}\right)^{L} \mathcal{A}_0.$$

If $(\lambda_n, \mathbf{x}_n)$ is a mode-n eigenpair of the tensor polynomial $\mathcal{P}$, then, we can choose a pair $\{\alpha_n; \beta_n\}$ such that

$$\mathcal{P}(\alpha_n, \beta_n) \times_1 \mathbf{x}_n^\top \cdots \times_{n-1} \mathbf{x}_n^\top \times_{n+1} \mathbf{x}_n^\top \cdots \times_N \mathbf{x}_n^\top = \mathbf{0}_I$$

and

$$\lambda_n = \begin{cases} \alpha_n/\beta_n, & \beta_n \neq 0, \\ \infty, & \beta_n = 0. \end{cases}$$

with $\{\alpha_n, \beta_n\} \neq \{0, 0\}$.

In Sect. 1.3.2, we introduce the definition of the singular tuples (i.e., singular values and associated mode-n singular vectors) of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The restricted singular tuples (i.e., restricted singular values and associated restricted mode-n singular vectors) of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ are defined below.

Definition 3.1.7 Let $\mathbf{B}_n \in \mathbb{R}^{I_n \times I_n}$ be symmetric positive definite matrices and $\mathbf{x}_n \in \mathbb{R}^{I_n}$ satisfy $\mathbf{x}_n^\top \mathbf{B}_n \mathbf{x}_n = 1$ with all n. For a given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\sigma \in \mathbb{R}$, if $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is a solution of the nonlinear equations

$$F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)_{-n} = \sigma \mathbf{B}_n \mathbf{x}_n, \tag{3.1.4}$$

where

$$F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)_{-n} = \mathcal{A} \times_1 \mathbf{x}_1^\top \cdots \times_{n-1} \mathbf{x}_{n-1}^\top \times_{n+1} \mathbf{x}_{n+1}^\top \cdots \times_N \mathbf{x}_N^\top,$$

then the unit vector $\mathbf{x}_n$ and σ are called the restricted mode-n singular vector and a restricted singular value of $\mathcal{A}$, respectively.

3.1.2 Symmetric and Mode-Symmetric Embeddings

Well known relationship exists between the singular value decomposition of $\mathbf{A} \in \mathbb{R}^{I \times J}$ and the Schur decomposition of its symmetric embedding $\mathbf{sym}(\mathbf{A}) = ([\mathbf{0}_{I \times I}\ \mathbf{A}; \mathbf{A}^\top\ \mathbf{0}_{J \times J}])$ (see [2, Chapter 8.6]). For a general tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, Ragnarsson et al. [16] derive a method for obtaining a symmetric embedding $\mathbf{sym}(\mathcal{A}) \in RT_{N,\hat{I}}$ from $\mathcal{A}$ with $\hat{I} = I_1 + I_2 + \cdots + I_N$.

We now consider how to obtain a mode-symmetric embedding $\mathbf{msym}(\mathcal{A}) \in RT_{N,\hat{I}}$ from $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The number of the nonzero entries of $\mathbf{msym}(\mathcal{A})$ is $1/(N-1)!$ of that of $\mathbf{sym}(\mathcal{A})$.

For a given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, we define $\mathcal{A}_n \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ as follows:

$$(\mathcal{A}_1)_{i_1, i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N}, \quad (\mathcal{A}_2)_{i_1 i_2 \ldots i_N} = a_{i_2 \ldots i_N i_1}, \quad \ldots, \quad (\mathcal{A}_N)_{i_1 i_2 \ldots i_N} = a_{i_N i_1 \ldots i_{N-1}}.$$

For all n, let $K_n = I_1 + I_2 + \cdots + I_n$. Then the nonzero entries of $\mathbf{msym}(\mathcal{A}) \in RT_{N,K_N}$ satisfy

$$\begin{cases} \mathbf{msym}(\mathcal{A})[1:K_1|K_1+1:K_2|\ldots|K_{N-1}+1:K_N] = \mathcal{A}_1, \\ \mathbf{msym}(\mathcal{A})[K_1+1:K_2|\ldots|K_{N-1}+1:K_N|1:K_1] = \mathcal{A}_2, \\ \qquad\qquad\vdots \qquad\qquad\qquad\vdots \\ \mathbf{msym}(\mathcal{A})[K_{N-1}+1:K_N|1:K_1|\ldots|K_{N-2}+1:K_{N-1}] = \mathcal{A}_N. \end{cases}$$

According to Definition 3.1.1, $\mathbf{msym}(\mathcal{A})$ is mode-symmetric. When $N = 3$, $\mathbf{msym}(\mathcal{A})$ is displayed in Fig. 3.1.

Given N nonzero vectors $\mathbf{x}_n \in \mathbb{C}^{I_n}$ with all n, let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$. Then

$$\mathbf{msym}(\mathcal{A})\mathbf{x}^N = N\mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top,$$

and

$$\mathbf{msym}(\mathcal{A})\mathbf{x}^{N-1} = \begin{pmatrix} \mathcal{A} \times_2 \mathbf{x}_2^\top \times_3 \mathbf{x}_3^\top \cdots \times_N \mathbf{x}_N^\top \\ \mathcal{A} \times_1 \mathbf{x}_1^\top \times_3 \mathbf{x}_3^\top \cdots \times_N \mathbf{x}_N^\top \\ \vdots \\ \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_{N-1} \mathbf{x}_{N-1}^\top \end{pmatrix}.$$

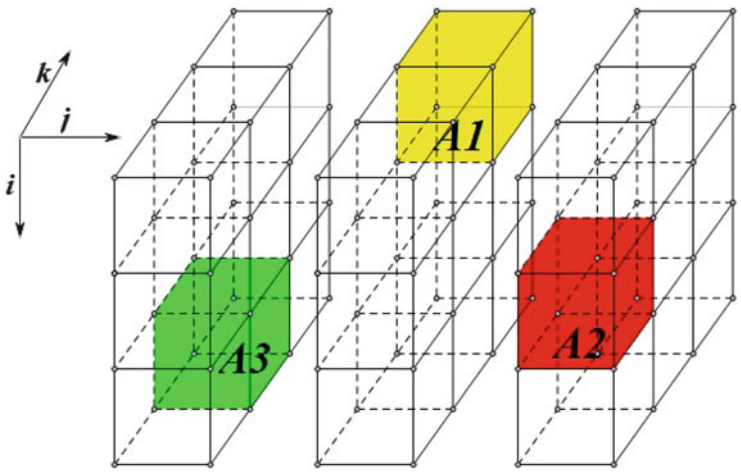

Fig. 3.1 $\mathbf{msym}(\mathcal{A})$; unlabeled parts are zero elements

By some manipulations, we have

$$\frac{\partial \mathbf{msym}(\mathcal{A})\mathbf{x}^N}{\partial \mathbf{x}} = N\mathbf{msym}(\mathcal{A})\mathbf{x}^{N-1}, \quad \frac{\partial^2 \mathbf{msym}(\mathcal{A})\mathbf{x}^N}{\partial \mathbf{x}^2} = N(N-1)\mathbf{msym}(\mathcal{A})\mathbf{x}^{N-2},$$

with

$$\mathbf{msym}(\mathcal{A})\mathbf{x}^{N-2} = \mathbf{msym}(\mathcal{A}) \times_3 \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top.$$

The interested readers are referred to [16, 17] for other details of $\mathbf{sym}(\mathcal{A})\mathbf{x}^N$ and $\mathbf{sym}(\mathcal{A})\mathbf{x}^{N-1}$.

3.2 Perturbation Bounds of Z- and H-Eigenvalues

We first characterize the properties of eigenvalues and Z-eigenvalues of a mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$. Based on these properties and the perturbation of simple eigenvalues of $\mathbf{A} \in \mathbb{R}^{I\times I}$, we then investigate perturbation bounds of an algebraically simple eigenvalue and Z-eigenvalue of $\mathcal{A}$. Lastly, for a given $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, based on the symmetric or mode-symmetric embedding from $\mathcal{A}$, perturbation bounds of an algebraically simple singular value (or restricted singular value) are obtained.

Wilkinson [3], Demmel [18], and Stewart et al. [1] study the condition number of a *simple* eigenvalue of $\mathbf{A} \in \mathbb{C}^{I\times I}$.

Lemma 3.2.1 ([18, Theorem 4.4], [1, Theorem 2.3]) *Let λ be a simple eigenvalue of $\mathbf{A} \in \mathbb{C}^{I\times I}$ with the right and left eigenvectors $\mathbf{x} \in \mathbb{C}^I$ and $\mathbf{y} \in \mathbb{C}^I$, respectively, normalized so that $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$. Let $\lambda + \delta\lambda$ be the corresponding eigenvalue of $\mathbf{A} + \delta\mathbf{A}$. Then*

$$\delta\lambda = \frac{\mathbf{y}^*\delta\mathbf{A}\mathbf{x}}{\mathbf{y}^*\mathbf{x}} + O(\|\delta A\|_2^2),$$

or

$$|\delta\lambda| \le \frac{\|\delta\mathbf{A}\|_2}{|\mathbf{y}^*\mathbf{x}|} + O(\|\delta\mathbf{A}\|_2^2) = \sec\Theta(\mathbf{y},\mathbf{x})\|\delta\mathbf{A}\|_2 + O(\|\delta\mathbf{A}\|_2^2),$$

where $\Theta(\mathbf{y},\mathbf{x})$ is the acute angle between $\mathbf{y}$ and $\mathbf{x}$ and $\|\mathbf{A}\|_2$ is the largest singular value of $\mathbf{A}$. In other words, $\sec\Theta(\mathbf{y},\mathbf{x}) = 1/|\mathbf{y}^*\mathbf{x}|$ is a condition number of λ.

3.2.1 Properties of Eigenvalues and Z-Eigenvalues

In 2005, Qi [8] derives the basic properties of eigenvalues and Z-eigenvalues of a symmetric tensor $\mathcal{A} \in RT_{N,I}$. Those results also hold with any mode-symmetric tensor. In order to prove Theorems 3.2.4 and 3.2.6, we need the following theorem.

Theorem 3.2.1 *Suppose that $\mathcal{A} \in RT_{N,I}$ is mode-symmetric with an even N.*

(a) *$\mathcal{A}$ always has* H*-eigenvalues. $\mathcal{A}$ is positive definite (positive semi-definite) if and only if all of its* H*-eigenvalues are positive (nonnegative).*
(b) *$\mathcal{A}$ always has* Z*-eigenvalues. $\mathcal{A}$ is positive definite (positive semi-definite) if and only if all of its* Z*-eigenvalues are positive (nonnegative).*

Proof We see (3.1.2) is the optimality condition of

$$\max \left\{ \mathcal{A}\mathbf{x}^N : \sum_{i=1}^{I} x_i^N = 1, \mathbf{x} \in \mathbb{R}^I \right\} \tag{3.2.1}$$

and

$$\min \left\{ \mathcal{A}\mathbf{x}^N : \sum_{i=1}^{I} x_i^N = 1, \mathbf{x} \in \mathbb{R}^I \right\}. \tag{3.2.2}$$

Since the feasible set is compact and the objective function is continuous, the global maximizer and minimizer always exist. This shows that (3.1.2) has real solutions, i.e., $\mathcal{A}$ always has H-eigenvalues. Since $\mathcal{A}$ is positive definite (resp. positive semi-definite) if and only if the optimal value of (3.2.1) is positive (resp. nonnegative), we draw the second result in (a).

The proof of (b) is similar. □

Chang et al. [14] define the geometric multiplicity of an eigenvalue λ. Hu et al. [13] consider the algebraic multiplicity of an eigenvalue λ. Similarly, we define the geometric and algebraic multiplicities of a Z-eigenvalue. A Z-eigenvalue is algebraically simple if its algebraic multiplicity is one. This definition is applicable to a tensor generalized eigenvalue, an eigenvalue, a tensor singular value and a tensor restricted singular value.

Next, we introduce a basic result of algebraic function theory [19].

Lemma 3.2.2 *Let*

$$f(x, y) = y^L + p_{L-1}(x)y^{L-1} + \cdots + p_1(x)y + p_0(x),$$

where $p_t(x)$ with $t = 0, 1, \ldots, L-1$ is a polynomial respect to x. For a given $x \in \mathbb{C}$, the nonlinear equation $f(x, y) = 0$ has L roots, denoted by $y_t(x)$ with $t = 1, 2, \ldots, L$. Specially, the roots of $f(0, y) = 0$ are denoted by $y_t(0)$, respectively.

Suppose that $y_t(0)$ is a single root of $f(0, y) = 0$. Then there exists a positive scalar δ_t such that $f(x, y) = 0$ has a single root satisfying

$$y_t(x) = y_t(0) + p_{t1}x + p_{t2}x^2 + \ldots, \tag{3.2.3}$$

and the right-hand side of (3.2.3) is convergent with $|x| < \delta_t$.

For a symmetric tensor $\mathcal{A} \in RT_{N,I}$ and all nonzero vectors $\mathbf{x} \in \mathbb{R}^I$, we have

$$f_2(\mathbf{x}) = \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_2^N}.$$

Theorem 3.2.2 *Assume that $(\lambda_*; \mathbf{x}_*)$ is an* E*-eigenpair of the symmetric tensor $\mathcal{A} \in RT_{N,I}$.*

(a) *If $(\lambda_*; \mathbf{x}_*)$ is a* Z*-eigenpair of $\mathcal{A}$, then for a small perturbation $\delta\mathbf{x} \in \mathbb{R}^I$, we have*

$$f_2(\mathbf{x}_* + \delta\mathbf{x}) = f_2(\mathbf{x}_*) + O(\|\delta\mathbf{x}\|_2^2) = \lambda_* + O(\|\delta\mathbf{x}\|_2^2).$$

(b) *If $\lambda_* \in \mathbb{C}$ is an algebraically simple* E*-eigenvalue of $\mathcal{A}$, then for any symmetric tensor $\mathcal{B} \in RT_{N,I}$, there exists an algebraically simple* E*-eigenvalue $\hat{\lambda}_* \in \mathbb{C}$ of $\mathcal{A} + \mu\mathcal{B}$ such that $\hat{\lambda}_* = \lambda_* + O(\mu)$, where $\mu \in \mathbb{R}$ satisfies $|\mu| \le \epsilon$ with a sufficiently small $\epsilon > 0$.*

Proof The proof of the first part can be found in [20, Theorem 4.1]. In following, we prove the second part.

Since N is even, let $N = 2K$. Define $\mathcal{E} \in RT_{N,I}$ as $e_{i_1 i_2 \ldots, i_N} = \delta_{i_1 i_2}\delta_{i_3 i_4} \ldots \delta_{i_{2K-1} i_{2K}}$, where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$$

According to [8, Theorem 2], the E-eigenvalues λ of the symmetric tensor $\mathcal{A}$ satisfies $\det(\mathcal{A} - \lambda\mathcal{E}) = 0$. Let $\phi(z) = \det(\mathcal{A} - z\mathcal{E})$. If λ_* is an algebraically simple E-eigenvalue of $\mathcal{A}$, then $\phi(\lambda_*) = \det(\mathcal{A} - \lambda_*\mathcal{E}) = 0$ and $\phi'(\lambda_*) = \frac{d\phi}{d\lambda}|_{\lambda=\lambda_*} = 0$.

Similarly, we define $\phi_\mu(z) = \det(z\mathcal{E} - \mathcal{A} - \mu\mathcal{B})$. The degree of $\phi_\mu(z)$ in z is equal to that of $\phi(z)$, and the coefficients of $\phi_\mu(z)$ are polynomials in μ. According to Lemma 3.2.2, there exists a simple root $\hat{\lambda}_*$ of $\phi_\mu(z)$ such that

$$\hat{\lambda}_* = \lambda_* + O(\mu),$$

where $\mu \in \mathbb{R}$ satisfies $|\mu| \le \epsilon$ with a sufficiently small $\epsilon > 0$. □

Let $\mathcal{A} \in RT_{N,I}$ be symmetric with an even N, and $\|\mathbf{x}\|_N^N = x_1^N + x_2^N + \cdots + x_I^N$ with all nonzero vectors $\mathbf{x} \in \mathbb{R}^I$, we define

$$f_N(\mathbf{x}) = \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_N^N}.$$

Theorem 3.2.3 *Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric with an even N. Let $(\lambda_*; \mathbf{x}_*)$ be an eigenpair of $\mathcal{A}$.*

(a) *If $(\lambda_*; \mathbf{x}_*)$ is an* H*-eigenpair, then for a small perturbation $\delta\mathbf{x}$, we have*

$$f_N(\mathbf{x}_* + \delta\mathbf{x}) = f_N(\mathbf{x}_*) + O(\|\delta\mathbf{x}\|_2^2) = \lambda_* + O(\|\delta\mathbf{x}\|_2^2).$$

(b) *If λ_* is an algebraically simple eigenvalue and $\mathcal{B} \in RT_{N,I}$ is symmetric, then there exists an algebraically simple eigenvalue $\tilde{\lambda}_*$ of $\mathcal{A} + \mu\mathcal{B}$ such that $\tilde{\lambda}_* = \lambda_* + O(\mu)$, where real ε satisfies $|\mu| \le \epsilon$ with a sufficiently small $\epsilon > 0$.*

Proof To prove part (a), we consider the first-order Taylor expansion of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}_0$:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \nabla f(\mathbf{x}_0) + O(\|\mathbf{x} - \mathbf{x}_0\|_2^2),$$

where $f : \mathbb{R}^I \to \mathbb{R}$ is continuously differentiable and

$$\nabla f(\mathbf{x}) = (\partial f/\partial x_1, \partial f/\partial x_2, \ldots, \partial f/\partial x_I)^\top.$$

Since $(\lambda_*; \mathbf{x}_*)$ is an H-eigenpair of $\mathcal{A}$, then

$$f_N(\mathbf{x}_*) = \lambda_* = \frac{\mathcal{A}\mathbf{x}_*^N}{\|\mathbf{x}_*\|_N^N}.$$

Hence

$$f_N(\mathbf{x}_* + \delta\mathbf{x}) = f_N(\mathbf{x}_*) + \delta\mathbf{x}^\top \nabla f_N(\mathbf{x}_*) + O(\|\delta\mathbf{x}\|_2^2).$$

Since

$$\nabla f_N(\mathbf{x}_*) = \frac{N}{\|\mathbf{x}_*\|_N^N}\left(\mathcal{A}\mathbf{x}_*^{N-1} - f_N(\mathbf{x}_*)\mathbf{x}_*^{[N-1]}\right) = \mathbf{0}_I,$$

Part (a) has been proved.

Part (b) is similar to part (b) of Theorem 3.2.2, from the properties of the characteristic polynomial of $\mathcal{A}$ and Lemma 3.2.2. □

Note that Theorems 3.2.2 and 3.2.3 are also valid for real mode-symmetric tensors.

3.2.2 Algebraically Simple Z-Eigenvalues

It is observed [21] that the complex E-eigenpairs of the symmetric tensor $\mathcal{A} \in RT_{N,I}$ form an equivalence class under a multiplicative transformation [22]. If $(\lambda; \mathbf{x})$ is an E-eigenpair of $\mathcal{A}$ and $\mathbf{y} = e^{\iota\varphi}\mathbf{x}$ with $\varphi \in (-\pi, \pi]$, then $\mathbf{y}^*\mathbf{y} = \mathbf{x}^*\mathbf{x} = 1$ and

$$\mathcal{A}\mathbf{y}^{N-1} = e^{\iota(N-1)\varphi}\mathcal{A}\mathbf{x}^{N-1} = e^{\iota(N-1)\varphi}\lambda\mathbf{x} = e^{\iota(N-2)\varphi}\lambda\mathbf{y}. \tag{3.2.4}$$

Therefore $(e^{\iota(N-2)\varphi}\lambda; e^{\iota\varphi}\mathbf{x})$ is also an E-eigenpair of $\mathcal{A}$ for any $\varphi \in \mathbb{R}$. We can choose $\varphi_* \in (-\pi, \pi]$ such that $e^{\iota(N-2)\varphi_*}\lambda \in \mathbb{R}$. We next consider the perturbation bounds of a Z-eigenvalue of any symmetric or mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$.

Theorem 3.2.4 *Let $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ be mode-symmetric, $\varepsilon \in \mathbb{R}$ and $(\lambda; \mathbf{x})$ be an algebraically simple Z-eigenpair of $\mathcal{A}$. Then there exist $\varepsilon_0 > 0$ and an analytic function $\lambda(\varepsilon)$ with $|\varepsilon| \le \varepsilon_0$ such that*

$$\lambda(0) = \lambda, \quad \lambda'(0) = \left.\frac{d\lambda}{d\varepsilon}\right|_{\varepsilon=0} = \mathcal{B}\mathbf{x}^N,$$

where $\mathbf{x}^\top\mathbf{x} = 1$. Therefore, $\lambda(\varepsilon)$ is an algebraically simple Z-eigenvalue of $\mathcal{A} + \varepsilon\mathcal{B}$ over $|\varepsilon| \le \varepsilon_0$, and

$$\lambda(\varepsilon) = \lambda + \varepsilon\mathcal{B}\mathbf{x}^N + O(\varepsilon^2).$$

Proof When $\mathcal{A}$ is symmetric, we know that $\mathcal{A}$ always has Z-eigenvalues [8, Theorem 5]. This also holds when $\mathcal{A}$ is mode-symmetric (see Theorem 3.2.1).

According to [8], the E-characteristic polynomial of $\mathcal{A} + \varepsilon\mathcal{B}$ is

$$\varphi_\varepsilon(z) = \det(z\mathcal{E} - \mathcal{A} - \varepsilon\mathcal{B}).$$

It is obvious that $\varphi_\varepsilon(z)$ is an analytic function with respect to ε and z. Define $\mathfrak{D}_r := \{z \in \mathbb{C} : |z - \lambda| \le r\}$. Let r be sufficiently small such that $Z(\mathcal{A}) \cap \mathfrak{D}_r = \{\lambda\}$. Denote the boundary of $\mathfrak{D}_r$ as $\partial\mathfrak{D}_r$. Then

$$\min_{z\in\partial\mathfrak{D}_r} |\varphi_0(z)| = \gamma > 0.$$

Since $\varphi_\varepsilon(z)$ is a continuous function of ε, there exists $\varepsilon_0 > 0$, such that for all ε with $|\varepsilon| \le \varepsilon_0$, $\varphi_\varepsilon(z)$ has only one zero point in $\mathfrak{D}_r$ and

$$\min_{z\in\partial\mathfrak{D}_r,\ |\varepsilon|\le\varepsilon_0} |\varphi_\varepsilon(z)| > 0.$$

It follows from the Residue theorem [23] that the zero point $\lambda(\varepsilon)$ of $\varphi_\varepsilon(z)$ in $\mathfrak{D}_r$ can be represented as $\lambda(\varepsilon) = \frac{1}{2\pi}\oint_{\partial\mathfrak{D}_r} \frac{z\varphi_\varepsilon'(z)}{\varphi_\varepsilon(z)}dz$, where $\varphi_\varepsilon'(z) = d\varphi_\varepsilon(z)/dz$.

Both $\frac{z\varphi_\varepsilon'(z)}{\varphi_\varepsilon(z)}$ and $\frac{d}{dz}\left(\frac{z\varphi_\varepsilon'(z)}{\varphi_\varepsilon(z)}\right)$ are continuous on $\partial\mathfrak{D}_r$, by the differential and integral order exchange theorem, $\lambda(\varepsilon)$ is an analytic function, if $|\varepsilon| \le \varepsilon_0$. Hence

$$\lambda(\varepsilon) = \lambda(0) + \lambda'(0)\varepsilon + O(\varepsilon^2), \quad \lambda(0) = \lambda, \ |\varepsilon| \le \varepsilon_0.$$

For an algebraically simple Z-eigenvalue λ of $\mathcal{A}$, there exist $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^I$ such that

$$\mathcal{A}\mathbf{x}_i^{N-1} = \lambda\mathbf{x}_i, \quad \mathbf{x}_i^\top\mathbf{x}_i = 1, \quad i = 1, 2.$$

If $\mathbf{x}_1 = c\mathbf{x}_2$, then c satisfies that $c^2 = 1$ and $c^{N-2} = 1$. In this case, we can see that $\mathbf{x}_1$ and $\mathbf{x}_2$ are the same vector. Hence we define

$$\delta = \min\{\|\mathbf{y} - \mathbf{x}\|_2 : \ \mathbf{y} \text{ and } \mathbf{x} \text{ are the different eigenvectors associated with } \lambda\}.$$

Then over $\{\mathbf{z} \in \mathbb{C}^I : \|\mathbf{z} - \mathbf{x}\|_2 < \delta\}$, there exists a unique eigenvector $\mathbf{x}$ of $\mathcal{A}$ associated with λ. (For an algebraically simple Z-eigenvalue λ, if its geometric multiplicity is also 1, then set $\delta \le \varepsilon_0$.)

For an algebraically simple Z-eigenvalue $\lambda(\varepsilon)$ of $\mathcal{A}+\varepsilon\mathcal{B}$, since $\lambda(\varepsilon)$ is an analytic function with $|\varepsilon| \le \varepsilon_0$, there exists $\tilde{\delta} = \min\{\delta, \varepsilon_0\}$ such that $\|\mathbf{x}(\varepsilon) - \mathbf{x}\|_2 < \tilde{\delta}$ and $\mathbf{x}(\varepsilon)$ is the unique eigenvector of $\mathcal{A}+\varepsilon\mathcal{B}$ corresponding to $\lambda(\varepsilon)$. According to results on algebraic functions [19], we derive that $\mathbf{x}(\varepsilon)$ is an analytic function, where $|\varepsilon| \le \tilde{\delta}$ and $\mathbf{x}(0) = \mathbf{x}$.

As $(\mathcal{A}+\varepsilon\mathcal{B})\mathbf{x}(\varepsilon)^{N-1} = \lambda(\varepsilon)\mathbf{x}(\varepsilon)$, by differentiation with respect to ε, and setting $\varepsilon = 0$, we have

$$\mathcal{A}\tilde{\mathbf{x}} + \mathcal{B}\mathbf{x}(0)^{N-1} = \lambda'(0)\mathbf{x}(0) + \lambda(0)\mathbf{x}'(0),$$

where

$$\mathcal{A}\tilde{\mathbf{x}} = \mathcal{A}(\times_2\mathbf{x}'(0)^\top \times_3 \mathbf{x}(0)^\top \cdots \times_N \mathbf{x}(0)^\top + \cdots + \times_2\mathbf{x}(0)^\top \cdots \times_{N-1}\mathbf{x}(0)^\top \times_N \mathbf{x}'(0)^\top).$$

Then

$$\lambda'(0) = \mathcal{B}\mathbf{x}(0)^N = \mathcal{B}\mathbf{x}^N,$$

and the theorem is proved. □

Gohberg and Koltracht [24] explore condition numbers of maps in finite-dimensional spaces $F : \mathbb{R}^P \to \mathbb{R}^Q$. The condition number of F at a point $\mathbf{a} \in D_F$[1] characterizes the instantaneous rate of change in $F(\mathbf{a})$ with respect to $\mathbf{a}$.

Another perturbation result of a nonzero algebraically simple Z-eigenvalue λ of the mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$ is considered, whose associated eigenvector $\mathbf{x} \in \mathbb{R}^I$ satisfies $\|\mathbf{x}\|_2 = 1$. It is well known that the map $F_{\mathcal{E}} : \varepsilon \to \lambda(\mathcal{A} + \varepsilon\mathcal{B})$ is analytic in a neighborhood of 0 [13]. Therefore, the map $F : \mathcal{A} \to \lambda(\mathcal{A})$ has continuous partial derivatives with respect to each entry of $\mathcal{A}$ and

$$\frac{\partial F}{\partial_{i_1 i_2 \ldots i_N}}(\mathcal{A}) := \lim_{t \to 0} \frac{F(\mathcal{A} + t\mathcal{S}) - F(\mathcal{A})}{t} = x_{i_1} x_{i_2} \ldots x_{i_N},$$

where $\mathcal{S} \in RT_{N,I}$ is the zero tensor except for $s_{i_1 i_2 \ldots i_N} = 1$ with all i_n and n.

We illustrate the meaning of $\mathcal{S}$ via a special case of $N = 4$. Without loss of generality, $(i_1, i_2, i_3, i_4) = (1, 2, 3, 4)$, then $\mathcal{S}$ can be written as $\mathcal{J} = \mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3 \otimes \mathbf{e}_4$ where $\mathbf{e}_i$ is the ith column of $\mathbf{I}_I \in \mathbb{R}^{I \times I}$ for $i = 1, 2, 3, 4$. Then, all entries of $\mathcal{S}$ are given in the following:

$$s_{j_1 j_2 j_3 j_4} = \mathbf{e}_1(j_1)\mathbf{e}_2(j_2)\mathbf{e}_3(j_3)\mathbf{e}_4(j_4),$$

where $j_n = 1, 2, \ldots, I$ and $\mathbf{e}_n(j_n)$ is the j_nth element of $\mathbf{e}_n$ with $n = 1, 2, 3, 4$. In general, $\mathcal{S} = \mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \cdots \otimes \mathbf{e}_{i_N}$ where $\mathbf{e}_{i_n}$ is the i_nth column of $\mathbf{I}_I \in \mathbb{R}^{I \times I}$ for all n.

Hence, as a map from $\mathbb{R}^{I_N} \to \mathbb{R}$, F is differentiable at $\mathcal{A}$ and

$$F'(\mathcal{A}) = \left[\frac{\partial F}{\partial_{11\ldots1}}, \ldots, \frac{\partial F}{\partial a_{11\ldots1I}}, \frac{\partial F}{\partial a_{21\ldots1}}, \ldots, \frac{\partial F}{\partial a_{II\ldots I}} \right].$$

According to a formula by Gohberg and Koltracht [24], for relatively small componentwise perturbations in $\mathcal{A}$, i.e., $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon |a_{i_1 i_2 \ldots i_N}|$, for all i_n and n, where $\mathcal{E} \in RT_{N,I}$ is mode-symmetric and $\varepsilon > 0$ is arbitrarily small, the sensitivity of $F(\mathcal{A})$ is characterized by the componentwise condition number of F at $\mathcal{A}$:

$$c(F, \mathcal{A}) = \frac{\|F'(\mathcal{A})\mathbf{D}_{\mathcal{A}}\|_\infty}{\|F(\mathcal{A})\|_\infty},$$

with $\mathbf{D}_{\mathcal{A}} = \operatorname{diag}(a_{11\ldots1}, \ldots, a_{11\ldots1I}, a_{21\ldots1}, \ldots, a_{II\ldots I})$.

For $\lambda \neq 0$, it indicates that

$$\begin{aligned} c(F, \mathcal{A}) = \|(&a_{11\ldots1} x_1 x_1 \ldots x_1, \ldots, a_{11\ldots1I} x_1 \ldots x_1 x_I, \\ &a_{21\ldots1} x_2 x_1 \ldots x_1, \ldots, a_{II\ldots I} x_I x_I \ldots x_n)\|_\infty / |\lambda|, \end{aligned}$$

[1] D_F is denoted by the domain of the map $F : \mathbb{R}^P \to \mathbb{R}^Q$.

where the infinity norm of the map $F'(\mathcal{A})\mathbf{D}_{\mathcal{A}}$ is, in fact, the 1-norm of the row vector that represents it. Thus

$$c(F,\mathcal{A}) = \sum_{i_1 i_2 \ldots i_N = 1}^{I} |a_{i_1 i_2 \ldots i_N} x_{i_1} x_{i_2} \ldots x_{i_N}|/|\lambda| = |\mathcal{A}||\mathbf{x}|^N/|\lambda|,$$

where

$$|\mathcal{A}||\mathbf{x}|^N = \sum_{i_1, i_2, \ldots, i_N = 1}^{I} |a_{i_1 i_2 \ldots i_N}||x_{i_1}||x_{i_2}| \ldots |x_{i_n}|.$$

Theorem 3.2.5 *Let $\mathcal{A} \in RT_{N,I}$ be mode-symmetric. For an arbitrary small $\varepsilon > 0$, suppose that the mode-symmetric tensor $\mathcal{E} \in RT_{N,I}$ satisfies $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon |a_{i_1 i_2 \ldots i_N}|$ for all i_n and n. Then for an algebraically simple Z-eigenvalue $\lambda \neq 0$ of $\mathcal{A}$, there exists a Z-eigenvalue $\hat{\lambda}$ of $\mathcal{A} + \mathcal{E}$ such that*

$$\frac{|\hat{\lambda} - \lambda|}{|\lambda|} \leq c(F,\mathcal{A})\varepsilon + O(\varepsilon^2).$$

If $(\lambda; \mathbf{z})$ is a Z-eigenpair of the mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$, then $(\lambda; -\mathbf{z})$ is also a Z-eigenpair of $\mathcal{A}$ with an even N. Hence, if ϱ is the Z-spectral radius of $\mathcal{A}$, then the associated Z-eigenvalue is ϱ or $-\varrho$. In particular, if $\mathcal{A}$ is nonnegative and irreducible, according to Theorem 3.2.5, we can derive the perturbation bound of the Z-spectral radius of $\mathcal{A}$.

Corollary 3.2.1 *Suppose that $\mathcal{A}, \mathcal{E} \in RT_{N,I}$ are mode-symmetric with an even N such that $\mathcal{A}$ is nonnegative and irreducible. For an arbitrary small $\varepsilon > 0$, suppose that the mode-symmetric tensor $\mathcal{E} \in RT_{N,I}$ satisfies $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon a_{i_1 i_2 \ldots i_N}$ for all i_n and n. Let ϱ and ϱ_ε denote, respectively, the Z-spectral radii of $\mathcal{A}$ and $\mathcal{A} + \mathcal{E}$. If $\mathcal{A}$ is nonnegative and irreducible, then*

$$\frac{|\varrho_\varepsilon - \varrho|}{\varrho} \leq \varepsilon.$$

Assume that $(\lambda; \mathbf{x})$ is an algebraically simple E-eigenpair of the mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$. If there exists $\varphi_* \in (-\pi, \pi]$ such that $e^{\iota(N-1)\varphi_*}\lambda \in \mathbb{R}$, then according to Theorem 3.2.4, we can derive the perturbation bounds of λ. Furthermore, Theorems 3.2.4 and 3.2.5, and Corollary 3.2.1 hold for symmetric tensors in $RT_{N,I}$.

3.2.3 Algebraically Simple Eigenvalues

Let $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ be mode-symmetric. For any sufficiently small $\varepsilon > 0$, we provide the explicit expression of an algebraically simple eigenvalue $\lambda(\varepsilon) \in \mathbb{C}$ of $\mathcal{A} + \varepsilon\mathcal{B}$, as a generalization of the classical results in [1, Theorem 2.3].

Theorem 3.2.6 *Suppose that $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ are mode-symmetric with an even N. Let $(\lambda; \mathbf{x})$ be an algebraically simple eigenpair of $\mathcal{A}$ with $\|\mathbf{x}\|_N^N = x_1^N + \cdots + x_N^N \neq 0$, there exist $\varepsilon_0 > 0$ and an analytic function $\lambda(\varepsilon)$ with $|\varepsilon| \leq \varepsilon_0$ such that*

$$\lambda(0) = \lambda, \quad \lambda'(0) = \left.\frac{d\lambda}{d\varepsilon}\right|_{\varepsilon=0} = \frac{\mathcal{B}\mathbf{x}^N}{\|\mathbf{x}\|_N^N}.$$

Therefore, $\lambda(\varepsilon)$ is an algebraically simple eigenvalue of $\mathcal{A} + \varepsilon\mathcal{B}$ over $|\varepsilon| \leq \varepsilon_0$, and

$$\lambda(\varepsilon) = \lambda + \varepsilon\frac{\mathcal{B}\mathbf{x}^N}{\|\mathbf{x}\|_N^N} + O(\varepsilon^2).$$

In particular, if $(\lambda; \mathbf{x})$ is an algebraically simple H*-eigenpair of $\mathcal{A}$ such that $\|\mathbf{x}\|_N^N = 1$, then there exist $\varepsilon_0 > 0$ and an analytic function $\lambda(\varepsilon)$ with $|\varepsilon| \leq \varepsilon_0$ such that*

$$\lambda(0) = \lambda, \quad \lambda'(0) = \left.\frac{d\lambda}{d\varepsilon}\right|_{\varepsilon=0} = \mathcal{B}\mathbf{x}^N.$$

Then $\lambda(\varepsilon)$ is an algebraically simple H*-eigenvalue of $\mathcal{A} + \varepsilon\mathcal{B}$ over $|\varepsilon| \leq \varepsilon_0$, and*

$$\lambda(\varepsilon) = \lambda + \varepsilon\mathcal{B}\mathbf{x}^N + O(\varepsilon^2). \tag{3.2.5}$$

The proof of the above theorem is similar to that of Theorem 3.2.4 and is omitted.

Remark 3.2.1 For a given $\varepsilon > 0$ if $\mathcal{A}$ and $\mathcal{B}$ in Theorem 3.2.6 are irreducible and symmetric nonnegative tensors, then formula (3.2.5) reduces to the result by Li et al. [25, Theorem 5.2].

Suppose that $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ are mode-symmetric. We now consider the first relative perturbation bound of the nonzero algebraically simple H-eigenvalues of $\mathcal{A}$ for componentwise perturbations. For a sufficiently small $\varepsilon \in \mathbb{R}$, it is well known that the map $F_{\mathcal{E}} : \varepsilon \to \lambda(\mathcal{A}+\varepsilon\mathcal{B})$ is analytic in a neighborhood of 0 [13]. Therefore, the map $F : \mathcal{A} \to \lambda(\mathcal{A})$ has continuous partial derivatives with respect to each entry of $\mathcal{A}$ and

$$\frac{\partial F}{\partial_{i_1 i_2 \ldots i_N}}(\mathcal{A}) := \lim_{t\to 0} \frac{F(\mathcal{A} + t\mathcal{S}) - F(\mathcal{A})}{t} = x_{i_1} x_{i_2} \ldots x_{i_N},$$

where $\mathcal{S} = \mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \cdots \otimes \mathbf{e}_{i_N}$ where $\mathbf{e}_{i_n}$ is the i_nth column of $\mathbf{I}_I$ for all i_n and n. Thus, as a map from $\mathbb{R}^{I^N} \to \mathbb{R}$, F is differentiable at $\mathcal{A}$ and

$$F'(\mathcal{A}) = \left[\frac{\partial F}{\partial a_{11\ldots1}}, \ldots, \frac{\partial F}{\partial a_{11\ldots1I}}, \frac{\partial F}{\partial a_{21\ldots1}}, \ldots, \frac{\partial F}{\partial a_{II\ldots I}} \right].$$

Based on a formula by Gohberg and Koltracht [24], for the *componentwise perturbations* in $\mathcal{A}$, i.e., $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon |a_{i_1 i_2 \ldots i_N}|$ for all i_n and n, where $\mathcal{E} \in RT_{N,I}$ is mode-symmetric and $\varepsilon > 0$ is arbitrarily small, the sensitivity of $F(\mathcal{A})$ is characterized by the componentwise condition number of F at $\mathcal{A}$:

$$c(F, \mathcal{A}) = \frac{\|F'(\mathcal{A})\mathbf{D}_{\mathcal{A}}\|_\infty}{\|F(\mathcal{A})\|_\infty},$$

with $\mathbf{D}_{\mathcal{A}} = \operatorname{diag}(a_{11\ldots1}, \ldots, a_{11\ldots1I}, a_{21\ldots1}, \ldots, a_{II\ldots I})$.

For $\lambda \neq 0$, it indicates that

$$\begin{aligned} c(F, \mathcal{A}) = \|&(a_{11\ldots1}x_1x_1\ldots x_1, \ldots, a_{11\ldots1I}x_1\ldots x_1x_I, \\ &a_{21\ldots1}x_2x_1\ldots x_1, \ldots, a_{II\ldots I}x_Ix_I\ldots x_n)\|_\infty/|\lambda|, \end{aligned}$$

where the infinity norm of the map $F'(\mathcal{A})\mathbf{D}_{\mathcal{A}}$ is the 1-norm of the row vector that represents it. Thus

$$c(F, \mathcal{A}) = \frac{\sum_{i_1 i_2 \ldots i_N = 1}^{I} |a_{i_1 i_2 \ldots i_N} x_{i_1} x_{i_2} \ldots x_{i_N}|}{|\lambda| \|\mathbf{x}\|_N^N} = \frac{|\mathcal{A}||\mathbf{x}|^N}{|\lambda| \|\mathbf{x}\|_N^N}.$$

Theorem 3.2.7 *Let $\mathcal{A}, \mathcal{E} \in RT_{N,I}$ be mode-symmetric with an even N. If for a sufficiently small $\varepsilon > 0$, the entries of $\mathcal{E}$ satisfy $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon |a_{i_1 i_2 \ldots i_N}|$ for all i_n and n, then for an algebraically simple* H*-eigenvalue $\lambda \neq 0$ of $\mathcal{A}$, there exists an eigenvalue $\hat{\lambda}$ of $\mathcal{A} + \mathcal{E}$ such that*

$$\frac{|\hat{\lambda} - \lambda|}{|\lambda|} \leq c(F, \mathcal{A})\varepsilon + O(\varepsilon^2).$$

If a mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$ is irreducible and nonnegative, then for any $0 < \varepsilon < 1$, we can derive the perturbation bound of the spectral radius of $\mathcal{A}$ from Theorem 3.2.7.

Corollary 3.2.2 *Suppose that $\mathcal{A}, \mathcal{E} \in RT_{N,I}$ are mode-symmetric with an even N such that $\mathcal{A}$ is nonnegative and irreducible. For a sufficiently small $\varepsilon > 0$, the entries of $\mathcal{E}$ satisfy $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon a_{i_1 i_2 \ldots i_N}$ for all i_n and n. Let ρ and ρ_ε be the*

spectral radius of $\mathcal{A}$ and $\mathcal{A}+\mathcal{E}$, respectively. Then

$$\frac{|\rho_\varepsilon - \rho|}{\rho} \leq \varepsilon.$$

Proof Since $|e_{i_1,\ldots,i_N}| \leq \varepsilon a_{i_1,\ldots,i_N}$, for all i_n and n, we can write

$$0 \leq \mathcal{A} - \varepsilon\mathcal{A} \leq \mathcal{A} + \mathcal{E} \leq \mathcal{A} + \varepsilon\mathcal{A}.$$

Since the spectral radius $\rho(\cdot)$ of $\mathcal{A}$ is monotone [26], it follows that $\rho(\mathcal{A} - \varepsilon\mathcal{A}) \leq \rho(\mathcal{A} + \mathcal{E}) \leq \rho(\mathcal{A} + \varepsilon\mathcal{A})$. As $\rho(\mathcal{A} \pm \varepsilon\mathcal{A}) = (1 \pm \varepsilon)\rho(\mathcal{A})$, we obtain

$$(1 - \varepsilon)\rho \leq \rho_\varepsilon \leq (1 + \varepsilon)\rho.$$

As $\rho > 0$, the last inequality is equivalent to the result. □

Remark 3.2.2 If $\mathcal{A}$ is an irreducible nonnegative tensor, then $c(F, \mathcal{A}) = 1$ for the spectral radius. The perturbation bound for the Perron root of nonnegative matrices has been discussed by Elsner et al. [27].

3.2.4 Singular Values

For a given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, Ragnarsson et al. [16] explore the singular tuples of $\mathcal{A}$ through its symmetric embedding $\mathbf{sym}(\mathcal{A})$. Chen et al. [17] further develop the connection between the singular values $\mathcal{A}$ and the Z-eigenvalues of $\mathbf{sym}(\mathcal{A})$. We now consider the connection between the singular values of $\mathcal{A}$ and the Z-eigenvalues of its mode-symmetric embedding $\mathbf{msym}(\mathcal{A})$.

For the case of $N = 3$, if $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ is a singular tuple of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, then

$$\mathcal{A} \times_2 \mathbf{x}_2^\top \times_3 \mathbf{x}_3^\top = \sigma\mathbf{x}_1, \quad \mathcal{A} \times_1 \mathbf{x}_1^\top \times_3 \mathbf{x}_3^\top = \sigma\mathbf{x}_2, \quad \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top = \sigma\mathbf{x}_3. \tag{3.2.6}$$

We have the elementwise,

$$\sum_{j,k=1}^{I} a_{ijk}x_{2,j}x_{3,k} = \sigma x_{1,i}, \quad \sum_{i,k=1}^{I} a_{ijk}x_{1,i}x_{3,k} = \sigma x_{2,j}, \quad \sum_{i,j=1}^{I} a_{ijk}x_{1,i}x_{2,j} = \sigma x_{3,k}. \tag{3.2.7}$$

In (3.2.7), if we change the summation order, then we have

$$\sum_{j,k=1}^{I} a_{ijk}x_{2,j}x_{3,k} = \sigma x_{1,i}, \quad \sum_{k,i=1}^{I} a_{jki}x_{3,k}x_{1,i} = \sigma x_{2,j}, \quad \sum_{i,j=1}^{I} a_{kij}x_{1,i}x_{2,j} = \sigma x_{3,k}.$$

Let $\mathbf{x} = \frac{1}{\sqrt{3}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, then $\|\mathbf{x}\|_2 = 1$. Then (3.2.6) can be transformed into

$$\begin{cases} \frac{1}{\sqrt{3}}\mathbf{msym}(\mathcal{A}) \times_2 \mathbf{x}^\top \times_3 \mathbf{x}^\top = \sigma\mathbf{x}, \\ \frac{1}{\sqrt{3}}\mathbf{msym}(\mathcal{A}) \times_1 \mathbf{x}^\top \times_3 \mathbf{x}^\top = \sigma\mathbf{x}, \\ \frac{1}{\sqrt{3}}\mathbf{msym}(\mathcal{A}) \times_1 \mathbf{x}^\top \times_2 \mathbf{x}^\top = \sigma\mathbf{x}. \end{cases}$$

Furthermore, let $\tilde{\mathcal{A}} = \frac{1}{\sqrt{3}}\mathbf{msym}(\mathcal{A})$, then the above set of formulae is equivalent to

$$\tilde{\mathcal{A}}\mathbf{x}^2 = \sigma\mathbf{x}, \quad \|\mathbf{x}\|_2 = 1.$$

Generally, suppose that $(\sigma; \mathbf{x}_1, \ldots, \mathbf{x}_N)$ is a singular tuple of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, with $\|\mathbf{x}_n\|_2 = 1$ for all n. Let $\mathbf{x} = \frac{1}{\sqrt{N}}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, then we have $\|\mathbf{x}\|_2 = 1$ and

$$\begin{cases} \left(\frac{1}{\sqrt{N}}\right)^{N-2} \mathbf{msym}(\mathcal{A}) \times_2 \mathbf{x}^\top \times_3 \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top = \sigma\mathbf{x}, \\ \left(\frac{1}{\sqrt{N}}\right)^{N-2} \mathbf{msym}(\mathcal{A}) \times_1 \mathbf{x}^\top \times_3 \mathbf{x}^\top \cdots \times_N \mathbf{x}^\top = \sigma\mathbf{x}, \\ \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ \left(\frac{1}{\sqrt{N}}\right)^{N-2} \mathbf{msym}(\mathcal{A}) \times_1 \mathbf{x}^\top \times_2 \mathbf{x}^\top \cdots \times_{N-1} \mathbf{x}^\top = \sigma\mathbf{x}. \end{cases}$$

Furthermore, let $\tilde{\mathcal{A}} = \left(\frac{1}{\sqrt{N}}\right)^{N-2} \mathbf{msym}(\mathcal{A})$, then $(\sigma; \mathbf{x})$ is a Z-eigenpair of $\tilde{\mathcal{A}}$, that is, $(\sigma; \mathbf{x})$ is the solution of the nonlinear equations

$$\tilde{\mathcal{A}}\mathbf{x}^{N-1} = \sigma\mathbf{x}, \quad \|\mathbf{x}\|_2 = 1.$$

Both $\mathbf{msym}(\mathcal{A})$ and $\tilde{\mathcal{A}}$ are mode-symmetric. The value σ is called an algebraically simple singular value, if σ is an algebraically simple Z-eigenvalue of $\tilde{\mathcal{A}}$.

According to the mode-symmetric embedding of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and the perturbation bounds of an algebraically simple Z-eigenvalue of $\mathbf{msym}(\mathcal{A})$, it is trivial to derive the following results for the perturbation of an algebraically simple singular value.

Theorem 3.2.8 *Let $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ be a singular tuple of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ such that σ is the algebraically simple singular value and $\mathbf{x}_n \in \mathbb{R}^{I_n}$ satisfy $\|\mathbf{x}_n\|_2 = 1$ with all n. Then there exist $\varepsilon_0 > 0$ and an analytic function $\sigma(\varepsilon)$ with $|\varepsilon| \leq \varepsilon_0$ such that*

$$\sigma(0) = \lambda, \quad \sigma'(0) = \left.\frac{d\sigma}{d\varepsilon}\right|_{\varepsilon=0} = \mathcal{B} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top,$$

with $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. *Thus* $\sigma(\varepsilon)$ *is an algebraically simple singular value of* $\mathcal{A} + \varepsilon\mathcal{B}$ *over* $|\varepsilon| \leq \varepsilon_0$, *and*

$$\sigma(\varepsilon) = \sigma + \varepsilon\mathcal{B} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top + O(\varepsilon^2).$$

Theorem 3.2.9 *Let* $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ *be a singular tuple of* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *such that* $\sigma \neq 0$ *is the algebraically simple singular value and* $\mathbf{x}_n \in \mathbb{R}^{I_n}$ *satisfy* $\|\mathbf{x}_n\|_2 = 1$ *with all* n. *If for* $0 < \varepsilon < 1$, *the entries of* $\mathcal{E} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *satisfy* $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon |a_{i_1 i_2 \ldots i_N}|$ *with all* i_n *and* n, *then there exists a singular value* $\hat{\sigma}$ *of* $\mathcal{A} + \mathcal{E}$ *such that*

$$\frac{|\hat{\sigma} - \sigma|}{|\sigma|} \leq c(F, \mathcal{A})\varepsilon + O(\varepsilon^2),$$

where $c(F, \mathcal{A}) = \frac{1}{|\sigma|}|\mathcal{A}| \times_1 |\mathbf{x}_1|^\top \times_2 |\mathbf{x}_2|^\top \cdots \times_N |\mathbf{x}_N|^\top$.

Corollary 3.2.3 *Let* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *be nonnegative and irreducible. Suppose that for* $0 < \varepsilon < 1$, *the entries of* $\mathcal{E} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *satisfy* $|e_{i_1 i_2 \ldots i_N}| \leq \varepsilon |a_{i_1 i_2 \ldots i_N}|$ *with all* i_n *and* n. *Let* $\bar{\sigma}$ *and* $\bar{\sigma}_\varepsilon$ *denote, respectively, the modulus largest singular values of* $\mathcal{A}$ *and* $\mathcal{A} + \mathcal{E}$. *Then*

$$\frac{|\bar{\sigma}_\varepsilon - \bar{\sigma}|}{\bar{\sigma}} \leq \varepsilon.$$

Suppose that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\mathbf{B}_n \in \mathbb{R}^{I_n \times I_n}$ is symmetric positive definite with all n. Let $(\sigma; \mathbf{x}_1, \ldots, \mathbf{x}_N)$ be a restricted singular tuple of $\mathcal{A}$. If $\mathbf{B}_n = \mathbf{G}_n \mathbf{G}_n^\top$ where $\mathbf{G}_n \in \mathbb{R}^{I_n \times I_n}$ is a nonsingular upper triangle matrix, then

$$F(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)_{-n} = \sigma \mathbf{y}_n$$

where

$$F(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)_{-n} = \mathcal{B} \times_1 \mathbf{y}_1^\top \cdots \times_{n-1} \mathbf{y}_{n-1}^\top \times_{n+1} \mathbf{y}_{n+1}^\top \cdots \times_N \mathbf{y}_N^\top$$

with $\mathcal{B} = \mathcal{A} \times_1 \mathbf{G}_1^{-1} \times_2 \mathbf{G}_2^{-1} \cdots \times_N \mathbf{G}_N^{-1}$ and $\|\mathbf{y}_n\|_2 = 1$. We see that $(\sigma; \mathbf{y}_1, \ldots, \mathbf{y}_N)$ is a singular tuple of $\mathcal{B}$.

We can also consider the perturbation results of algebraically simple restricted singular value of $\mathcal{A}$ with respect to $\mathbf{B}_n$ according to Theorems 3.2.8 and 3.2.9, and Corollary 3.2.3.

3.3 Perturbation for the Tensor Polynomial Eigenvalue Problem

Let $\mathcal{A}_l \in RT_{N,I}$ be mode-symmetric with all l. Suppose that $\mathcal{P}$ is a tensor polynomial defined in (3.1.3), we define

$$\mathcal{G}(z) = \mathcal{A}_1 + 2z\mathcal{A}_2 + \cdots + Lz^{L-1}\mathcal{A}_L.$$

Suppose that the perturbation tensor $\Delta\mathcal{A}_l \in RT_{N,I}$ are mode-symmetric with all l. For any $z \in \mathbb{C}$, denote

$$\Delta\mathcal{P}(z) = \Delta\mathcal{A}_0 + z\Delta\mathcal{A}_1 + \cdots + z^L\Delta\mathcal{A}_L.$$

We study the first-order perturbation bounds of the algebraically simple eigenvalues of $\mathcal{P}$ for normwise and componentwise perturbations. We consider two special cases of the tensor polynomial eigenvalue problem determined by $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$: the tensor generalized and quadratic eigenvalue problems, respectively for $L = 1, 2$. Then we investigate the general case with $L > 2$. We assume that $\mathcal{A}_L$ is nonsingular, which implies that $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_L\}$ is regular.

3.3.1 Tensor Generalized Eigenvalue Problem

Suppose that $L = 1$. For any $z \in \mathbb{C}$, $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1$. The tensor polynomial eigenvalue problem determined by $\{\mathcal{A}_0, \mathcal{A}_1\}$ reduces to the tensor generalized eigenvalue problem determined by $\{\mathcal{A}_0, \mathcal{A}_1\}$. Denote

$$\mathcal{A} = \mathcal{A}_0, \quad \mathcal{B} = -\mathcal{A}_1.$$

Since $\mathcal{A}_1$ is nonsingular and mode-symmetric, so is $\mathcal{B}$. Hence the generalized spectrum $\Lambda(\mathcal{A}, \mathcal{B})$ of the tensor generalized eigenvalue problem determined by $\{\mathcal{A}, \mathcal{B}\}$ is a finite subset of $\mathbb{C}$ and the number of the generalized eigenvalues is $I(N-1)^{I-1}$, counting multiplicity. If the algebraic multiplicity of $\lambda \in \Lambda(\mathcal{A}, \mathcal{B})$ is one, we have the following theorem.

Theorem 3.3.1 *Assume that $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ are mode-symmetric and $\mathcal{B}$ is nonsingular. If $\lambda \in \mathbb{C}$ is an algebraically simple generalized eigenvalue of $\{\mathcal{A}, \mathcal{B}\}$, then there exists an algebraically simple generalized eigenvalue $\tilde{\lambda} \in \mathbb{C}$ of $\{\mathcal{A} + \epsilon\mathcal{E}, \mathcal{B} + \epsilon\mathcal{F}\}$ such that*

$$\tilde{\lambda} = \lambda + O(\epsilon),$$

where $|\epsilon| \leq \epsilon_0$ for any sufficiently small $\epsilon_0 > 0$ and $\mathcal{E}, \mathcal{F} \in RT_{N,I}$ are mode-symmetric.

Proof From [13], we have $\Lambda(\mathcal{A}, \mathcal{B}) = \{\lambda \in \mathbb{C} : \det(\mathcal{A} - \lambda\mathcal{B}) = 0\}$, where $\det(\mathcal{A} - \lambda\mathcal{B})$ is a $I(N-1)^{I-1}$th polynomial with respect to λ and its leading coefficient is $\det(\mathcal{B}) \neq 0$. There exists $\hat{\epsilon} > 0$ such that $\det(\mathcal{B} + \hat{\epsilon}\mathcal{B}_1) \neq 0$. Hence, when $|\epsilon| \leq \hat{\epsilon}$, the set of all generalized eigenvalues of $\{\mathcal{A} + \epsilon\mathcal{E}, \mathcal{B} + \epsilon\mathcal{F}\}$ satisfies

$$\Lambda_\epsilon(\mathcal{A}, \mathcal{B}) = \{\lambda \in \mathbb{C} : \det((\mathcal{A} - \lambda\mathcal{B}) + \epsilon(\mathcal{E} - \lambda\mathcal{F})) = 0\}.$$

Meanwhile, $\det((\mathcal{A} - \lambda\mathcal{B}) + \epsilon(\mathcal{E} - \lambda\mathcal{F}))$ is also a $I(N-1)^{I-1}$th polynomial with respect to λ and its leading coefficient is $\det(\mathcal{B} + \epsilon\mathcal{F})$. Note that $\det(\mathcal{B} + \epsilon\mathcal{F})$ is a $I(N-1)^{I-1}$th polynomial with respect to ϵ, whose constant term is $\det(\mathcal{B})$. Hence $\det(\mathcal{B} + \epsilon\mathcal{F}) \not\equiv 0$.

According to theorems of algebraic functions [19], if $\lambda \in \mathbb{C}$ is an algebraically simple generalized eigenvalue, then there exist an algebraically simple generalized eigenvalue $\tilde{\lambda} \in \mathbb{C}$ of $\{\mathcal{A} + \epsilon\mathcal{E}, \mathcal{B} + \epsilon\mathcal{F}\}$ and $\epsilon_0 > 0$ such that $\tilde{\lambda} = \lambda + O(\epsilon)$, where $|\epsilon| \leq \min\{\hat{\epsilon}, \epsilon_0\}$. □

A generalization of Theorem 3.3.1 is stated as follows.

Theorem 3.3.2 *Suppose that $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ such that $\mathcal{B}$ is nonsingular. For a given n, if $\lambda \in \mathbb{C}$ is an algebraically simple mode-n generalized eigenvalue of $\{\mathcal{A}, \mathcal{B}\}$, then there exists an algebraically simple mode-n generalized eigenvalue $\tilde{\lambda} \in \mathbb{C}$ of $\{\mathcal{A} + \epsilon\mathcal{E}, \mathcal{B} + \epsilon\mathcal{F}\}$ such that*

$$\tilde{\lambda} = \lambda + O(\epsilon),$$

where $|\epsilon| \leq \epsilon_0$ for a sufficiently small $\epsilon_0 > 0$ and $\mathcal{E}, \mathcal{F} \in RT_{N,I}$.

The choice of $\{\mathcal{E}, \mathcal{F}\}$ in Theorem 3.3.1 is not unique. In practice, $\{\mathcal{E}, \mathcal{F}\} = \{\mathcal{A}, \mathcal{B}\}$ for the relative perturbation case and all entries of $\mathcal{E}$ and $\mathcal{F}$ are unit for the absolute perturbation case. Theorem 3.3.1 only states the first-order perturbation of an algebraically simple generalized eigenvalue, but does not present the coefficient of the first-order perturbation term. For some nonzero algebraically simple generalized eigenvalues, we present an expression of this coefficient in the following theorems.

Theorem 3.3.3 *Suppose that $\alpha, \beta > 0$. Let $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ be mode-symmetric and $\mathcal{B}$ be nonsingular. If $(\lambda; \mathbf{x})$ is an algebraically simple generalized* H*-eigenpair of $\{\mathcal{A}, \mathcal{B}\}$ with $\lambda \neq 0$ and $\mathbf{x} \in \mathbb{R}^I$, then there exists an algebraically simple generalized* H*-eigenvalue $\hat{\lambda} \in \mathbb{R}$ of $\{\mathcal{A} + \Delta\mathcal{A}, \mathcal{B} + \Delta\mathcal{B}\}$ such that*

$$\frac{|\tilde{\lambda} - \lambda|}{|\lambda|} \leq \epsilon \frac{(\alpha + |\lambda|\beta)\|\mathbf{x}\|_2^N}{|\lambda||\mathcal{B}\mathbf{x}^N|} + O(\epsilon^2),$$

where two mode-symmetric tensors $\Delta\mathcal{A}, \Delta\mathcal{B} \in RT_{N,I}$ satisfy $\|\Delta\mathcal{A}\|_F \leq \epsilon\alpha$ and $\|\Delta\mathcal{B}\|_F \leq \epsilon\beta$ with $\epsilon < 1$.

Proof Since $\det(\mathcal{B}) \neq 0$, then $\mathcal{B}\mathbf{y}^N \neq 0$ for all nonzero vectors $\mathbf{y} \in \mathbb{R}^I$. Let $(\lambda; \mathbf{x})$ be a generalized H-eigenvalue of $\{\mathcal{A}, \mathcal{B}\}$ and let $\lambda \neq 0$ be algebraically simple, then a normwise condition number of λ can be defined as follows,

$$\kappa(\lambda) := \limsup_{\epsilon \to 0} \left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} : (\mathcal{A} + \Delta\mathcal{A})(\mathbf{x} + \Delta\mathbf{x})^{N-1} = (\lambda + \Delta\lambda)(\mathcal{B} + \Delta\mathcal{B})(\mathbf{x} + \Delta\mathbf{x})^{N-1}, \right.$$
$$\left. \|\Delta\mathcal{A}\|_F \leq \epsilon\alpha, \ \|\Delta\mathcal{B}\|_F \leq \epsilon\beta \right\}.$$

We prove that the following formula holds:

$$\kappa(\lambda) = \frac{(\alpha + |\lambda|\beta)\|\mathbf{x}\|_2^N}{|\lambda||\mathcal{B}\mathbf{x}^N|}.$$

The given expression is clearly an upper bound for $\kappa(\lambda)$. We now show that the bound is attainable. From the definition of a normwise condition number of λ, we have

$$\Delta\lambda = \frac{\Delta\mathcal{A}\mathbf{x}^N - \lambda\Delta\mathcal{B}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N} + O(\epsilon^2). \tag{3.3.1}$$

Let $\mathcal{G} = \mathbf{x}^{\otimes N}/\|\mathbf{x}\|_2^N$. Then $\|\mathcal{G}\|_F = 1$ and $\mathcal{G}\mathbf{x}^N = \|\mathbf{x}\|_2^N$. Define $\Delta\mathcal{A}_0 = \epsilon\alpha\mathcal{G}$ and $\Delta\mathcal{B}_0 = -\text{sign}(\lambda)\epsilon\beta\mathcal{G}$, where $\text{sign}(\lambda)$ is defined in (2.3.7). Then $\|\Delta\mathcal{A}\|_F \leq \epsilon\alpha$ and $\|\Delta\mathcal{B}\|_F \leq \epsilon\beta$. Hence the modulus of the first-order term of (3.3.1) is $\epsilon\|\mathbf{x}\|_2^N(\alpha + |\lambda|\beta)/|\mathcal{B}\mathbf{x}^N|$; dividing (3.3.1) by $\epsilon|\lambda|$ and taking the limit as $\epsilon \to 0$ gives the desired equality.

From the definition of $\kappa(\lambda)$ we have, for the perturbation system in (3.3.1),

$$\frac{|\Delta\lambda|}{|\lambda|} \leq \kappa(\lambda)\epsilon + O(\epsilon^2).$$

The proof is complete. □

Theorem 3.3.4 *Suppose that $\mathcal{E}, \mathcal{F} \in RT_{N,I}$ are mode-symmetric with positive entries. Let $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ be mode-symmetric and $\mathcal{B}$ be nonsingular. If $(\lambda; \mathbf{x})$ is an algebraically simple generalized* H*-eigenpair of $\{\mathcal{A}, \mathcal{B}\}$ with $\lambda \neq 0$ and $\mathbf{x} \in \mathbb{R}^I$, then there exists an algebraically simple generalized* H*-eigenvalue $\hat{\lambda} \in \mathbb{R}$ of $\{\mathcal{A} + \Delta\mathcal{A}, \mathcal{B} + \Delta\mathcal{B}\}$ such that*

$$\frac{|\tilde{\lambda} - \lambda|}{|\lambda|} \leq \epsilon\frac{(\mathcal{E} + |\lambda|\mathcal{F})|\mathbf{x}|^N}{|\lambda||\mathcal{B}\mathbf{x}^N|} + O(\epsilon^2),$$

where two mode-symmetric tensors $\Delta\mathcal{A}, \Delta\mathcal{B} \in RT_{N,I}$ satisfy $|\Delta\mathcal{A}| \leq \epsilon\mathcal{E}$ and $|\Delta\mathcal{B}| \leq \epsilon\mathcal{F}$ with $\epsilon > 0$.

Proof Since $\det(\mathcal{B}) \neq 0$, it is obvious that $\mathcal{B}\mathbf{y}^N \neq 0$ for all nonzero vectors $\mathbf{y} \in \mathbb{R}^I$. Let $(\lambda; \mathbf{x})$ be a generalized H-eigenvalue of $\{\mathcal{A}, \mathcal{B}\}$ and let $\lambda \neq 0$ be algebraically simple, then a componentwise condition number for an algebraically simple generalized H-eigenvalue λ is defined by

$$\operatorname{cond}(\lambda) := \limsup_{\epsilon \to 0} \left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} : (\mathcal{A} + \Delta\mathcal{A})(\mathbf{x} + \Delta\mathbf{x})^{N-1} = (\lambda + \Delta\lambda)(\mathcal{B} + \Delta\mathcal{B})(\mathbf{x} + \Delta\mathbf{x})^{N-1}, \right.$$
$$\left. |\Delta\mathcal{A}| \leq \epsilon\mathcal{E},\ |\Delta\mathcal{B}| \leq \epsilon\mathcal{F} \right\}.$$

It follows from the above definition that

$$\frac{|\Delta\lambda|}{|\lambda|} \leq \operatorname{cond}(\lambda)\epsilon + O(\epsilon^2).$$

According to the definition of a componentwise condition number for an algebraically simple generalized H-eigenvalue λ, we have

$$\operatorname{cond}(\lambda) \geq \frac{(\mathcal{E} + |\lambda|\mathcal{F})|\mathbf{x}|^N}{|\lambda||\mathcal{B}\mathbf{x}^N|}.$$

The lower bound of $\operatorname{cond}(\lambda)$ is attained when $\Delta\mathcal{A}_0 = \epsilon\mathcal{E} \times_1 \mathbf{D} \times_2 \mathbf{D} \cdots \times_N \mathbf{D}$ and $\Delta\mathcal{B}_0 = -\operatorname{sign}(\lambda)\epsilon\mathcal{F} \times_1 \mathbf{D} \times_2 \mathbf{D} \cdots \times_N \mathbf{D}$, where $\mathbf{D} = \operatorname{diag}(\operatorname{sign}(\mathbf{x}))$ with $\operatorname{sign}(\mathbf{x}) = (\operatorname{sign}(x_1), \operatorname{sign}(x_2) \ldots, \operatorname{sign}(x_I)$ and $\operatorname{sign}(\cdot)$ is defined in (2.3.7). The proof is complete. □

Remark 3.3.1 According to [13, Theorem 3.1], for all nonzero vectors $\mathbf{y} \in \mathbb{C}^I$, $\mathcal{B}\mathbf{y}^N \neq 0$, where $\mathcal{B} \in RT_{N,I}$ is mode-symmetric and nonsingular. Hence if $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ are mode-symmetric and $\mathcal{B}$ is nonsingular, then the above two theorems also hold for any nonzero algebraically simple generalized eigenvalue of $\{\mathcal{A}, \mathcal{B}\}$.

Corollary 3.3.1 *Suppose that the mode-symmetric tensor $\mathcal{A} \in RT_{N,I}$ is nonnegative irreducible and $\mathcal{B} \in RT_{N,I}$ is a diagonal tensor with positive diagonal entries. Let $\mathbf{B} \in \mathbb{R}^{I \times I}$ be a diagonal matrix such that its main diagonal entries equal to the main diagonal entries of $\mathcal{B}$. If λ_n is the mode-n Perron root of $\mathcal{A} \times_n \mathbf{B}^{-1}$, then the following statements hold.*

(1) *All Perron roots λ_n are equal, denoted by λ.*
(2) *Let λ be simple, $\mathcal{E} = \mathcal{A}$ and $\mathcal{F} = \mathcal{B}$. Then cond$(\lambda) = 2$.*
(3) *Moreover, if $\lambda + \Delta\lambda$ is the mode-n Perron root of the pair $(\mathcal{A} + \Delta\mathcal{A}, \mathcal{B} + \Delta\mathcal{B})$, then for $0 \leq \epsilon < 1$, we have*

$$\frac{|\Delta\lambda|}{|\lambda|} \leq \frac{2\epsilon}{1 - \epsilon}.$$

Proof Parts (1) and (2) are direct to verify. We only prove Part (3) in the case of $n = 1$, and other cases are similar. Since $|\Delta\mathcal{B}| \leq \epsilon\mathcal{B}$, with the diagonal tensor $\mathcal{B}$,

and $|\Delta\mathcal{A}| \leq \epsilon\mathcal{A}$, we have

$$\left(\frac{1-\epsilon}{1+\epsilon}\right)\mathcal{A} \times_1 \mathbf{D}^{-1} \leq \mathcal{A} \times_1 \mathbf{D}^{-1} \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)\mathcal{A} \times_1 \mathbf{D}^{-1}.$$

Since the spectral radius ρ_1 is monotonic on the nonnegative tensors [26], then

$$\left(\frac{1-\epsilon}{1+\epsilon}\right)\rho_1(\mathcal{A} \times_1 \mathbf{D}^{-1}) \leq \rho_1(\mathcal{A} \times_1 \mathbf{D}^{-1}) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)\rho_1(\mathcal{A} \times_1 \mathbf{D}^{-1}).$$

Hence, Part (3) is proved for $n = 1$. □

If $\mathcal{B}$ is the identity tensor and $\Delta\mathcal{B}$ is the zero tensor, Part (3) of Corollary 3.3.1 gives a perturbation bound of the spectral radius of a mode-symmetric nonnegative irreducible tensor in $RT_{N,I}$.

According to [15, Theorem 2.1], $\det(\mathcal{B}) \neq 0$ implies that all generalized eigenvalues are finite. Then a generalized eigenvalue λ can be represented as $\lambda = \alpha/\beta$ with $\beta \neq 0$. Hence, a generalized eigenpair $(\lambda; \mathbf{x})$ can also be represented as $(\alpha, \beta; \mathbf{x})$. Denote a generalized eigenvalue λ by (α, β) or $\langle\alpha, \beta\rangle$, where $\langle\alpha, \beta\rangle = \tau(\alpha, \beta)$ with $\tau \neq 0$. A property of the pair $\langle\alpha, \beta\rangle$ is given below.

Theorem 3.3.5 *Let $\mathcal{A}, \mathcal{B} \in RT_{N,I}$ be mode-symmetric and $\mathcal{B}$ be nonsingular. If $\langle\alpha, \beta\rangle$ is an algebraically simple generalized eigenvalue of the pair $\{\mathcal{A}, \mathcal{B}\}$ with corresponding generalized eigenvector $\mathbf{x} \in \mathbb{R}^I$, then*

$$\langle\alpha, \beta\rangle = \langle\mathcal{A}x^N, \mathcal{B}x^N\rangle.$$

Proof For the vector $\mathbf{x} \in \mathbb{R}^I$, there exists a Householder matrix [2] $\mathbf{P} \in \mathbb{R}^{I\times I}$ such that $\mathbf{Px} = \|\mathbf{x}\|\mathbf{e}_1$, where $\mathbf{e}_1$ is the first column of $\mathbf{I}_I$. According to [15], we have $\Lambda(\mathcal{A}, \mathcal{B}) = \Lambda(\hat{\mathcal{A}}, \hat{\mathcal{B}})$, where

$$\hat{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{P} \times_2 \mathbf{P} \cdots \times_N \mathbf{P}, \quad \hat{\mathcal{B}} = \mathcal{B} \times_1 \mathbf{P} \times_2 \mathbf{P} \cdots \times_N \mathbf{P}.$$

Thus the pair $(\alpha, \beta; \mathbf{e}_1)$ satisfies $\beta\hat{\mathcal{A}}\mathbf{e}_1^{N-1} = \alpha\hat{\mathcal{B}}\mathbf{e}_1^{N-1}$, that is, $\langle\alpha, \beta\rangle = \langle\hat{\mathcal{A}}\mathbf{e}_1^N, \hat{\mathcal{B}}\mathbf{e}_1^N\rangle$. □

3.3.2 Tensor Quadratic Eigenvalue Problem

For $z \in \mathbb{C}$, let $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + z^2\mathcal{A}_2$. Suppose that $(\lambda; \mathbf{x})$ is an eigenpair of $\mathcal{P}$. Define $\mu = \lambda^{1/(N-1)}$ and $\mathbf{y} = \mu\mathbf{x}$, then $\mathcal{P}(\lambda)\mathbf{x}^{N-1} = \mathbf{0}_I$ can be represented as

$$\lambda\mathcal{A}_2\mathbf{y}^{N-1} + \mathcal{A}_1\mathbf{y}^{N-1} + \mathcal{A}_0\mathbf{x}^{N-1} = \mathbf{0}_I, \quad \mathbf{y}^{[N-1]} - \lambda\mathbf{x}^{[N-1]} = \mathbf{0}_I.$$

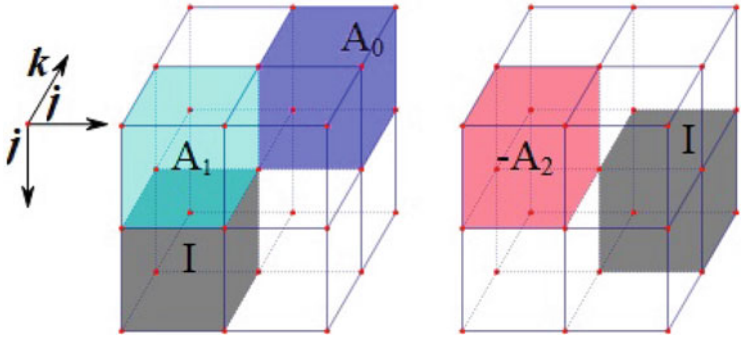

Fig. 3.2 Left for $\tilde{\mathcal{A}}$ and right for $\tilde{\mathcal{B}}$. Unlabeled parts are zero

Define $\tilde{\mathbf{x}} = (\mathbf{y}, \mathbf{x})$, then

$$\tilde{\mathcal{A}}\tilde{\mathbf{x}}^{N-1} = \lambda\tilde{\mathcal{B}}\tilde{\mathbf{x}}^{N-1},$$

where $\tilde{\mathcal{A}} \in RT_{N,2I}$ is

$$\begin{cases} \tilde{\mathcal{A}}[1:I|1:I|\ldots|1:I] = \mathcal{A}_1, \\ \tilde{\mathcal{A}}[1:I|I+1:2I|\ldots|I+1:2I] = \mathcal{A}_0, \\ \tilde{\mathcal{A}}[I+1:2I|1:I|\ldots|1:I] = \mathcal{I}; \end{cases}$$

and $\tilde{\mathcal{B}} \in RT_{N,2I}$ is given by

$$\tilde{\mathcal{B}}[1:I|1:I|\ldots|1:I] = -\mathcal{A}_2, \quad \tilde{\mathcal{B}}[I+1:2I|I+1:2I|\ldots|I+1:2I] = \mathcal{I}.$$

Other entries of $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ are zero, as displayed in Fig. 3.2. $\tilde{\mathcal{A}}$ is not mode-symmetric. It follows from [13, Theorem 4.2] that

$$\det(\tilde{\mathcal{B}}) = \det(\mathcal{A}_2)^{(N-1)^I} \neq 0.$$

Since $\mathcal{A}_2$ is mode-symmetric, hence $\tilde{\mathcal{B}}$ is mode-symmetric and nonsingular. The number of mode-1 generalized eigenvalues of $\{\tilde{\mathcal{A}}, \tilde{\mathcal{B}}\}$ is $2I(N-1)^{2I-1}$ and $\Lambda(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$ is a finite subset of $\mathbb{C}$. Thus, the number of eigenvalues of $\mathcal{P}$ is less than or equal to $2I(N-1)^{2I-1}$ and $\Lambda(\mathcal{P})$ is a finite subset of $\mathbb{C}$.

The first-order perturbation of an algebraically simple eigenvalue, a special case of Theorem 3.3.7, is given in the following theorem without the proof.

Theorem 3.3.6 *For given mode-symmetric tensors $\mathcal{A}_l \in RT_{N,I}$ $(l = 0, 1, 2)$, suppose that $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + z^2\mathcal{A}_2$ where $z \in \mathbb{C}$, and $\mathcal{A}_2$ is nonsingular. If $(\lambda; \mathbf{x})$ is an algebraically simple eigenpair of $\mathcal{P}$ with $\lambda \in \mathbb{C}$ and $\mathbf{x} \in \mathbb{C}^I$, then*

there exists an algebraically simple eigenvalue $\tilde{\lambda} \in \mathbb{C}$ *of* $\mathcal{P} + \Delta\mathcal{P}$ *such that*

$$\tilde{\lambda} = \lambda + O(\epsilon),$$

where $|\epsilon| \leq \epsilon_0$ *with* $\epsilon_0 > 0$. *Here* $\Delta\mathcal{A}_l \in RT_{N,I}$ *with* $l = 0, 1, 2$ *in* $\Delta\mathcal{P}$ *are mode-symmetric.*

Remark 3.3.2 In the above theorem, we assume without loss of generality that $\mathcal{A}_2$ is nonsingular. This assumption eliminates infinite eigenvalues.

3.3.3 Tensor Polynomial Eigenvalue Problem

In this section, we suppose that $L > 2$. Let $(\lambda; \mathbf{x})$ be an eigenpair of $\mathcal{P}$. Define $\mu = \lambda^{1/(N-1)}$ and $\mathbf{y}_l = \mu^l \mathbf{x}$ with $l = 0, 1, \ldots, L-1$, then $\mathcal{P}(\lambda)\mathbf{x}^{N-1} = \mathbf{0}_I$ can be rewritten as the system of nonlinear equations:

$$\begin{cases} \lambda\mathcal{A}_L\mathbf{y}_{L-1}^{N-1} + \cdots + \mathcal{A}_1\mathbf{y}_1^{N-1} + \mathcal{A}_0\mathbf{x}^{N-1} = \mathbf{0}_I, \\ \mathbf{y}_l = \mu\mathbf{y}_{l-1}, \quad l = 2, 3, \ldots, L-1, \\ \mathbf{y}_1 = \mu\mathbf{x}. \end{cases}$$

Let $\tilde{\mathbf{x}} = (\mathbf{y}_{L-1}, \ldots, \mathbf{y}_1, \mathbf{x})$, then $(\lambda; \tilde{\mathbf{x}})$ is a solution of the system of nonlinear equations:

$$\tilde{\mathcal{A}}\tilde{\mathbf{x}}^{N-1} = \lambda\tilde{\mathcal{B}}\tilde{\mathbf{x}}^{N-1},$$

where for $l = 0, 1, \ldots, L-1$, $\tilde{\mathcal{A}} \in RT_{N,LI}$ is given by

$$\begin{cases} \tilde{\mathcal{A}}[1:I|lI+1:(l+1)I|\ldots|lI+1:(l+1)I] = \mathcal{A}_{L-1-l}, \\ \tilde{\mathcal{A}}[(l+1)I+1:(l+2)I|lI+1:(l+1)I|\ldots|lI+1:(l+1)I] = \mathcal{I}; \end{cases}$$

and the entries of $\tilde{\mathcal{B}} \in RT_{N,LI}$ are given by

$$\begin{cases} \tilde{\mathcal{B}}[1:I|1:I|\ldots|1:I] = -\mathcal{A}_L, \\ \tilde{\mathcal{B}}[Il+1:(l+1)I|Il+1:(l+1)I|\ldots|Il+1:(l+1)I] = \mathcal{I}. \end{cases}$$

Non-specified entries of $\tilde{\mathcal{A}}$ and $\hat{\mathcal{B}}$ are zero. $\tilde{\mathcal{A}}$ is not mode-symmetric. According to [13, Theorem 4.2], we have

$$\det(\tilde{\mathcal{B}}) = \det(\mathcal{A}_2)^{(N-1)^{(L-1)I}} \neq 0.$$

Since $\mathcal{A}_L$ is mode-symmetric, $\tilde{\mathcal{B}}$ is mode-symmetric and nonsingular. As we know, the number of mode-1 generalized eigenvalues of $\{\tilde{\mathcal{A}}, \tilde{\mathcal{B}}\}$ is $LI(N-1)^{LI-1}$ and $\Lambda(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$ is a finite subset of $\mathbb{C}$. Thus, the number of eigenvalues of $\mathcal{P}$ is less than or equal to $LI(N-1)^{LI-1}$ and $\Lambda(\mathcal{P})$ is finite in $\mathbb{C}$.

Theorem 3.3.7 *For mode-symmetric tensors $\mathcal{A}_l \in RT_{N,I}$ with all l, suppose that $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^L\mathcal{A}_L$ where $z \in \mathbb{C}$ and $\mathcal{A}_L$ is nonsingular. If $(\lambda; \mathbf{x})$ is an algebraically simple eigenpair of $\mathcal{P}$ with $\lambda \in \mathbb{C}$ and $\mathbf{x} \in \mathbb{C}^I$, there exists an algebraically simple eigenvalue $\tilde{\lambda} \in \mathbb{C}$ of $\mathcal{P} + \Delta\mathcal{P}$ such that*

$$\tilde{\lambda} = \lambda + O(\epsilon),$$

where $|\epsilon| \le \epsilon_0$ with $\epsilon_0 > 0$ and $\Delta\mathcal{A}_l \in RT_{N,I}$ with all l in $\Delta\mathcal{P}$ are mode-symmetric.

Proof If $(\lambda; \mathbf{x})$ is an eigenpair of $\mathcal{P}$, there exist a vector $\tilde{\mathbf{x}} = (\mathbf{y}_{L-1}, \ldots, \mathbf{y}_1, \mathbf{x})$ and $\tilde{\mathcal{A}}, \tilde{\mathcal{B}} \in RT_{N,LI}$ such that $\tilde{\mathcal{A}}\tilde{\mathbf{x}}^{N-1} = \lambda(-\tilde{\mathcal{B}})\tilde{\mathbf{x}}^{N-1}$. If $(\lambda; \mathbf{x})$ is an algebraically simple eigenpair, then it is an algebraically simple mode-1 generalized eigenpair of $\{\tilde{\mathcal{A}}, -\tilde{\mathcal{B}}\}$.

According to Theorem 3.3.1, for the pair $\{\tilde{\mathcal{A}}, -\tilde{\mathcal{B}}\}$, there exists an algebraically simple mode-1 generalized eigenvalue $\hat{\lambda}$ of $\{\tilde{\mathcal{A}}+\epsilon\hat{\mathcal{A}}, -(\tilde{\mathcal{B}}+\epsilon\hat{\mathcal{B}})\}$ such that $\tilde{\lambda} = \lambda + O(\epsilon)$. For these two tensors $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$, there exist mode-symmetric tensors $\Delta\mathcal{A}_l \in RT_{N,I}$ with all l such that $\tilde{\lambda}$ is a algebraically simple eigenvalue of $\mathcal{P} + \Delta\mathcal{P}$. □

Theorem 3.3.8 *Let $\alpha_l > 0$ and $\mathcal{A}_l \in RT_{N,I}$ be mode-symmetric with all l. Suppose that $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^L\mathcal{A}_L$ where $z \in \mathbb{C}$ and $\mathcal{A}_L$ is nonsingular. Let $(\lambda; \mathbf{x})$ be an algebraically simple* H*-eigenpair of $\mathcal{P}$ with $\lambda \neq 0$ and $\mathbf{x} \in \mathbb{R}^I$. When $\mathcal{G}(\lambda)\mathbf{x}^N \neq 0$, there exists an algebraically simple* H*-eigenvalue $\hat{\lambda}$ of $\mathcal{P} + \Delta\mathcal{P}$ such that*

$$\frac{|\hat{\lambda} - \lambda|}{|\lambda|} \le \epsilon \frac{\left(\sum_{l=0}^{L} |\lambda|^l \alpha_l\right) \|\mathbf{x}\|_2^N}{|\lambda||\mathcal{G}(\lambda)\mathbf{x}^N|} + O(\epsilon^2),$$

where for $0 < \epsilon < 1$, the mode-symmetric tensors $\Delta\mathcal{A}_l \in RT_{N,I}$ with all l in $\Delta\mathcal{P}$ satisfy $\|\Delta\mathcal{A}_l\|_F \le \epsilon\alpha_l$.

Theorem 3.3.9 *For all l, let $\mathcal{E}_l \in RT_{N,I}$ be mode-symmetric, with positive entries, and $\mathcal{A}_l \in RT_{N,I}$ be mode-symmetric. Suppose that $\mathcal{P}(z) = \mathcal{A}_0 + z\mathcal{A}_1 + \cdots + z^L\mathcal{A}_L$ where $z \in \mathbb{C}$ and $\mathcal{A}_L$ is nonsingular. Let $(\lambda; \mathbf{x})$ be an algebraically simple* H*-eigenpair of $\mathcal{P}$ with $\lambda \neq 0$ and $\mathbf{x} \in \mathbb{R}^I$. When $\mathcal{G}(\lambda)\mathbf{x}^N \neq 0$, there exists an algebraically simple* H*-eigenvalue $\hat{\lambda}$ of $\mathcal{P} + \Delta\mathcal{P}$ such that*

$$\frac{|\hat{\lambda} - \lambda|}{|\lambda|} \le \epsilon \frac{\left(\sum_{l=0}^{L} |\lambda|^l \mathcal{E}_l\right) |\mathbf{x}|^N}{|\lambda||\mathcal{G}(\lambda)\mathbf{x}^N|} + O(\epsilon^2),$$

where the mode-symmetric tensors $\Delta\mathcal{A}_l \in RT_{N,I}$ with all l in $\Delta\mathcal{P}$ satisfy $|\Delta\mathcal{A}_l| \le \epsilon\mathcal{E}_l$ and $\epsilon < 1$.

The omitted proofs of Theorems 3.3.8 and 3.3.9 are similar to that of Theorems 3.3.3 and 3.3.4, respectively. In Theorems 3.3.8 and 3.3.9, we assume that $\mathcal{G}(\lambda)\mathbf{x}^N \neq 0$, where $(\lambda; \mathbf{x})$ is an algebraically simple H-eigenpair of $\mathcal{P}(\lambda)$. However, in Theorems 3.3.3 and 3.3.4, if $\mathcal{B}$ is nonsingular, then $\mathcal{B}\mathbf{y}^N \neq 0$ holds for all nonzero vectors $\mathbf{y} \in \mathbb{C}^I$.

3.4 The Smallest Eigenvalue of Nonsingular $\mathcal{M}$-Tensors

Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric. If $\mathcal{A} = s\mathcal{I} - \mathcal{B}$, where the symmetric tensor $\mathcal{B} \in RT_{N,I}$ is nonnegative irreducible and $s > \rho(\mathcal{B})$, then $\mathcal{A}$ is an irreducible symmetric nonsingular $\mathcal{M}$-tensor [15, 28]. In this section, we focus on the perturbation theory for the smallest eigenvalue (i.e., the smallest H-eigenvalue) of an irreducible symmetric nonsingular $\mathcal{M}$-tensor in $RT_{N,I}$. We need an inequality for the Perron root of any nonnegative irreducible symmetric tensor in $RT_{N,I}$.

3.4.1 An Inequality for the Perron Root

We first introduce the weighted arithmetic mean and weighted geometric mean (AM-GM) inequality. Let $x_i, w_i \geq 0$ with all i. Set $w = w_1 + w_2 + \cdots + w_I$. If $w > 0$, then the inequality

$$\frac{w_1x_1 + w_2x_2 + \cdots + w_Ix_I}{w} \geq \sqrt[w]{x_1^{w_1}x_2^{w_2}\ldots x_I^{w_I}}.$$

holds, with equality if and only if all the x_i with $w_i > 0$ are equal, where $\sqrt[w]{x_1^{w_1}x_2^{w_2}\ldots x_I^{w_I}}$ is the wth root of $x_1^{w_1}x_2^{w_2}\ldots x_I^{w_I}$. Here the convention $0^0 = 1$ is used.

Assume that the symmetric tensor $\mathcal{A} \in RT_{N,I}$ is nonnegative irreducible. Use $(\rho(\mathcal{A}); \mathbf{u})$ to denote the Perron pair (defined in Sect. 1.3.1) of $\mathcal{A}$ with $\rho(\mathcal{A}) > 0$ and $\mathbf{u} \in \mathbb{R}^I_+$. We have the following theorem.

Theorem 3.4.1 *Suppose that the symmetric tensor $\mathcal{A} \in RT_{N,I}$ is nonnegative irreducible. Let $(\rho(\mathcal{A}); \mathbf{u})$ be the Perron pair of $\mathcal{A}$ with $\rho(\mathcal{A}) > 0$ and $\mathbf{u} \in \mathbb{R}^I_+$. Then for any positive diagonal matrix $\mathbf{D} \in \mathbb{R}^{I\times I}$, we have the following results.*

(1) *The inequality $\mathcal{A}\mathbf{u}^N \leq \hat{\mathcal{A}}\mathbf{u}^N$ holds, where $\hat{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{D}^{1-N} \times_2 \mathbf{D} \cdots \times_N \mathbf{D}$, and the equality is attained if and only if $\mathbf{D} = \alpha\mathbf{I}_I$ for all $\alpha > 0$.*
(2) *$\rho(\mathcal{A})$ can be expressed by*

$$\max_{\mathbf{z}>\mathbf{0}_I} \min_{\mathbf{y}>\mathbf{0}_I} \frac{\mathbf{y}^\top(\mathcal{A}\mathbf{z}^{N-1})}{\mathbf{y}^\top\mathbf{z}^{[N-1]}} = \rho(\mathcal{A}) = \min_{\mathbf{z}>\mathbf{0}_I} \max_{\mathbf{y}>\mathbf{0}_I} \frac{\mathbf{y}^\top(\mathcal{A}\mathbf{z}^{N-1})}{\mathbf{y}^\top\mathbf{z}^{[N-1]}}.$$

In above theorem, replace $\mathbf{D}$ by $\mathbf{D}^{-1}$, part (1) of Theorem 3.4.1 has another form.

(1)' The inequality $\mathcal{A}\mathbf{u}^N \leq \hat{\mathcal{A}}\mathbf{u}^N$ holds, where $\hat{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{D}^{N-1} \times_2 \mathbf{D}^{-1} \cdots \times_N \mathbf{D}^{-1}$, and the equality is attained if and only if $\mathbf{D} = \alpha \mathbf{I}_I$ for all $\alpha > 0$.

Proof Since the symmetric tensor $\mathcal{A} \in RT_{N,I}$ is nonnegative irreducible, the Perron pair $(\rho(\mathcal{A}); \mathbf{u})$ is unique, and we have

$$\begin{aligned}\mathcal{A} \times_2 \mathbf{u}^\top \cdots \times_N \mathbf{u}^\top &= \mathcal{A} \times_1 \mathbf{u}^\top \times_3 \mathbf{u}^\top \cdots \times_N \mathbf{u}^\top \\ &= \cdots = \mathcal{A} \times_1 \mathbf{u}^\top \cdots \times_{N-1} \mathbf{u}^\top \\ &= \rho(\mathcal{A})\mathbf{u}^{[N-1]}.\end{aligned}$$

Let $\mathbf{u} = (u_1, u_2, \ldots, u_I)^\top \in \mathbb{R}^I$. Define $\mathbf{U} = \mathrm{diag}(u_1, u_2, \ldots, u_I) \in \mathbb{R}^{I \times I}$, and

$$\tilde{\mathcal{D}} = \mathrm{tendiag}\left(\rho(\mathcal{A})u_1^N, \rho(\mathcal{A})u_2^N, \ldots, \rho(\mathcal{A})u_I^N\right),$$

$$\sigma = \left(1 + \max_{i=1,2,\ldots,I}\left[(\rho(\mathcal{A}) - a_{ii\ldots i})u_i^N\right]\right)^{-1},$$

where 'tendiag($\mathbf{v}$)' is the function in the MATLAB Tensor Toolbox [29] that creates a diagonal tensor whose main diagonal entries equal to the entries of $\mathbf{v}$. It is easy to verify that the tensor

$$\mathcal{D} = \sigma \cdot \left(\mathcal{A} \times_1 \mathbf{U} \times_2 \mathbf{U} \cdots \times_N \mathbf{U} - \tilde{\mathcal{D}}\right) + \mathcal{I}.$$

is plane stochastic (cf. Definition 5.1.1). For all i, let two positive vectors $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^I$ satisfy

$$\mathbf{y}_{2,i}\mathbf{y}_{1,i}^{N-1} = u_i^N.$$

Without loss of generality, we define a positive vector $\mathbf{z} \in \mathbb{R}^I$ such that for all i, $\mathbf{y}_{1,i} = u_i z_i$ and $\mathbf{y}_{2,i} = u_i z_i^{1-N}$. Hence, for all i, we have

$$\mathbf{y}_{2,i}\mathbf{y}_{1,i}^{N-1} = u_i^N.$$

We next show that

$$\mathcal{D} \times_1 \left(\mathbf{z}^{[1-N]}\right)^\top \times_2 \mathbf{z}^\top \cdots \times_N \mathbf{z}^\top \geq \mathcal{D} \times_1 \mathbf{e}^\top \times_2 \mathbf{e}^\top \cdots \times_N \mathbf{e}^\top. \tag{3.4.1}$$

Since $\mathcal{D}$ is plane stochastic, we have

$$\mathcal{D} \times_1 \mathbf{e}^\top \times_2 \mathbf{e}^\top \cdots \times_N \mathbf{e}^\top = I.$$

Let $\mathbf{z} = (z_1, z_2, \ldots, z_I)^\top \in \mathbb{R}^I$. According to the AM-GM inequality, we have

$$\begin{aligned}\mathcal{D} \times_1 \left(\mathbf{z}^{[1-N]}\right)^\top \times_2 \mathbf{z}^\top \cdots \times_N \mathbf{z}^\top &= \sum_{i_1,i_2,\ldots,i_N=1}^{I} d_{i_1 i_2 \ldots i_N} \cdot \frac{z_{i_2} \cdots z_{i_N}}{z_{i_1}^{N-1}} \\ &= I \left(\sum_{i_1,i_2,\ldots,i_N=1}^{I} \frac{d_{i_1 i_2 \ldots i_N}}{N} \cdot \frac{z_{i_2} \cdots z_{i_N}}{z_{i_1}^{N-1}} \right) \\ &\geq I \sqrt[I]{\prod_{i_1,i_2,\ldots,i_N=1}^{I} \left(\frac{z_{i_2} \cdots z_{i_N}}{z_{i_1}^{N-1}} \right)^{d_{i_1 i_2 \ldots i_N}}}. \quad (3.4.2)\end{aligned}$$

Let

$$\frac{Q(z_1, z_2, \ldots, z_I)}{P(z_1, z_2, \ldots, z_I)} = \prod_{i_1,i_2,\ldots,i_N=1}^{I} \left(\frac{z_{i_2} \cdots z_{i_N}}{z_{i_1}^{N-1}} \right)^{d_{i_1 i_2 \ldots i_N}}.$$

Since $\mathcal{D}$ is plane stochastic, then, for all i_1, the number of z_{i_1} in $P(z_1, z_2, \ldots, z_I)$ is

$$(N-1) \left(\sum_{i_2,\ldots,i_N=1}^{I} d_{i_1 i_2 \ldots i_N} \right) = N - 1.$$

Tedious manipulation yields that z_i in the numerator $Q(z_1, z_2, \ldots, z_I)$:

$$\sum_{i_1,i_3,\ldots,i_N=1}^{I} d_{i_1 i i_3 \ldots i_N} + \cdots + \sum_{i_1,\ldots,i_{N-2},i_N=1}^{I} d_{i_1 \ldots i_{N-2} i i_N} + \sum_{i_1,i_2,\ldots,i_{N-1}=1}^{n} d_{i_1 i_2 \ldots i_{N-1} i} = N-1.$$

Hence

$$\breve{\mathcal{A}} \times_1 \left(\mathbf{z}^{[1-N]}\right)^\top \times_2 \mathbf{z}^\top \cdots \times_N \mathbf{z}^\top \geq \breve{\mathcal{A}} \times_1 \mathbf{e}^\top \times_2 \mathbf{e}^\top \cdots \times_N \mathbf{e}^\top,$$

where $\breve{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{U} \times_2 \mathbf{U} \cdots \times_N \mathbf{U}$. It is easy to check that

$$\begin{aligned}\mathcal{A} \times_1 \mathbf{u}^\top \times_2 \mathbf{u}^\top \cdots \times_N \mathbf{u}^\top &\leq \mathcal{A} \times_1 \mathbf{y}_2^\top \times_2 \mathbf{y}_1^\top \cdots \times_N \mathbf{y}_1^\top \\ &= \mathcal{A} \times_1 (\mathbf{u}. * \mathbf{z}^{[1-N]})^\top \times_2 (\mathbf{u}. * \mathbf{z})^\top \cdots \times_N (\mathbf{u}. * \mathbf{z})^\top,\end{aligned}$$

where $\mathbf{x}. * \mathbf{y} = (x_1 y_1, x_2 y_2 \ldots, x_I y_I)^\top \in \mathbb{R}^I$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$.

It remains to prove Part (2). For positive a_i, b_i and d_i with n, we prove that

$$\max\left\{\frac{b_i}{d_i} : i = 1, 2, \ldots, I\right\} \geq \frac{\sum\limits_{i=1}^{I} a_i b_i}{\sum\limits_{i=1}^{I} a_i d_i} \geq \min\left\{\frac{b_i}{d_i} : i = 1, 2, \ldots, I\right\}.$$

Assume that $\min\left\{\frac{b_i}{d_i} : i = 1, 2, \ldots, I\right\} = \frac{b_k}{d_k}$, we have

$$\sum_{i=1}^{I} a_i b_i - \min\left\{\frac{b_i}{d_i} : i = 1, 2, \ldots, I\right\} \sum_{i=1}^{I} a_i d_i = \sum_{i=1}^{I} a_i d_i \left(\frac{b_i}{d_i} - \frac{b_k}{d_k}\right) \geq 0,$$

and the equality holds if $a_i = \delta_{ik}$. Thus

$$\frac{\sum\limits_{i=1}^{I} a_i b_i}{\sum\limits_{i=1}^{I} a_i d_i} \geq \min\left\{\frac{b_i}{d_i} : i = 1, 2, \ldots, I\right\},$$

and similarly,

$$\frac{\sum\limits_{i=1}^{I} a_i b_i}{\sum\limits_{i=1}^{I} a_i d_i} \leq \max\left\{\frac{b_i}{d_i} : i = 1, 2, \ldots, I\right\}.$$

According to [14, Theorem 4.2], we have

$$\max_{\mathbf{z}>\mathbf{0}_I} \min_{z_i>0} \frac{(\mathcal{A}\mathbf{z}^{N-1})_i}{z_i^{N-1}} = \lambda = \min_{\mathbf{z}>\mathbf{0}_I} \max_{z_i>0} \frac{(\mathcal{A}\mathbf{z}^{N-1})_i}{z_i^{N-1}}. \tag{3.4.3}$$

Since $\mathbf{z} > \mathbf{0}_I$, and $\mathcal{A}$ is an irreducible nonnegative tensor, we obtain

$$\max_{z_i>0} \frac{(\mathcal{A}\mathbf{z}^{N-1})_i}{z_i^{N-1}} = \max_{\mathbf{y}>\mathbf{0}_I} \frac{\mathbf{y}^\top(\mathcal{A}\mathbf{z}^{N-1})}{\mathbf{y}^\top\mathbf{z}^{[N-1]}}.$$

We prove the second part from (3.4.3). □

Remark 3.4.1 The number of z_i in $Q(z_1, z_2, \ldots, z_I)$ is $N-1$. Let $N = 3$ and $i = 1$ for illustration. Then the number of z_1 in $Q(z_1, z_2, z_3)$ is

$$\begin{aligned} &2d_{111} + d_{121} + d_{112} + \cdots + d_{1I1} + d_{11I} + 2d_{211} + d_{221} + d_{212} + \cdots + d_{2I1} + d_{21I} \\ &\quad + \cdots + 2d_{I11} + d_{I21} + d_{I12} + \cdots + d_{II1} + d_{I1I} \\ &= \sum_{i_1,i_2=1}^{I} d_{i_1 i_2 1} + \sum_{i_1,i_3=1}^{I} d_{i_1 1 i_3} = 2, \end{aligned}$$

the last equality holds due to the plane stochastic tensor $\mathcal{D} \in RT_{3,I}$.

A more general form for Part (1) of Theorem 3.4.1 is given below.

Theorem 3.4.2 *Suppose that the symmetric tensor $\mathcal{A} \in RT_{N,I}$ is nonnegative irreducible. Let $(\rho(\mathcal{A}); \mathbf{u})$ be the Perron pair of $\mathcal{A}$ with $\rho(\mathcal{A}) > 0$ and $\mathbf{u} \in \mathbb{R}^I_+$. Then for any positive diagonal matrix $\mathbf{D} \in \mathbb{R}^{I\times I}$, we have the following results.*

(a) *The inequality $\mathcal{A}\mathbf{u}^N \le \hat{\mathcal{A}}_n \mathbf{u}^N$ holds, where*

$$\hat{\mathcal{A}}_n = \mathcal{A} \times_1 \mathbf{D} \cdots \times_{n-1} \mathbf{D} \times_n \mathbf{D}^{1-N} \times_{n+1} \mathbf{D} \cdots \times_N \mathbf{D},$$

and the equality is attained if and only if $\mathbf{D} = \alpha \mathbf{I}_I$ for all $\alpha > 0$.

(b) *The inequality $\mathcal{A}\mathbf{u}^N \le \hat{\mathcal{A}}_n \mathbf{u}^N$ holds, where*

$$\hat{\mathcal{A}}_n = \mathcal{A} \times_1 \mathbf{D}^{\alpha_1} \times_2 \mathbf{D}^{\alpha_2} \cdots \times_N \mathbf{D}^{\alpha_N}, \quad \prod_{n=1}^{N} \alpha_n = 1, \quad \alpha > 0,$$

and the equality is attained if and only if $\mathbf{D} = \alpha \mathbf{I}_I$ for all $\alpha > 0$.

3.4.2 Perturbation of the Smallest Eigenvalue of $\mathcal{M}$-Tensors

Suppose that $\mathcal{A} \in RT_{N,I}$ is an irreducible symmetric nonsingular $\mathcal{M}$-tensor, the perturbation of the smallest eigenvalue of $\mathcal{A}$ is derived. For the perturbation bound, we need the following theorem, generalizing the matrix case [30].

Theorem 3.4.3 *Let $\mathcal{A} \in RT_{N,I}$ be an irreducible symmetric nonsingular $\mathcal{M}$-tensor with $\mathcal{A} = \mathcal{D} - \mathcal{N}$, where $\mathcal{D} \in RT_{N,I}$ is diagonal and the symmetric nonnegative tensor $\mathcal{N} \in RT_{N,I}$ has zero main diagonal. Let $\mathbf{D} \in \mathbb{R}^{I\times I}$ be a diagonal matrix, whose main diagonal entries equal to the main diagonal entries of $\mathcal{A}$. Let $(\lambda; \mathbf{x})$ be the smallest eigenpair of $\mathcal{A}$ with $\mathbf{x} \in \mathbb{R}^I_{++}$ and $\lambda > 0$. Let $\rho(\mathcal{N} \times_n \mathbf{D}^{-1})$ be the mode-n spectral radius of $\mathcal{N} \times_n \mathbf{D}^{-1}$ for all n. Then we have*

$$\mathcal{N}\mathbf{x}^N \le \rho(\mathcal{N} \times_n \mathbf{D}^{-1}) \mathcal{D}\mathbf{x}^N. \tag{3.4.4}$$

The equality holds if and only if $\mathcal{D} = \alpha\mathcal{I}$ for $\alpha > 0$. Denote $\rho = \min\{\rho(\mathcal{N}\times_n \mathbf{D}^{-1}) : n = 1, 2, \ldots, N\}$, (3.4.4) can be simplified as $\mathcal{N}\mathbf{x}^N \leq \rho\mathcal{D}\mathbf{x}^N$.

Proof Without loss of generality, we only prove the case of $n = 1$. Since $\mathcal{A} \in RT_{N,I}$ is an irreducible symmetric nonsingular $\mathcal{M}$-tensor and $(\lambda; \mathbf{x})$ is the eigenpair corresponding to the smallest eigenvalue, we have

$$\mathcal{A}\times_2\mathbf{x}^\top\cdots\times_N\mathbf{x}^\top = \mathcal{A}\times_1\mathbf{x}^\top\times_3\mathbf{x}^\top\cdots\times_N\mathbf{x}^\top = \ldots = \mathcal{A}\times_1\mathbf{x}^\top\cdots\times_{N-1}\mathbf{x}^\top = \rho(\mathcal{A})\mathbf{x}^{[N-1]}.$$

We apply Theorem 3.4.1 to $\mathcal{B} = t\mathcal{I} - \mathcal{D} + \mathcal{N}$, which for sufficiently large t satisfies the assumptions. For any diagonal $\mathbf{\Delta} \in \mathbb{R}^{I\times I}$, it holds $\mathcal{B}\mathbf{x}^N \leq \tilde{\mathcal{B}}\mathbf{x}^N$, where $\tilde{\mathcal{B}} = \mathcal{B} \times_1 \mathbf{\Delta}^{N-1} \times_2 \mathbf{\Delta}^{-1} \cdots \times_N \mathbf{\Delta}^{-1}$, from which we infer that

$$\mathcal{N}\mathbf{x}^N \leq \tilde{\mathcal{N}}\mathbf{x}^N,$$

where $\tilde{\mathcal{N}} = \mathcal{N} \times_1 \mathbf{\Delta}^{N-1} \times_2 \mathbf{\Delta}^{-1} \cdots \times_N \mathbf{\Delta}^{-1}$.

Let $\rho(\mathcal{N} \times_1 \mathbf{D}^{-1})$ be the mode-1 spectral radius of $\mathcal{N} \times_1 \mathbf{D}^{-1}$, then

$$(\mathcal{N} \times_1 \mathbf{D}^{-1}) \times_2 \mathbf{y}_1^\top \cdots \times_m \mathbf{y}_1^\top = \rho(\mathcal{N} \times_1 \mathbf{D}^{-1})\mathbf{y}_1^{[N-1]},$$

We choose $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^I_{++}$ such that $\mathbf{x} = \mathbf{\Delta} \cdot \mathbf{y}_1$ and $\mathbf{x} = \mathbf{\Delta}^{1-N} \cdot \mathbf{y}_2$, then

$$\begin{aligned}
\mathcal{N}\mathbf{x}^N &\leq ((\mathcal{N} \times_1 \mathbf{D}^{-1}) \times_2 \mathbf{y}_1^\top \cdots \times_N \mathbf{y}_1^\top) \times_1 \mathbf{D} \times_1 \mathbf{y}_2^\top \\
&= \rho(\mathcal{N} \times_1 \mathbf{D}^{-1})\mathcal{I} \times_1 \left(\mathbf{y}_1^{[N-1]}\right)^\top \times_1 \mathbf{D} \times_1 \mathbf{y}_2^\top \\
&= \rho(\mathcal{N} \times_1 \mathbf{D}^{-1})\mathcal{D}\mathbf{x}^N.
\end{aligned}$$

The equality is attained if and only if $\mathbf{\Delta} = \alpha\mathbf{I}_I$ for $\alpha > 0$. Thus $\mathbf{x}$ is also a H-eigenvector of $\mathcal{N} \times_1 \mathbf{D}^{-1}$. Hence, we have $\mathcal{A}\mathbf{x}^{N-1} = (\mathcal{D} - \mathcal{N})\mathbf{x}^{N-1} = \lambda\mathbf{x}^{[N-1]}$.

According to the assumptions, we have

$$\mathbf{D}\mathbf{x}^{[N-1]} - \mathcal{N}\mathbf{x}^{N-1} = \lambda\mathbf{x}^{[N-1]},$$

i.e.,

$$\mathbf{x}^{[N-1]} - \mathbf{D}^{-1}\mathcal{N}\mathbf{x}^{N-1} = \lambda\mathbf{D}^{-1}\mathbf{x}^{[N-1]}.$$

Since $\mathbf{D}^{-1}\mathcal{N}\mathbf{x}^{N-1} = (\mathcal{N} \times_1 \mathbf{D}^{-1})\mathbf{x}^{N-1}$, we have $\mathcal{D}\mathbf{x}^{N-1} = \tau\mathbf{x}^{[N-1]}$ for some constant τ. Thus, $\mathcal{D} = \alpha\mathcal{I}$ for $\alpha > 0$. Similar, for all n, we have $\mathcal{N}\mathbf{x}^N \leq \rho(\mathcal{N} \times_n \mathbf{D}^{-1})\mathcal{D}\mathbf{x}^N$. Then, for the smallest $\rho(\mathcal{N}\times_n D^{-1})$ denoted by ρ, $\mathcal{N}\mathbf{x}^N \leq \rho\mathcal{D}\mathbf{x}^N$ holds. □

From Theorem 3.4.3, the following corollary is obvious.

Corollary 3.4.1 *Let $\mathcal{A}$ be the same as in Theorem 3.4.3. We have $0 < \rho(\mathcal{N} \times_n \mathbf{D}^{-1}) < 1$ for all n.*

Theorem 3.4.4 *Suppose that $\mathcal{A}$ and ρ are as defined in Theorem 3.4.3. Let $\mathcal{E} \in RT_{N,I}$ be symmetric such that $|\mathcal{E}| \le \eta|\mathcal{A}|$ and $0 < \eta < (1-\rho)/(1+\rho)$. If λ and λ' are the smallest eigenvalues of $\mathcal{A}$ and $\mathcal{A}+\mathcal{E}$, respectively, then*

$$\frac{|\lambda - \lambda'|}{\lambda} \le \frac{1+\rho}{1-\rho}\eta.$$

Proof It is obvious that $\mathcal{A} - \eta|\mathcal{A}| \le \mathcal{A} + \mathcal{E} \le \mathcal{A} + \eta|\mathcal{A}|$. Since $0 < \eta < (1-\gamma)/(1+\gamma)$, $\mathcal{A}+\mathcal{E}$ is also an irreducible symmetric nonsingular $\mathcal{M}$-tensor. Let $\lambda_1(\eta)$ and $\lambda_2(\eta)$ be the smallest eigenvalues of $\mathcal{A}+\eta|\mathcal{A}|$ and $\mathcal{A}-\eta|\mathcal{A}|$, respectively. We obtain $\lambda_2(\eta) \le \lambda' \le \lambda_1(\eta)$ and the algebraic multiplicity of $\lambda_1(\eta)$ and $\lambda_2(\eta)$ is 1.

For $t \in [-\eta, \eta]$, let $(\lambda_1(t); \mathbf{x}(t))$ be the smallest eigenpair of $\mathcal{A} + t|\mathcal{A}|$ with $\mathbf{x}(t) \in \mathbb{R}^I_{++}$ and $\lambda_1(t) > 0$, we have

$$\begin{aligned}\mathcal{A}(t) \times_2 \mathbf{x}(t)^\top \cdots \times_N \mathbf{x}(t)^\top &= \mathcal{A}(t) \times_1 \mathbf{x}(t)^\top \times_3 \mathbf{x}(t)^\top \cdots \times_N \mathbf{x}(t)^\top \\ &= \cdots = \mathcal{A}(t) \times_1 \mathbf{x}(t)^\top \cdots \times_{N-1} \mathbf{x}(t)^\top \\ &= \lambda_1(t)\mathbf{x}(t)^{[N-1]}.\end{aligned}$$

Let λ (> 0) be the smallest eigenvalue of $\mathcal{A}$. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^I_+$ such that

$$\mathcal{A}\mathbf{x}_i^{N-1} = \lambda \mathbf{x}_i^{[N-1]}, \quad i = 1, 2.$$

If $\mathbf{x}_1 = c\mathbf{x}_2$, where $c \neq 0$, then $\mathbf{x}_1$ and $\mathbf{x}_2$ are the same H-eigenvector. Hence we define

$$\delta = \min\{\|\mathbf{y} - \mathbf{x}\|_2 : \mathbf{y} \text{ and } \mathbf{x} \text{ are the different H-eigenvectors associated with } \lambda\}.$$

Over $\{\mathbf{z} \in \mathbb{C}^I : \|\mathbf{z} - \mathbf{x}\|_2 < \delta\}$, there exists a unique eigenvector $\mathbf{x}$ of $\mathcal{A}$ associated with λ. (For the smallest eigenvalue λ, if its geometric multiplicity is also 1, let $\delta \le \eta$)

For the smallest eigenvalue $\lambda_1(t)$ of $\mathcal{A}+t|\mathcal{A}|$, since $\lambda_1(t)$ is an analytic function with $|t| \le \eta$, there exists $\tilde{\delta} = \min\{\delta, \eta\}$ such that $\|\mathbf{x}(t) - \mathbf{x}\|_2 < \tilde{\delta}$ and $\mathbf{x}(t)$ is the unique eigenvector of $\mathcal{A} + t|\mathcal{A}|$ corresponding to $\lambda_1(t)$, with $0 \le t \le \eta$. According to Theorem 3.2.3 by differentiating this equation with respect to t, we obtain

$$|\mathcal{A}|\mathbf{x}(t)^{N-1} + \mathcal{A}(t)\tilde{\mathbf{x}}(t) = \dot{\lambda}_1(t)\mathbf{x}(t)^{[N-1]} + (N-1)\lambda_1(t)\hat{\mathbf{x}}(t),$$

where

$$\begin{cases} \mathcal{A}(t)\tilde{\mathbf{x}}(t) = \mathcal{A}(t) \times_2 \mathbf{x}'(t)^\top \times_3 \mathbf{x}(t)^\top \cdots \times_N \mathbf{x}(t)^\top \\ \quad + \cdots + \mathcal{A}(t) \times_2 \mathbf{x}(t)^\top \times_3 \mathbf{x}(t)^\top \cdots \times_N \mathbf{x}'(t)^\top) \\ (\hat{\mathbf{x}}(t))_i = x_1(t) \ldots x_{i-1}(t)\dot{x}_i(t)x_{i+1}(t) \ldots x_n(t), \quad i = 1, 2, \ldots, I. \end{cases}$$

Applying $\mathbf{x}(t)^\top$ with $\|\mathbf{x}(t)\|_N = 1$, we have

$$\dot{\lambda}_1(t) = |\mathcal{A}|\mathbf{x}(t)^N.$$

Consequently, we have

$$\frac{d}{dt}(\ln(\lambda_1(t))) = \frac{\dot{\lambda}_1(t)}{\lambda_1(t)} = \frac{(\mathcal{D} + \mathcal{N})\mathbf{x}(t)^N}{[(1+t)\mathcal{D} - (1-t)\mathcal{N}]\mathbf{x}(t)^m}. \tag{3.4.5}$$

It follows from Theorem 3.4.3 that

$$\frac{((1-t)\mathcal{N})\mathbf{x}(t)^N}{((1+t)\mathcal{D})\mathbf{x}(t)^N} \leq \rho(((1-t)\mathcal{N}) \times_n ((1+t)\mathbf{D})^{-1}),$$

i.e., $\mathcal{N}\mathbf{x}(t)^N/\mathcal{D}\mathbf{x}(t)^N \leq \rho(\mathcal{N} \times_n \mathbf{D}^{-1})$. So we obtain

$$\frac{d}{dt}(\ln(\lambda_1(t))) \leq \frac{(1 + \rho(\mathcal{N} \times_n \mathbf{D}^{-1}))}{(1 - \rho(\mathcal{N} \times_n \mathbf{D}^{-1})) + (1 + \rho(\mathcal{N} \times_n \mathbf{D}^{-1}))t}.$$

Noting that $\lambda_1(0) = \lambda$ and integrating (3.4.5) from 0 to η yields

$$\begin{aligned} \ln\lambda_1(\eta) - \ln\lambda_1 \leq & \ln((1 - \rho(\mathcal{N} \times_n \mathbf{D}^{-1})) + (1 + \rho(\mathcal{N} \times_n \mathbf{D}^{-1}))\eta) \\ & - \ln(1 - \rho(\mathcal{N} \times_n \mathbf{D}^{-1})), \end{aligned}$$

i.e., $(\lambda_1(\eta) - \lambda)/\lambda \leq \frac{1+\rho(\mathcal{N}\times_n\mathbf{D}^{-1})}{1-\rho(\mathcal{N}\times_n\mathbf{D}^{-1})}\eta$. Similarly, we have $(\lambda - \lambda_2(\eta))/\lambda \leq (1 + \rho(\mathcal{N} \times_n \mathbf{D}^{-1}))/(1 - \rho(\mathcal{N} \times_n \mathbf{D}^{-1}))\eta$, and

$$\frac{|\lambda - \lambda'|}{\lambda} \leq \frac{1 + \rho(\mathcal{N} \times_n \mathbf{D}^{-1})}{1 - \rho(\mathcal{N} \times_n \mathbf{D}^{-1})}\eta.$$

Finally, let $f(x) = (1+x)/(1-x)$ with $0 < x < 1$, then $f'(x) = 2/(x^2-2x+1)$. Since $f'(x) > 0$ over $0 < x < 1$, $f(x)$ is monotonically increasing. □

3.5 Numerical Examples

We verify the inequalities related to the Perron pair of any irreducible symmetric nonnegative tensor in $RT_{N,I}$ in the first part of Theorems 3.4.1, 3.4.3 and Corollary 3.4.1. Then we illustrate the perturbation bounds via several test tensors: Theorem 3.4.4 for the smallest eigenvalue of any irreducible symmetric nonsingular $\mathcal{M}$-tensor in $RT_{N,I}$, Corollaries 3.2.1 and 3.2.2 for the Z-spectral radius and the spectral radius of any irreducible symmetric nonnegative tensor in $RT_{N,I}$, respectively, and Corollary 3.2.3 for the modulus largest singular value of any irreducible nonnegative tensor in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$.

3.5.1 Verification of Inequalities

We compute the approximate Perron pair of a given irreducible symmetric tensor by the NQZ method [31].

Example 3.5.1 We verify Theorem 3.4.1 by 100 randomly generated irreducible symmetric nonnegative tensors. All entries of the test tensors are uniformly distributed over (0, 1). Let $\mathbf{D} \in \mathbb{R}^{I \times I}$ be diagonal and positive definite. For a given test tensor, we select $\mathbf{D} = \text{diag}(\mathbf{d})$, where $\mathbf{d} \in \mathbb{R}^I_{++}$ satisfies $s_1 \neq s_2 \neq \cdots \neq s_I$, or $\mathbf{D} = \alpha \mathbf{I}_I$ with $\alpha > 0$.

Let $\mathcal{A} \in RT_{4,10}$ be irreducible and symmetric nonnegative, $(\rho(\mathcal{A}); \mathbf{x})$ be the Perron pair of $\mathcal{A}$.

For each part of Fig. 3.3, the vertical axis represents the value of $\hat{\mathcal{A}}\mathbf{x}^4 - \mathcal{A}\mathbf{x}^4$, where $\hat{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{D}^{-3} \times_2 \mathbf{D} \times_3 \mathbf{D} \times_4 \mathbf{D}$ and the diagonal matrix $\mathbf{D} \in \mathbb{R}^{10 \times 10}$ is

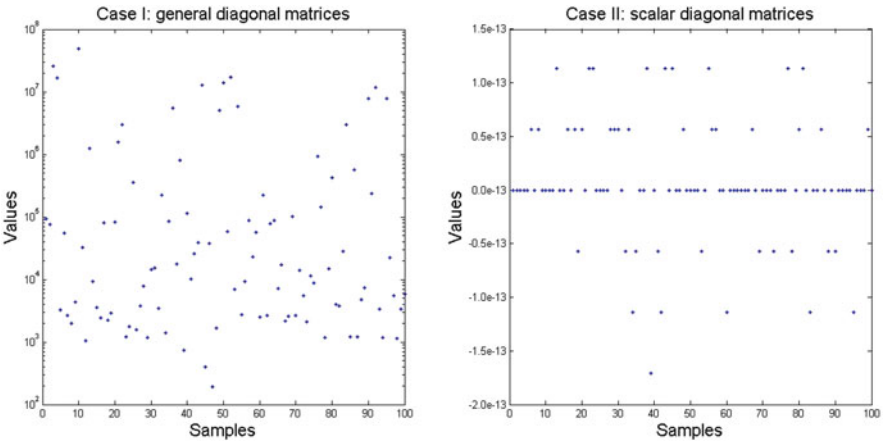

Fig. 3.3 Verification of Theorem 3.4.1 with irreducible symmetric nonnegative tensors

positive definite, and the horizontal axis is the indices of the test tensors. For the left of Fig. 3.3, the main diagonal entries of $\mathbf{D}$ are different, but for the right of Fig. 3.3, the main diagonal entries of $\mathbf{D}$ are the same. In Fig. 3.3, $\hat{\mathcal{A}}\mathbf{x}^4 \geq \mathcal{A}\mathbf{x}^4$ holds for any positive diagonal matrix $\mathbf{D} \in \mathbb{R}^{10\times 10}$.

Example 3.5.2 ([12, Example 2]) Consider the symmetric tensor $\mathcal{A} \in RT_{3,3}$ such that

$$\begin{cases} a_{111} = 0.0517,\ a_{112} = 0.3579,\ a_{113} = 0.5298, a_{122} = 0.7544,\ a_{123} = 0.2156, \\ a_{133} = 0.3612,\ a_{222} = 0.3943,\ a_{223} = 0.0146, a_{233} = 0.6718,\ a_{333} = 0.9723. \end{cases}$$

By the NQZ method, the approximated Perron pair of $\mathcal{A}$ is calculated by

$$\rho(\mathcal{A}) = 3.6167; \quad \mathbf{u} = (0.6674, 0.6655, 0.7417)^\top.$$

Let $\mathbf{D} \in \mathbb{R}^{3\times 3}$ be positive diagonal, we define

$$d_1 = (\mathcal{A} \times_1 \mathbf{D}^{-2} \times_2 \mathbf{D} \times_3 \mathbf{D} - \mathcal{A})\mathbf{u}^3,\ d_2 = (\mathcal{A} \times_1 \mathbf{D}^2 \times_2 \mathbf{D}^{-1} \times_3 \mathbf{D}^{-1} - \mathcal{A})\mathbf{u}^3.$$

For 100 randomly generated positive diagonal matrices $\mathbf{D} \in \mathbb{R}^{3\times 3}$, d_1 and d_2 are shown in Fig. 3.4.

According to Fig. 3.4, when $\mathbf{D} = \text{diag}(\mathbf{s})$, where $\mathbf{s} \in \mathbb{R}^I_{++}$ satisfies $s_1 \neq s_2 \neq \cdots \neq s_I$, d_1 and d_2 are positive; and when $\mathbf{D} = \alpha\mathbf{I}_I$ with any given $\alpha > 0$, d_1 and d_2 are approximately zero.

To verify Corollary 3.4.1, we need a strategy for computing the smallest eigenvalue of an irreducible and symmetric nonsingular $\mathcal{M}$-tensor $\mathcal{A} \in RT_{N,I}$. It follows from [15, 28] that $\mathcal{A} = s\mathcal{I} - \mathcal{B}$ with $s > \rho(\mathcal{B})$. If $(\rho(\mathcal{B}); \mathbf{x})$ is a Perron pair

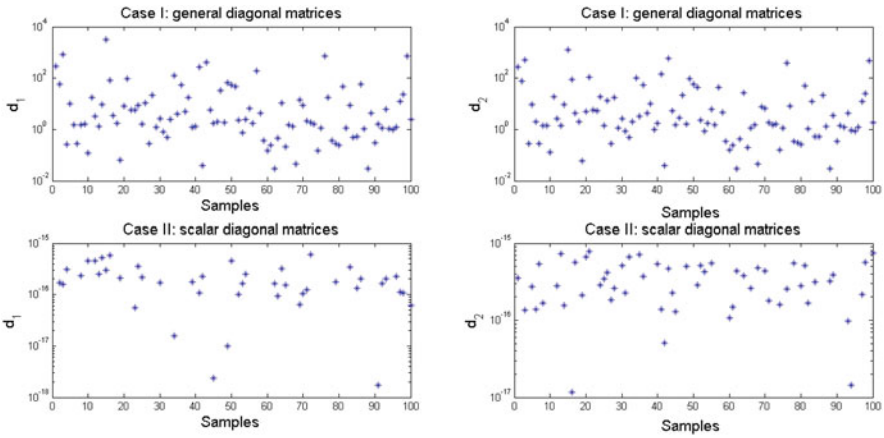

Fig. 3.4 Verification of Theorem 3.4.1 with positive diagonal matrices

of $\mathcal{B}$, then $(s - \rho(\mathcal{B}); \mathbf{x})$ is an eigenpair, corresponding to the smallest eigenvalue of $\mathcal{A}$.

$\mathcal{A}$ can be expressed as $\mathcal{A} = \mathcal{D} - \mathcal{N}$, where $\mathcal{D} \in RT_{N,I}$ is diagonal and the symmetric nonnegative tensor $\mathcal{N} \in RT_{N,I}$ has zero main diagonal. In this form, we cannot use the NQZ method to compute the smallest eigenvalue of $\mathcal{A}$. However, we can convert $\mathcal{A}$ to the form of $s\mathcal{I} - \mathcal{B}$ through as follows:

1. Compute the spectrum radius $\rho(\mathcal{N})$ of $\mathcal{N}$ by the NQZ method;
2. Find the largest element in $\mathcal{D}$ denoted as $d > 0$;
3. Let $s = d + \rho(\mathcal{N})$ and construct a new diagonal tensor $\mathcal{D}_1$ with its main diagonal element s;
4. Form $\mathcal{A} = s\mathcal{I} - \mathcal{B}$ such that $\mathcal{B} = \mathcal{D}_1 - \mathcal{D} + \mathcal{N}$.

Example 3.5.3 We verify Theorem 3.4.3 and Corollary 3.4.1 by 100 randomly generated irreducible symmetric nonsingular $\mathcal{M}$-tensors. Each test tensor is generated as follows:

(1) Generate a symmetric nonnegative tensor $\mathcal{B}$ such that whose entries are uniformly distributed over (0, 1);
(2) Compute the spectral radius $\rho(\mathcal{B})$ of $\mathcal{B}$ by the NQZ method; and
(3) Form $\mathcal{A} = (\rho(\mathcal{B}) + \text{rand}(1))\mathcal{I} - \mathcal{B}$, where the MATLAB function rand generates a random value uniformly distributed over (0, 1).

We select $\mathcal{A} \in RT_{4,10}$ being an irreducible symmetric nonsingular $\mathcal{M}$-tensors. Let $(\lambda; \mathbf{x})$ be an eigenpair, corresponding to the smallest eigenvalue of $\mathcal{A}$.

For each part in Fig. 3.5, "Values" means the value of $\rho(\mathcal{N} \times_k \mathbf{D}^{-1})\mathcal{D}\mathbf{x}^4 - \mathcal{N}\mathbf{x}^4$, where $(\rho(\mathcal{N} \times_k \mathbf{D}^{-1}); \mathbf{x})$ is computed by the NQZ algorithm and $\mathbf{D}$ is a diagonal matrix, with main diagonal entries equal to the main diagonal entries of $\mathcal{A}$. For each part in Fig. 3.6, "Values" display $1 - \rho(\mathcal{N} \times_k \mathbf{D}^{-1})$, where $k = 1, 2, 3, 4$.

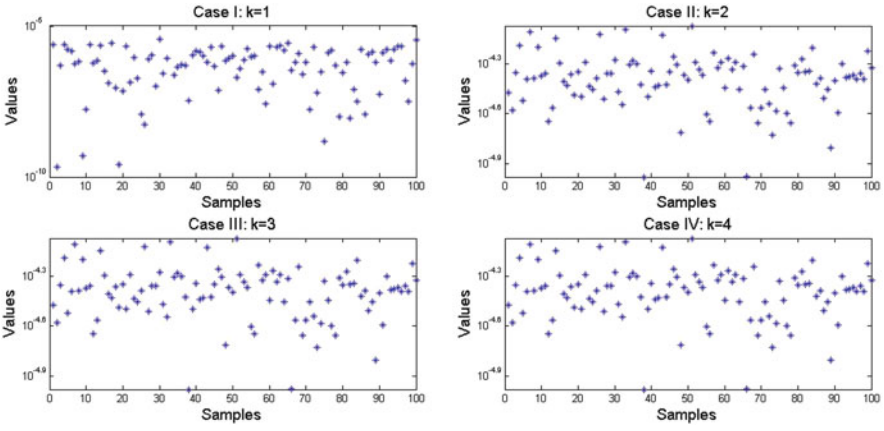

Fig. 3.5 Verification of Theorem 3.4.3 with irreducible symmetric nonsingular $\mathcal{M}$-tensors

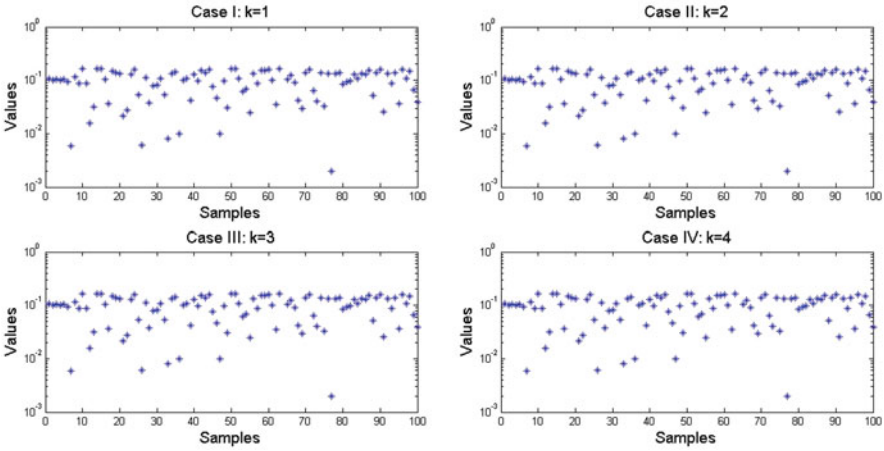

Fig. 3.6 Verification of Corollary 3.4.1 with irreducible symmetric nonsingular $\mathcal{M}$-tensors

3.5.2 *Verification of Perturbation Bounds*

In the following three examples, let $\mathcal{A} \in RT_{4,10}$ be symmetric. Suppose that $\mathcal{E} \in RT_{4,10}$ is symmetric such that $|\mathcal{E}| \leq \varepsilon|\mathcal{A}|$ for $\varepsilon \in (0, 1)$. For a given $\mathcal{A}$, we consider these two cases: $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$ and $\mathcal{E} = \varepsilon\mathcal{A}$.

Example 3.5.4 Let $\mathcal{A} = s\mathcal{I} - \mathcal{B} \in RT_{4,10}$ be an irreducible symmetric nonsingular $\mathcal{M}$-tensor, where $\mathcal{B} \in RT_{4,10}$ is a randomly generated irreducible symmetric nonnegative tensors and $s = \rho(\mathcal{B}) + 100$. The entries of $\mathcal{B}$ are uniformly distributed over (0, 1). We verify Theorem 3.4.4 with different symmetric perturbation tensors $\mathcal{E} \in RT_{4,10}$.

Let $(\lambda; \mathbf{x})$ be an eigenpair, corresponding to the smallest eigenvalue λ of $\mathcal{A}$. We use $\hat{\lambda}_+$ and $\hat{\lambda}_-$ to denote the smallest eigenvalues of $\mathcal{A} + \mathcal{E}$ and $\mathcal{A} - \mathcal{E}$, respectively. The values of $|\hat{\lambda}_+ - \lambda|/\lambda$ and $|\hat{\lambda}_- - \lambda|/\lambda$ are denoted by δ_+ and δ_-, respectively. Similar notations are used in the examples below.

Let $\varepsilon \in \text{linspace}(1e-5, 1-1e-5, 50)(1-\rho)/(1+\rho)$ and $\varepsilon(\rho) = (1+\rho)/(1-\rho)\varepsilon$. For two different $\alpha, \beta \in \mathbb{R}$ and a positive integer K, linspace(α, β, K) generates K equally spaced points between α and β. When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, the relationship between $\varepsilon(\rho)$ and δ_+, or $\varepsilon(\rho)$ and δ_-, are shown in Fig. 3.7; for $\mathcal{E} = \varepsilon\mathcal{A}$, see Fig. 3.8.

Example 3.5.5 $\mathcal{A} \in RT_{4,10}$ is an irreducible symmetric nonnegative tensor, whose entries are uniformly distributed over (0, 1). We verify Corollaries 3.2.1 and 3.2.2 with different symmetric perturbation tensors $\mathcal{E} \in RT_{4,10}$.

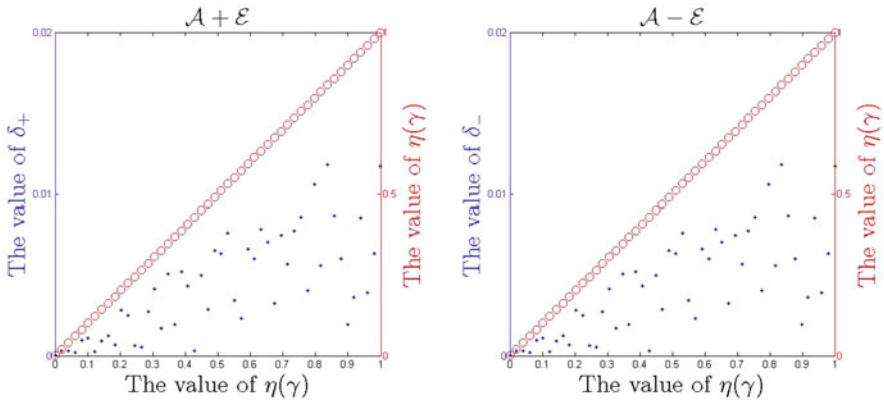

Fig. 3.7 When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, verification of Theorem 3.4.4 with different ε. Left for $\varepsilon(\rho)$ and δ_+; right for $\varepsilon(\rho)$ and δ_-

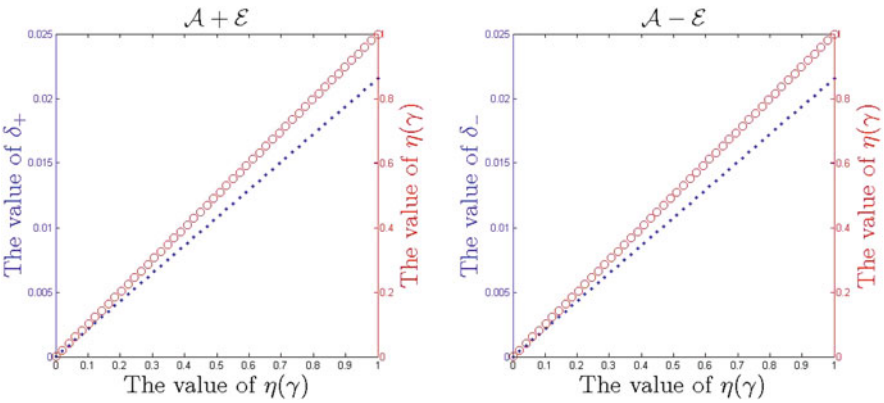

Fig. 3.8 When $\mathcal{E} = \varepsilon\mathcal{A}$, verification of Theorem 3.4.4 with different ε. Left for $\varepsilon(\rho)$ and δ_+; right for $\varepsilon(\rho)$ and δ_-

We compare the perturbation bounds of the Z-spectral radius of $\mathcal{A}$. Let $(\varrho; \mathbf{x})$ be the Z-spectral pair of $\mathcal{A}$. We use $\hat{\varrho}_+$ and $\hat{\varrho}_-$ to denote the Z-spectral radius of $\mathcal{A}+\mathcal{E}$ and $\mathcal{A}-\mathcal{E}$, respectively. Let $\varepsilon \in \text{linspace}(0.001, 0.1, 50)$. When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, the relationship between ε and δ_+, or ε and δ_-, are shown in Fig. 3.9; for $\mathcal{E} = \varepsilon\mathcal{A}$, see Fig. 3.10.

We also consider the perturbation bound of the spectral radius of $\mathcal{A}$. Let $(\rho; \mathbf{x})$ be the Perron pair of $\mathcal{A}$. We denote $\hat{\rho}_+$ and $\hat{\rho}_-$ the Z-spectral radiuses of $\mathcal{A}+\mathcal{E}$ and $\mathcal{A}-\mathcal{E}$, respectively. Let $\varepsilon = \text{linspace}(0.001, 0.999, 50)$. When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, the

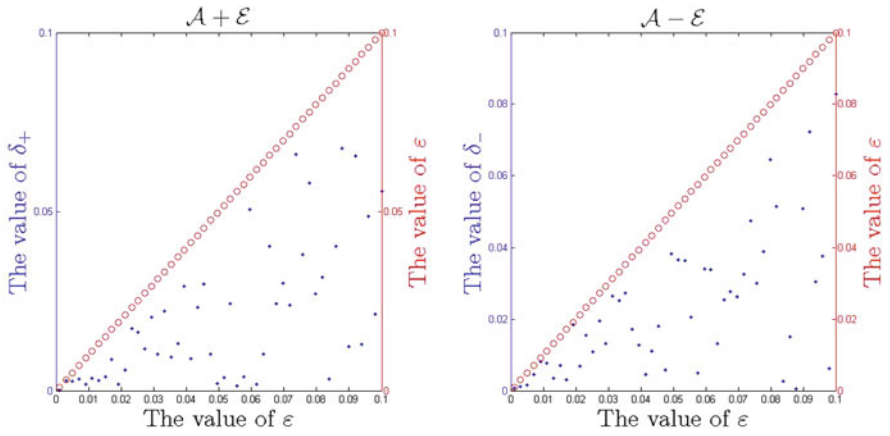

Fig. 3.9 When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, verification of Corollary 3.2.1 with different ε. Left for ε and δ_+; right for ε and δ_-

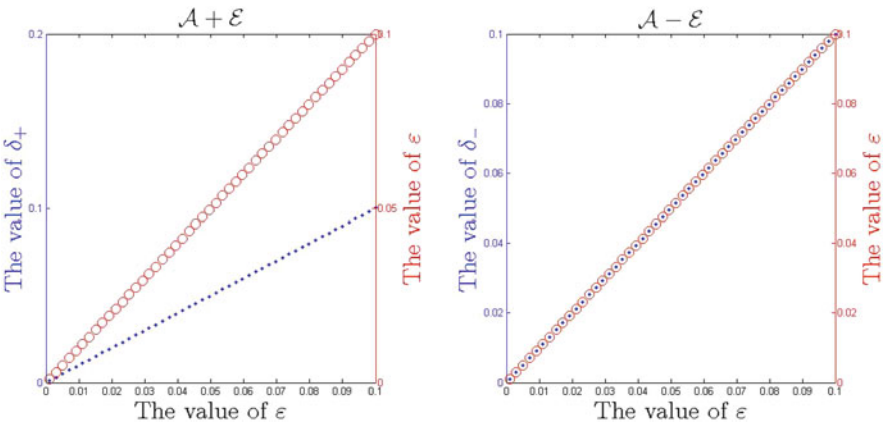

Fig. 3.10 When $\mathcal{E} = \varepsilon\mathcal{A}$, verification of Corollary 3.2.1 with different ε. Left for ε and δ_+; right for ε and δ_-

relationship between ε and δ_+, or ε and δ_-, are shown in Fig. 3.11; for $\mathcal{E} = \varepsilon\mathcal{A}$, see Fig. 3.12.

Example 3.5.6 Suppose that $\mathcal{A} \in \mathbb{R}_+^{I_1\times I_2\times\cdots\times I_N}$ is irreducible, whose entries are uniformly distributed over $(0, 1)$. We verify Corollary 3.2.3 with different symmetric perturbation tensors $\mathcal{E} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, where $|\mathcal{E}| \leq \varepsilon|\mathcal{A}|$ with $0 < \varepsilon < 1$.

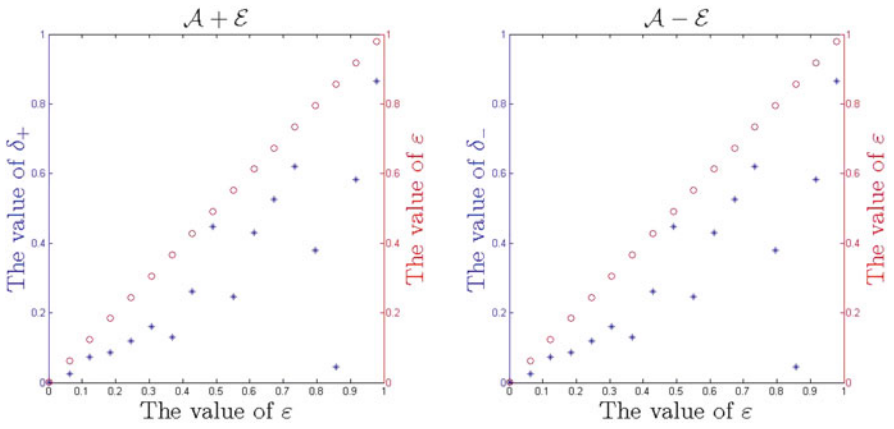

Fig. 3.11 When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, verification of Corollary 3.2.2 with different ε. Left for ε and δ_+; right for ε and δ_-

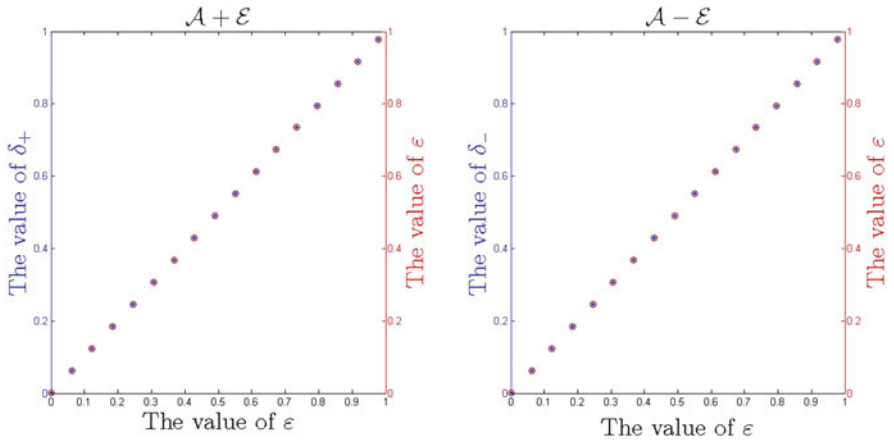

Fig. 3.12 When $\mathcal{E} = \varepsilon\mathcal{A}$, verification of Corollary 3.2.2 with different ε. Left for ε and δ_+; right for ε and δ_-

Let $N = 4$ and $\{I_1, I_2, I_3, I_4\} = \{10, 12, 14, 16\}$. For a given $\mathcal{A}$, we consider two cases: $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$ and $\mathcal{E} = \varepsilon\mathcal{A}$. Let σ be the largest singular value in modulus of $\mathcal{A}$. We use $\hat{\sigma}_+$ and $\hat{\sigma}_-$ to denote the Z-spectral radiuses of $\mathcal{A}+\mathcal{E}$ and $\mathcal{A}-\mathcal{E}$, respectively. Let $\varepsilon = \text{linspace}(0.001, 0.1, 50)$. When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, the relationship between ε and δ_+, or ε and δ_-, are shown in Fig. 3.13; for $\mathcal{E} = \varepsilon\mathcal{A}$, see Fig. 3.14.

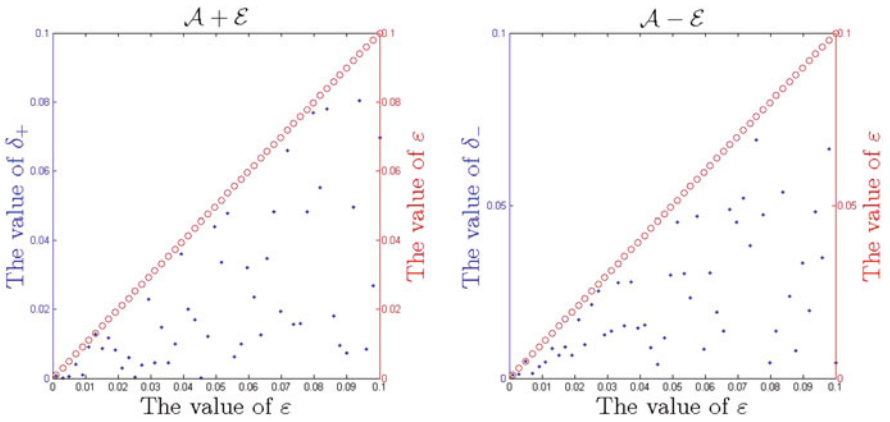

Fig. 3.13 When $\mathcal{E} = \text{rand}(1)\varepsilon\mathcal{A}$, verification of Corollary 3.2.3 with different ε. Left for ε and δ_+; right for ε and δ_-

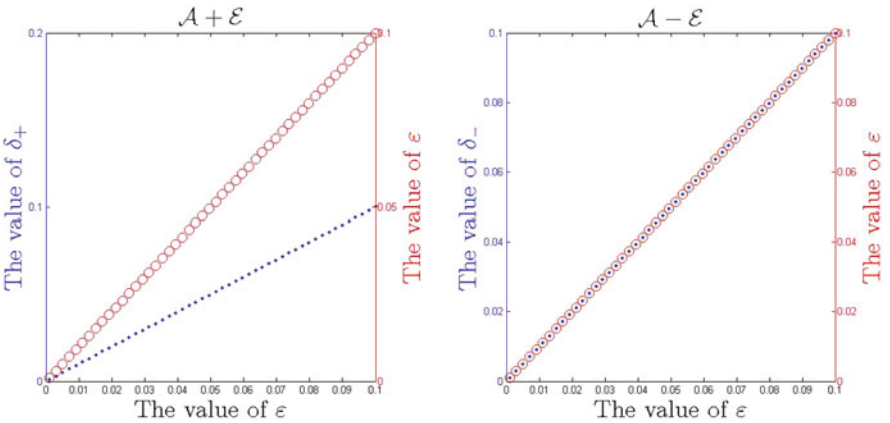

Fig. 3.14 When $\mathcal{E} = \varepsilon\mathcal{A}$, verification of Corollary 3.2.3 with different ε. Left for ε and δ_+; right for ε and δ_-

References

1. G. Stewart, J. Sun, *Matrix Perturbation Theory*. Computer Science and Scientific Computing (Academic Press, Boston, 1990)
2. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)
3. J. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford University, Oxford, 1965)
4. B. Kågström, A. Ruhe (eds.) *Matrix Pencils*. Lecture Notes in Mathematics, vol. 973 (Springer, New York, 1983)
5. E.K.-W. Chu, Exclusion theorems and the perturbation analysis of the generalized eigenvalue problem. SIAM J. Numer. Anal. **24**(5), 1114–1125 (1987)

6. N. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd edn. (Society for Industrial and Applied Mathematics, Philadelphia, 2002)
7. D. Kressner, *Numerical Methods for General and Structured Eigenvalue Problems* (Springer, New York, 2005)
8. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
9. L. Lim, Singular values and eigenvalues of tensors: a variational approach, in *IEEE CAMSAP 2005: First International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (IEEE, Piscataway, 2005), pp. 129–132
10. L. De Lathauwer, First-order perturbation analysis of the best rank-(r_1, r_2, r_3) approximation in multilinear algebra. J. Chemometrics **18**(1), 2–11 (2004)
11. L. Qi, *Transposes, l-Eigenvalues and Invariants of Third Order Tensors* (2017). ArXiv preprint:1704.01327v5
12. L. Qi, The best rank-one approximation ratio of a tensor space. SIAM J. Matrix Anal. Appl. **32**(2), 430–442 (2011)
13. S. Hu, Z. Huang, C. Ling, L. Qi, On determinants and eigenvalue theory of tensors. J. Symb. Comput. **50**, 508–531 (2013)
14. K. Chang, K. Pearson, T. Zhang, Perron-Frobenius theorem for nonnegative tensors. Commun. Math. Sci. **6**(2), 507–520 (2008)
15. W. Ding, L. Qi, Y. Wei, $\mathcal{M}$-tensors and nonsingular $\mathcal{M}$-tensors. Linear Algebra Appl. **439**(10), 3264–3278 (2013)
16. S. Ragnarsson, C. Van Loan, Block tensors and symmetric embeddings. Linear Algebra Appl. **438**(2), 853–874 (2013)
17. Z. Chen, L. Lu, A tensor singular values and its symmetric embedding eigenvalues. J. Comput. Appl. Math. **250**(1), 217–228 (2013)
18. J. Demmel, *Applied Numerical Linear Algebra* (Society for Industrial and Applied Mathematics, Philadelphia, 1997)
19. C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable* (American Mathematical Society, New York, 1951)
20. T. Zhang, G. Golub, Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. **23**(2), 534–550 (2001)
21. D. Cartwright, B. Sturmfels, The number of eigenvalues of a tensor. Linear Algebra Appl. **438**(2), 942–952 (2013)
22. T. Kolda, J. Mayo, Shifted power method for computing tensor eigenpairs. SIAM J. Matrix Anal. Appl. **32**(4), 1095–1124 (2011)
23. L. Ahlfors, *Complex Analysis: An Introduction of the Theory of Analytic Functions of One Complex Variable* (McGraw-Hill Book Co., New York, 1966)
24. I. Gohberg, I. Koltracht, Mixed, componentwise, and structured condition numbers. SIAM J. Matrix Anal. Appl. **14**(3), 688–704 (1993)
25. W. Li, M. Ng, Some bounds for the spectral radius of nonnegative tensors. Numer. Math. **130**(2), 315–335 (2015)
26. Y. Yang, Q. Yang, Further results for Perron-Frobenius theorem for nonnegative tensors. SIAM J. Matrix Anal. Appl. **31**(5), 2517–2530 (2010)
27. L. Elsner, I. Koltracht, M. Neumann, D. Xiao, On accurate computations of the Perron root. SIAM J. Matrix Anal. Appl. **14**(2), 456–467 (1993)
28. L. Zhang, L. Qi, G. Zhou, M-tensors and some applications. SIAM J. Matrix Anal. Appl. **35**(2), 437–452 (2014)
29. B. Bader, T. Kolda, *Matlab Tensor Toolbox Version 2.6* (2015). http://www.sandia.gov/~tgkolda/TensorToolbox/
30. J. Xue, Computing the smallest eigenvalue of an M-matrix. SIAM J. Matrix Anal. Appl. **17**(4), 748–762 (1996)
31. M. Ng, L. Qi, G. Zhou, Finding the largest eigenvalue of a nonnegative tensor. SIAM J. Matrix Anal. Appl. **31**(3), 1090–1099 (2009)

Chapter 4
Tensor Complementarity Problems

4.1 Preliminaries

The research of finite-dimensional variational inequality and complementarity problems [1–6] have been rapidly developed in the theory of existence, uniqueness and sensitivity of solutions, algorithms, and the application of these techniques to transportation planning, regional science, socio-economic analysis, energy modeling and game theory.

Linear complementarity problems can be solved by the algorithm [7]. The criss-cross algorithm terminates only if the associated matrix is a sufficient matrix [7]. A sufficient matrix is a generalization both of a positive definite matrix [8, Section 4.2] and a P-matrix [9], whose principal minors are positive. Qi [10] gives the definition of the symmetric positive definite tensor and derives a method for checking whether a symmetric tensor is positive definite. The concept of copositive matrices [11] is important in applied mathematics, with applications in control theory, optimization modeling, linear complementarity problems, graph theory and linear evolution variational inequalities [12]. Qi [13] extends the concept to tensors.

The tensor complementarity problem is a class of nonlinear complementarity problems with the involved function being defined by a tensor, which is also a direct and natural extension of the linear complementarity problem. In the last few years, the tensor complementarity problem has attracted a lot of attention, and has been studied extensively, from theory to solution methods and applications. Recently, Huang and Qi [14–16] survey the state-of-the-art of studies for the tensor complementarity problem and related models. Wang et al. [17] introduce tensor variational inequalities, which are a natural extension of the affine variational inequality and the tensor complementarity problem.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_4

4.1.1 Notation and Definitions

Given a mapping $F : \mathbb{X} \subset \mathbb{R}^K \to \mathbb{R}^I$, we assume that $F(\mathbf{x}) \in \mathbb{R}^I$ is a vector for $\mathbf{x} \in \mathbb{X}$. Our next definition is motivated by the class of copositive matrices [9].

Definition 4.1.1 ([3]) A mapping $F : \mathbb{X} \to \mathbb{R}^I$ is said to be

(a) copositive with respect to $\mathbb{X}$ if and only if for all $\mathbf{x} \in \mathbb{X}$, we have

$$\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)) \geq 0;$$

(b) strictly copositive with respect to $\mathbb{X}$ if and only if for all nonzero vectors $\mathbf{x} \in \mathbb{X}$, we have

$$\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)) > 0;$$

(c) strongly copositive with respect to $\mathbb{X}$ if and only if there exists a scalar $\alpha > 0$ such that for all nonzero $\mathbf{x} \in \mathbb{X}$, we have

$$\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)) \geq \alpha \|\mathbf{x}\|_2^2.$$

The mapping

$$G(\mathbf{x}) = F(\mathbf{x}) - F(\mathbf{0}_I) \tag{4.1.1}$$

plays a fundamental role in the nonlinear complementarity problem, motivated by the linear version. The strict copositivity of a mapping can be relaxed through the introduction of the class of **d**-regular mappings with a given positive $\mathbf{d} \in \mathbb{R}^I$.

Definition 4.1.2 ([3]) For any vector $\mathbf{x} \in \mathbb{R}^I_+$, we define the index sets

$$\mathbb{I}_+(\mathbf{x}) = \{i : x_i > 0\}, \quad \mathbb{I}_0(\mathbf{x}) = \{i : x_i = 0\}.$$

Let $\mathbf{d} \in \mathbb{R}^I$ be an arbitrary positive vector. A mapping $G : \mathbb{R}^I \to \mathbb{R}^I$ is said to be **d**-regular, if the following system of nonlinear equations has no solution $(\mathbf{x}, t) \in \mathbb{R}^I_+ \times \mathbb{R}_+$ except for $\mathbf{x} = \mathbf{0}_I$,

$$\begin{cases} G_i(\mathbf{x}) + t d_i = 0, & i \in \mathbb{I}_+(\mathbf{x}), \\ G_i(\mathbf{x}) + t d_i \geq 0, & i \in \mathbb{I}_0(\mathbf{x}). \end{cases} \tag{4.1.2}$$

Equivalently, G is **d**-regular if, for any scalar $r > 0$, the augmented nonlinear complementarity problem NCP(H) defined by $H : \mathbb{R}^{I+1} \to \mathbb{R}^{I+1}$,

$$H \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix} = \begin{pmatrix} G(\mathbf{x}) + t\mathbf{d} \\ r - \mathbf{d}^\top \mathbf{x} \end{pmatrix},$$

has no solution $(\mathbf{x}, t) \in \mathbb{R}^I_+ \times \mathbb{R}_+$ except for $\mathbf{x} = \mathbf{0}_I$.

The definition of the diagonalizable tensors [18, 19] is similar to the diagonalizable matrices [8] as follows.

Definition 4.1.3 Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric. It is diagonalizable if and only if $\mathcal{A} = \mathcal{D} \times_1 \mathbf{B} \times_2 \mathbf{B} \cdots \times_N \mathbf{B}$, where $\mathbf{B} \in \mathbb{R}^{I \times I}$ is nonsingular and $\mathcal{D}$ is a diagonal tensor. Denote the set of all diagonalizable tensors by $D_{N,I}$.

In particular, in the case of $N = 2$, $\mathcal{A}$ is congruent to $\mathcal{D}$ ([20] for more details on the congruent transformation).

4.2 Lemmas and Problem Description

For $F : \mathbb{R}^I \to \mathbb{R}^I$, the corresponding nonlinear complementarity problem NCP(F) is to find a vector $\mathbf{x}^* \in \mathbb{R}^I_+$ such that

$$F(\mathbf{x}^*) \in \mathbb{R}^I_+, \quad (\mathbf{x}^*)^\top F(\mathbf{x}^*) = 0.$$

If $F(\mathbf{x})$ is an affine function of $\mathbf{x}$, $F(\mathbf{x}) = \mathbf{q} + \mathbf{M}\mathbf{x}$ for $\mathbf{q} \in \mathbb{R}^I$ and $\mathbf{M} \in \mathbb{R}^{I \times I}$, NCP($F$) reduces to the linear complementarity problem LCP($\mathbf{q}$, $\mathbf{M}$). Some results of the linear complementarity problem can be found in [9, 21].

A further generalization of NCP(F) is the variational inequality, given a mapping $F : \mathbb{R}^I \to \mathbb{R}^I$ and $\emptyset \neq \mathbb{K} \subseteq \mathbb{R}^I$, find a $\mathbf{x}^* \in \mathbb{K}$ satisfying

$$(\mathbf{y} - \mathbf{x}^*)^\top F(\mathbf{x}^*) \geq 0,$$

for all $\mathbf{y} \in \mathbb{K}$, abbreviated as VI($\mathbb{K}$, F). If $\mathbb{K} = \{\mathbf{x} : \mathbf{x} \geq \mathbf{0}_I\}$, then a solution $\mathbf{x}^*$ of VI($\mathbb{K}$, F) solves the NCP(F).

It is well-known that $\mathbf{A} \in \mathbb{R}^{I \times I}$ is a P-matrix [9] if and only if the following linear complementarity problem has a unique solution for all $\mathbf{q} \in \mathbb{R}^I$: find $\mathbf{z} \in \mathbb{R}^I$ such that

$$\mathbf{z} \geq \mathbf{0}_I, \quad \mathbf{q} + \mathbf{A}\mathbf{z} \geq \mathbf{0}_I, \quad \mathbf{z}^\top(\mathbf{q} + A\mathbf{z}) = 0.$$

Given a P-tensor[1] [22] $\mathcal{A} \in RT_{N,I}$ ($N > 2$), for $\mathbf{q} \in \mathbb{R}^I$, the following nonlinear complementarity problem has a unique solution: find $\mathbf{x} \in \mathbb{R}^I$ such that

$$\mathbf{x} \geq \mathbf{0}_I, \quad \mathbf{q} + \mathcal{A}\mathbf{x}^{N-1} \geq \mathbf{0}_I, \quad \mathbf{x}^\top(\mathbf{q} + \mathcal{A}\mathbf{x}^{N-1}) = 0.$$

[1]For a given tensor $\mathcal{A} \in RT_{N,I}$, if $\max_{i=1,2,\ldots,I} x_i(\mathcal{A}\mathbf{x})_i > 0$ with all nonzero $\mathbf{x} \in \mathbb{R}^I$, then $\mathcal{A}$ is a P-tensor.

We consider a special kind of NCP(F), where $F_i(\mathbf{x})$ is a multivariate polynomial and the degree of $F_i(\mathbf{x})$ is K_i. Note that $F(\mathbf{x})$ can be expressed by

$$F(\mathbf{x}) = \sum_{k=1}^{K} \mathcal{A}_k \mathbf{x}^{k-1}, \quad K = \max_{1 \le l \le I} K_l.$$

where $\mathcal{A}_k \in RT_{k,I}$ with $k \ge 3$, $\mathcal{A}_1 \in \mathbb{R}^I$ and $\mathcal{A}_2 \in \mathbb{R}^{I \times I}$.

4.2.1 Lemmas

The following lemma is an existence and uniqueness theorem by Cottle [1]. It involves the notion of positively bounded Jacobians, and the original proof is constructive with an algorithm to compute the unique solution.

Lemma 4.2.1 ([1, 3]) *Let $F : \mathbb{R}^I_+ \to \mathbb{R}^I$ be continuously differentiable and $\delta \in (0, 1)$, such that all principal minors of the Jacobian matrix $\nabla F(\mathbf{x})$ are bounded between δ and δ^{-1}, for all $\mathbf{x} \in \mathbb{R}^I_+$. Then* NCP($F$) *has a unique solution.*

If $F(\cdot)$ is strictly copositive, then the following result holds.

Lemma 4.2.2 ([5]) *Let $F : \mathbb{R}^I_+ \to \mathbb{R}^I$ be continuous and strictly copositive with respect to $\mathbb{R}^I_+$. If there exists a mapping $c : \mathbb{R}_+ \to \mathbb{R}$ such that $c(\lambda) \to \infty$ as $\lambda \to \infty$, and for all $\lambda \ge 1$ and $\mathbf{x} \ge \mathbf{0}_I$,*

$$\mathbf{x}^\top (F(\lambda \mathbf{x}) - F(\mathbf{0}_I)) \ge c(\lambda) \mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)), \tag{4.2.1}$$

then NCP(F) *has a nonempty, compact solution set.*

For the linear complementarity problem, $G(\cdot) = F(\cdot) - F(\mathbf{0}_I)$, given in (4.1.1), is obviously linear. Thus condition (4.2.1) is satisfied with $c(\lambda) = \lambda$. More generally, the same condition will hold with $c(\lambda) = \lambda^\alpha$, if G is positively homogeneous of degree $\alpha > 0$; i.e., if $G(\lambda \mathbf{x}) = \lambda^\alpha G(\mathbf{x})$ for $\lambda > 0$.

If $F(\cdot)$ is strictly copositive with respect to $\mathbb{R}^I_+$, then $G(\cdot)$ from (4.1.1) is $\mathbf{d}$-regular for any $\mathbf{d} > \mathbf{0}_I$. The following lemma presents an existence result for NCP(F) with a $\mathbf{d}$-regular mapping $G(\cdot)$.

Lemma 4.2.3 ([4]) *Let $F(\cdot)$ be a continuous mapping from $\mathbb{R}^I$ into itself and $G(\cdot)$ defined by (4.1.1). Suppose that $G(\cdot)$ is positively homogeneous of degree $\alpha > 0$ and G is $\mathbf{d}$-regular for some $\mathbf{d} > \mathbf{0}_I$. Then* NCP($F$) *has a nonempty, compact solution set.*

The main characterization theorem for copositive tensors is quoted.

Lemma 4.2.4 ([13, Theorem 5]) *Let $\mathcal{A} \in RT_{N,I}$ be symmetric. Then $\mathcal{A}$ is copositive if and only if*

$$\min\left\{\mathcal{A}\mathbf{x}^N : \mathbf{x} \in \mathbb{R}^I_+, \ \sum_{i=1}^{I} x_i^N = 1\right\} \geq 0.$$

Also, $\mathcal{A}$ is strictly copositive if and only if

$$\min\left\{\mathcal{A}\mathbf{x}^N : \mathbf{x} \in \mathbb{R}^I_+, \ \sum_{i=1}^{I} x_i^N = 1\right\} > 0.$$

4.2.2 Problem Description

Next, we present two research problems.

Problem 4.2.1 ([22]) Given $\mathcal{A} \in RT_{N,I}$ and $\mathbf{q} \in \mathbb{R}^I$. The NCP($\mathbf{q}, \mathcal{A}$) seeks a vector $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$F(\mathbf{x}) = \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q} \in \mathbb{R}^I_+, \quad \mathcal{A}\mathbf{x}^N + \mathbf{x}^\top \mathbf{q} = 0.$$

Problem 4.2.2 For a given even integer $N > 0$, let $\mathcal{A}_k \in RT_{N-(2k-2),I}$ and $\mathbf{q} \in \mathbb{R}^I$ with $k = 1, 2, \ldots, N/2$. The NCP($\mathbf{q}, \{\mathcal{A}_k\}$) seeks a vector $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$F(\mathbf{x}) = \sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}^{N-(2k-1)} + \mathbf{q} \in \mathbb{R}^I_+, \quad \sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}^{N-2k+2} + \mathbf{x}^\top \mathbf{q} = 0,$$

where $\mathcal{A}_{N/2} \in \mathbb{R}^{I \times I}$.

We call Problem 4.2.1 the tensor complementarity problem (TCP), which is a natural generalization of LCP and a special case of NCP. It is a new topic, inspired by the growing interests in structured tensors. The TCP [14–16] is widely used in nonlinear compressing sensing, commutations, DNA microarrays and multi-person games. Problem 4.2.2 is a special case of the polynomial complementarity problem. Gowda [23] presents Karamardian type results for the polynomial complementarity problem (PCP), as a special case of the NCP and a generalization of the TCP. In [23], Gowda also introduces the definition of degree of an R_0 tensor and shows that it is one.

Let the feasible set FEA($\mathbf{q}, \mathcal{A}$) = $\{\mathbf{x} \in \mathbb{R}^I_+ : \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q} \in \mathbb{R}^I_+\}$. If FEA($\mathbf{q}, \mathcal{A}$) $\neq \emptyset$, then we see that NCP($\mathbf{q}, \mathcal{A}$) is feasible. It is obvious that Problem 4.2.1 is a special case of Problem 4.2.2. For simplicity, we only consider the solvability of Problem 4.2.1.

4.3 Main Results

Let $\mathcal{A} \in RT_{N,I}$ be symmetric positive definite. If $\mathbf{q}$ is zero, then the solution of Problem 4.2.1 is zero. To avoid this triviality, let $\mathbf{q} \in \mathbb{R}^I$ in Problems 4.2.1 and 4.2.2 be nonzero.

If the zero vector $\mathbf{0}_I$ solves Problems 4.2.1 and 4.2.2, then we derive that $\mathbf{q} \in \mathbb{R}^I$ is a nonnegative vector. We only consider a nonzero solution $\mathbf{x} \in \mathbb{R}^I$ of Problems 4.2.1 and 4.2.2.

4.3.1 Necessary Conditions

We first consider the necessary conditions for Problem 4.2.1. The cornerstone for the *necessary* conditions is its nonlinear programming formulation:

$$\begin{array}{ll} \min & \mathcal{A}\mathbf{x}^N + \mathbf{x}^\top \mathbf{q} \\ s.t. & \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q} \in \mathbb{R}^I_+, \ \mathbf{x} \in \mathbb{R}^I_+. \end{array} \tag{4.3.1}$$

If $\mathbf{x}_*$ minimizes the nonlinear program in (4.3.1) satisfying $\mathcal{A}\mathbf{x}_*^N + \mathbf{q}^\top \mathbf{x}_* = 0$, then $\mathbf{x}_*$ is a solution of Problem 4.2.1. According to first-order necessary conditions given in [24], we obtain the following theorem.

Theorem 4.3.1 *Suppose $\mathcal{A} \in RT_{N,I}$ is symmetric. If* FEA$(\mathbf{q}, \mathcal{A}) \neq \emptyset$ *and $\mathbf{x}_*$ is a local solution of (4.3.1), then there exists a vector $\mathbf{u}_*$ of multipliers satisfying the conditions,*

$$\begin{aligned} \mathbf{q} + N\mathcal{A}\mathbf{x}_*^{N-1} - (N-1)\mathcal{A}\mathbf{x}_*^{N-2}\mathbf{u}_* &\geq \mathbf{0}_I, \\ \mathbf{x}_*^\top(\mathbf{q} + N\mathcal{A}\mathbf{x}_*^{N-1} - (N-1)\mathcal{A}\mathbf{x}_*^{N-2}\mathbf{u}_*) &= 0, \\ \mathbf{u}_* &\geq \mathbf{0}_I, \\ \mathbf{u}_*^\top(\mathbf{q} + \mathcal{A}\mathbf{x}_*^{N-1}) &= 0. \end{aligned} \tag{4.3.2}$$

For all i, the vectors $\mathbf{x}_*$ and $\mathbf{u}_*$ satisfy

$$(N-1)(\mathbf{x}_* - \mathbf{u}_*)_i(\mathcal{A}\mathbf{x}_*^{N-2}(\mathbf{x}_* - \mathbf{u}_*))_i \leq 0. \tag{4.3.3}$$

Proof Since FEA$(\mathbf{q}, \mathcal{A}) \neq \emptyset$, the nonlinear program (4.3.1) is feasible. An optimal solution $\mathbf{x}_*$ and a suitable vector $\mathbf{u}_*$ of multipliers satisfy the KKT conditions (4.3.2). To prove (4.3.3), we examine the inner product

$$\mathbf{x}_*^\top(\mathbf{q} + N\mathcal{A}\mathbf{x}_*^{N-1} - (N-1)\mathcal{A}\mathbf{x}_*^{N-2}\mathbf{u}_*) = 0,$$

componentwise and deduce for all i that,

$$(N-1)(\mathbf{x}_*)_i(\mathcal{A}\mathbf{x}_*^{N-2}(\mathbf{x}_*-\mathbf{u}_*))_i \leq 0, \tag{4.3.4}$$

using the fact that $\mathbf{x}_* \in \text{FEA}(\mathbf{q}, \mathcal{A})$. Multiplying the ith component in

$$\mathbf{q} + N\mathcal{A}\mathbf{x}_*^{N-1} - (N-1)\mathcal{A}\mathbf{x}_*^{N-2}\mathbf{u}_* \geq \mathbf{0}_I,$$

by $\mathbf{u}_*$ and then invoking the complementarity condition

$$(\mathbf{u}_*)_i(\mathbf{q} + \mathcal{A}\mathbf{x}_*^{N-1})_i = 0$$

which is implied by $\mathbf{u}_* \geq \mathbf{0}_I$, $\mathbf{u}_*^\top(\mathbf{q} + \mathcal{A}\mathbf{x}_*^{N-1}) = 0$ and the feasibility of of $\mathbf{x}_*$, we obtain

$$-(N-1)(\mathbf{u}_*)_i(\mathcal{A}\mathbf{x}_*^{N-2}(\mathbf{x}_*-\mathbf{u}_*))_i \leq 0. \tag{4.3.5}$$

Thus, (4.3.3) follows from adding (4.3.4) and (4.3.5). □

Remark 4.3.1 Theorem 4.3.1 is a special case of the result given in Cottle [1, Theorem 3].

From Theorem 4.3.1, we have the following conjecture, which has been disproved by Hieu et al. [25],

Conjecture 4.3.1 If $\text{FEA}(\mathbf{q}, \mathcal{A}) \neq \emptyset$, then the nonlinear programming problem (4.3.1) has an optimal solution, $\mathbf{x}_*$. Moreover, there exists a vector $\mathbf{u}_*$ of multipliers satisfying the conditions,

$$\begin{aligned}
\mathbf{q} + N\mathcal{A}\mathbf{x}_*^{N-1} - (N-1)\mathcal{A}\mathbf{x}_*^{N-2}\mathbf{u}_* &\geq \mathbf{0}_I,\\
\mathbf{x}_*^\top(\mathbf{q} + N\mathcal{A}\mathbf{x}_*^{N-1} - (N-1)\mathcal{A}\mathbf{x}_*^{N-2}\mathbf{u}_*) &= 0,\\
\mathbf{u}_* &\geq \mathbf{0}_I,\\
\mathbf{u}_*^\top(\mathbf{q} + \mathcal{A}\mathbf{x}_*^{N-1}) &= 0.
\end{aligned}$$

For all i, $\mathbf{x}_*$ and $\mathbf{u}_*$ satisfy

$$(N-1)(\mathbf{x}_*-\mathbf{u}_*)_i(\mathcal{A}\mathbf{x}_*^{N-2}(\mathbf{x}_*-\mathbf{u}_*))_i \leq 0.$$

When $N = 2$, the existence result for the quadratic programming problem associated with LCP is given in [21]. However, Cottle [1] presents some counter examples to show that this conjecture is not true for the general nonlinear program problem.

With Theorem 4.3.1, we prove the following existence result for NCP($\mathbf{q}, \mathcal{A}$).

Theorem 4.3.2 *Let a nonzero $\mathbf{x}_* \in \mathbb{R}_I$ be a local solution of (4.3.1). If $\mathcal{A}\mathbf{x}^{N-2}$ is positive definite for all $\mathbf{x} \in \mathbb{R}^I$, then $\mathbf{x}_*$ solves* NCP($\mathbf{q}, \mathcal{A}$).

Proof According to Theorem 4.3.1, for all i, there exists a nonnegative vector $\mathbf{u}_*$ such that

$$(N-1)(\mathbf{x}_* - \mathbf{u}_*)_i (\mathcal{A}\mathbf{x}_*^{N-2}(\mathbf{x}_* - \mathbf{u}_*))_i \le 0,$$

that is,

$$(\mathbf{x}_* - \mathbf{u}_*)^\top (\mathcal{A}\mathbf{x}_*^{N-2}(\mathbf{x}_* - \mathbf{u}_*)) \le 0.$$

Then we know that $\mathbf{x}_* = \mathbf{u}_*$. Based on (4.3.2), $\mathbf{x}_*$ solves NCP($\mathbf{q}, \mathcal{A}$). □

Remark 4.3.2 If $\mathbf{x}_* = \mathbf{0}_I$ is a local solution of (4.3.1), then $\mathbf{x}_*$ solves NCP($\mathbf{q}, \mathcal{A}$) for any $\mathbf{q} \in \mathbb{R}^I_+$.

We then consider the necessary conditions for Problem 4.2.2. Consider its nonlinear programming formulation:

$$\begin{array}{ll} \min & \displaystyle\sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}^{N-2k+2} + \mathbf{q}^\top \mathbf{x} \\ s.t. & \displaystyle\sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}^{N-(2k-1)} + \mathbf{q} \in \mathbb{R}^I_+, \ \mathbf{x} \in \mathbb{R}^I_+. \end{array} \tag{4.3.6}$$

With FEA($\mathbf{q}, \{\mathcal{A}_k\}$) denoting the feasible set of (4.3.6), if $\mathbf{x}_*$ minimizes the nonlinear programming in (4.3.6) satisfying $\sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}_*^{N-2k+2} + \mathbf{q}^\top \mathbf{x}_* = 0$, then $\mathbf{x}_*$ is a solution of Problem 4.2.2.

Theorem 4.3.3 *Suppose $\mathcal{A}_k \in RT_{N-(2k-2),I}$ is symmetric for $k = 1, 2, \ldots, N/2$ with an even $N > 3$. If $\mathbf{x}_*$ is a local solution of (4.3.6) and* FEA($\mathbf{q}, \{\mathcal{A}_k\}$) $\neq \emptyset$, *there exists $\mathbf{u}_*$ of multipliers satisfying the conditions*

$$\mathbf{q} + \sum_{k=1}^{N/2} (N-2k+2)\mathcal{A}_k \mathbf{x}_*^{N-2k+1} - \sum_{k=1}^{N/2} (N-2k+2)(N-2k+1)\mathcal{A}_k \mathbf{x}_*^{N-2k} \mathbf{u}_* \ge \mathbf{0}_I,$$

$$\mathbf{x}_*^\top \left(\mathbf{q} + \sum_{k=1}^{N/2} (N-2k+2)\mathcal{A}_k \mathbf{x}_*^{N-2k+1} - \sum_{k=1}^{N/2} (N-2k+2)(N-2k+1)\mathcal{A}_k \mathbf{x}_*^{N-2k} \mathbf{u}_* \right) = 0,$$

$$\mathbf{u}_* \ge \mathbf{0}_I,$$

$$\mathbf{u}_*^\top \left(\sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}_*^{N-(2k-1)} + \mathbf{q} \right) = 0.$$

For all i, $\mathbf{x}_*$ and $\mathbf{u}_*$ satisfy

$$(\mathbf{x}_* - \mathbf{u}_*)_i \left(\sum_{k=1}^{N/2} (N - 2k + 1) \mathcal{A}_k \mathbf{x}_*^{N-2k} (\mathbf{x}_* - \mathbf{u}_*) \right)_i \leq 0.$$

4.3.2 Solving Problem 4.2.1

In Problem 4.2.1, let $F(\mathbf{x}) = \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q}$. We first consider some properties of $F(\mathbf{x})$ when $\mathcal{A}$ is structured.

Theorem 4.3.4 *Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric and $\mathbf{x} \in \mathbb{R}^I$ is nonnegative.*

(a) *If $\mathcal{A}$ is copositive (resp. strictly copositive), then $F(\mathbf{x})$ is copositive (resp. strictly copositive) with respect to $\mathbb{R}^I_+$.*
(b) *For an even N, if $\mathcal{A}$ is positive definite, then $F(\mathbf{x})$ is strongly copositive with respect to $\mathbb{R}^I_+$ when $\alpha \leq \lambda_{\min} \|\mathbf{x}\|_2^{N-2}$ (resp. $\lambda_{\min} \|\mathbf{x}\|_N^N / \|\mathbf{x}\|_2^2$), where $\lambda_{\min}$ is the smallest Z-eigenvalue (resp. H-eigenvalue) of $\mathcal{A}$.*

Proof According to Definition 4.1.1 and $F(\mathbf{x}) = \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q}$, we have $\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)) = \mathcal{A}\mathbf{x}^N$. Since $\mathcal{A}$ is copositive (resp. strictly copositive), that is, $\mathcal{A}\mathbf{x}^N \geq 0$ for all $\mathbf{x} \in \mathbb{R}^I_+$ (resp. $\mathcal{A}\mathbf{x}^N > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^I_+$). Part (a) is proved.

If $\mathcal{A}$ is positive definite, then it follows from [10, Theorem 5] that the smallest Z-eigenvalue (resp. H-eigenvalue) of $\mathcal{A}$ satisfies $\lambda_{\min} > 0$.

If $\lambda_{\min}$ is the smallest Z-eigenvalue of $\mathcal{A}$, then

$$\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)) = \mathcal{A}\mathbf{x}^N \geq \lambda_{\min} \|\mathbf{x}\|_2^N = \lambda_{\min} \|\mathbf{x}\|_2^{N-2} \|\mathbf{x}\|_2^2.$$

In this case, part (b) is proved. Similarly, we can prove the case where $\lambda_{\min}$ is the smallest H-eigenvalue of $\mathcal{A}$. □

When $\mathcal{A} \in D_{N,I}$ is positive semi-definite, we have the following theorem on the Jacobian matrix $\nabla F(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^I$ is nonzero.

Theorem 4.3.5 *Let $\mathcal{A} \in D_{N,I}$ be positive semi-definite with an even N. Then the Jacobian matrix $\nabla F(\mathbf{x})$ is positive semi-definite for nonzero $\mathbf{x} \in \mathbb{R}^I$.*

Proof As $\mathcal{A}$ is diagonalizable, for $\mathbf{x} \in \mathbb{R}^I$, according to Definition 4.1.3, we have

$$\mathcal{A}\mathbf{x}^N = (\mathcal{D} \times_1 \mathbf{B} \times_2 \mathbf{B} \cdots \times_N \mathbf{B})\mathbf{x}^N = \mathcal{D}(\mathbf{B}^\top \mathbf{x})^N$$

$$= \mathcal{D}\mathbf{y}^N \ (\mathbf{y} \triangleq \mathbf{B}^\top \mathbf{x}) = \sum_{i=1}^{I} d_i y_i^N,$$

where d_i is the ith main diagonal entry of $\mathcal{D}$. With $d_i \geq 0$, we have $\mathcal{A}\mathbf{x}^N \geq 0$ for all $\mathbf{x} \in \mathbb{R}^I$.

Since the Jacobian matrix $\nabla F(\mathbf{x}) = (N-1)\mathcal{A}\mathbf{x}^{N-2}$, for any vector $\mathbf{z} \in \mathbb{R}^I$, we obtain

$$\mathbf{z}^\top \nabla F(\mathbf{x})\mathbf{z} = (N-1)\sum_{i=1}^{I} d_i y_i^{N-2} \tilde{z}_i^2 \geq 0,$$

where $\tilde{\mathbf{z}} = \mathbf{B}^\top \mathbf{z} = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_I)^\top \in \mathbb{R}^I$. The Jacobian matrix $\nabla F(\mathbf{x})$ is positive semi-definite with $\mathbf{x} \in \mathbb{R}^I$. □

From Theorem 4.3.5, we have an open question in the following.

Question 4.3.1 Suppose $\mathcal{A} \in RT_{N,I}$ with an even N and a nonzero $\mathbf{x} \in \mathbb{R}^I$. Under what conditions, there exists $\delta \in (0, 1)$, such that all principal minors of matrix $\mathcal{A}\mathbf{x}^{N-2}$ are bounded between δ and δ^{-1}, for all $\mathbf{x} \in \mathbb{R}^I_+$?

For a given symmetric tensor $\mathcal{A} \in RT_{N,I}$, the existence theorems for Problem 4.2.1 are given below.

Theorem 4.3.6 *Suppose that $\mathcal{A} \in RT_{N,I}$ is symmetric. For Problem 4.2.1, the following results hold.*

(a) *For an even N, if $\mathcal{A}$ is positive definite, then* NCP($\mathbf{q}, \mathcal{A}$) *has a nonempty, compact solution set.*
(b) *If $\mathcal{A}$ is strictly copositive with respect to $\mathbb{R}^I_+$, then* NCP($\mathbf{q}, \mathcal{A}$) *has a nonempty, compact solution set.*

Proof If $\mathcal{A}$ is positive definite, according to Theorem 4.3.4, $F(\mathbf{x}) = \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q}$ is strictly copositive. Let $c(\lambda) = \lambda^\alpha$ with $0 < \alpha \leq N-1$ and $\lambda \geq 1$, we know that $c(\lambda) \to \infty$ as $\lambda \to \infty$ and $\mathbf{x}^\top(F(\lambda\mathbf{x}) - F(\mathbf{0}_I)) \geq c(\lambda)\mathbf{x}^\top(F(\mathbf{x}) - F(\mathbf{0}_I))$. By Lemma 4.2.2, the NCP($\mathbf{q}, \mathcal{A}$) has a nonempty, compact solution set.

By Theorem 4.3.4, $F(\mathbf{x}) = \mathcal{A}\mathbf{x}^{N-1} + \mathbf{q}$ is strictly copositive. From [4], if $F(\mathbf{x})$ is strictly copositive with respect to $\mathbb{R}^I_+$, then the mapping G in (4.1.1) is $\mathbf{d}$-regular for any positive vector $\mathbf{d} \in \mathbb{R}^I$ and $G(\lambda\mathbf{x}) = \lambda^{N-1}G(\mathbf{x})$ with $\lambda > 0$. By Lemma 4.2.3, if $\mathcal{A}$ is strictly copositive with respect to $\mathbb{R}^n_+$, then NCP($\mathbf{q}, \mathcal{A}$) has a nonempty, compact solution set. □

From Theorem 4.3.6, we have the following conjecture.

Conjecture 4.3.2 If $\mathcal{A} \in D_{N,I}$ is positive definite with an even N, then NCP($\mathbf{q}, \mathcal{A}$) has a unique solution.

In the following theorem, we consider how to transform Problem 4.2.1 to a system of nonlinear equations.

Theorem 4.3.7 *Let $\mathcal{A} \in RT_{N,I}$, and $\mathbf{\Omega}, \mathbf{\Gamma} \in \mathbb{R}^{I\times I}$ be nonsingular diagonal matrices. For* NCP($\mathbf{q}, \mathcal{A}$)*, the following statements hold:*

(i) *If* $(\mathbf{w}, \mathbf{z})$ *is a solution of* NCP($\mathbf{q}, \mathcal{A}$)*, then* $\mathbf{x} = \frac{1}{2}(\mathbf{\Gamma}^{-1}\mathbf{z} - \mathbf{\Omega}^{-1}\mathbf{w})$ *satisfies the implicit fixed-point equation*

$$F(\mathbf{x}, |\mathbf{x}|) := \hat{\mathcal{A}}(|\mathbf{x}| + \mathbf{x})^{N-1} - \mathbf{\Omega}(|\mathbf{x}| - \mathbf{x}) + \mathbf{q} = \mathbf{0}_I, \tag{4.3.7}$$

where $\hat{\mathcal{A}} = \mathcal{A} \times_2 \mathbf{\Gamma} \cdots \times_N \mathbf{\Gamma}$*;*

(ii) *If* $\mathbf{x}$ *satisfies the implicit fixed-point equation* $F(\mathbf{x}) = \mathbf{0}_I$*, then*

$$\mathbf{z} = \mathbf{\Gamma}(|\mathbf{x}| + \mathbf{x}), \quad \mathbf{w} = \mathbf{\Omega}(|\mathbf{x}| - \mathbf{x}) \tag{4.3.8}$$

is a solution of NCP($\mathbf{q}, \mathcal{A}$).

Proof As $\mathbf{z}$ is a solution of NCP($\mathbf{q}, \mathcal{A}$), it is a nonnegative vector and can be expressed in the form

$$\mathbf{z} = \mathbf{\Gamma}(|\mathbf{x}| + \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^I.$$

Define another nonnegative vector

$$\mathbf{w} = \mathbf{\Omega}(|\mathbf{x}| - \mathbf{x}).$$

Then we have $\mathbf{z}^\top \mathbf{w} = 0$ and $\mathbf{w} = \mathcal{A}\mathbf{z}^{N-1} + \mathbf{q}$ if and only if

$$\mathbf{\Omega}(|\mathbf{x}| - \mathbf{x}) = \hat{\mathcal{A}}(|\mathbf{x}| + \mathbf{x})^{N-1} + \mathbf{q},$$

where $\hat{\mathcal{A}} = \mathcal{A} \times_2 \mathbf{\Gamma} \cdots \times_N \mathbf{\Gamma}$. It proves the validity of (i).

The implicit fixed-point equations (4.3.7) can be rewritten as

$$\mathbf{\Omega}(|\mathbf{x}| - \mathbf{x}) = \hat{\mathcal{A}}(|\mathbf{x}| + \mathbf{x})^{N-1} + \mathbf{q},$$

with $\hat{\mathcal{A}} = \mathcal{A} \times_2 \Gamma \cdots \times_N \mathbf{\Gamma}$.

For the nonnegative vectors $\mathbf{w}$ and $\mathbf{z}$ defined in (4.3.8), we have

$$\mathbf{w} = \mathbf{\Omega}(|\mathbf{x}| - \mathbf{x}).$$

In addition, we can observe that:

1. if $x_i > 0$, then $z_i > 0$ and $w_i = 0$;
2. if $x_i = 0$, then $z_i = w_i = 0$; and
3. if $x_i < 0$, then $z_i = 0$ and $w_i > 0$.

It then follows that $\mathbf{z}^\top \mathbf{w} = 0$. Therefore, $(\mathbf{w}, \mathbf{z})$ is a solution of NCP$(\mathbf{q}, \mathcal{A})$. □

If Ω and Γ are the identity matrices, then the above theorem is the same as Theorem 4.3 in [26]. When $N = 2$, the above theorem reduces to [27, Theorem 2.1]. Based on this, Bai [27] designs a modulus matrix splitting iterative method for solving the linear complementarity problem.

4.3.3 Solving Problem 4.2.2

In the previous subsection, we have considered the solvability of Problem 4.2.1. Analogously, the following theorems describe the solvability of Problem 4.2.2.

Theorem 4.3.8 *For a given even integer $N > 3$, suppose that $\mathcal{A}_k \in RT_{N-(2k-2),I}$ is symmetric, with $k = 1, 2, \ldots, N/2$ and $\mathcal{A}_{N/2} \in \mathbb{R}^{I\times I}$. For Problem 4.2.2, the following results hold.*

(a) *If $\mathcal{A}_k$ with $k = 1, 2, \ldots, N/2 - 1$ is diagonalizable and positive semi-definite and $\mathcal{A}_{N/2}$ is positive definite, then* NCP$(\mathbf{q}, \{\mathcal{A}_k\})$ *has a unique solution.*
(b) *If $\mathcal{A}_k$ is positive semi-definite and there exists at least one $k_0 \in \{1, 2, \ldots, N/2\}$ such that $\mathcal{A}_{k_0}$ is positive definite, then* NCP$(\mathbf{q}, \{\mathcal{A}_k\})$ *has a nonempty compact solution set.*
(c) *If $\mathcal{A}_k$ is strictly copositive with respect to $\mathbb{R}^I_+$, then* NCP$(\mathbf{q}, \{\mathcal{A}_k\})$ *has a nonempty compact solution set.*

Proof For part (a), according to the assumptions, $\mathcal{A}_k\mathbf{x}^{N-2k}$ with $k = 1, 2, \ldots, N/2 - 1$ are symmetric and positive semi-definite for all $\mathbf{x} \in \mathbb{R}^I$. When $\mathcal{A}_{N/2}$ is symmetric positive definite, then $\nabla F(\mathbf{x})$ is symmetric and positive definite, where $F(\mathbf{x})$ is defined in Problem 4.2.2. Then by Lemma 4.2.1, NCP$(\mathbf{q}, \{\mathcal{A}_k\})$ has a unique solution.

For part (b), from the assumptions and if $k_0 = N/2$, we have

$$\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I)) = \sum_{k=1}^{N/2} \mathcal{A}_k \mathbf{x}^{N-(2k)} \geq \lambda \|\mathbf{x}\|_2^2 > 0,$$

for all nonzero $\mathbf{x} \in \mathbb{R}^I_+$, where λ is the smallest eigenvalue of $\mathcal{A}_{N/2}$. If $k_0 \in \{1, 2, \ldots, N/2-1\}$, then the same result holds, where λ is the smallest Z-eigenvalue of $\mathcal{A}_{k_0}$. (We can also consider the case when λ is the smallest H-eigenvalue of $\mathcal{A}_{k_0}$) Thus $F(\mathbf{x})$ is strictly copositive. Let $c(\lambda) = \lambda$ with $\alpha = 1$ and $\lambda \geq 1$, we know that $c(\lambda) \to \infty$ as $\lambda \to \infty$ and $\mathbf{x}^\top (F(\lambda\mathbf{x}) - F(\mathbf{0}_I)) \geq c(\lambda)\mathbf{x}^\top (F(\mathbf{x}) - F(\mathbf{0}_I))$.

From Lemma 4.2.2, if $\mathcal{A}_k$ is positive semi-definite and there exists at least one $k_0 \in \{1, 2, \ldots, N/2\}$ such that $\mathcal{A}_{k_0}$ is positive definite, then NCP($\mathbf{q}$, $\{\mathcal{A}_k\}$) has a nonempty compact solution set.

Part (c) is similar part (b) without the proof. □

The constraints of $\mathcal{A}_k$ in part (a) of Theorem 4.3.8 can be weakened. A more general result is given below.

Theorem 4.3.9 *For a given even integer $N > 3$, suppose that $\mathcal{A}_k \in RT_{N-(2k-2),I}$ is symmetric, with $k \in \{1, 2, \ldots, N/2 - 1\}$ and $\mathcal{A}_{N/2}$ is a square matrix. If $\mathcal{A}_k$ are diagonalizable and positive semi-definite and there exists $\delta \in (0, 1)$, such that all principal minors of $\mathcal{A}_{N/2}$ are bounded between δ and δ^{-1}, then* NCP($\mathbf{q}$, $\{\mathcal{A}_k\}$) *has a unique solution.*

Proof Since there exists $\delta \in (0, 1)$, such that all principal minors of $\mathcal{A}_{N/2}$ are bounded between δ and δ^{-1}, then the real part of every eigenvalue of $\mathcal{A}_{N/2}$ is positive. Hence, for all nonzero $\mathbf{x}$, we can derive $\mathbf{x}^\top \mathcal{A}_{N/2}\mathbf{x} > 0$. According to the assumptions, the Jacobian matrix $\Delta F(\mathbf{x})$ in Problem 4.2.2 is positive definite. Hence, NCP($\mathbf{q}$, $\{\mathcal{A}_k\}$) has a unique solution. □

4.4 Generalizations

Huang and Qi [28] reformulate the multilinear game (a class of N-person noncooperative games) as the TCP and show that finding a Nash equilibrium point of the multilinear game is equivalent to finding a solution of the corresponding TCP. The readers may consult [22, 28–32] for a thorough survey of the existence of the solution set of the TCP.

Recently, some researchers focus on the numerical algorithms for TCPs [26, 28, 33–38]. Che et al. [39] consider the stochastic tensor complementarity problem via the theory of stochastic R_0 tensors. Barbagallo et al. [40] study some variational inequalities on a class of structured tensors. Wang et al. [17] introduce the tensor variational inequality, where the involved function is the sum of an arbitrary given vector and a homogeneous polynomial defined by a tensor. The interested readers are referred to [14–16] for the basic theory, solution methods and applications of TCPs.

We next consider the stochastic TCPs [39] and generalized order TCPs [41].

4.4.1 Stochastic TCPs

Assume that $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space with the underlying sample space $\Omega \in \mathbb{R}^I$, where the probability distribution $\mathcal{P}$ is known. The stochastic linear

complementarity problem SLCP($\mathbf{M}(\omega)$, $\mathbf{q}(\omega)$) [42–46] is to find $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$\mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega) \in \mathbb{R}^I_+, \quad \mathbf{x}^\top(\mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega)) = 0, \quad \text{a.e. } \omega \in \Omega,$$

or

$$\mathscr{P}\{\omega \in \Omega : \mathbf{x} \in \mathbb{R}^I_+, \quad \mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega) \in \mathbb{R}^I_+, \quad \mathbf{x}^\top(\mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega)) = 0\} = 1,$$

where $\omega \in \Omega$ is a random vector with a given probability distribution $\mathscr{P}$, $\mathbf{M}(\omega) \in \mathbb{R}^{I\times I}$ and $\mathbf{q}(\omega) \in \mathbb{R}^I$. Here "a.e." is the abbreviation of "almost everywhere".

For a given probability space $(\Omega, \mathscr{F}, \mathscr{P})$, the stochastic nonlinear complementarity problem SNCP($F(\mathbf{x}, \omega)$) [47–52]) is to find $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$F(\mathbf{x}, \omega) \in \mathbb{R}^I_+, \quad \mathbf{x}^\top F(\mathbf{x}, \omega) = 0, \quad \text{a.e. } \omega \in \Omega,$$

or

$$\mathscr{P}\{\omega \in \Omega : \mathbf{x} \in \mathbb{R}^I_+, \quad F(\mathbf{x}, \omega) \in \mathbb{R}^I_+, \quad \mathbf{x}^\top F(\mathbf{x}, \omega) = 0\} = 1,$$

where $F : \mathbb{R}^I \times \Omega \to \mathbb{R}^I$ is a given vector-valued function.

Gürkan et al. [53] consider the sample-path approach for stochastic variational inequalities and provide the convergence theory and applications for the approach. When applied to SLCP($\mathbf{M}(\omega)$, $\mathbf{q}(\omega)$) and SNCP($F(\mathbf{x}, \omega)$), the approach is the same as the expected value method. This uses (i) the expectation of the random function $\mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega)$, from the deterministic problem

$$\mathbf{x} \in \mathbb{R}^I_+, \quad \mathscr{E}\{\mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega)\} \in \mathbb{R}^I_+, \quad \mathscr{E}\{\mathbf{x}^\top(\mathbf{M}(\omega)\mathbf{x} + \mathbf{q}(\omega))\} = 0,$$

and (ii) the expectation of $F(\mathbf{x}, \omega)$, from the deterministic problem

$$\mathbf{x} \in \mathbb{R}^I_+, \quad \mathscr{E}\{F(\mathbf{x}, \omega)\} \in \mathbb{R}^I_+, \quad \mathscr{E}\{\mathbf{x}^\top F(\mathbf{x}, \omega)\} = 0.$$

In particular, the stochastic tensor complementarity problem STCP($\mathcal{M}(\omega)$, $\mathbf{q}(\omega)$) [54] is to find $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$\mathcal{A}(\omega)\mathbf{x}^{N-1} + \mathbf{q}(\omega) \in \mathbb{R}^I_+, \quad \mathbf{x}^\top(\mathcal{A}(\omega)\mathbf{x}^{N-1} + \mathbf{q}(\omega)) = 0, \quad \text{a.e. } \omega \in \Omega,$$

or

$$\mathscr{P}\{\omega \in \Omega : \mathbf{x} \in \mathbb{R}^I_+, \quad \mathcal{A}(\omega)\mathbf{x}^{N-1} + \mathbf{q}(\omega) \in \mathbb{R}^I_+, \quad \mathbf{x}^\top(\mathcal{A}(\omega)\mathbf{x}^{N-1} + \mathbf{q}(\omega)) = 0\} = 1,$$

where $\mathcal{A}(\omega) \in RT_{N,I}$ and $\mathbf{q}(\omega) \in \mathbb{R}^I$. The STCP($\mathcal{M}(\omega), \mathbf{q}(\omega)$) is equivalent to finding a vector $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$\mathscr{E}\{\mathcal{A}(\omega)\mathbf{x}^{N-1} + \mathbf{q}(\omega)\} \in \mathbb{R}^I_+, \quad \mathscr{E}\{\mathbf{x}^\top(\mathcal{A}(\omega)\mathbf{x}^{N-1} + \mathbf{q}(\omega))\} = 0.$$

4.4.2 Generalized Order TCPs

For given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$, the ith components of $\mathbf{x} \wedge \mathbf{y}$ and $\mathbf{x} \vee \mathbf{y}$ are $\min\{x_i, y_i\}$ and $\max\{x_i, y_i\}$, respectively, for all i. Also, $\mathbf{x} > \mathbf{y}$ and $\mathbf{x} \geq \mathbf{y}$ mean $x_i > y_i$ and $x_i \geq y_i$ for all i, respectively. Similarly, we define $\mathbf{x} < \mathbf{y}$ and $\mathbf{x} \leq \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$.

Let $K \geq 1$ be any positive number. Given $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K \in \mathbb{R}^{I\times I}$ and $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K \in \mathbb{R}^I$, the generalized order linear complementarity problem (GOLCP) [55] is to find a vector $\mathbf{x} \in \mathbb{R}^I$ such that

$$\mathbf{x} \wedge (\mathbf{A}_1\mathbf{x} + \mathbf{q}_1) \wedge (\mathbf{A}_2\mathbf{x} + \mathbf{q}_2) \wedge \cdots \wedge (\mathbf{A}_K\mathbf{x} + \mathbf{q}_K) = \mathbf{0}_I. \tag{4.4.1}$$

For $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_K \in T_{N,I}$ and $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K \in \mathbb{R}^I$, we denote

$$\hat{\mathcal{A}} := \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_K\}, \quad \hat{\mathbf{q}} := \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K\}. \tag{4.4.2}$$

The corresponding GOTCP($\hat{\mathcal{A}}, \hat{\mathbf{q}}$) is to find a vector $\mathbf{x} \in \mathbb{R}^I$ such that

$$\mathbf{x} \wedge (\mathcal{A}_1\mathbf{x}^{N-1} + \mathbf{q}_1) \wedge (\mathcal{A}_2\mathbf{x}^{N-1} + \mathbf{q}_2) \wedge \cdots \wedge (\mathcal{A}_K\mathbf{x}^{N-1} + \mathbf{q}_K) = \mathbf{0}_I.$$

We introduce the notations

$$\begin{aligned}
\mathbf{x} \wedge (\hat{\mathcal{A}}\mathbf{x}^{N-1} + \hat{\mathbf{q}}) &:= \mathbf{x} \wedge (\mathcal{A}_1\mathbf{x}^{N-1} + \mathbf{q}_1) \wedge (\mathcal{A}_2\mathbf{x}^{N-1} + \mathbf{q}_2) \wedge \cdots \wedge (\mathcal{A}_K\mathbf{x}^{N-1} + \mathbf{q}_K),\\
\mathbf{x} \wedge \hat{\mathcal{A}}\mathbf{x}^{N-1} &:= \mathbf{x} \wedge (\mathcal{A}_1\mathbf{x}^{N-1}) \wedge (\mathcal{A}_2\mathbf{x}^{N-1}) \wedge \cdots \wedge (\mathcal{A}_K\mathbf{x}^{N-1}),\\
\mathbf{x} \vee \hat{\mathcal{A}}\mathbf{x}^{N-1} &:= \mathbf{x} \vee (\mathcal{A}_1\mathbf{x}^{N-1}) \vee (\mathcal{A}_2\mathbf{x}^{N-1}) \vee \cdots \vee (\mathcal{A}_K\mathbf{x}^{N-1}).
\end{aligned}$$

Also $\hat{\mathcal{A}}\mathbf{x}^{N-1} + \hat{\mathbf{q}} \geq \mathbf{0}_I$ means $\mathcal{A}_k\mathbf{x}^{N-1} + \mathbf{q}_k \geq \mathbf{0}_I$ for $k = 1, 2, \ldots, K$. A vector $\mathbf{x} \in \mathbb{R}^I_+$ with $\hat{\mathcal{A}}\mathbf{x}^{N-1} + \hat{\mathbf{q}} \geq \mathbf{0}_I$ ($> \mathbf{0}_I$) is called a feasible (respectively, strictly feasible) vector for GOTCP($\hat{\mathcal{A}}, \hat{\mathbf{q}}$). If there is such a vector, we say that GOTCP($\hat{\mathcal{A}}, \hat{\mathbf{q}}$) is feasible (respectively, strictly feasible). Let

$$\begin{cases}
\mathbb{F}(\hat{\mathcal{A}}) = \{\hat{\mathbf{q}} : \hat{\mathcal{A}}\mathbf{x}^{N-1} + \hat{\mathbf{q}} \geq \mathbf{0}_I \text{ for some } \mathbf{x} \in \mathbb{R}^I_+\},\\
\mathbb{K}(\hat{\mathcal{A}}) = \{\hat{\mathbf{q}} : \mathrm{SOL}(\hat{\mathcal{A}}, \hat{\mathbf{q}}) \neq \emptyset\},
\end{cases}$$

where SOL($\hat{\mathcal{A}}, \hat{\mathbf{q}}$) denotes the solution set of GOTCP($\hat{\mathcal{A}}, \hat{\mathbf{q}}$). It is obvious that $\mathbb{K}(\hat{\mathcal{A}}) \subseteq \mathbb{F}(\hat{\mathcal{A}})$. Borrowing from LCP, we say that these sets are, respectively (when

$\hat{\mathcal{A}}$ is fixed), the set of all "feasible" $\hat{\mathbf{q}}$'s and "solvable" $\hat{\mathbf{q}}$'s. Note that

$$\text{int}\mathbb{F}(\hat{\mathcal{A}}) = \{\hat{\mathbf{q}} : \hat{\mathcal{A}}\mathbf{x}^{N-1} + \hat{\mathbf{q}} > \mathbf{0}_I \text{ for some } \mathbf{x} \in \mathbb{R}^I_+\}.$$

Consider $\mathcal{B} \in \mathbb{R}^{J\times I\times\cdots\times I}$ with $J \geq I$ and $\mathbf{p} \in \mathbb{R}^J$. Let $J = J_1 + J_2 + \cdots + J_I$ and $J_0 = 0$. Suppose that $\mathcal{B}_i \in \mathbb{R}^{J_i\times I\times\cdots\times I}$ and $\mathbf{p}_i \in \mathbb{R}^{J_i}$ satisfy

$$\begin{cases} \mathcal{B}_i = \mathcal{B}(J_1 + \cdots + J_{i-1} + 1 : J_1 + \cdots + J_i, :, \ldots, :) \in \mathbb{R}^{J_i\times I\times\cdots\times I}, \\ \mathbf{p}_i = \mathbf{p}(J_1 + \cdots + J_{i-1} + 1 : J_1 + \cdots + J_i) \in \mathbb{R}^{I_i}, \end{cases}$$

for all i. Then VTCP($\mathcal{B}$, $\mathbf{p}$) is to find a vector $\mathbf{x} \in \mathbb{R}^I_+$ such that

$$\mathcal{B}\mathbf{x}^{N-1} + \mathbf{p} \in \mathbb{R}^J_+, \quad x_i \prod_{j=1}^{J_i} \left(\mathcal{B}_i\mathbf{x}^{N-1} + \mathbf{p}_i\right)_j = 0,$$

for all i, where the jth component of $\mathcal{B}\mathbf{x}^{N-1}$ is defined by

$$(\mathcal{B}\mathbf{x}^{N-1})_j = \sum_{i_2,\ldots,i_N=1}^{I} b_{ji_2\ldots i_N} x_{i_2}\ldots x_{i_N}, \quad j = 1, 2, \ldots, J.$$

We show that this problem can be formulated as a GOTCP. Let $K = \max\{J_i : i = 1, 2, \ldots, I\}$. Let $\mathcal{B}_i^j := \mathcal{B}_i(j, :, \ldots, :)$ is the jth mode-1 slice of $\mathcal{B}_i$ where $j = 1, 2, \ldots, J_i$ and $i = 1, 2, \ldots, I$. We define tensors $\hat{\mathcal{B}}_1, \hat{\mathcal{B}}_2, \ldots, \hat{\mathcal{B}}_I \in \mathbb{R}^{K\times I\times\cdots\times I}$ of order N in the following way. For each i, the jth mode-1 slice of $\hat{\mathcal{B}}_i$ is $\mathcal{B}_i^j$ if $j \leq J_i$ and $\mathcal{B}_i^1$ if $j > J_i$. Similarly, we define vectors $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \ldots, \hat{\mathbf{p}}_I \in \mathbb{R}^K$. For each i, the jth element of $\hat{\mathbf{p}}_i$ is the jth element of $\mathbf{p}_i$ if $j \leq J_i$ and the first element of $\mathbf{p}_i$ if $j > J_i$.

It is clear that VTCP($\hat{\mathcal{B}}$, $\hat{\mathbf{p}}$) is equivalent to VTCP($\mathcal{B}$, $\mathbf{p}$). Let the jth mode-1 slice of $\mathcal{A}_i \in T_{N,I}$ be the ith mode-1 slice of $\hat{\mathcal{B}}_j$ with $i = 1, 2, \ldots, K$ and all j. For example, $\mathcal{A}_1$ is formed by considering $\hat{\mathcal{B}}_j^1$ for all j. Similarly, let $\mathbf{q}_i$ be the vector of size $I \times 1$ whose jth component is the ith component in the vector $\hat{\mathbf{p}}_j$. Hence we can verify that VTCP($\hat{\mathcal{B}}$, $\hat{\mathbf{p}}$) is equivalent to GOTCP($\hat{\mathcal{A}}$, $\hat{\mathbf{q}}$).

We shall show that every GOTCP can be formulated as a VTCP. For a given GOTCP($\hat{\mathcal{A}}$, $\hat{\mathbf{q}}$), let the ith mode-1 slice of $\mathcal{B}_j \in \mathbb{R}^{K\times I\times\cdots\times I}$ be the jth mode-1 slice of $\mathcal{A}_i$ with $i = 1, 2, \ldots, K$ and all j. For example, $\mathcal{B}_1$ is formed by considering $\hat{\mathcal{A}}_i^1$ with $i = 1, 2, \ldots, K$. We define $\mathbf{p}_i \in \mathbb{R}^K$ for all i. This construction leads to the pair ($\mathcal{B}$, $\mathbf{p}$), and the corresponding VTCP is easily seen to be equivalent to GOTCP($\hat{\mathcal{A}}$, $\hat{\mathbf{q}}$).

References

1. R. Cottle, Nonlinear programs with positively bounded Jacobians. SIAM J. Appl. Math. **14**(1), 147–158 (1966)
2. F. Facchinei, J. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*, vol. I and II (Springer, New York, 2003)
3. P. Harker, J. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Math. Program. **48**(Series B), 161–220 (1990)
4. S. Karamardian, The complementarity problem. Math. Program. **2**(1), 107–129 (1972)
5. J. Moré, Classes of functions and feasibility conditions in nonlinear complementarity problems. Math. Program. **6**(1), 327–338 (1974)
6. M. Noor, On the nonlinear complementarity problem. J. Math. Anal. Appl. **123**(2), 455–460 (1987)
7. D. den Hertog, C. Roos, T. Terlaky, The linear complementarity problem, sufficient matrices, and the criss-cross method. Linear Algebra Appl. **187**, 1–14 (1993)
8. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)
9. A. Berman, R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Society for Industrial and Applied Mathematics, Philadelphia, 1994)
10. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
11. L. Baumert, Extreme copositive quadratic forms II. Pac. J. Math. **20**, 1–20 (1967)
12. J. Hiriart-Urruty, A. Seeger, A variational approach to copositive matrices. SIAM Rev. **52**(4), 593–629 (2010)
13. L. Qi, Symmetric nonnegative tensors and copositive tensors. Linear Algebra Appl. **439**(1), 228–238 (2013)
14. Z. Huang, L. Qi, Tensor complementarity problems-part I: basic theory. J. Optim. Theory Appl. **183**(1), 1–23 (2019)
15. Z. Huang, L. Qi, Tensor complementarity problems-part III: applications. J. Optim. Theory Appl. **183**(3), 771–791 (2019)
16. L. Qi, Z. Huang, Tensor complementarity problems-part II: solution methods. J. Optim. Theory Appl. **183**(2), 365–385 (2019)
17. Y. Wang, Z. Huang, L. Qi, Global uniqueness and solvability of tensor variational inequalities. J. Optim. Theory Appl. **177**(1), 137–152 (2018)
18. W. Ding, L. Qi, Y. Wei, Fast Hankel tensor-vector products and application to exponential data fitting. Numer. Linear Algebra Appl. **22**(5), 814–832 (2015)
19. W. Ding, Y. Wei, Generalized tensor eigenvalue problems. SIAM J. Matrix Anal. Appl. **36**(3), 1073–1099 (2015)
20. J. Demmel, *Applied Numerical Linear Algebra* (Society for Industrial and Applied Mathematics, Philadelphia, 1997)
21. R. Cottle, J. Pang, R. Stone, *The Linear Complementarity Problem* (Academic Press, Boston (1992)
22. Y. Song, L. Qi, Properties of some classes of structured tensors. J. Optim. Theory Appl. **165**(3), 854–873 (2015)
23. M. Gowda, *Polynomial Complementarity Problems* (2016). ArXiv preprint:1609.05267
24. J. Nocedal, S. Wright, *Numerical Optimization* (Springer, New York, 1999)
25. V. Hieu, Y. Wei, J. Yao, Notes on the optimization problems corresponding to polynomial complementarity problems. J. Optim. Theory Appl. **184**(2), 687–695 (2020)
26. D. Liu, W. Li, S. Vong, Tensor complementarity problems: the GUS-property and an algorithm. Linear and Multilinear Algebra **66**, 1726–1749 (2018)
27. Z. Bai, Modulus-based matrix splitting iteration methods for linear complementarity problems. Numer. Linear Algebra Appl. **17**(6), 917–933 (2010)

28. Z. Huang, L. Qi, Formulating an n-person noncooperative game as a tensor complementarity problem. Comput. Optim. Appl. **66**, 557–576 (2017)
29. X. Bai, Z. Huang, Y. Wang, Global uniqueness and solvability for tensor complementarity problems. J. Optim. Theory Appl. **170**(1), 72–84 (2016)
30. M. Gowda, Z. Luo, L. Qi, N. Xiu, *Z-tensors and Complementarity Problems* (2015). ArXiv preprint:1510.07933
31. Z. Luo, L. Qi, N. Xiu, The sparsest solutions to Z-tensor complementarity problems. Optim. Lett. **11**, 471–482 (2017)
32. Y. Song, L. Qi, Tensor complementarity problem and semi-positive tensors. J. Optim. Theory Appl. **169**(3), 1069–1078 (2016)
33. S. Du, L. Zhang, A mixed integer programming approach to the tensor complementarity problem. J. Glob. Optim. **73**, 789–800 (2019)
34. S. Du, L. Zhang, C. Chen, L. Qi, Tensor absolute value equations. Sci. China Mater. **61**(9), 1695–1710 (2018)
35. L. Han, A continuation method for tensor complementarity problems. J. Optim. Theory Appl. **180**, 949–963 (2019)
36. X. Wang, M. Che, L. Qi, Y. Wei, Modified gradient dynamic approach to the tensor complementarity problem. *Optimization Methods and Software* **35**(2), 394–415 (2020)
37. X. Wang, M. Che, Y. Wei, Global uniqueness and solvability of tensor complementarity problems for $\mathcal{H}_+$-tensors. Numer. Algor. (2019). https://doi.org/10.1007/s11075-019-00769-9
38. S. Xie, D. Li, H. Xu, An iterative method for finding the least solution to the tensor complementarity problem. J. Optim. Theory Appl. **175**(1), 119–136 (2017)
39. M. Che, L. Qi, Y. Wei, Stochastic R_0 tensors to stochastic tensor complementarity problems. Optim. Lett. **13**(2), 261–279 (2019)
40. A. Barbagallo, S. Guarino, L. Bianco, Variational inequalities on a class of structured tensors. J. Nonlinear Convex Anal. **19**, 711–729 (2018)
41. M. Che, L. Qi, Y. Wei, The generalized order tensor complementarity problems. Numer. Math. Theory Methods Appl. **13**(1), 131–149 (2020).
42. X. Chen, C. Zhang, M. Fukushima, Robust solution of monotone stochastic linear complementarity problems. Math. Program. **117**(Ser. B), 51–80 (2009)
43. X. Chen, M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res. **30**(4), 1022–1038 (2005)
44. H. Fang, X. Chen, M. Fukushima, Stochastic R_0 matrix linear complementarity problems. SIAM J. Optim. **18**(2), 482–506 (2007)
45. C. Ling, L. Qi, G. Zhou, L. Caccetta, The SC^1 property of an expected residual function arising from stochastic complementarity problems. Oper. Res. Lett. **36**(4), 456–460 (2008)
46. G. Zhou, L. Caccetta, Feasible semismooth Newton method for a class of stochastic linear complementarity problems. J. Optim. Theory Appl. **139**(2), 379–392 (2008)
47. S. Gabriel, J. Zhuang, R. Egging, Solving stochastic complementarity problems in energy market modeling using scenario reduction. Eur. J. Oper. Res. **197**(3), 1028–1040 (2009)
48. G. Lin, Combined Monte Carlo sampling and penalty method for stochastic nonlinear complementarity problems. Math. Comput. **78**(267), 1671–1686 (2009)
49. G. Lin, M. Fukushima, New reformulations for stochastic nonlinear complementarity problems. Optim. Methods Softw. **21**(4), 551–564 (2006)
50. M. Wang, M. Ali, Stochastic nonlinear complementarity problems: stochastic programming reformulation and penalty-based approximation method. J. Optim. Theory Appl. **144**(3), 597–614 (2010)
51. C. Zhang, X. Chen, Stochastic nonlinear complementarity problem and applications to traffic equilibrium under uncertainty. J. Optim. Theory Appl. **137**(2), 277–295 (2008)
52. L. Xu, B. Yu, CVaR-constrained stochastic programming reformulation for stochastic nonlinear complementarity problems. Comput. Optim. Appli. **58**(2), 483–501 (2014)
53. G. Gürkan, A. Özge, S. Robinson, Sample-path solution of stochastic variational inequalities. Math. Program. **84**(Ser. A), 313–333 (1999)

54. S. Du, M. Che, Y. Wei, Stochastic structured tensors to stochastic complementarity problems. Comput. Optim. Appl. (2019). https://doi.org/10.1007/s10589-019-00144-3
55. M.S. Gowda, R. Sznajder, The generalized order linear complementarity problem. SIAM J. Matrix Anal. Appl. **25**, 148–162 (1993)

Chapter 5
Plane Stochastic Tensors

We study combinatorial properties of nonnegative tensors. We make the following contributions: (1) we obtain equivalent conditions for sign nonsingular tensors and relationships between the combinatorial determinant and the permanent of nonnegative tensors, in Theorems 5.2.1 and 5.2.2; (2) the sets of plane stochastic tensors and totally plane stochastic tensors are closed, bounded and convex sets, and an nonnegative tensor has a plane stochastic pattern if and only if its positive entries are contained in a positive diagonal, in Lemma 5.3.1 and Theorem 5.3.2; (3) from a nonnegative tensor, we propose a normalization algorithm which converges to a plane stochastic tensor, in Theorem 5.3.8; (4) we discuss the boundlessness of the diagonal products of any nonnegative tensor and obtain a probabilistic algorithm after Theorem 5.4.4 for locating a positive diagonal in a $(0, 1)$-tensor[1]; (5) we explore the axial N-index assignment problem via the set of plane stochastic tensors in Sect. 5.5.

5.1 Preliminaries

For i_n and n, the i_n*th mode-n slice* [1–5] of $\mathcal{A} \in RT_{N,I}$ is defined as a tensor in $RT_{N-1,I}$, by fixing the mode-n index of $\mathcal{A}$ to i_n: $\mathcal{A}(:, \ldots, :, i_n, :, \ldots, :)$. Similar to the row and the column sums of a matrix, for all i_n and n, the i_nth mode-n sum of $\mathcal{A} \in RT_{N,I}$ is

$$f(\mathcal{A}; i_n, n) = \sum_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N=1}^{I} a_{i_1\ldots i_{n-1}i_n i_{n+1}\ldots i_N}.$$

[1] A tensor is a $(0, 1)$-tensor [1], if its entries are chosen from the set $\{0, 1\}$.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_5

For both i_n and n, we denote by $\mathcal{A}(i_1, i_2, \ldots, i_N)$ the tensor in $RT_{N,I-1}$, derived from $\mathcal{A}$ by deleting its i_nth mode-n slices.

5.1.1 Plane Stochastic Tensors

Similar to a stochastic matrix (see [6, Chapter 2] and [7]), a *plane stochastic tensor* is defined as follows.

Definition 5.1.1 For a given n, a tensor $\mathcal{A} \in NT_{N,I}$ is mode-n plane stochastic provided that $f(\mathcal{A}; i_n, n) = 1$ for all i_n. Particularly, if $f(\mathcal{A}; i_n, n) = 1$, for all i_n and n, then $\mathcal{A}$ is called a *plane stochastic tensor*. We denote the set of all plane stochastic tensors in $NT_{N,I}$ by $\Omega_{N,I}$.

Yang and Yang [8] give an alternative proof of the minimax theorem for nonnegative tensors with positive eigenvectors corresponding to the spectral radius [9] by using mode-1 plane stochastic tensors. Jurkat and Ryser [4] study the basic combinatorial properties of tensors and these properties also hold for plane stochastic tensors. We can refer to [1–4, 10] for the details on plane stochastic tensors.

Cui et al. [11] define *multi-stochastic tensors* (or called line stochastic tensors [3]), and *permutation tensors*. We make the difference via an example. For $N = 3$, Definition 5.1.1 indicates that a plane stochastic tensor $\mathcal{A} \in NT_{3,I}$ satisfies

$$\sum_{i_2,i_3=1}^{I} a_{i_1 i_2 i_3} = \sum_{i_1,i_3=1}^{I} a_{i_1 i_2 i_3} = \sum_{i_1,i_2=1}^{I} a_{i_1 i_2 i_3} = 1.$$

In [11], $\mathcal{B} \in NT_{3,I}$ is called *multi-stochastic*, provided

$$\sum_{i_1=1}^{I} b_{i_1 i_2 i_3} = \sum_{i_2=1}^{I} b_{i_1 i_2 i_3} = \sum_{i_3=1}^{I} b_{i_1 i_2 i_3} = 1.$$

Let a nonnegative vector $\mathbf{x} \in \mathbb{R}^I$ satisfy $x_1 + x_2 + \cdots + x_I = 1$. Then

$$\sum_{i_1,i_2,i_3=1}^{I} b_{i_1 i_2 i_3} x_{i_2} x_{i_3} = \sum_{i_1,i_2,i_3=1}^{I} b_{i_1 i_2 i_3} x_{i_1} x_{i_3} = \sum_{i_1,i_2,i_3=1}^{I} b_{i_1 i_2 i_3} x_{i_1} x_{i_2} = 1,$$

$$\sum_{i_1,i_2,i_3=1}^{I} a_{i_1 i_2 i_3} x_{i_2} x_{i_3} = \sum_{i_1,i_2,i_3=1}^{I} a_{i_1 i_2 i_3} x_{i_1} x_{i_3} = \sum_{i_1,i_2,i_3=1}^{I} a_{i_1 i_2 i_3} x_{i_1} x_{i_2} \neq 1.$$

The tensor $\mathcal{B}/I$ is plane stochastic. For $\mathcal{C} \in RT_{N,I}$ and a nonzero $\alpha \in \mathbb{R}$, each element of the tensor $\mathcal{C}/\alpha$ is $b_{i_1 i_2 \ldots i_N}/\alpha$ for i_n and n. Furthermore, if $\mathcal{A} \in NT_{N,I}$ is multi-stochastic, then $\mathcal{A}/I^{N-2}$ is plane stochastic.

Mode-1 multi-stochastic tensors have been studied in higher order Markov chains [12], under the name "transition probability tensors" [12–16]. Christensen and Fischer [17] use the plane stochastic tensors in error-correcting codes [18]. In Sect. 5.5, we will present a lower bound for the minimum of the axial N-index assignment problem by means of plane stochastic tensors.

The following definition is a generalization of the permutation matrix [19, Chapter 1].

Definition 5.1.2 A tensor $\mathcal{P} \in NT_{N,I}$ is a plane permutation tensor, if for $(N-1)$ given $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$, we have $p_{i_1 i_2 \ldots i_N} = \delta_{i_1 \pi_2(i_2) \ldots \pi_N(i_N)}$, where

$$\delta_{i_1 \pi_2(i_2) \ldots \pi_N(i_N)} = \begin{cases} 1, & \text{if} \quad i_1 = \pi_2(i_2) = \cdots = \pi_N(i_N); \\ 0, & \text{otherwise.} \end{cases}$$

The number of the plane permutation tensors in $NT_{N,I}$ is $\tilde{N} = (I!)^{N-1}$.

We denote the set of plane permutation tensors in $NT_{N,I}$ by $P_{N,I}$. $P_{N,I}$ is a subset of $\Omega_{N,I}$, which is the set of all plane stochastic tensors in $NT_{N,I}$. A plane permutation tensor can also be characterized by the system of linear equations:

$$\sum_{i_1, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N = 1}^{I} p_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} = 1,$$

with the constraints $p_{i_1 i_2 \ldots i_N} \in \{0, 1\}$, for all i_n and n.

In Fig. 5.1, we list four plane permutation tensors in $\Omega_{3,2}$ with all their entries.

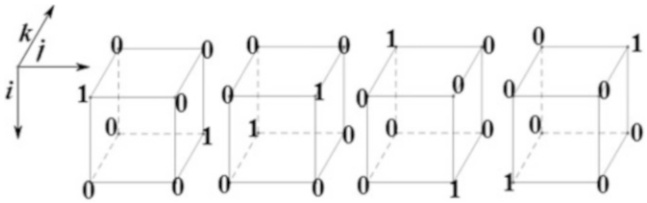

Fig. 5.1 Plane permutation tensors in $\Omega_{3,2}$

5.1.2 Combinatorial Determinant and Permanent of Tensors

There are two ways to extend the determinant of matrices to tensors. One is to extend the usual expression of an $I \times I$ matrix determinant as a sum of $I!$ monomials in the entries of the matrix, which we call the *matrix combinatorial determinant*. Another we call the *determinant* [20, 21] uses the characterization that a matrix $\mathbf{A} \in \mathbb{R}^{I\times I}$ has $\det(\mathbf{A}) = 0$ if and only if $\mathbf{Ax} = \mathbf{0}_I$ has nonzero solutions.

Definition 5.1.3 (Combinatorial Determinant of Tensors [22]) The combinatorial determinant of $\mathcal{A} \in RT_{N,I}$, denoted by $\det_c(\mathcal{A})$, is defined as

$$\det_c(\mathcal{A}) = \sum_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \operatorname{sign}(\pi_{N-K+1})\ldots\operatorname{sign}(\pi_N)\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}, \tag{5.1.1}$$

where $\operatorname{sign}(\pi)$ is the sign of $\pi \in \mathbb{S}_I$ [23] and the positive integer K satisfies

$$K = (N+1)/2, \text{ for an odd } N; \quad \text{or} \quad K = N/2, \text{ for an even } N. \tag{5.1.2}$$

Cayley begins to study the combinatorial determinant of hypermatrices [24]. The fundamental properties and applications of the higher-order determinant (our tensor combinatorial determinant) are considered by Rice [22], Oldenburger [25] and Vein and Dale [26]. The combinatorial determinant in Definition 5.1.3 is also called a P-sign determinant [22]. More generally, we refer to [22] for the P-sign determinant of $\mathcal{A} \in RT_{N,I}$ with an even $P \leq N$. Lim [27] considers the P-sign determinant with an even $K = N$.

Analogous to the definition of the tensor combinatorial determinant, we define the tensor permanent as follows.

Definition 5.1.4 ([1]) The permanent of the tensor $\mathcal{A} \in RT_{N,I}$, denoted by $\operatorname{perm}(\mathcal{A})$, is defined as

$$\operatorname{perm}(\mathcal{A}) = \sum_{\pi_2,\ldots,\pi_N\in S_I} \prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}.$$

Dow and Gibson [1] extend properties from the matrix permanent to the tensor permanent (or called multi-dimensional permanent) and investigate some inequalities for the permanent of $(0, 1)$-tensors. Dow and Gibson deduce that the permanent of a $(0, 1)$-tensor is equivalent to the sum of of all perfect matchings of the N-partite N-uniform hypergraph [28, 29]. We refer to [30] for the theory of hypergraphs. Recently, Taranenko [31] investigates the properties and applications for the permanent of multi-dimensional matrices. Barvinok [32] presents a deterministic algorithm to compute the matrix permanent and extends this method to computing the tensor permanent.

Avgustinovich [33] indicates that the problem of counting the number of distinct 1-perfect binary codes reduces to compute the permanent of a particularly constructed multi-dimensional matrix (tensor). Taranenko [34] derives inequalities for the permanent of the multi-stochastic tensors. Cui et al. [11] utilize the tensor permanent to study the extreme points for the set of multi-stochastic tensors. The tensor permanent defined by Cui et al. [11] is identical to the tensor S-permanent [1].

We give an example to illustrate the permanent and the combinatorial determinant of $\mathcal{A} \in RT_{N,I}$.

Example 5.1.1 The permanent of $\mathcal{A} \in RT_{3,2}$ is

$$\operatorname{perm}(\mathcal{A}) = a_{111}a_{222} + a_{112}a_{221} + a_{121}a_{212} + a_{122}a_{211}.$$

According to [22], $P = 2$ and the combinatorial determinant of $\mathcal{A} \in RT_{3,2}$ is

$$\det_c(\mathcal{A}) = a_{111}a_{222} - a_{112}a_{221} - a_{121}a_{212} + a_{122}a_{211}.$$

Furthermore, the permanent of $\mathcal{A} \in RT_{4,2}$ is

$$\begin{aligned}\operatorname{perm}(\mathcal{A}) =& a_{1111}a_{2222} + a_{1112}a_{2221} + a_{1121}a_{2212} + a_{1211}a_{2122} \\ &+ a_{2111}a_{1222} + a_{1122}a_{2211} + a_{1212}a_{2121} + a_{1221}a_{2112}.\end{aligned}$$

It follow from [22] that $P = 2$ and the combinatorial determinant of $\mathcal{A} \in RT_{4,2}$ is

$$\begin{aligned}\det_c(\mathcal{A}) =& a_{1111}a_{2222} - a_{1112}a_{2221} - a_{1121}a_{2212} + a_{1211}a_{2122} \\ &+ a_{2111}a_{1222} + a_{1122}a_{2211} - a_{1212}a_{2121} - a_{1221}a_{2112}.\end{aligned}$$

Both the combinatorial determinant and the permanent of $\mathcal{A} \in RT_{N,I}$ are special cases of a more general function of tensors called the *tensor immanant*, generalized from that of a matrix [35].

Definition 5.1.5 Let $\lambda = (\lambda_1, \lambda_2, \dots)$ be a partition of I and χ_λ be the corresponding irreducible representation-theoretic character of the symmetric group $\mathbb{S}_I$, the immanant of $\mathcal{A} \in RT_{N,I}$ associated with the character χ_λ is defined as

$$\operatorname{Imm}_\lambda(\mathcal{A}) = \sum_{\pi_2 \in \mathbb{S}_I} \cdots \sum_{\pi_N \in \mathbb{S}_I} \chi_\lambda(\pi_2) \dots \chi_\lambda(\pi_N) \prod_{i=1}^{I} a_{i\pi_2(i)\dots\pi_N(i)}.$$

For the combinatorial determinant, $\chi_\lambda(\pi_n)$ is the alternating character of $\mathbb{S}_I$, defined by the parity of a permutation with $n \geq N - P + 1$, and the other χ_λ are identically 1. The permanent has χ_λ identically equal to 1.

More generally, let $\mathbb{G}$ be a subgroup of the symmetric group $\mathbb{S}_I$ of degree I and $\chi_n \in \mathbb{S}_I$ with $n = 2, 3, \dots, N$ be a linear character of $\mathbb{G}$. The generalized tensor

function associated with $\mathbb{G}$ and χ_n (also known as the $\mathbb{G}$-immanant) of $\mathcal{A} \in RT_{N,I}$ is defined as

$$d^{\mathbb{G}}_{\chi_2,\ldots,\chi_N}(\mathcal{A}) = \sum_{\pi_2,\ldots,\pi_N \in \mathbb{G}} \chi_2(\pi_2)\ldots\chi_N(\pi_N) \prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}.$$

For simplicity, we denote $d(\mathcal{A}) = d^{\mathbb{G}}_{\chi_2,\ldots,\chi_N}(\mathcal{A})$. When $N = 2$, $d^{\mathbb{G}}_{\chi_2,\ldots,\chi_N}(\mathcal{A})$ reduces to the generalized matrix function [36, 37]. Macrcus and Minc [38] obtain a relationship between the generalized matrix function and a function involving the eigenvalues of normal matrices. They reveal the relationship between the generalized matrix function and an appropriate function of the singular values. Berndt and Sra [39] obtain the generalized Hlawka and Popoviciu inequalities for generalized matrix function. Huang et al. [40] derive inequalities on the non-integer power of products of the generalized matrix function on the sum of positive semi-definite matrices. Chang et al. [41] present an inequality for the Kronecker product (or called tensor product) of positive operators on Hilbert spaces and then apply that to the generalized matrix function. Paksoy et al. [42] obtain some inequalities for the generalized matrix function of positive semi-definite matrices by an embedding and the Kronecker product.

It is known that there is a decomposable vector $\mathbf{x} \in \mathbb{C}^{I^N}$ such that $d(\mathcal{A}) = \left(\otimes^I \mathcal{A}\right)\mathbf{x}^N$. The Kronecker product of $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ and $\mathcal{B} \in \mathbb{R}^{J_1\times J_2\times\cdots\times J_N}$ yields a tensor $\mathcal{C} \in \mathbb{R}^{I_1J_1\times I_2J_2\times\cdots\times I_NJ_N}$ with entries [43]

$$C(j_1 + (i_1 - 1)I_1, j_2 + (i_2 - 1)I_2, \ldots, j_N + (i_N - 1)I_N) = a_{i_1i_2\ldots i_N}b_{j_1j_2\ldots j_N}$$

for all i_n, j_n and n. When $\mathcal{A} \in RT_{N,I}$ is diagonalizable and symmetric, Che et al. [44] investigate the inequalities on $d^{\mathbb{G}}_{\chi_2,\ldots,\chi_N}(\mathcal{A})$.

5.2 Sign Nonsingular Tensors

If there exists $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$ such that

$$\alpha = \operatorname{sign}(\pi_{N-P+1})\ldots\operatorname{sign}(\pi_N)\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} \neq 0,$$

where K is given in (5.1.2), then α is called a nonzero term in (5.1.1). The sign pattern of $\mathcal{A} \in RT_{N,I}$, denoted by $\operatorname{sign}(\mathcal{A})$, belongs to the set $RT_{N,I}$ with entries

$$(\operatorname{sign}(\mathcal{A}))_{i_1i_2\ldots i_N} = \begin{cases} 1, & \text{if } a_{i_1i_2\ldots i_N} > 0, \\ 0, & \text{if } a_{i_1i_2\ldots i_N} = 0, \\ -1, & \text{if } a_{i_1i_2\ldots i_N} < 0. \end{cases}$$

Let $Q(\mathcal{A}) = \{\mathcal{B} \in RT_{N,I} : \text{sign}(\mathcal{B}) = \text{sign}(\mathcal{A})\}$ be the set of tensors with the same sign pattern of $\mathcal{A}$. If $\det_c(\tilde{\mathcal{A}})$ and $\det_c(\mathcal{A})$ have the same sign for any $\tilde{\mathcal{A}} \in Q(\mathcal{A})$, then we say that $\mathcal{A}$ has a signed determinant. If $\det_c(\tilde{\mathcal{A}}) \neq 0$ for any $\tilde{\mathcal{A}} \in Q(\mathcal{A})$, then $\mathcal{A}$ is called a *sign nonsingular tensor* (differently defined in [45]). Note that singularity and sign-singularity of tensors do not imply each other. The following lemma extends [46, Lemma 1.2.4] from matrices to tensors.

Lemma 5.2.1 *Suppose that $\mathcal{A} \in RT_{N,I}$ with an even N. Then the tensor $\mathcal{A}$ has a signed determinant if and only if all nonzero terms in (5.1.1) have the same sign.*

Proof If all nonzero terms have the same sign, then $\mathcal{A}$ has a signed determinant. We shall prove that all nonzero terms in (5.1.1) have the same sign. Let

$$\alpha = \text{sign}(\pi_{N-K+1}) \dots \text{sign}(\pi_N) \prod_{i=1}^{I} a_{i\pi_2(i)\dots\pi_N(i)}$$

be a nonzero term in (5.1.1). The sign of $\det_c(\mathcal{A})$ is the same as $\text{sign}(\alpha)$ if we multiply each entry of $\mathcal{A}$ not occurring in α by a sufficiently small positive number. The converse also follows similarly. □

The following theorem illustrates the relationship between the sign nonsingular tensors and the signed determinant, which extends [46, Theorem 1.2.5] from matrices to tensors.

Theorem 5.2.1 *Assume that $\mathcal{A} \in RT_{N,I}$ with an even N. The following are equivalent:*

(1) *$\mathcal{A}$ is a sign nonsingular tensor.*
(2) *$\det_c(\mathcal{A}) \neq 0$ and $\mathcal{A}$ has a signed determinant.*
(3) *All nonzero terms in (5.1.1) have the same sign.*

Proof Since $Q(\mathcal{A})$ is a connected set and $\det_c(\mathcal{A})$ is continuous, then (1) is equivalent to (2). By Lemma 5.2.1, we have (2) being equivalent to (3). □

We reveal the relationship between the combinatorial determinant and the permanent of nonnegative tensors, summarized in the following theorem.

Theorem 5.2.2 *Let $\mathcal{A} \in RT_{N,I}$ be nonnegative with an even N and its combinatorial determinant is nonzero. Then $|\det_c(\mathcal{A})| \leq \text{perm}(\mathcal{A})$, with equality if and only if $\mathcal{A}$ is sign nonsingular.*

Proof Since $\det_c(\mathcal{A}) \neq 0$ with an even N and there exist nonzero terms in (5.1.1), then we get $|\det_c(\mathcal{A})| \leq \text{perm}(\mathcal{A})$, with equality if and only if all nonzero terms in (5.1.1) have the same sign. By Theorem 5.2.1, we have $|\det_c(\mathcal{A})| = \text{perm}(\mathcal{A})$ if and only if $\mathcal{A}$ is a sign nonsingular nonnegative tensor. □

5.3 Properties of Plane Stochastic Tensors

Christensen and Fischer [17] indicate that when considering error-correcting codes, we need to study the extremal points for the set of all T-flat stochastic tensors [3] with a fixed positive integer $1 \leq T \leq N-1$. A doubly stochastic matrix always has a positive diagonal [47, Lemma 2.1.5]. However, for a plane stochastic tensor, this result is not true. The following example illustrates the fact.

Example 5.3.1 Let $\mathcal{A} \in NT_{3,2}$ be a plane stochastic tensor with

$$a_{111} = 0.5, \quad a_{122} = 0.5, \quad a_{221} = 0.5, \quad a_{212} = 0.5,$$

and zero elsewhere. For diagonals of $\mathcal{A}$, we have

$$\{a_{111}, a_{222}\} = \{0.5, 0\}, \quad \{a_{122}, a_{211}\} = \{0.5, 0\},$$
$$\{a_{112}, a_{221}\} = \{0, 0.5\}, \quad \{a_{121}, a_{212}\} = \{0, 0.5\}.$$

Then $\mathcal{A}$ has no positive diagonal.

Generally speaking, all elements in $P_{N,I}$ are extremal points of $\Omega_{N,I}$, and the set of all plane stochastic tensors, which also has other extremal points [3]. Please refer to [1–4, 10, 48–54] on the extremal points for the set of all T-flat stochastic tensors. Now we consider some interesting properties of a special subset of $\Omega_{N,I}$:

$$\Omega_{N,I}^{+} = \left\{ \mathcal{A} \in \Omega_{N,I} : \mathcal{A} = \sum_{i=1}^{\tilde{M}} \lambda_i \mathcal{P}_i, \sum_{i=1}^{\tilde{M}} \lambda_i = 1, 0 \leq \lambda_i \leq 1, i = 1, 2, \ldots, \tilde{M} \right\}, \tag{5.3.1}$$

where $\mathcal{P}_i \in P_{N,I}$ and $\tilde{M} \leq (I!)^{N-1}$.

We call a tensor in $\Omega_{N,I}^{+}$ *totally plane stochastic*. A tensor $\mathcal{A} \in RT_{N,I}$ has a totally plane stochastic pattern, if there exists a totally plane stochastic tensor in $\Omega_{N,I}^{+}$ with the same pattern as $\mathcal{A}$.

For a given $\mathcal{A} \in NT_{N,I}$, if perm($\mathcal{A}$) is positive, according to the definition of the tensor permanent, then there exists a positive diagonal of $\mathcal{A}$. It is obvious that any totally plane stochastic tensor has a positive diagonal. A plane stochastic tensor with a positive permanent does not always belong to $\Omega_{N,I}^{+}$, as illustrated below.

Example 5.3.2 Let $\mathcal{A} \in NT_{3,2}$ be plane stochastic with

$$\mathcal{A} = \sum_{i=1}^{4} \lambda_i \mathcal{P}_i + \lambda_5 \mathcal{Q}_1 + \lambda_6 \mathcal{Q}_2, \quad \sum_{i=1}^{6} \lambda_i = 1, 0 < \lambda_i \leq 1,$$

where $\mathcal{P}_i \in P_{3,2}$, and all entries of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are zero except

$$(\mathcal{Q}_1)_{111} = 0.5, \quad (\mathcal{Q}_1)_{122} = 0.5, \quad (\mathcal{Q}_1)_{221} = 0.5, \quad (\mathcal{Q}_1)_{212} = 0.5,$$
$$(\mathcal{Q}_2)_{121} = 0.5, \quad (\mathcal{Q}_2)_{112} = 0.5, \quad (\mathcal{Q}_2)_{211} = 0.5, \quad (\mathcal{Q}_2)_{222} = 0.5.$$

Manipulation yields

$$\text{perm}(\mathcal{P}_i) = 1, \quad \text{perm}(\mathcal{Q}_1) = \text{perm}(\mathcal{Q}_2) = 0.$$

Thus $\text{perm}(\mathcal{A}) = \sum_{i=1}^{4} \lambda_i > 0$. However, by (5.3.1), it is obvious that $\mathcal{A} \notin \Omega_{3,2}^{+}$.

5.3.1 Totally Plane Stochastic Tensors

For the plane stochastic and totally plane stochastic tensors, we have the following result.

Lemma 5.3.1 *The set $\Omega_{N,I}$, viewed as a subset of $\mathbb{R}^{I^N}$, is closed, bounded and convex.*

Proof Let $\mathcal{A}, \mathcal{B} \in \Omega_{N,I}$. For any $0 \leq \lambda \leq 1$, the entries of $\lambda\mathcal{A} + (1-\lambda)\mathcal{B}$ are nonnegative. For any i_1, we have

$$\lambda \sum_{i_2,\ldots,i_N=1}^{I} a_{i_1 i_2 \ldots i_N} + (1-\lambda) \sum_{i_2,\ldots,i_N=1}^{I} b_{i_1 i_2 \ldots i_N} = 1.$$

Similarly, its i_nth mode-n sums are unity for all n. Hence $\Omega_{N,I}$ is a convex set. Because $0 \leq a_{i_1 i_2 \ldots i_N} \leq 1$ for all i_n and n, $\Omega_{N,I}$ is bounded.

To show that $\Omega_{N,I}$ is closed, without loss of generality, we assume that $N = 3$. Let

$$\begin{cases} \mathbb{S}_1 = \{\mathcal{A} \in NT_{3,I} : \sum\limits_{i_2,i_3=1}^{I} a_{i_1 i_2 i_3} = 1, i_1 = 1, 2, \ldots, I\}, \\ \mathbb{S}_2 = \{\mathcal{A} \in NT_{3,I} : \sum\limits_{i_1,i_3=1}^{I} a_{i_1 i_2 i_3} = 1, i_2 = 1, 2, \ldots, I\}, \\ \mathbb{S}_3 = \{\mathcal{A} \in NT_{3,I} : \sum\limits_{i_1,i_2=1}^{I} a_{i_1 i_2 i_3} = 1, i_3 = 1, 2, \ldots, I\}. \end{cases}$$

Note that $\Omega_{3,I} = \mathbb{S}_1 \cap \mathbb{S}_2 \cap \mathbb{S}_3$. Next we prove that $\mathbb{S}_n$ $(n = 1, 2, 3)$ are closed sets. For $\mathcal{A} \in \mathbb{S}_1$, let

$$\mathbb{U}_\delta(\mathcal{A}) = \{\mathcal{B} = \mathcal{A} + \mathcal{E} \in NT_{N,I} : \sum_{i_1,i_2,i_3=1}^{I} |e_{i_1 i_2 i_3}| < \delta\}.$$

Then $\mathbb{U}_\delta(\mathcal{A})$ contains neighborhood elements of $\mathcal{A}$. We denote the complement of $\mathbb{S}_1$ as $\mathbb{S}_1^c$:

$$\mathbb{S}_1^c = \{\mathcal{A} \in NT_{3,I} : \sum_{i_2,i_3=1}^{I} a_{i_1 i_2 i_3} \neq 1, \text{ for some } i_1\}.$$

Assume that $\mathcal{A} \in \mathbb{S}_1^c$ and $\sum_{i_2,i_3=1}^{I} a_{i_1^0 i_2 i_3} \neq 1$, let $\epsilon = |\sum_{i_2,i_3=1}^{I} a_{i_1^0 i_2 i_3} - \delta|$. Let us set $0 < \delta_0 < \epsilon$. We find that $\mathbb{U}_{\delta_0}(\mathcal{A})$ contains neighborhood elements of $\mathcal{A}$ and $\mathbb{U}_{\delta_0}(\mathcal{A}) \subseteq \mathbb{S}_1^c$. Then $\mathbb{S}_1^c$ is an open set, thus $\mathbb{S}_1$ is closed. Similarly, $\mathbb{S}_2$, $\mathbb{S}_3$ and $\Omega_{N,I}$ are closed sets. □

Since $\Omega_{N,I}^+$ is a subset of $\Omega_{N,I}$, it is closed, bounded and convex. A point $\mathbf{x}_0$ of a convex set $\mathbb{S} \subset \mathbb{R}^I$ is called an *extreme point* of $\mathbb{S}$, if $\mathbb{S}\backslash\{\mathbf{x}_0\}$ is also convex. The next result identifies the extreme points of $\Omega_{N,I}^+$.

Lemma 5.3.2 *Any plane permutation tensor $\mathcal{P} \in P_{N,I}$ is an extreme point of $\Omega_{N,I}^+$.*

Proof Let $\mathcal{P} \in P_{N,I}$. If $\mathcal{P}$ is not an extreme point of $\Omega_{N,I}^+$, then we have $\mathcal{P} = (\mathcal{A} + \mathcal{B})/2$ with $\mathcal{A} \neq \mathcal{B}$ and $\mathcal{A}, \mathcal{B} \in \Omega_{N,I}^+$. As shown in [47, Lemma 2.1.2], for all i_n and n, the values of $p_{i_1 i_2 \ldots i_N}$ are 0 or 1; and $0 \leq a_{i_1 i_2 \ldots i_N}, b_{i_1 i_2 \ldots i_N} \leq 1$. If $p_{i_1 i_2 \ldots i_N} = 0$, then $a_{i_1 i_2 \ldots i_N} = b_{i_1 i_2 \ldots i_N} = 0$ and if $p_{i_1 i_2 \ldots i_N} = 1$, then $a_{i_1 i_2 \ldots i_N} = b_{i_1 i_2 \ldots i_N} = 1$. This contradicts the assumption that $\mathcal{A} \neq \mathcal{B}$. □

In following lemma, we show that the upper bound of $\tilde{M}$ in (5.3.1) can be reduced.

Lemma 5.3.3 *$\tilde{M}$ in (5.3.1) satisfies $\tilde{M} \leq I^N - NI + N$.*

Proof For any given $\mathcal{A} \in \Omega_{N,I}$, by Definition 5.1.1, we have

$$f(\mathcal{A}; i_n, n) = 1, \tag{5.3.2}$$

for all i_n and n, and

$$\sum_{i_1,i_2,\ldots,i_N=1}^{I} a_{i_1 i_2 \ldots i_N} = I. \tag{5.3.3}$$

Substituting (5.3.3) into (5.3.2), we transform (5.3.2) to a linear system with I^N unknowns and $NI-N+1$ equations. This linear system has at most $I^N-NI+N-1$ free variables. □

By Lemma 5.3.3, the set $\Omega^+_{N,I}$ is given in (5.3.1) with $\mathcal{P}_i \in P_{N,I}$ and $\tilde{M} \leq I^N - NI + N$.

Lemma 5.3.4 *Suppose that $\mathcal{A} \in NT_{3,I}$ satisfies all i_nth mode-n sums not exceeding 1 for all i_n and $n = 1, 2, 3$. If $f(\mathcal{A}; 1, 1) < 1$ and $f(\mathcal{A}; i, 1) = 1$ for all i, then there exist $j, k \in \{1, 2, \ldots, I\}$ such that $f(\mathcal{A}; j, 2) < 1$ and $f(\mathcal{A}; k, 3) < 1$.*

Remark 5.3.1 Let $\mathcal{A}$ be the same as in Lemma 5.3.4, we have the following results.

(a) If $f(\mathcal{A}; 1, 2) < 1$ and $f(\mathcal{A}; j, 2) = 1$ for $j = 2, 3, \ldots, I$, then there exist $i, k \in \{1, 2, \ldots, I\}$ such that $f(\mathcal{A}; i, 1) < 1$ and $f(\mathcal{A}; k, 3) < 1$.
(b) If $f(\mathcal{A}; 1, 3) < 1$ and $f(\mathcal{A}; k, 3) = 1$ for $k = 2, 3, \ldots, I$, then there exist $i, j \in \{1, 2, \ldots, I\}$ such that $f(\mathcal{A}; i, 1) < 1$ and $f(\mathcal{A}; j, 2) < 1$.

From Lemma 5.3.4, it is obvious to derive the following corollary.

Corollary 5.3.1 *Assume that $\mathcal{A} \in NT_{N,I}$ satisfies all i_nth mode-n sums not exceeding 1 for all i_n and n. For a given i_n, if $f(\mathcal{A}; i_n, n) < 1$ and $f(\mathcal{A}; i_m, m) = 1$ for all $i_m \neq i_n$, then for all $t \in \{1, \ldots, n-1, n+1, \ldots, N\}$, there exists $i_t \in \{1, 2, \ldots, I\}$ such that $f(\mathcal{A}; i_t, t) < 1$.*

Theorem 5.3.1 *Suppose that the i_nth mode-n sum of $\mathcal{A} \in NT_{N,I}$ does not exceed 1 for all i_n and n. Then there exists a $\mathcal{D} \in \Omega_{N,I}$ such that $\mathcal{D} \geq \mathcal{A}$.*

Proof If $\mathcal{A} \in NT_{N,I}$ is plane stochastic, then the result is trivial. If $\mathcal{A} \in T_{N,I}$ is not plane stochastic, according to Corollary 5.3.1, for all n, then there exists $r_n \in \{1, 2, \ldots, I\}$ such that $f(\mathcal{A}; r_n, n) < 1$. Let $u_{r_n} := f(\mathcal{A}; r_n, n) < 1$ and $\delta = \min\{1 - u_{r_1}, 1 - u_{r_2}, \ldots, 1 - u_{r_N}\}$. Adding δ to $a_{r_1 r_2 \ldots r_N}$ and keeping other entries unchanged, we derive a new tensor $\mathcal{A}^{(1)}$ such that $\mathcal{A} \leq \mathcal{A}^{(1)}$. Thus $\mathcal{A}^{(1)}$ is nonnegative with all the i_nth mode-n sums at most 1. Furthermore, there exists $r_n \in \{1, 2, \ldots, I\}$ for all n such that the r_n-th mode-n sum is less than 1 in $\mathcal{A}$, but equals to 1 in $\mathcal{A}^{(1)}$.

If $\mathcal{A}^{(1)}$ is plane stochastic, then stop; otherwise, the above procedure can be repeated until we get a plane stochastic tensor $\mathcal{D} \geq \mathcal{A}$. □

The following theorem without the proof indicates that $\mathcal{A} \in NT_{N,I}$ has a totally plane stochastic pattern under certain conditions.

Theorem 5.3.2 *Let $\mathcal{A} \in RT_{N,I}$ be a nonzero nonnegative tensor. Then $\mathcal{A}$ has a totally plane stochastic pattern if and only if any positive entry of $\mathcal{A}$ is in a positive diagonal.*

5.3.2 A Self Map on Totally Plane Stochastic Tensors

Let $\mathcal{A} \in RT_{N,I}$ be a nonnegative tensor with $\text{perm}(\mathcal{A}) > 0$. We define the tensor $f(\mathcal{A}) \in RT_{N,I}$ by

$$(f(\mathcal{A}))_{i_1 i_2 \ldots i_N} = \frac{a_{i_1 i_2 \ldots i_N} \text{perm}(\mathcal{A}(i_1, i_2, \ldots, i_N))}{\text{perm}(\mathcal{A})}, \tag{5.3.4}$$

for all i_n and n. According to Definition 5.1.1 and the definition of the permanent, we have $f(\mathcal{A}) \in \Omega^+_{N,I}$ for all $\mathcal{A} \in \Omega^+_{N,I}$. The map f and its restriction to $\Omega^+_{N,I}$ have interesting properties.

Lemma 5.3.5 *If $\mathcal{A} \in \Omega^+_{N,I}$, then $\mathcal{A}$ and $f(\mathcal{A})$ have the same totally plane stochastic pattern.*

Proof If $\mathcal{A} \in \Omega^+_{N,I}$, then $\text{perm}(\mathcal{A}) > 0$ and $f(\mathcal{A})$ is well defined. If $a_{i_1 i_2 \ldots i_N} = 0$, then obviously $f_{i_1 i_2 \ldots i_N}(\mathcal{A}) = 0$. If $a_{i_1 i_2 \ldots i_N} > 0$, then by Theorem 5.3.2, there exists $\pi_n \in \mathbb{S}_I$ $(n = 2, 3, \ldots, N)$ such that $\pi_n(i) = i_n$ and $\prod_{i=1}^{I} a_{i \pi_2(i) \ldots \pi_N(i)} > 0$. Thus $\text{perm}(\mathcal{A}(i_1, i_2, \ldots, i_N)) > 0$ and $\text{perm}(\mathcal{A}) > 0$. This shows that $(f(\mathcal{A}))_{i_1 i_2 \ldots i_N} > 0$. □

Theorem 5.3.3 *Let $K_{N,I} \subset \Omega^+_{N,I}$ be the set of all totally plane stochastic tensors whose entries satisfy a given set of equality constraints. Suppose $f(\mathcal{A}) \in K_{N,I}$ whenever $\mathcal{A} \in K_{N,I}$, where f is the map defined by (5.3.4). Then f maps $K_{N,I}$ onto $K_{N,I}$. In particular, each of the following sets is mapped onto itself by f:*

(i) $\Omega^+_{N,I}$.
(ii) *The set of symmetric tensors in $\Omega^+_{N,I}$.*
(iii) *The set of tensors in $\Omega^+_{N,I}$ with the first mode-1 slice equal to the second mode-1 slice.*

Proof By Lemma 2.7.3 in [47], Theorem 5.3.2 and Lemma 5.3.5, for any $\mathcal{A} \in \Omega^+_{N,I}$, $a_{i_1 i_2 \ldots i_N} = 0$ if and only if $(f(\mathcal{A}))_{i_1 i_2 \ldots i_N} = 0$ for all i_n and n. The sets in (i)–(iii) are all defined by imposing equality constraints on the entries of the tensor. The results are easily verified. □

Theorem 5.3.4 *The map $f : \Omega^+_{N,I} \to \Omega^+_{N,I}$ is one-to-one.*

Proof Let $f(\mathcal{A}) = f(\mathcal{B})$ for $\mathcal{A}$ and $\mathcal{B}$ in $\Omega^+_{N,I}$. Since $\text{perm}(\mathcal{A}) > 0$ and $\text{perm}(\mathcal{B}) > 0$, then the tensor C is well defined with

$$c_{i_1 i_2 \ldots i_N} = \left(\frac{\text{perm}(\mathcal{A})}{\text{perm}(\mathcal{B})} \right)^{1/I} b_{i_1 i_2 \ldots i_N}.$$

From $f(\mathcal{A}) = f(\mathcal{B})$, we have

$$\frac{a_{i_1 i_2 \ldots i_N} \text{perm}(\mathcal{A}(i_1, i_2, \ldots, i_N))}{\text{perm}(\mathcal{A})} = \frac{b_{i_1 i_2 \ldots i_N} \text{perm}(\mathcal{B}(i_1, i_2, \ldots, i_N))}{\text{perm}(\mathcal{B})}.$$

Therefore

$$a_{i_1 i_2 \ldots i_N} \mathrm{perm}(\mathcal{A}(i_1, i_2, \ldots, i_N)) = c_{i_1 i_2 \ldots i_N} \mathrm{perm}(\mathcal{C}(i_1, i_2, \ldots, i_N))$$

for all i_n and n.

For $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$, let

$$\alpha_{\pi_2,\ldots,\pi_N} = \prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}, \quad \beta_{\pi_2,\ldots,\pi_N} = \prod_{i=1}^{I} c_{i\pi_2(i)\ldots\pi_N(i)},$$

we have

$$\begin{aligned}
\prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \alpha_{\pi_2,\ldots,\pi_N}^{\alpha_{\pi_2,\ldots,\pi_N}} &= \prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \left(\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} \right)^{\alpha_{\pi_2,\ldots,\pi_N}} \\
&= \prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)}^{\alpha_{\pi_2,\ldots,\pi_N}} \\
&= \prod_{i=1}^{I} \prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} a_{i\pi_2(i)\ldots\pi_N(i)}^{\alpha_{\pi_2,\ldots,\pi_N}}. \qquad (5.3.5)
\end{aligned}$$

Recall the convention of $0^0 = 1$. Let $\mathbb{K}_{ii_2\ldots i_N} = \{(\pi_2, \pi_3, \ldots, \pi_N) : \pi_2(i) = i_2, \pi_3(i) = i_3, \ldots, \pi_N(i) = i_N\}$. Then

$$\begin{aligned}
\prod_{i=1}^{I} \prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} a_{i\pi_2(i)\ldots\pi_N(i)}^{\alpha_{\pi_2,\ldots,\pi_N}} &= \prod_{i,i_2,\ldots,i_N=1}^{I} \prod_{(\pi_2,\ldots,\pi_N) \in \mathbb{K}_{ii_2\ldots i_N}} a_{i\pi_2(i)\ldots\pi_N(i)}^{\alpha_{\pi_2,\ldots,\pi_N}} \\
&= \prod_{i,i_2,\ldots,i_N=1}^{I} a_{ii_2\ldots i_N}^{\sum_{(\pi_2,\ldots,\pi_N) \in \mathbb{K}_{ii_2\ldots i_N}} \alpha_{\pi_2,\ldots,\pi_N}} \\
&= \prod_{i,i_2,\ldots,i_N=1}^{I} a_{ii_2\ldots i_N}^{a_{ii_2\ldots i_N} \mathrm{perm}(\mathcal{A}(i,i_2,\ldots,i_N))} \\
&= \prod_{i,i_2,\ldots,i_N=1}^{I} a_{ii_2\ldots i_N}^{c_{ii_2\ldots i_N} \mathrm{perm}(\mathcal{C}(i,i_2,\ldots,i_N))} \\
&= \prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \left(\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} \right)^{\beta_{\pi_2,\ldots,\pi_N}} \\
&= \prod_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \alpha_{\pi_2,\ldots,\pi_N}^{\beta_{\pi_2,\ldots,\pi_N}}.
\end{aligned}$$

From (5.3.5), we obtain

$$\prod_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \alpha_{\pi_2,\ldots,\pi_N}^{\alpha_{\pi_2,\ldots,\pi_N}} = \prod_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \alpha_{\pi_2,\ldots,\pi_N}^{\beta_{\pi_2,\ldots,\pi_N}}, \tag{5.3.6}$$

we prove in a similar way,

$$\prod_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \beta_{\pi_2,\ldots,\pi_N}^{\alpha_{\pi_2,\ldots,\pi_N}} = \prod_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \beta_{\pi_2,\ldots,\pi_N}^{\beta_{\pi_2,\ldots,\pi_N}}. \tag{5.3.7}$$

Combining (5.3.6) and (5.3.7), we obtain

$$\prod_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \left(\frac{\alpha_{\pi_2,\ldots,\pi_N}}{\beta_{\pi_2,\ldots,\pi_N}}\right)^{\alpha_{\pi_2,\ldots,\pi_N}} \cdot \left(\frac{\beta_{\pi_2,\ldots,\pi_N}}{\alpha_{\pi_2,\ldots,\pi_N}}\right)^{\beta_{\pi_2,\ldots,\pi_N}} = 1.$$

Note that

$$\operatorname{perm}(\mathcal{A}) = \sum_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \alpha_{\pi_2,\ldots,\pi_N} = \sum_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \beta_{\pi_2,\ldots,\pi_N}.$$

By [47, Lemma 2.6.2], we have $\alpha_{\pi_2,\ldots,\pi_N} = \beta_{\pi_2,\ldots,\pi_N}$ for all $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$, i.e.,

$$\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} = \prod_{i=1}^{I} c_{i\pi_2(i)\ldots\pi_N(i)} = \frac{\operatorname{perm}(\mathcal{A})}{\operatorname{perm}(\mathcal{B})} \prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)},$$

By Corollary 5.4.1, this is possible only if $\mathcal{A} = \mathcal{B}$. □

We say that tensors $\mathcal{A}$ and $\mathcal{B}$ have *proportional diagonal products*, if for all $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$, there exists a nonzero constant α such that

$$\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} = \alpha \prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)}.$$

By Theorems 5.3.3, 5.3.4 and Corollary 5.4.1, we have the following theorem.

Theorem 5.3.5 *If $\mathcal{A} \in RT_{N,I}$ is a nonnegative tensor with a totally plane stochastic pattern, then there exists a unique totally plane stochastic tensor $\mathcal{B}$ such that $\mathcal{A}$ and $\mathcal{B}$ have proportional diagonal products.*

5.3.3 Relationship Between Nonnegative and Plane Stochastic Tensors

A tensor $\mathcal{A} \in RT_{N,I}$ has a plane stochastic pattern if there exists a plane stochastic tensor $\mathcal{B} \in \Omega_{N,I}$ such that

$$a_{i_1 i_2 \ldots i_N} = 0 \quad \Leftrightarrow \quad b_{i_1 i_2 \ldots i_N} = 0,$$

for all i_n and n.

Sinkhorn [55] derives a relationship between arbitrary positive matrices and doubly stochastic matrices. For a positive matrix $\mathbf{A} \in \mathbb{R}^{I \times I}$, we can derive a matrix sequence by alternately normalizing the rows and columns of $\mathbf{A}$ and show that the limit of the sequence is a doubly stochastic matrix. Sinkhorn's theorem can be generalized from positive matrices to certain nonnegative matrices [56, 57]. Furthermore, Bapat [58] generalizes Sinkhorn's theorem from matrices to multi-dimensional matrices (i.e., tensors). Also, Raghavan [59] and Franklin and Lorenz [60] consider the relationship between nonnegative and plane stochastic tensors.

Theorem 5.3.6 ([60, Theorem 3]) *Let $\mathcal{A} \in RT_{N,I}$ be nonnegative with a plane stochastic pattern. Then there exists N positive diagonal matrices $\mathbf{D}_n \in \mathbb{R}^{I \times I}$ such that $\mathcal{A} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \cdots \times_N \mathbf{D}_N$ is plane stochastic.*

In the following, we consider how to numerically implement the normalization process. Starting from a symmetric nonnegative tensor with certain conditions, Shashua et al. [61] propose a normalization algorithm which converges to a symmetric plane stochastic tensor.

Theorem 5.3.7 ([61, Proposition 2]) *For any symmetric $\mathcal{A}^{(0)} \in NT_{N,I}$ without vanishing slices, the iterative process:*

$$a^{(k+1)}_{i_1 i_2 \ldots i_N} = \frac{a^{(k)}_{i_1 i_2 \ldots i_N}}{(a^{(k)}_{i_1} a^{(k)}_{i_2} \ldots a^{(k)}_{i_N})^{1/N}}, \tag{5.3.8}$$

where for all i,

$$a^{(k)}_{i} = \sum_{i_2, \ldots, i_N = 1}^{I} a^{(k)}_{i i_2 \ldots i_N} \tag{5.3.9}$$

converges to a symmetric plane stochastic tensor.

As shown in Example 5.3.1, the permanent of all plane stochastic tensors in $\Omega_{N,I}$ cannot be always positive. Hence, the original proof of Theorem 5.3.7 in [61] is incomplete. We have the following more general theorem.

Theorem 5.3.8 *For any $\mathcal{A}^{(0)} \in NT_{N,I}$, without vanishing slices, the iterative process:*

$$a^{(k+1)}_{i_1 i_2 \ldots i_N} = \frac{a^{(k)}_{i_1 i_2 \ldots i_N}}{(a^{(k)}_{1,i_1} a^{(k)}_{2,i_2} \ldots a^{(k)}_{N,i_N})^{1/N}}, \quad k = 0, 1, \ldots, \tag{5.3.10}$$

where for all i_n and n,

$$a^{(k)}_{n,i_n} = \sum_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N=1}^{I} a^{(k)}_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} > 0 \tag{5.3.11}$$

converges to a plane stochastic tensor.

Proof Suppose that $\mathrm{perm}(\mathcal{A}^{(0)}) > 0$. By the definition of the permanent and one step of the normalization step described in (5.3.10), we have

$$\begin{aligned}
\mathrm{perm}(\mathcal{A}^{(k+1)}) &= \sum_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \prod_{i=1}^{I} a^{(k+1)}_{i \pi_2(i) \ldots \pi_N(i)} \\
&= \sum_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \prod_{i=1}^{I} \frac{a^{(k)}_{i_1 i_2 \ldots i_N}}{(a^{(k)}_{1,i_1} a^{(k)}_{2,i_2} \ldots a^{(k)}_{N,i_N})^{1/N}} \cdot \\
&= \frac{1}{\left(\prod_{i=1}^{I} a^{(k)}_{1,i} a^{(k)}_{2,i} \ldots a^{(k)}_{N,i} \right)^{1/N}} \mathrm{perm}(\mathcal{A}^{(k)}).
\end{aligned}$$

To show that the normalization scheme produces non-decreasing the permanents we need to prove $\prod_{i=1}^{I} a^{(k)}_{1,i} a^{(k)}_{2,i} \ldots a^{(k)}_{N,i} \leq 1$. From the arithmetic-geometric means inequality it is sufficient to show that $\sum_{i=1}^{I} (a^{(k)}_{1,i} + a^{(k)}_{2,i} + \cdots + a^{(k)}_{N,i}) \leq NI$. From the definition of $a^{(k)}_{n,i}$, we have

$$\begin{aligned}
&\sum_{i=1}^{I} (a^{(k)}_{1,i} + a^{(k)}_{2,i} + \cdots + a^{(k)}_{N,i}) \\
&\quad = \sum_{i,i_2,\ldots,i_N=1}^{I} a^{(k-1)}_{i i_2 \ldots i_N} \frac{1}{\left(a^{(k-1)}_{1,i} a^{(k-1)}_{2,i_2} \ldots a^{(k-1)}_{N,i_N} \right)^{1/N}} \\
&\qquad + \sum_{i_1,i,\ldots,i_N=1}^{I} a^{(k-1)}_{i_1 i \ldots i_N} \frac{1}{\left(a^{(k-1)}_{1,i_1} a^{(k-1)}_{2,i} \ldots a^{(k-1)}_{N,i_N} \right)^{1/N}} + \ldots \\
&\qquad + \sum_{i_1,\ldots,i_{N-1},i=1}^{I} a^{(k-1)}_{i_1 \ldots i_{N-1} i} \frac{1}{\left(a^{(k-1)}_{1,i_1} \ldots a^{(k-1)}_{N-1,i_{N-1}} a^{(k-1)}_{N,i} \right)^{1/N}}.
\end{aligned}$$

From the inequality $(\prod_{j=1}^{J} x_j)^{1/J} \le (1/J)\sum_{j=1}^{J} x_j$, replacing x_j with $1/a_j$ (recall that $a_j > 0$), we obtain

$$\frac{1}{(a_1 a_2 \dots a_J)^{1/J}} \le \frac{1}{J}\left(\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_J}\right).$$

We obtain $\sum_{i=1}^{I}(a_{1,i}^{(k)} + a_{2,i}^{(k)} + \dots + a_{N,i}^{(k)}) \le NI$. Therefore, we conclude that each step of the normalization scheme does not decrease the tensor of the previous step. Furthermore, we have

$$\operatorname{perm}(\mathcal{A}^{(k+1)}) \le \left(\prod_{i=1}^{I} a_{1,i}^{(k)} a_{2,i}^{(k)} \dots a_{N,i}^{(k)}\right)^{1/N} \le 1.$$

For the case of $\operatorname{perm}(\mathcal{A}^{(0)}) = 0$, we consider $\mathcal{B}^{(0)} := \mathcal{A}^{(0)} + \epsilon\mathcal{I}$ with any arbitrary small $\epsilon > 0$. Note that $\operatorname{perm}(\mathcal{B}^{(0)}) > 0$. The entries of $\mathcal{B}^{(k)}$ are given by

$$b_{i_1 i_2 \dots i_N}^{(k+1)} = \frac{b_{i_1 i_2 \dots i_N}^{(k)}}{(b_{1,i_1}^{(k)} b_{2,i_2}^{(k)} \dots b_{N,i_N}^{(k)})^{1/N}}, \quad k = 1, 2, \dots,$$

where for all i_n and n

$$b_{n,i_n}^{(k)} = \sum_{i_1,\dots,i_{n-1},i_{n+1},\dots,i_N=1}^{I} b_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N}^{(k)} > 0.$$

The sequence $\{\mathcal{B}^{(k)}\}$ $(k = 0, 1, 2, \dots)$ converges to a plane stochastic tensor. Furthermore, as $\epsilon \to 0$, we have $b_{i_1 i_2 \dots i_N}^{(1)} \to a_{i_1 i_2 \dots i_N}^{(1)}$ and $b_{i_1}^{(0)} \to a_{i_1}^{(0)}$ for all i_n and n, where $a_{i_1 i_2 \dots i_N}^{(1)}$ and $a_{i_1}^{(0)}$ are respectively generated from (5.3.10) and (5.3.11) with $k = 0$.

For each $k = 0, 1, \dots$, as $\epsilon \to 0$, we have $b_{i_1 i_2 \dots i_N}^{(k+1)} \to a_{i_1 i_2 \dots i_N}^{(k+1)}$ and $b_{i_1}^{(k)} \to a_{i_1}^{(k)}$ for all i_n and n, where $a_{i_1 i_2 \dots i_N}^{(k+1)}$ and $a_{i_1}^{(k)}$ are respectively generated from (5.3.10) and (5.3.11). The sequence $\{\mathcal{A}^{(k)}\}$ $(k = 0, 1, 2, \dots)$, generated by (5.3.10), converges to a plane stochastic tensor. □

For a given $\mathcal{A} \in \mathbb{R}_{+}^{I_1 \times I_2 \times \dots \times I_N}$ and all n, if we have

$$\sum_{i_1=1}^{I_1} \dots \sum_{i_{n-1}=1}^{I_{n-1}} \sum_{i_{n+1}=1}^{I_{n+1}} \dots \sum_{i_N=1}^{I_N} a_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} = 1,$$

then we call $\mathcal{A}$ a plane stochastic tensor. We have the following conjecture.

Conjecture 5.3.1 For any $\mathcal{A}^{(0)} \in \mathbb{R}_{+}^{I_1\times I_2\times\cdots\times I_N}$ without vanishing slices, the iterative process:

$$a_{i_1i_2\ldots i_N}^{(k+1)} = \frac{a_{i_1i_2\ldots i_N}^{(k)}}{(a_{1,i_1}^{(k)} a_{2,i_2}^{(k)} \ldots a_{N,i_N}^{(k)})^{1/N}}, \quad k = 0, 1, \ldots$$

where

$$a_{n,i_n}^{(k)} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_{n-1}=1}^{I_{n-1}} \sum_{i_{n+1}=1}^{I_{n+1}} \cdots \sum_{i_N=1}^{I_N} a_{i_1\ldots i_{n-1}i_ni_{n+1}\ldots i_N}^{(k)} > 0,$$

for all i_n and n, converges to a plane stochastic tensor.

From [62], consider the symmetric $\mathcal{A} \in NT_{3,3}$ with entries

$$\begin{cases} a_{111} = 0.0517, & a_{112} = 0.3579, & a_{113} = 0.5298, & a_{122} = 0.7544, & a_{123} = 0.2156, \\ a_{133} = 0.3612, & a_{222} = 0.3943, & a_{223} = 0.0146, & a_{233} = 0.6718, & a_{333} = 0.9723. \end{cases}$$

and the positive $\mathcal{B} \in NT_{3,3}$ with

$$\begin{cases} b_{111} = 0.4333, & b_{121} = 0.4278, & b_{131} = 0.4140, & b_{211} = 0.8154, & b_{221} = 0.0199, \\ b_{231} = 0.5598, & b_{311} = 0.0643, & b_{321} = 0.3815, & b_{331} = 0.8834, & b_{112} = 0.4866, \\ b_{122} = 0.8087, & b_{132} = 0.2073, & b_{212} = 0.7641, & b_{222} = 0.9924, & b_{232} = 0.8752, \\ b_{312} = 0.6708, & b_{322} = 0.8296, & b_{332} = 0.1325, & b_{113} = 0.3871, & b_{123} = 0.0769, \\ b_{133} = 0.3151, & b_{213} = 0.1355, & b_{223} = 0.7727, & b_{233} = 0.4089, & b_{313} = 0.9715, \\ b_{323} = 0.7726, & b_{333} = 0.5526. \end{cases}$$

We illustrate Theorems 5.3.7 and 5.3.8 via $\mathcal{A}$ and $\mathcal{B}$, respectively.

Suppose that $\mathbf{e} = (1, 1, 1)^\top$. At the kth step, we define $\gamma^{(k)} = \|\mathbf{e} - \mathbf{a}^{(k)}\|_2$ for $\mathcal{A}^{(0)} := \mathcal{A}$ and $\gamma_n^{(k)} = \|\mathbf{e} - \mathbf{a}_n^{(k)}\|_2$ for $\mathcal{B}^{(0)} := \mathcal{B}$ with $n = 1, 2, 3$, where the entries of $\mathbf{a}^{(k)}$ and $\mathbf{a}_n^{(k)}$ are given by (5.3.9) and (5.3.11), respectively. The related results are shown in Fig. 5.2.

Consider $\mathcal{A} \in NT_{2,2}$ with

$$a_{111} = 1,\ a_{112} = 0,\ a_{121} = 0,\ a_{211} = 0,\ a_{122} = 2,\ a_{212} = 3,\ a_{221} = 4,\ a_{222} = 0,$$

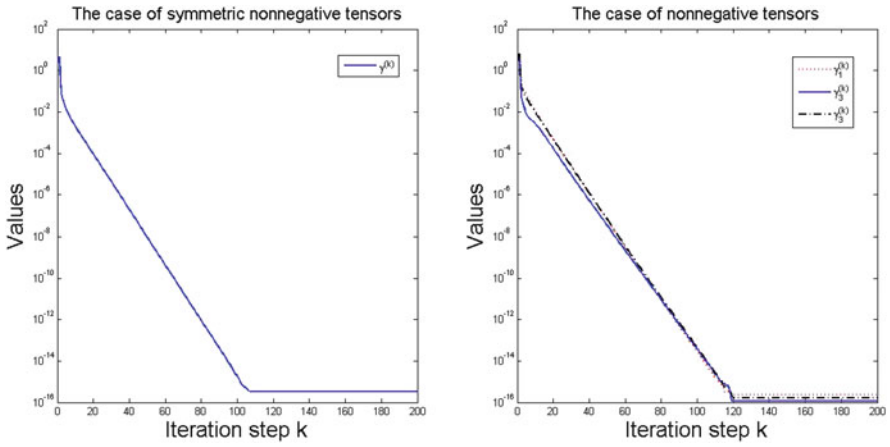

Fig. 5.2 The values of $\gamma^{(k)}$ and $\gamma_n^{(k)}$ as k increases

with perm$(\mathcal{A}) = 0$. Starting from $\mathcal{A}$, the sequence generated by (5.3.10) converges to $\mathcal{B} \in \Omega_{2,2}$:

$$\begin{cases} b_{111} = 0.5,\ b_{112} = 0,\ b_{121} = 0,\ b_{211} = 0, \\ b_{122} = 0.5,\ b_{212} = 0.5,\ b_{221} = 0.5,\ b_{222} = 0. \end{cases}$$

For a given $\mathcal{A} \in NT_{N,I}$, for all i_n and n, if

$$\sum_{i_1=1}^{I} a_{i_1 i_2 \ldots i_N} = \sum_{i_2=1}^{I} a_{i_1 i_2 \ldots i_N} = \cdots = \sum_{i_N=1}^{I} a_{i_1 i_2 \ldots i_N} = 1,$$

we call $\mathcal{A}$ line stochastic. Hence we propose the following conjecture.

Conjecture 5.3.2 For any $\mathcal{A}^{(0)} \in NT_{N,I}$ without vanishing fibers,[2] the process:

$$a_{i_1 i_2 \ldots i_N}^{(k+1)} = \frac{a_{i_1 i_2 \ldots i_N}^{(k)}}{(a_{1,i_2 i_3 \ldots i_N}^{(k)} a_{2,i_1 i_3 \ldots i_N}^{(k)} \cdots a_{N,i_1 i_2 \ldots i_{N-1}}^{(k)})^{1/N}}, \quad k = 0, 1, \ldots$$

where

$$a_{n,i_1 \ldots i_{n-1} i_{n+1} \ldots i_N}^{(k)} = \sum_{i_n=1}^{I} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N}^{(k)} > 0,$$

[2]For any $\mathcal{A} \in T_{N,I}$ and a given n, the mode-n $(i_1, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N)$-fiber [63] of $\mathcal{A}$ is defined by $\mathcal{A}(i_1, \ldots, i_{n-1}, :, i_{n+1}, \ldots, i_N) \in \mathbb{R}^I$ for all $i_1, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N$.

converges to a line stochastic tensor.

In general, for all i_n and n, if the entries of $\mathcal{A} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ satisfy

$$\sum_{i_1=1}^{I_1} a_{i_1 i_2 \ldots i_N} = \sum_{i_2=1}^{I_2} a_{i_1 i_2 \ldots i_N} = \cdots = \sum_{i_N=1}^{I_N} a_{i_1 i_2 \ldots i_N} = 1,$$

then we call $\mathcal{A}$ line stochastic. We propose the following conjecture.

Conjecture 5.3.3 For any $\mathcal{A}^{(0)} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ without vanishing fibers, the process:

$$a_{i_1 i_2 \ldots i_N}^{(k+1)} = \frac{a_{i_1 i_2 \ldots i_N}^{(k)}}{(a_{1, i_2 i_3 \ldots i_N}^{(k)} a_{2, i_1 i_3 \ldots i_N}^{(k)} \cdots a_{N, i_1 i_2 \ldots i_{N-1}}^{(k)})^{1/N}}, \quad k = 0, 1, \ldots$$

where

$$a_{n, i_1 \ldots i_{n-1} i_{n+1} \ldots i_N}^{(k)} = \sum_{i_n=1}^{I_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N}^{(k)} > 0,$$

converges to a line stochastic tensor.

To define the T-flat stochastic tensors with $1 \leq T \leq N-1$, we need the following notations [64]. Let N index sets $\mathbb{P}_n = \{1, 2, \ldots, I\}$ be given with all n. For a fixed $T \in \{1, 2, \ldots, N-1\}$, let $\mathbb{Q}_T$ be the class of all subsets of $\mathbb{K} = \{1, 2, \ldots, I\}$ with cardinality T, i.e., $\mathfrak{Q}_T = \{\mathbb{Q} : \mathbb{Q} \subset \mathbb{K}, |\mathbb{Q}| = T\}$. A set $\mathbb{Q} \in \mathfrak{Q}_T$ determines "fixed" indices. Therefore, we call the indices in $\mathbb{K} \backslash \mathbb{Q}$ "free" indices. For every fixed T-tuple $(i_{q_1}, i_{q_2}, \ldots, i_{q_T}) \in \mathbb{P}_{q_1} \times \mathbb{P}_{q_2} \times \cdots \times \mathbb{P}_{q_T}$ and for any $\mathbb{Q} = \{q_1, q_2, \ldots, q_T\} \in \mathfrak{Q}_T$, we define $\mathbb{J}_1^{\mathbb{Q}} \times \mathbb{J}_2^{\mathbb{Q}} \times \cdots \times \mathbb{J}_N^{\mathbb{Q}}$ by

$$\mathbb{J}_r^{\mathbb{Q}} := \begin{cases} \{1, 2, \ldots, I\}, & \text{if } r \text{ is a free index,} \\ \{i_{q_l}\}, & \text{if } r \text{ is the fixed index } q_l. \end{cases}$$

This definition is independent of the sequence $(q_1, q_2, \ldots, q_T)$ but depends only on $\mathbb{Q}$. A tensor $\mathcal{A}$ is called T-flat stochastic if its entries satisfy

$$\sum_{r=1}^{\mathbb{K}} \sum_{i_r \in \mathbb{J}_r^{\mathbb{Q}}} a_{i_1 i_2 \ldots i_N} = 1,$$

with all $(i_{q_1}, i_{q_2}, \ldots, i_{q_T}) \in \mathbb{P}_{q_1} \times \mathbb{P}_{q_2} \times \cdots \times \mathbb{P}_{q_T}$ and $\mathbb{Q} \in \mathfrak{Q}_T$.

We raise the following question.

Question 5.3.1 From any $\mathcal{A} \in NT_{N,I}$ without vanishing flats, how can we generate a sequence $\{\mathcal{A}_k\}$ $(k = 1, 2, \dots)$ such that, $\{\mathcal{A}_k\}$ converges to a T-flat stochastic tensor as $k \to \infty$?

5.4 Some Results for Diagonals

In this section, we investigate a relationship between the entries of a nonnegative tensor and its permanent. We obtain a probabilistic algorithm for locating a positive diagonal in a (0, 1)-tensor, if such a diagonal exists.

5.4.1 Diagonal Products

If there is an upper bound on the diagonal products of $\mathcal{A} \in NT_{N,I}$, then its entries are also bounded.

Theorem 5.4.1 *Suppose* $\mathcal{A} \in NT_{N,I}$ *satisfies* perm $(\mathcal{A}) > 0$. *Then for a given* $c > 0$, *the following conditions are equivalent:*

(i) $\prod_{i=1}^{I} a_{i\pi_2(i)\dots\pi_N(i)} \le c$ *for any* $\pi_n \in \mathbb{S}_I$ *with* $n = 2, 3, \dots, N$.

(ii) *There exists a positive matrix* $\mathbf{X} \in \mathbb{R}^{I\times N}$ *with* $\prod_{i=1}^{I} x_{i1}x_{i2}\dots x_{iN} \le c$ *such that* $a_{i_1 i_2\dots i_N} \le x_{i_1 1}x_{i_2 2}\dots x_{i_N N}$ *for all* i_n *and* n.

Proof Suppose that (ii) holds. Then for $(N-1)$ given $\pi_n \in \mathbb{S}_I$, (i) holds as

$$\prod_{i=1}^{I} a_{i\pi_2(i)\dots\pi_N(i)} \le \prod_{i=1}^{I} x_{i1}x_{\pi_2(i)2}\dots x_{\pi_N(i)N} \le c.$$

Suppose that (i) holds. Define a tensor $\mathcal{B} \in T_{N,I}$ with entries

$$b_{i_1 i_2\dots i_N} = \begin{cases} \ln(a_{i_1 i_2\dots i_N}), & \text{if } a_{i_1 i_2\dots i_N} > 0, \\ -M, & \text{if } a_{i_1 i_2\dots i_N} = 0, \end{cases}$$

where $M > 0$ is chosen as follows: for two different $\pi_n, \tau_n \in \mathbb{S}_I$ such that

$$\prod_{i=1}^{I} a_{i\pi_2(i)\dots\pi_N(i)} > 0, \quad \prod_{i=1}^{I} a_{i\tau_2(i)\dots\tau_N(i)} = 0,$$

we have

$$\prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)} > \prod_{i=1}^{I} b_{i\tau_2(i)\ldots\tau_N(i)}.$$

Since the cardinality of $\mathbb{S}_I$ is $I!$, such a choice of M is possible. By (5.3.1), we have

$$\max_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \sum_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)} = \max_{\mathcal{P}\in P_{N,I}} \sum_{i_1,i_2,\ldots,i_N=1}^{I} b_{i_1 i_2\ldots i_N} p_{i_1 i_2\ldots i_N} = \max_{\mathcal{P}\in P_{N,I}} \langle \mathcal{B}, \mathcal{P}\rangle.$$

Any totally plane stochastic tensor can be represented as a convex combination of some plane permutation tensors. Hence, we have

$$\max_{\mathcal{P}\in P_{N,I}} \langle \mathcal{B}, \mathcal{P}\rangle = \max_{C\in\Omega^{+}_{N,I}} \langle \mathcal{B}, C\rangle \le \max_{C\in\Omega_{N,I}} \langle \mathcal{B}, C\rangle.$$

Next we consider the maximal programming problem $\max_{C\in\Omega_{N,I}} \langle \mathcal{B}, C\rangle$. By the duality Theorem, there must be an optimal solution to the dual problem

$$\min\left(\sum_{i_1=1}^{I} x_{i_1 1} + \sum_{i_2=1}^{I} x_{i_2 2} + \cdots + \sum_{i_N=1}^{I} x_{i_N N}\right)$$

subject to $x_{i_1 1} + x_{i_2 2} + \cdots + x_{i_N N} \ge b_{i_1 i_2\ldots i_N}$ for all i_n and n.

Let $\{\overline{x}_{i_1 1}, \overline{x}_{i_2 2}, \ldots, \overline{x}_{i_N N}\}$ with given i_n and all n be the optimal solution with $\sum_{i_1=1}^{I} \overline{x}_{i_1 1} + \sum_{i_2=1}^{I} \overline{x}_{i_2 2} + \cdots + \sum_{i_N=1}^{I} \overline{x}_{i_N N} := \theta$. Hence, we have

$$\max_{\mathcal{P}\in P_{N,I}} \langle \mathcal{B}, \mathcal{P}\rangle \le \max_{C\in\Omega_{N,I}} \langle \mathcal{B}, C\rangle = \theta.$$

According to the hypothesis, the target tensor $\mathcal{A}$ has at least one positive diagonal, then

$$\max_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} = \max_{\pi_2,\ldots,\pi_N\in\mathbb{S}_I} \prod_{i=1}^{I} e^{b_{i\pi_2(i)\ldots\pi_N(i)}} = \exp(\theta) \le c.$$

Let $x_{i_n n} = \exp(\overline{x}_{i_n n})$ with given i_n and all n. From $a_{i_1 i_2\ldots i_N} > 0$ and the definition of $b_{i_1 i_2\ldots i_m}$, we have

$$b_{i_1 i_2\ldots i_N} = \ln(a_{i_1 i_2\ldots i_N}) \le \overline{x}_{i_1 1} + \overline{x}_{i_2 2} + \cdots + \overline{x}_{i_N N}.$$

If $a_{i_1 i_2 \ldots i_N} > 0$, then we have $a_{i_1 i_2 \ldots i_N} \leq x_{i_1 1} x_{i_2 2} \ldots x_{i_N N}$ with given i_n and all n. Trivially,

$$a_{i_1 i_2 \ldots i_N} \leq x_{i_1 1} x_{i_2 2} \ldots x_{i_N N}, \quad \text{if} \quad a_{i_1 i_2 \ldots i_N} = 0.$$

Furthermore,

$$\prod_{i=1}^{I} x_{i1} x_{i2} \ldots x_{iN} = \exp\left(\sum_{i_1=1}^{I} \overline{x}_{i_1 1} + \sum_{i_2=1}^{I} \overline{x}_{i_2 2} + \cdots + \sum_{i_N=1}^{I} \overline{x}_{i_N N} \right) = \exp(\theta) \leq c.$$

This completes the proof. □

Theorem 5.4.2 *Let $\mathcal{A} \in RT_{N,I}$ be a nonnegative tensor with a totally plane stochastic pattern. Let every diagonal product take the value of either 0 or some constant α. Then, there exists a rank-one tensor $C \in RT_{N,I}$ such that the positive entries of $\mathcal{A}$ coincide with the corresponding entries of C.*

Proof Suppose that $\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} = 0$ or α for any $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$. Since $\mathcal{A}$ has a totally plane stochastic pattern, there exists $\tau_n \in \mathbb{S}_I$ such that $\prod_{i=1}^{I} a_{i\tau_2(i)\ldots\tau_N(i)} = \alpha > 0$. By Theorem 5.4.1, there exists a positive matrix $\mathbf{X} \in \mathbb{R}^{I \times N}$ with $\prod_{i=1}^{I} x_{i1} x_{i2} \ldots x_{iN} \leq c$ such that $a_{i_1 i_2 \ldots i_N} \leq x_{i_1 1} x_{i_2 2} \ldots x_{i_N N}$ for all i_n and n. If $a_{i_1 i_2 \ldots i_N} > 0$, we know from Theorem 5.3.2 that $\pi_n(i) = i_n$ and $\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} = \alpha$ for $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$.

We also have

$$\alpha = \prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} \leq \prod_{i=1}^{I} x_{i1} x_{i2} \ldots x_{iN} \leq \alpha.$$

Then we obtain $a_{i_1 i_2 \ldots i_N} = x_{i_1 1} x_{i_2 2} \ldots x_{i_N N}$ if $a_{i_1 i_2 \ldots i_N} > 0$. Define $C \in T_{N,I}$ with $c_{i_1 i_2 \ldots i_N} = x_{i_1 1} x_{i_2 2} \ldots x_{i_N N}$. Then C is a rank-one tensor, which satisfies $a_{i_1 i_2 \ldots i_N} = c_{i_1 i_2 \ldots i_N}$ if $a_{i_1 i_2 \ldots i_N} > 0$. □

Theorem 5.4.3 *Let $\mathcal{A}$ and $\mathcal{B}$ be two distinct totally plane stochastic tensors. There exists $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$ such that*

$$\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} > \prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)}.$$

Proof We assume that $b_{i_1 i_2 \ldots i_N} = 0 \Rightarrow a_{i_1 i_2 \ldots i_N} = 0$, otherwise, $a_{i_1 i_2 \ldots i_N} > 0$ and $b_{i_1 i_2 \ldots i_N} = 0$ for all i_n and n. The diagonal through $a_{i_1 i_2 \ldots i_N}$ (which exists by

Theorem 5.3.2) is positive. With the convention $\ln\left(\frac{0}{0}\right) = 1$, for $\mathcal{X} \in \Omega_{N,I}$, let

$$\phi(\mathcal{X}) = \sum_{i_1,i_2,\ldots,i_N=1}^{I} x_{i_1 i_2 \ldots i_N} \ln\left(\frac{a_{i_1 i_2 \ldots i_N}}{b_{i_1 i_2 \ldots i_N}}\right).$$

The linear function $\phi(\mathcal{X})$ attains its maximum at an extreme point of the domain $\Omega^{+}_{N,I}$. By the Information Inequality [47, Theorem 2.6.2], we have

$$\phi(\mathcal{A}) = \sum_{i_1,i_2,\ldots,i_N=1}^{I} a_{i_1 i_2 \ldots i_N} \ln\left(\frac{a_{i_1 i_2 \ldots i_N}}{b_{i_1 i_2 \ldots i_N}}\right) > 0.$$

Thus, there exists $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$ such that the corresponding plane permutation tensor $\mathcal{P} = (\delta_{i\pi_2(i)\ldots\pi_N(i)})$ satisfies

$$\phi(\mathcal{P}) = \sum_{i_1,i_2,\ldots,i_N=1}^{I} p_{i_1 i_2 \ldots i_N} \ln\left(\frac{a_{i_1 i_2 \ldots i_N}}{b_{i_1 i_2 \ldots i_N}}\right) = \ln\left[\prod_{i=1}^{I} \frac{a_{i_1 i_2 \ldots i_N}}{b_{i_1 i_2 \ldots i_N}}\right] > 0.$$

It follows that $\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} > \prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)}$. □

From Theorem 5.4.3, we have the following corollary.

Corollary 5.4.1 *For a given constant α and any $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$, suppose that two totally plane stochastic tensors $\mathcal{A}$ and $\mathcal{B}$ satisfy*

$$\prod_{i=1}^{I} a_{i\pi_2(i)\ldots\pi_N(i)} = \alpha \prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)},$$

then $\mathcal{A} = \mathcal{B}$.

5.4.2 Finding a Positive Diagonal

First, we have the following theorem.

Theorem 5.4.4 *Let each entry of $\mathcal{W} \in RT_{N,I}$ be a nonnegative integer. Suppose that*

$$\min_{\pi_2,\ldots,\pi_N \in \mathbb{S}_I} \sum_{i=1}^{I} w_{i\pi_2(i)\ldots\pi_N(i)}$$

is attained at the unique $\tau_n \in \mathbb{S}_I$ *with* $n = 2, 3, \ldots, N$ *and let the minimum value be* θ. *Let* $\mathcal{B} \in RT_{N,I}$ *be defined as* $b_{ii_2\ldots i_N} = 0$, *if* $w_{ii_2\ldots i_N} = 0$; *and* $b_{ii_2\ldots i_N} = 2^{w_{ii_2\ldots i_N}}$ *otherwise. Then* $(i, i_2, \ldots, i_N)$ *lies on the diagonal associated with* $\pi_n \in \mathbb{S}_I$ *with* $n = 2, 3, \ldots, N$, *i.e.,* $i_n = \pi_n(i)$, *if and only if*

$$\frac{2^{w_{ii_2\ldots i_N}} \det_c(\mathcal{B}(i, i_2, \ldots, i_N))}{2^\theta} \tag{5.4.1}$$

is odd.

Proof Fix $(i, i_2, \ldots, i_N)$ such that $w_{ii_2\ldots i_N}$ is nonzero. Let $\mathbb{P} = \{(\pi_2, \pi_3, \ldots, \pi_N) : \tau_n(i) = i_n, n = 2, 3, \ldots, N\}$. The numerator in (5.4.1) can be expressed, unique up to sign, as a sum of those diagonal products of $\mathcal{B}$ corresponding to $(\pi_2, \pi_3 \ldots, \pi_N) \in \mathbb{P}$. Let us write

$$\begin{aligned}&(-1)^{i+i_2+\cdots+i_N} 2^{w_{ii_2\ldots i_N}} \det_c(\mathcal{B}(i, i_2, \ldots, i_N)) \\ &= \sum_{(\pi_2,\ldots,\pi_N)\in\mathbb{P}} \operatorname{sign}(\tau_{N-P+1}) \ldots \operatorname{sign}(\tau_N) \prod_{i=1}^{I} b_{i\pi_2(i)\ldots\pi_N(i)},\end{aligned} \tag{5.4.2}$$

where P is given in (5.1.2). Each term in the summation in (5.4.2) is, unique up to a sign, either zero or 2^α for some $\alpha \geq \theta$.

In view of the uniqueness of τ_n, there exists precisely one 2^θ term in the sum if and only if $(\tau_2, \tau_3, \ldots, \tau_N) \in \mathbb{P}$. Thus, the expression in (5.4.2) is of the term $2^\theta(1+u)$ for an even integer u if $(\tau_2, \tau_3, \ldots, \tau_N) \in \mathbb{P}$, or $2^\theta v$ for some even integer v otherwise. It follows that the ratio (5.4.1) is odd if and only if $(\tau_2, \tau_3, \ldots, \tau_N) \in \mathbb{P}$. □

Let $\mathcal{A} \in RT_{N,I}$ be a $(0, 1)$-tensor such that $\operatorname{perm}(\mathcal{A}) > 0$. From Theorem 5.4.4, an algorithm for locating a positive diagonal is given in the following steps:

Step 1: Let the number of nonzero entries in $\mathcal{A}$ be β. For each $a_{i_1i_2\ldots i_N} = 1$, we choose an integer $w_{i_1i_2\ldots i_N}$ at random from $\{1, 2, \ldots, 2\beta\}$. Define $\mathcal{B} \in T_{N,I}$ by $b_{i_1i_2\ldots i_N} = 0$, if $a_{i_1i_2\ldots i_N} = 0$ and $b_{i_1i_2\ldots i_N} = 2^{w_{i_1i_2\ldots i_N}}$, otherwise.

Step 2: Compute $\det_c(\mathcal{B})$ and let θ be the highest power of 2 such that 2^θ divides $\det_c(\mathcal{B})$. (Observe that θ is precisely the minimum diagonal sum in $\mathcal{W}$)

Step 3: For each $(i_1, i_2, \ldots, i_N)$, compute $2^{w_{i_1i_2\cdots i_N}} \det_c(\mathcal{B}(i_1, i_2, \ldots, i_N))/2^\theta$. Select $(i_1, i_2, \ldots, i_N)$, if this number is odd.

To analyze the convergence of the algorithm, we need the following lemma.

Lemma 5.4.1 ([47, Lemma 2.5.1]) *Let* $\mathbb{F} = \{\mathbb{S}_1, \mathbb{S}_2, \ldots, \mathbb{S}_K\}$ *be a family of nonempty subsets of the finite set* $\mathbb{S} = \{x_1, x_2, \ldots, x_I\}$. *Let the elements of* $\mathbb{S}$ *be assigned integer weights chosen at random, uniformly and independently from* $\{1, 2, \ldots, 2I\}$. *Then the probability that there is a unique set in* $\mathbb{F}$ *with minimum weights is at least* $1/2$.

By the above lemma, with probability at least 1/2, there are unique $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$ such that the corresponding diagonal sum of $\mathcal{W}$ is minimum. In this case, the hypotheses of Lemma 5.4.1 are satisfied and the particular $(i_1, i_2, \ldots, i_N)$ selected in Step 3 corresponds to a positive diagonal of the tensor $\mathcal{A}$.

If we repeat the algorithm T times, then the probability that it fails every time is at most $(1/2)^T$; therefore, we successfully locate a positive diagonal with probability at least $1 - (1/2)^T$. The main computational cost in the algorithm involves calculating the combinatorial determinant of the integer tensor $\mathcal{B}$.

5.5 Axial N-Index Assignment Problems

Multi-index assignment problems are introduced by Pieskalla [65] in 1968 as a natural extension of linear assignment problems [64]. In this section, we only consider the axial N-index assignment problems. We investigate how to convert the axial N-index assignment problem to an integer linear programming problem based on the set of totally plane stochastic tensors. In particular, the 3-index assignment problem has been introduced by Schell [66] in 1955 and applied in quite a number of situations, such as the investment of capital into different possible physical locations over some time horizon [67]. Qi and Sun [68] mention the following application in a rolling mill: ingots are to be scheduled through soaking pits (temperture stabilizing baths) so as to minimize the idle time for the rolling mill. Minimizing the maximum cost instead of a sum of costs leads to the bottleneck objective functions. Axial 3-index assignment problems with a bottleneck objective function have been considered by Malhotra et al. [69] as well as Geetha and Vartk [70].

The axial N-index assignment problem can be stated in the following way. Let I^N cost coefficients $c_{i_1 i_2 \ldots i_N}$ for all i_n and n be given. We seek $\pi_n^* \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$ such that $\sum_{i=1}^{I} c_{i\pi_2^*(i)\ldots\pi_N^*(i)}$ is minimum, i.e.,

$$\min_{\pi_2,\ldots,\pi_N \in \in \mathbb{S}_I} \sum_{i=1}^{I} c_{i\pi_2(i)\ldots\pi_N(i)} = \sum_{i=1}^{I} c_{i\pi_2^*(i)\ldots\pi_N^*(i)}.$$

Since $\pi_n \in \mathbb{S}_I$ with $n = 2, 3, \ldots, N$ which describe a feasible solution can be chosen arbitrarily, the axial N-index assignment problem has $(I!)^{N-1}$ feasible solutions. If we treat the coefficients $c_{i_1 i_2 \ldots i_N}$ for all i_n and n as all entries of the tensor C and introduce plane permutation tensors, then we can rewrite the axial N-index assignment problem as an integer linear programming problem:

$$\min\langle C, \mathcal{P}\rangle, \quad \text{s.t. } \mathcal{P} \in P_{N,I}.$$

We present this problem as follows.

Problem 5.5.1 Suppose that $C \in RT_{N,I}$ is positive. Find $\mathcal{A}_* \in \Omega_{N,I}^+$ such that

$$\langle C, \mathcal{A}_* \rangle = \min_{\mathcal{A} \in \Omega_{N,I}^+} \langle C, \mathcal{A} \rangle.$$

Any $\mathcal{A} \in \Omega_{N,I}^+$ is of the form $\mathcal{A} = \sum_{j=1}^{\tilde{M}} \lambda_j \mathcal{P}_j$, where $\sum_{j=1}^{\tilde{M}} \lambda_j = 1$ with $\lambda_j \geq 0$ and $\mathcal{P}_j \in P_{N,I}$ with $j = 1, 2, \ldots, \tilde{M}$). Thus $\langle C, \mathcal{A} \rangle = \sum_{j=1}^{\tilde{M}} \lambda_j \langle C, \mathcal{P}_j \rangle$ and Problem 5.5.1 reduces to minimizing $\sum_{j=1}^{\tilde{M}} \lambda_j \langle C, \mathcal{P}_j \rangle$ subject to $\lambda_j \geq 0$ and $\sum_{j=1}^{\tilde{M}} \lambda_j = 1$. The axial N-index assignment problem is clearly equivalent to finding the diagonal of C such that the corresponding sum of the entries on the diagonal is minimum. Thus, we have

$$\min_{\pi_2, \ldots, \pi_N \in \mathbb{S}_I} \sum_{i=1}^{I} c_{i\pi_2(i)\ldots\pi_N(i)} = \min_{\mathcal{P} \in P_{N,I}} \langle C, \mathcal{P} \rangle = \min_{\mathcal{A} \in \Omega_{N,I}^+} \langle C, \mathcal{A} \rangle \geq \min_{\mathcal{A} \in \Omega_{N,I}} \langle C, \mathcal{A} \rangle.$$

Finding $\min_{\mathcal{A} \in \Omega_{N,I}} \langle C, \mathcal{A} \rangle$ is a linear programming problem and the corresponding dual problem is given as follows.

Problem 5.5.2 Suppose that $C \in RT_{N,I}$ is positive. Find a positive $\mathbf{X} \in \mathbb{R}^{I \times N}$ such that

$$\max \left(\sum_{i_1=1}^{I} x_{i_1 1} + \sum_{i_2=1}^{I} x_{i_2 2} + \cdots + \sum_{i_N=1}^{I} x_{i_N N} \right)$$

with for all i_n and n

$$x_{i_1 1} + x_{i_2 2} + \cdots + x_{i_N N} \geq c_{i_1 i_2 \ldots i_N}.$$

Suppose that the maximum of Problem 5.5.2 is denoted by θ, then we have

$$\min_{\pi_2, \ldots, \pi_N \in \mathbb{S}_I} \sum_{i=1}^{I} c_{i\pi_2(i)\ldots\pi_N(i)} = \min_{\mathcal{P} \in P_{N,I}} \langle C, \mathcal{P} \rangle \geq \theta.$$

After we solve Problem 5.5.2, we derive a lower bound for the minimum of Problem 5.5.1, or a lower bound for the solution of the axial N-index assignment problem.

References

1. S. Dow, P. Gibson, Permanents of d-dimensional matrices. Linear Algebra Appl. **90**, 133–145 (1987)
2. R. Brualdi, J. Csima, Extremal plane stochastic matrices of dimension three. Linear Algebra Appl. **11**(2), 105–133 (1975)
3. J. Csima, Multidimensional stochastic matrices and patterns. J. Algebra **14**(2), 194–202 (1970)
4. W. Jurkat, H. Ryser, Extremal configurations and decomposition theorems. I. J. Algebra **8**(2), 194–222 (1968)
5. H. Lu, K. Plataniotis, A. Venetsanopoulos, *Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data* (CRC, Boca Raton, 2013)
6. A. Berman, R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Society for Industrial and Applied Mathematics, Philadelphia, 1994)
7. R. Brualdi, P. Gibson, Convex polyhedra of doubly stochastic matrices. I. Applications of the permanent function. J. Comb. Theor. A **22**(2), 194–230 (1977)
8. Q. Yang, Y. Yang, Further results for Perron-Frobenius theorem for nonnegative tensors II. SIAM J. Matrix Anal. Appl. **32**(4), 1236–1250 (2011)
9. K. Chang, K. Pearson, T. Zhang, Perron-Frobenius theorem for nonnegative tensors. Commun. Math. Sci. **6**(2), 507–520 (2008)
10. R. Brualdi, J. Csima, Stochastic patterns. J. Comb. Theor. A **19**, 1–12 (1975)
11. L. Cui, W. Li, M. Ng, Birkhoff–von Neumann theorem for multistochastic tensors. SIAM J. Matrix Anal. Appl. **35**(3), 956–973 (2014)
12. A. Raftery, A model for high-order markov chains. J. R. Stat. Soc. Ser. B **47**, 528–539 (1985)
13. K. Chang, T. Zhang, On the uniqueness and non-uniqueness of the positive Z-eigenvector for transition probability tensors. J. Math. Anal. Appl. **408**(2), 525–540 (2013)
14. K. Chang, L. Qi, T. Zhang, A survey on the spectral theory of nonnegative tensors. Numer. Linear Algebra Appl. **20**(6), 891–912 (2013)
15. D. Gleich, L. Lim, Y. Yu, Multilinear PageRank. SIAM J. Matrix Anal. Appl. **36**(4), 1507–1541 (2015)
16. S. Hu, L. Qi, Convergence of a second order Markov chain. Appl. Math. Comput. **241**, 183–192 (2014)
17. J. Christensen, P. Fischer, Multidimensional stochastic matrices and error-correcting codes. Linear Algebra Appl. **183**, 255–276 (1993)
18. V. Pless, *Introduction to the Theory of Error-correcting Codes* (Wiley, New York, 1998)
19. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)
20. S. Hu, Z. Huang, C. Ling, L. Qi, On determinants and eigenvalue theory of tensors. J. Symb. Comput. **50**, 508–531 (2013)
21. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
22. L. Rice, Introduction to higher determinants. J. Math. Phys. **9**(1), 47–70 (1930)
23. J. Rotman, *An Introduction to the Theory of Groups*, 4th edn. (Springer, New York, 1995)
24. A. Cayley, On the theory of determinants. Trans. Camb. Philos. Soc. **8**, 1–16 (1843)
25. R. Oldenburger, Higher dimensional determinants. Am. Math. Mon. **47**(1), 25–33 (1940)
26. R. Vein, P. Dale, *Determinants and their Applications in Mathematical Physics* (Springer, New York, 1999)
27. L. Lim, Tensors and hypermatrices, in *Handbook of Linear Algebra*, 2nd edn. (CRC Press, Boca Raton, 2013)
28. L. Lim, *Eigenvalues of Tensors and Spectral Hypergraph Theory* (*Talk*) (University of California, Berkeley, 2008)
29. K. Pearson, T. Zhang, On spectral hypergraph theory of the adjacency tensor. Graphs and Combinatorics **30**(5), 1233–1248 (2014)
30. C. Berge, Graphs and Hypergraphs, in *North-Holland Mathematical Library*, vol. 45 (North-Holland, Amsterdam, 1976)

31. A. Taranenko, Permanents of multidimensional matrices: Properties and applications. J. Appl. Ind. Math. **10**(4), 567–604 (2016)
32. A. Barvinok, Computing the permanent of (some) complex matrices. Found. Comput. Math. **16**(2), 329–342 (2016)
33. S. Avgustinovich, Multidimensional permanents in enumeration problems. J. Appl. Ind. Math. **4**(1), 19–20 (2010)
34. A. Taranenko, Multidimensional permanents and an upper bound on the number of transversals in Latin squares. J. Comb. Des. **23**(7), 305–320 (2015)
35. D. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford University Press, Oxford, 1940)
36. M. Marcus, Finite dimensional multilinear algebra. Part II, in *Pure and Applied Mathematics*, vol. 23 (Marcel Dekker, Inc., New York, 1975)
37. R. Merris, Multilinear algebra, in *Algebra, Logic and Applications*, vol. 8 (Gordon and Breach Science, Amsterdam, 1997)
38. M. Marcus, H. Minc, Generalized matrix functions. Trans. Am. Math. Soc. **116**, 316–329 (1965)
39. W. Berndt, S. Sra, Hlawka-popoviciu inequalities on positive definite tensors. Linear Algebra Appl. **486**, 317–327 (2015)
40. S. Huang, C. Li, Y. Poon, Q. Wang, Inequalities on generalized matrix functions. Linear Multilinear Algebra **65**, 1947–1961 (2017)
41. X. Chang, V. Paksoy, F. Zhang, An inequality for tensor product of positive operators and its applications. Linear Algebra Appl. **498**, 99–105 (2014)
42. V. Paksoy, R. Turkmen, F. Zhang, Inequalities of generalized matrix functions via tensor products. Electron. J. Linear Algebr. **27**(1), 332–341 (2014)
43. N. Lee, A. Cichockia, *Fundamental Tensor Operations for Large-Scale Data Analysis in Tensor Train Formats* (2016). ArXiv preprint:1405.7786v2
44. M. Che, D.S. Cvetković, Y. Wei, Inequalities on generalized tensor functions associated to the diagonalizable symmetric tensors. Stat. Optim. Infor. Comput. **6**(4), 483–496 (2018)
45. C. Bu, W. Wang, L. Sun, J. Zhou, Minimum (maximum) rank of sign pattern tensors and sign nonsingular tensors. Linear Algebra Appl. **483**, 101–114 (2015)
46. R. Brualdi, B. Shader, *Matrices of Sign-solvable Linear System* (Cambridge University, Cambridge, 1995)
47. R. Bapat, T. Raghavan, *Nonnegative Matrices and Applications* (Cambridge University, Cambridge, 1997)
48. R. Brualdi, J. Csima, Small matrices of large dimension. Linear Algebra Appl. **150**, 227–241 (1991)
49. H. Chang, V. Paksoy, F. Zhang, Polytopes of stochastic tensors. Ann. Funct. Anal. **7**(3), 386–393 (2016)
50. P. Fischer, E. Swart, Three dimensional line stochastic matrices and extreme points. Linear Algebra Appl. **69**, 179–203 (1985)
51. Z. Li, F. Zhang, X. Zhang, On the number of vertices of the stochastic tensor polytope. Linear Multilinear Algebra **65**, 2064–2075 (2017)
52. E. Marchi, P. Tarazaga, About (k, n) stochastic matrices. Linear Algebra Appl. **26**, 15–30 (1979)
53. G. Schrage, Some inequalities for multidimensional (0,1)- matrices. Discret. Math. **23**(2), 169–175 (1978)
54. Q. Wang, F. Zhang, The permanent function of tensors. Acta Math. Vietnam **43**, 701–713 (2018)
55. R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices. Ann. Math. Stat. **35**, 876–879 (1964)
56. A. Marshall, I. Olkin, Scaling of matrices to achieve specified row and column sums. Numer. Math. **12**, 83–90 (1968)
57. R. Sinkhorn, P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices. Pac. J. Math. **21**, 343–348 (1967)

58. R. Bapat, D_1AD_2 theorems for multidimensional matrices. Linear Algebra Appl. **48**, 437–442 (1982)
59. T. Raghavan, On pairs of multidimensional matrices. Linear Algebra Appl. **62**, 263–268 (1984)
60. J. Franklin, J. Lorenz, On the scaling of multidimensional matrices. Linear Algebra Appl. **114/115**, 717–735 (1989)
61. A. Shashua, R. Zass, T. Hazan, Multi-way clustering using super-symmetric non-negative tensor factorization. Lect. Notes Comput. Sci. **3954**, 595–608 (2006)
62. L. Qi, The best rank-one approximation ratio of a tensor space. SIAM J. Matrix Anal. Appl. **32**(2), 430–442 (2011)
63. T. Kolda, B. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
64. R. Burkard, M. Dell'Amico, S. Martello, *Assignment Problems* (Society for Industrial and Applied Mathematics, Philadelphia, 2009)
65. W. Pierskalla, The multidimensional assignment problem. Oper. Res. **16**(2), 422–431 (1968)
66. E. Schell, Distribution of a product by several properties, in *Proceedings of the Second Symposium in Linear Programming*, vol. 615, pp. 642 (1955)
67. W. Pierskalla, The tri-substitution method for the three-dimensional assignment problem. CORS J. **5**, 71–81 (1967)
68. L. Qi, D. Sun, Polyhedral methods for solving three index assignment problems, in *Nonlinear Assignment Problems*. Combinatorial Optimization, vol. 7 (Kluwer Academic, Dordrecht, 2000), pp. 91–107
69. R. Malhotra, H. Bhatia, M. Puri, The three-dimensional bottleneck assignment problem and its variants. Optimization **16**(2), 245–256 (1985)
70. S. Geetha, M. Vartak, The three-dimensional bottleneck assignment problem with capacity constraints. Eur. J. Oper. Res. **73**(3), 562–568 (1994)

Chapter 6
Neural Networks

We focus on the rank-one approximation problem of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ by neural networks: finding a real scalar σ and N unit $\mathbf{x}_n \in \mathbb{R}^{I_n}$ to minimize

$$\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} [a_{i_1 i_2 \ldots i_N} - \sigma \cdot (x_{1,i_1} x_{2,i_2} \ldots x_{N,i_N})]^2,$$

where x_{n,i_n} is the i_nth element of $\mathbf{x}_n \in \mathbb{R}^{I_n}$ for all i_n and n, and $\sigma > 0$ is a scaling factor.

The rank-one approximation problem of a tensor can be viewed as the multi-parameter constrained optimization problems. Multi-parameter optimization, constrained or otherwise, can be accomplished by the associated dynamical gradient systems, whose state evolves in time towards the steady-state solution or a critical point [1–3]. In the *neurodynamic optimization* approach [2], they minimize an (nonnegative) energy function of a dynamical system, typically described by first-order ordinary differential equations (ODE). For example, Cichocki [4] proposes a massively parallel algorithm (neural network) for the matrix singular value decomposition. Further research on this topic can be found in [5–7].

The neural networks for the rank-one approximation problem of a tensor, described by ODEs, are generalizations of those in [5, 6]. We prove the locally asymptotic stability of the solution by establishing an appropriate Lyapunov function. Semi-definite relaxation methods often produce the best rank-one approximation. However, these methods are generally expensive. Furthermore, the best rank-one approximations can not be obtained by ALS or HOPM [8]. The neural network models in this chapter can compute all locally optimal rank-one approximations starting from different initial values and one of them is the best rank-one approximation. Unfortunately, selecting an optimal initial value is an open problem. In general, neural network models converge to a local minimum.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_6

Liu and Wu [9] propose a neural network to compute eigenvectors related to the largest or smallest generalized eigenvalues of a symmetric-definite matrix pair. Many results on the generalized matrix eigenvalue problem using neural networks have been presented, see [10, 11]. Similarly, we design a continuation algorithm (neural networks) for the generalized tensor eigenvalue problem for a symmetric-definite tensor pairs, i.e., the neural network given by (6.5.3).

Numerical methods for solving differential equations (6.2.1), (6.2.3) and (6.5.3) can be applied, e.g., the Runge-Kutta methods. We only consider the simple continuous gradient descent method as an illustration.

6.1 Preliminaries

6.1.1 Tensor Singular Values and the Rank-One Approximation

In this section, we consider only tensors in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. Let $\mathbf{x}_n \in \mathbb{R}^{I_n}$ be nonzero vectors and $\|\mathbf{x}_n\|_2 = 1$ with all n. If the $(N+1)$-tuple $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is a solution of the following nonlinear equations [12, 13]:

$$F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)_{-n} = \mathcal{A} \times_1 \mathbf{x}_1^\top \cdots \times_{n-1} \mathbf{x}_{n-1}^\top \times_{n+1} \mathbf{x}_{n+1}^\top \cdots \times_N \mathbf{x}_N^\top = \sigma \mathbf{x}_n, \tag{6.1.1}$$

then $\sigma \in \mathbb{R}$ and the unit vectors $\mathbf{x}_n$ are called the *singular value* of $\mathcal{A}$ and the corresponding *mode-n singular vector* for all n, respectively.

In order to better illustrate the singular values and the associated mode-n singular vectors, we need the signs of a scalar and a vector [14]. Given $\alpha \in \mathbb{R}$, the sign $\delta(\alpha)$ is defined as:

$$\delta(\alpha) \in \{-1, +1\}, \text{ if } \alpha \neq 0; \quad \text{and } \delta(\alpha) = +1, \text{ if } \alpha = 0.$$

Similarly, for a vector $\mathbf{x} \in \mathbb{R}^I$, the sign $\delta(\mathbf{x})$ is defined as:

$$\delta(\mathbf{x}) \in \{-1, +1\}, \text{ if } \mathbf{x} \neq \mathbf{0}_I; \quad \text{and } \delta(\mathbf{x}) = +1, \text{ if } \mathbf{x} = \mathbf{0}_I.$$

If σ is a singular value of $\mathcal{A}$ and $\mathbf{x}_n$ is the associated mode-n singular vectors, then $\delta(\sigma)\sigma$ is also the singular value of the tensor $\mathcal{A}$ and $\delta(\mathbf{x}_n)\mathbf{x}_n$ are the associated mode-n singular vectors with $\delta(\sigma) \prod_{n=1}^N \delta(\mathbf{x}_n) = 1$. Based on the definitions of the sign of $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^I$, we can ensure that all singular values of the tensor $\mathcal{A}$ are nonnegative by changing the sign of all mode-n singular vectors such that $\prod_{n=1}^N \delta(\mathbf{x}_n) = 1$ if $\sigma \geq 0$ or $\prod_{n=1}^N \delta(\mathbf{x}_n) = -1$ if $\sigma < 0$.

We can rewrite the best rank-one approximation of $\mathcal{A}$ as follows: to find $\sigma \in \mathbb{R}$ and the unit vectors $\mathbf{u}_n \in \mathbb{R}^{I_n}$ such that

$$\sigma \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N) = \operatorname{argmin} \|\mathcal{A} - \tau \cdot (\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_N)\|_F, \tag{6.1.2}$$

subject to $\tau \in \mathbb{R}$ and $\mathbf{x}_n \in \mathbb{R}^{I_n}$ with $\|\mathbf{x}_n\|_2 = 1$.

This optimization problem is equivalent to a homogeneous polynomial optimization problem [15]. For convenience, we consider the homogeneous polynomial

$$\begin{aligned} F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) &:= \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top \\ &= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \ldots i_N} x_{1,i_1} x_{2,i_2} \ldots x_{N,i_N}. \end{aligned} \tag{6.1.3}$$

Note that $F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is a multilinear form, since it is linear in each $\mathbf{x}_n$. De Lathauwer et al. [15] prove the following result.

Theorem 6.1.1 ([15]) *For a given $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the rank-one approximation minimization problem (6.1.2) is equivalent to the optimization problem*

$$\max \quad |F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)|, \quad \text{s.t.} \quad \|\mathbf{x}_n\|_2 = 1 \tag{6.1.4}$$

for all n.

A rank-one tensor $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N)$, with $\sigma_* \in \mathbb{R}$ and each $\|\mathbf{u}_n\|_2 = 1$, is the best rank-one approximation for $\mathcal{A}$ if and only if $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$ is a global maximizer of (6.1.4) and $\sigma_* = F(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$. Moveover, we also have

$$\|\mathcal{A} - \sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N)\|_F^2 = \|\mathcal{A}\|_F^2 - \sigma_*^2.$$

A well-known algorithm for the optimization problem (6.1.4) is ALS or HOPM, which is summarized in Algorithm 6.1.1.

It follows from [13, 16] that $(N+1)$-tuple $(\sigma_*; \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ is a solution of (6.1.1) and σ_* is the norm of the multilinear function $F : \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \cdots \times \mathbb{R}^{I_N} \to \mathbb{R}$ induced by the norm $\|\cdot\|_2$ and defined as

$$\|\mathcal{A}\|_2 := \sup \frac{|F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)|}{\|\mathbf{x}_1\|_2 \|\mathbf{x}_2\|_2 \ldots \|\mathbf{x}_N\|_2},$$

where the supremum is taken over all nonzero $\mathbf{x}_n \in \mathbb{R}^{I_n}$ for all n.

The nuclear norm [17] of $\mathcal{A}$ is defined as

$$\|\mathcal{A}\|_* := \min \left\{ \sum_{r=1}^{R} |\lambda_r| : \mathcal{A} = \sum_{r=1}^{R} \lambda_r \mathbf{x}_{1,r} \otimes \mathbf{x}_{2,r} \otimes \cdots \otimes \mathbf{x}_{N,r} \right\}$$

Algorithm 6.1.1 The alternating least squares algorithm for (6.1.4) [15, Algorithm 3.2]

Input: $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$.
Output: $\widehat{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$: estimator of the best rank-one approximation of $\mathcal{A}$.

1: Initial values: $\mathbf{x}_n(0)$ is the dominant left singular vector of $\mathbf{A}_{(n)}$ with $n = 2, 3, \ldots, N$ or repeat the algorithm for several initial values.
2: Iterate until convergence:

1. Compute $\widetilde{\mathbf{x}}_1(k+1) = F(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k))_1$, $\lambda_1(k+1) = \|\widetilde{\mathbf{x}}_1(k+1)\|_2$ and $\mathbf{x}_1(k+1) = \widetilde{\mathbf{x}}_1(k+1)/\lambda_1(k+1)$;
2. Compute $\widetilde{\mathbf{x}}_2(k+1) = F(\mathbf{x}_1(k+1), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k))_2$, $\lambda_2(k+1) = \|\widetilde{\mathbf{x}}_2(k+1)\|_2$ and $\mathbf{x}_2(k+1) = \widetilde{\mathbf{x}}_2(k+1)/\lambda_2(k+1)$;
3. $\cdots$
4. Compute $\widetilde{\mathbf{x}}_N(k+1) = F(\mathbf{x}_1(k+1), \ldots, \mathbf{x}_{N-1}(k+1), \mathbf{x}_N(k))_N$, $\lambda_N(k+1) = \|\widetilde{\mathbf{x}}_N(k+1)\|_2$ and $\mathbf{x}_N(k+1) = \widetilde{\mathbf{x}}_N(k+1)/\lambda_N(k+1)$.

3: Collect converged values: $\mathbf{x}_n$ and λ.
4: Compute $\widehat{\mathcal{A}} = \lambda(\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_N)$.

where $\mathbf{x}_{n,r} \in \mathbb{R}^{I_n}$ have unit norm for all n. The spectral norm $\|\cdot\|_2$ is dual to the nuclear norm $\|\cdot\|_*$, i.e.,

$$\begin{cases} \|\mathcal{A}\|_2 = \max\{|\langle \mathcal{A}, \mathcal{X}\rangle| : \mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, \|\mathcal{X}\|_* = 1\}; \\ \|\mathcal{A}\|_* = \max\{|\langle \mathcal{A}, \mathcal{Y}\rangle| : \mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, \|\mathcal{Y}\|_2 = 1\}. \end{cases}$$

Nie [18] considers how to compute symmetric tensor nuclear norms, depending on the order and ground field.

Since the KKT conditions of the optimization problem (6.1.4) is the system of nonlinear equations (6.1.1), then $\mathbf{u}_n$ are the *mode-n principal singular vectors* of $\mathcal{A}$, corresponding to the singular value σ_* with $|\sigma_*| = \max\{|\sigma| : \sigma \in \sigma(\mathcal{A})\}$. The rank-one tensor $(\delta(\sigma_*)\sigma_*) \cdot (\delta(\mathbf{u}_1)\mathbf{u}_1 \otimes \delta(\mathbf{u}_2)\mathbf{u}_2 \otimes \cdots \otimes \delta(\mathbf{u}_N)\mathbf{u}_N)$ is also the best rank-one approximation for the tensor $\mathcal{A}$ with $\delta(\sigma_*) \prod_{n=1}^{N} \delta(\mathbf{u}_n) = 1$.

Remark 6.1.1 We can force $\sigma_* \geq 0$ by altering the sign of $\mathbf{u}_n$ with

$$\prod_{n=1}^{N} \delta(\mathbf{u}_n) = \begin{cases} 1, & \sigma_* \geq 0, \\ -1, & \sigma_* < 0. \end{cases}$$

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ is a local solution of the optimization problem (6.1.4), where $\|\mathbf{v}_n\|_2 = 1$ for all n, then $\tau \cdot (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_N)$ is called a *locally optimal rank-one approximation* of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, where $\tau = F(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$. It is obvious that the best rank-one approximation of $\mathcal{A}$ is also a locally optimal rank-one approximation of $\mathcal{A}$.

6.1.2 Tensor Z-Eigenvalues and the Symmetric Rank-One Approximation

In this section, we assume that tensors are symmetric and real. Zhang et al. [19] prove that when $\mathcal{A} \in RT_{N,I}$ is symmetric, the best rank-one approximation of $\mathcal{A}$ is also the best symmetric rank-one approximation of $\mathcal{A}$.

For such a case, the system of nonlinear equations (6.1.1) can be simplified to

$$\mathcal{A}\mathbf{x}^{N-1} = \mathcal{A} \times_2 \mathbf{x}^\top \times_3 \mathbf{x}^\top \ldots \times_N \mathbf{x}^\top = \sigma \mathbf{x}, \quad \|\mathbf{x}\|_2 = 1. \tag{6.1.5}$$

If any pair $(\sigma; \mathbf{x})$ satisfies (6.1.5), then we call σ as a Z-eigenvalue of the tensor $\mathcal{A}$ and $\mathbf{x}$ is the associated Z-eigenvector [20, 21].

We review the symmetric higher-order power method (S-HOPM), summarized in Algorithm 6.1.2, for the comptation of Z-eigenpairs of $\mathcal{A}$, which is introduced by De Lathauwer et al. [15] and analyzed by Kofidis and Regalia [22].

Algorithm 6.1.2 The symmetric higher-order power method [23, Algorithm 1]

Input: Symmetric $\mathcal{A} \in RT_{N,I}$.
Require: $\mathbf{x}_0 \in \mathbb{R}^I$ with $\|\mathbf{x}_0\|_2 = 1$. Let $\sigma_0 = \mathcal{A}\mathbf{x}_0^N$.
1: **for** $k = 1, 2, \ldots$ **do**
2: Compute $\mathbf{y}_k = \mathcal{A}\mathbf{x}_k^{N-1}$.
3: Normalize $\mathbf{x}_{k+1} = \mathbf{y}_k / \|\mathbf{y}_k\|_2$.
4: Compute $\sigma_{k+1} = \mathcal{A}\mathbf{x}_{k+1}^N$.
5: **end for**

When k is larger than a given positive integer or $\|\mathcal{A}\mathbf{x}_k^{N-1} - \sigma_k \mathbf{x}\|_2$ is less than tolerance, we can terminate Algorithm 6.1.2. Kolda and Mayo [22] present SS-HOPM, summarized in Algorithm 6.1.3, which is guaranteed to converge to an Z-eigenpair of $\mathcal{A}$.

Algorithm 6.1.3 The shifted symmetric higher-order power method [23, Algorithm 2]

Input: Symmetric $\mathcal{A} \in RT_{N,I}$.
Require: $\mathbf{x}_0 \in \mathbb{R}^I$ with $\|\mathbf{x}_0\|_2 = 1$ and $\alpha \in \mathbb{R}$. Let $\sigma_0 = \mathcal{A}\mathbf{x}_0^N$.
1: **for** $k = 1, 2, \ldots$ **do**
2: **if** $\alpha \geq 0$ **then**
3: Compute $\mathbf{y}_k = \mathcal{A}\mathbf{x}_k^{N-1} + \alpha \mathbf{x}_k$.
4: **else**
5: Compute $\mathbf{y}_k = -\mathcal{A}\mathbf{x}_k^{N-1} - \alpha \mathbf{x}_k$.
6: **end if**
7: Normalize $\mathbf{x}_{k+1} = \mathbf{y}_k / \|\mathbf{y}_k\|_2$.
8: Compute $\sigma_{k+1} = \mathcal{A}\mathbf{x}_{k+1}^N$.
9: **end for**

It is obvious that the tensor Z-eigenvalue problem is a special case of the tensor singular value problem. As shown in Sect. 6.1.1, we can ensure that all singular values of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ are positive by changing the sign of all mode-n singular vectors. However, the statement is not true for Z-eigenvalues. If a pair $(\sigma; \mathbf{x})$ satisfies (6.1.5), the pair $(\sigma; -\mathbf{x})$ solves (6.1.5) with an even N, and the pair $(-\sigma; -\mathbf{x})$ solves (6.1.5) with an odd N [23]. Hence, for a symmetric tensor $\mathcal{A}$, we cannot make all its Z-eigenvalues nonnegative by changing the sign of the associated Z-eigenvector.

The best symmetric rank-one approximation of the symmetric tensor $\mathcal{A}$ can also be formulated as follows: to find $\sigma \in \mathbb{R}$ and a unit vector $\mathbf{u} \in \mathbb{R}^I$ such that

$$\sigma \cdot \mathbf{u}^{\otimes N} = \operatorname{argmin} \left\| \mathcal{A} - \tau \cdot \mathbf{x}^{\otimes N} \right\|_F, \tag{6.1.6}$$

subject to $\tau \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^I$ satisfying $\|\mathbf{x}\|_2 = 1$. In the spirit of Theorem 6.1.1 and the results in [19], the minimization problem (6.1.6) is equivalent to the optimization problem

$$\max \quad |f(\mathbf{x})| := |F(\underbrace{\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}}_{N})| \quad \text{s.t.} \quad \|\mathbf{x}\|_2 = 1. \tag{6.1.7}$$

If $\mathbf{u}$ is a global maximizer of the optimization problem (6.1.7) and $\sigma_* = f(\mathbf{u})$, then $\sigma_* \cdot \mathbf{u}^{\otimes N}$ is the best symmetric rank-one approximation of $\mathcal{A}$. Thus two symmetric rank-one tensors $\sigma_* \cdot \big(\delta(\mathbf{u})\mathbf{u}\big)^{\otimes N}$ (for an even N and $\delta(\mathbf{u}) = \pm 1$), and $(\delta(\sigma_*)\sigma_*) \cdot \big(\delta(\mathbf{u})\mathbf{u}\big)^{\otimes N}$ (for an odd N and $\delta(\sigma_*)\delta(\mathbf{u}) = 1$) are also the best symmetric rank-one approximation of the symmetric tensor $\mathcal{A}$.

Furthermore, if $\mathbf{v}$ is a local maximizer of the optimization problem (6.1.7) and $\tau = f(\mathbf{v})$, then $\tau \cdot \mathbf{v}^{\otimes N}$ is a local optimal symmetric rank-one approximation of $\mathcal{A}$. It is obvious that the best symmetric rank-one approximation of $\mathcal{A}$ is the locally optimal symmetric rank-one approximation of $\mathcal{A}$.

6.2 Neural Networks Models

We propose the neural network model for computing the locally optimal rank-one approximation of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and the locally optimal symmetric rank-one approximation of a symmetric tensor $\mathcal{A} \in RT_{N,I}$. We prove that the proposed neural networks are asymptotically stable in the sense of Lyapunov stability theory.

6.2.1 Tensor Rank-One Approximation

Suppose that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. We present the dynamics of the neural network model for computing the nonzero singular values and the associated mode-n singular vectors.

For the nonzero singular values of $\mathcal{A}$, the dynamics of the neural network model is described by

$$\frac{d\mathbf{x}_n(t)}{dt} = \frac{F\left(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\right)_{-n}}{F\left(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\right)} - \mathbf{x}_n(t), \tag{6.2.1}$$

for $t \geq 0$, where $F\left(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\right)_{-n}$ and $F\left(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\right)$ are defined in (6.1.1) and (6.1.3), respectively, and $\mathbf{x}_n(t) \in \mathbb{R}^{I_n}$ is the state of the neural network. For arbitrary nonzero initial values $\mathbf{x}_n(0) \in \mathbb{R}^{I_n}$ and all n, a simplest discrete-time iterative algorithm corresponding to the neural network in (6.2.1) is

$$\mathbf{x}_n(k+1) = \mathbf{x}_n(k) + \alpha \left(\frac{F\left(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\right)_{-n}}{F\left(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\right)} - \mathbf{x}_n(k) \right), \tag{6.2.2}$$

where $\alpha > 0$ is called the *time-step length* or *learning rate*.

In general, (6.2.2) is a special case of the following adaptive algorithm,

$$\mathbf{x}_n(k+1) = \mathbf{x}_n(k) + \alpha_k \left(\frac{F\left(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\right)_{-n}}{F\left(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\right)} - \mathbf{x}_n(k) \right),$$

where α_k is a decreasing gain sequence.

From [24], we have the assumption that the gain sequence $\{\eta_k \geq 0\}$ is decreasing such that $\sum_{k=0}^{\infty} \alpha_k = \infty$, $\sum_{k=0}^{\infty} \alpha_k^r < \infty$ for some $r > 1$ and $\lim_{k\to\infty}(\alpha_k^{-1} - \alpha_{k-1}^{-1}) < \infty$. For simplicity, we assume that α in (6.2.2) is constant in each step.

In particular, when $\alpha = 1$, the discrete-time iterative algorithm (6.2.2) for computing a locally optimal rank-one approximation of $\mathcal{A}$ is reduced to ALS or HOPM. However, for a general α, such an iterative scheme is not simply a generalization of ALS.

For the $(k+1)$th step, the key point of ALS is to find the maximizer of the optimization problem (6.1.4) on the subspace

$$\otimes_{n=1}^{N} \left\{ F\left(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\right)_{-n} \right\}.$$

However, in (6.2.2), this maximizer is found on the subspace

$$\otimes_{n=1}^{N} \operatorname{span} \left\{ \mathbf{x}_n(k), F\left(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\right)_{-n} \right\}.$$

span$\{\mathbf{x}, \mathbf{y}\} \subset \mathbb{R}^I$ is the subspace generated by $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$, and $\mathbb{U} \otimes \mathbb{V} = \{(\mathbf{u}, \mathbf{v}) : \mathbf{u} \in \mathbb{U}, \mathbf{v} \in \mathbb{V}\}$ for $\mathbb{U}, \mathbb{V} \subset \mathbb{R}^I$.

It is worth noting that several existing algorithms can be extended to solve the local optimal best rank-one approximation of a real tensor, but their optimal learning rate is difficult to determine. The main reason to adopt (6.2.1) is that the learning rate in (6.2.2) can be easily selected in advance. We establish the following theorem.

Theorem 6.2.1 *Given the neural network in* (6.2.1) *with arbitrary initial conditions* $\mathbf{x}_n(0)$ *for all* n*, then* $\|\mathbf{x}_n(t)\|_2$ *converges exponentially to* 1 *as* $t \to +\infty$*, and the convergence is independent of* $\mathcal{A}$*.*

Proof This theorem is similar to Lemma 1 in [6], we neglect the proof. □

Remark 6.2.1 We refer to $\|\mathbf{x}_n(t)\|_2^2 \equiv 1$ if $\|\mathbf{x}_n(0)\|_2^2 = 1$ as the unit-norm conservation. From Theorem 6.2.1, if $\|\mathbf{x}_n(0)\|_2^2 > 1$, then $\|\mathbf{x}_n(t)\|_2^2 > 1$ for $t > 0$. Similarly, if $\|\mathbf{x}_n(0)\|_2^2 < 1$, then $\|\mathbf{x}_n(t)\|_2^2 < 1$ for $t > 0$. According to Theorem 6.2.1, the suitable learning rate α in (6.2.2) satisfies $0 < \alpha < 0.5$ for the stability of the algorithm.

From Theorem 6.2.1, we deduce the following result.

Corollary 6.2.1 *For any bounded initial values of* $\mathbf{x}_n(0)$ *and all* n*, the state vectors of the neural network in* (6.2.1) *are bounded.*

By using $G_n(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t))$ to denote the right-hand side of the neural network in (6.2.1), we have the following lemma.

Lemma 6.2.1 *Suppose that there exist* N *unit vectors* $\mathbf{x}_n^*$ *such that* $G_n(\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_N^*) = \mathbf{0}_I$*. For the neural network in* (6.2.1)*, given any initial values* $\mathbf{x}_n(0) \in \mathbb{R}^{I_n}$ *with* $\|\mathbf{x}_n(0)\|_2 = 1$*, let* $\mathbf{x}_n(t) \to \mathbf{x}_n^*$ *as* $t \to +\infty$*. Then* σ^* *and* $\mathbf{x}_n^*$ *are the singular value and the corresponding mode-*n *singular vectors of the tensor* $\mathcal{A}$*, respectively, where* $\sigma^* = F(\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_N^*)$*.*

Proof The proof is given in [25] and we omit it. □

6.2.2 Symmetric Tensor Rank-One Approximation

We present a neural network in (6.2.1) to compute the local maximizers of the optimization problem (6.1.4). As shown in Sect. 6.1.2, there is a relationship between the tensor singular value problem and the symmetric tensor Z-eigenvalue problem. In this section, we focus on the symmetric rank-one approximation of a symmetric tensor, and assume that all tensors are real symmetric tensors.

We now state how to find a locally optimal symmetric rank-one approximation of the symmetric tensor $\mathcal{A}$ by a neural network, similar to (6.2.1) for the tensor rank-one approximation problem. The dynamics of the neural network can be described

by the ODE

$$\frac{d\mathbf{x}(t)}{dt} = \frac{\mathcal{A}\mathbf{x}(t)^{N-1}}{\mathcal{A}\mathbf{x}(t)^{N}} - \mathbf{x}(t), \tag{6.2.3}$$

for $t \geq 0$, where $\mathbf{x}(t) \in \mathbb{R}^{I_n}$ represents the state of the neural network.

We present several results related to (6.2.3), similar to Theorem 6.2.1, Corollary 6.2.1 and Lemma 6.2.1 for (6.2.1).

Theorem 6.2.2 *Given the neural network in* (6.2.3) *for an arbitrary initial value* $\mathbf{x}(0) \in \mathbb{R}^I$, $\|\mathbf{x}(t)\|_2$ *exponentially converges to* 1 *as* $t \to +\infty$, *and the convergence is independent of* $\mathcal{A}$.

For an arbitrary nonzero initial value $\mathbf{x}(0) \in \mathbb{R}^I$, a discrete-time iterative algorithm corresponding to (6.2.3) is

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \alpha \left(\frac{\mathcal{A}\mathbf{x}(k)^{N-1}}{\mathcal{A}\mathbf{x}(k)^{N}} - \mathbf{x}(k) \right), \tag{6.2.4}$$

where $\alpha > 0$ is the *learning rate*.

According to Theorem 6.2.2, a suitable learning rate α in (6.2.4) can be taken in $(0, 0.5)$.

Corollary 6.2.2 *For any bounded initial values* $\mathbf{x}(0)$, *the state vectors of the neural network in* (6.2.3) *are bounded.*

By using $G(\mathbf{x}(t))$ to denote the right-hand side of (6.2.3), we have the following lemma.

Lemma 6.2.2 *Suppose that there exists a unit vectors* $\mathbf{x}^*$ *such that* $G(\mathbf{x}^*) = \mathbf{0}_I$. *Consider the neural network in* (6.2.3), *given* $\mathbf{x}(0) \in \mathbb{R}^I$ *with* $\|\mathbf{x}(0)\|_2 = 1$. *If* $\mathbf{x}(t) \to \mathbf{x}^*$ *as* $t \to +\infty$, *then* $\sigma^* = f(\mathbf{x}^*)$ *and* $\mathbf{x}^*$ *are the* Z*-eigenvalue and the corresponding* Z*-eigenvector of* $\mathcal{A}$, *respectively.*

For the positive Z-eigenvalues of symmetric positive definite tensors, Samardzijia and Waterland [26] propose a dynamical ODE system to estimate their Z-eigenvectors:

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{A}\mathbf{x}(t)^{N-1} - f_p(\mathbf{x}(t))\mathbf{x}(t), \quad (p > N-1) \tag{6.2.5}$$

for $t \geq 0$, where $f_p(\mathbf{x}) > 0$ is a positive 1-dimensional p-form, that is, $f_p(\mathbf{x})$ is a p-degree homogeneous polynomial with real-coefficients for all nonzero $\mathbf{x} \in \mathbb{R}^I$. For $N \geq 2$, the nontrivial solutions, or fixed points, of (6.2.5) are the Z-eigenvectors of the symmetric positive definite tensor $\mathcal{A}$ associated with the Z-eigenvalues. Vegas and Zufiria [27] analyze the neural network in (6.2.5) for matrices and prove a local stability result.

It is interesting that the ODE system (6.2.5) can be used to compute the Z-eigenvectors of $\mathcal{A}$ with $f_p(\mathbf{x}) = \mathcal{A}\mathbf{x}^N$, and the associated neural network is

$$\frac{d}{dt}\mathbf{x}(t) = \mathcal{A}\mathbf{x}(t)^{N-1} - \mathcal{A}\mathbf{x}(t)^N\mathbf{x}(t), \quad t \geq 0.$$

We can generalize the ODE system (6.2.5) to compute the mode-n singular vectors of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ as follows:

$$\frac{d}{dt}\mathbf{x}_n(t) = F\big(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\big)_{-n} - F\big(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\big)\mathbf{x}_n(t), \tag{6.2.6}$$

for $t \geq 0$, where $\mathbf{x}_n(t) \in \mathbb{R}^{I_n}$ represents the state.

We do not prove that the neural networks in (6.2.6) is locally asymptotically stable in the sense of Lyapunov stability theory. Howevre, based on the KKT conditions, Che et al. [28] propose a neural network model for the low multilinear rank approximation of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with the known multilinear rank $\{R_1, R_2, \ldots, R_N\}$ and prove that the state of this model is locally stable. For the case of $R_n = 1$ and all n, the neural networks in (6.2.6) is a special case of that of [28].

The definition of a higher-order multi-partially symmetric tensor is presented as follows.

Definition 6.2.1 Suppose $\mathcal{A} \in \mathbb{C}^{J_1 \times J_2 \times \cdots \times J_{P_1 + \cdots + P_N}}$. Define $Q_0 = 1$ and $Q_n = P_1 + \cdots + P_n$ for all n and N given positive integers I_n, with $J_q = I_n$ for $q = Q_{n-1} + 1, Q_{n-1} + 2, \ldots, Q_n$. If the entries of $\mathcal{A}$ satisfy

$$\begin{aligned} a_{i_{1,1}\ldots i_{1,P_1} i_{2,1}\ldots i_{2,P_2}\ldots i_{N,1}\ldots i_{N,P_N}} &= a_{\pi_1(i_{1,1})\ldots\pi_1(i_{1,P_1})i_{2,1}\ldots i_{2,P_2}\ldots i_{N,1}\ldots i_{N,P_N}} \\ &= a_{i_{1,1}\ldots i_{1,P_1}\pi_2(i_{2,1})\ldots\pi_2(i_{2,P_2})\ldots i_{N,1}\ldots i_{N,P_N}} \\ &= \ldots \\ &= a_{i_{1,1}\ldots i_{1,P_1} i_{2,1}\ldots i_{2,P_2}\ldots\pi_N(i_{N,1})\ldots\pi_N(i_{N,P_N})}, \end{aligned}$$

where $\pi_n \in \mathbb{S}_{I_n}$, $\mathcal{A}$ is called higher-order multi-partially symmetric. We use $ST_{I_1 \times \cdots \times I_N}^{[P_1, \ldots, P_N]}$ to denote the set of all $(P_1 + P_2 + \cdots + P_N)$th-order multi-partially symmetric tensors.

Remark 6.2.2 In Sect. 1.3, we reveal the relationship between the biquadratic optimization problem and the rank-one approximation of fourth-order partially symmetric tensors. Wang et al. [29] define the higher-order multi-partially symmetric tensors, as a generalization of fourth-order partially symmetric tensors, and consider the computation of the multi-partially symmetric rank-one approximations for higher-order multi-partially symmetric tensors via neural networks. This is generalization of the partially symmetric rank-one approximation for fourth-order partially symmetric tensors, related to the biquadratic optimization problem.

For any $\mathcal{A} \in ST_{I_1 \times \cdots \times I_N}^{[P_1,\ldots,P_N]}$ and arbitrary $\mathbf{x}_n \in \mathbb{R}^{I_n}$, we define a polynomial

$$\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_N^{P_N} := \mathcal{A} \times_1 \mathbf{x}_1^\top \cdots \times_{P_1} \mathbf{x}_1^\top \cdots \times_{(P_1+\cdots+P_{N-1}+1)} \mathbf{x}_N^\top \cdots \times_{(P_1+\cdots+P_N)} \mathbf{x}_N^\top.$$

The partial derivative of $\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_N^{P_N}$ with respect to $\mathbf{x}_n$ is denoted by $P_n\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_n^{P_n-1} \ldots \mathbf{x}_N^{P_n}$, where $\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_n^{P_n-1} \ldots \mathbf{x}_N^{P_n}$ is defined as

$$\mathcal{A} \times_1 \mathbf{x}_1^\top \cdots \times_{Q_1} \mathbf{x}_1^\top \cdots \times_{Q_{n-1}} \mathbf{x}_{n-1}^\top \times_{Q_{n-1}+2} \mathbf{x}_n^\top \cdots \times_{Q_n} \mathbf{x}_n^\top \cdots \times_{Q_{N-1}+1} x_N^\top \cdots \times_{Q_N} \mathbf{x}_N^\top,$$

with $Q_n = P_1 + P_2 + \cdots + P_n$ for all n.

For a given $\sigma \in \mathbb{R}$ and unit vectors $\mathbf{x}_n \in \mathbb{R}^{I_n}$, if $(\sigma, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is a solution of the nonlinear equations

$$\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_n^{P_n-1} \ldots \mathbf{x}_N^{P_N} = \sigma \mathbf{x}_n; \quad \|\mathbf{x}_n\|_2 = 1,$$

then σ and $\mathbf{x}_n$ are called the M-singular value and the nth *M-singular vector* of $\mathcal{A} \in ST_{I_1 \times \cdots \times I_N}^{[P_1,\ldots,P_N]}$.

For any nonzero M-singular value of $\mathcal{A} \in ST_{I_1 \times \cdots \times I_N}^{[P_1,\ldots,P_N]}$, the dynamics of the neural network model can be described by

$$\frac{d\mathbf{x}_n(t)}{dt} = \frac{\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_{n-1}^{P_{n-1}} \mathbf{x}_n^{P_n-1} \mathbf{x}_{n+1}^{P_{n+1}} \ldots \mathbf{x}_N^{P_N}}{\mathcal{A}\mathbf{x}_1^{P_1} \ldots \mathbf{x}_N^{P_N}} - \mathbf{x}_n(t), \tag{6.2.7}$$

for $t \geq 0$, where $\mathbf{x}_n(t) \in \mathbb{R}^{I_n}$ represents the state of the network. As shown in [29], we shall prove that the solution of (6.2.7) is locally asymptotically stable in the sense of Lyapunov stability theory.

6.3 Asymptotic Stability

We shall show that the neural networks in (6.2.1) and (6.2.3) are locally asymptotically stable in the sense of Lyapunov stability theory.

6.3.1 Tensor Rank-One Approximation

We see that σ^* in Lemma 6.2.1 must be nonzero for the neural network described in (6.2.1). For all local maximizers of the optimization problem (6.1.4), two distinct subsets are

$$\begin{cases} \mathbb{S}_+ = \{\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} : \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} \text{ solves (6.1.4) and } F(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) \geq 0\}, \\ \mathbb{S}_- = \{\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} : \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} \text{ solves (6.1.4) and } F(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) < 0\}. \end{cases}$$

It is obvious that $\mathbb{S}_+ \cap \mathbb{S}_- = \emptyset$ and $\mathbb{S} = \mathbb{S}_+ \cup \mathbb{S}_-$ is the set of local maximizers of (6.1.4). Based on the relationship between the tensor singular value problem and the tensor rank-one approximation problems, it is easy to see that $\mathbf{v}_n$ for all n are nonzero with $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\} \in \mathbb{S}$.

Given any local maximizer $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} \in \mathbb{S}$, we define $\epsilon_0 = \min \|\mathbf{u} - \mathbf{v}\|_2$, where $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ and $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ with $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\} \in \mathbb{S}$ and $\mathbf{u}_n \neq \mathbf{v}_n$ for all n. We define a neighbourhood of $\mathbf{u}$ as

$$\mathbb{B}(\mathbf{u}; \hat{\epsilon}) := \mathbb{B}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N; \hat{\epsilon}) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{u}\|_2 \leq \hat{\epsilon}\} \tag{6.3.1}$$

where $\mathbf{x}_n \in \mathbb{R}^{I_n}$ and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ with $0 < \hat{\epsilon} \leq \epsilon_0$.

Since $F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is continuous and differentiable, $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ is a stationary point of $F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, then there exists $\epsilon \leq \hat{\epsilon}$ such that the sign of $F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is the same as the sign of $F(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ and

$$|F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)| \leq |F(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)|$$

for all $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \in \mathbb{B}(\mathbf{u}; \epsilon)$.

By (6.3.1), there exists a unique local maximizer of the optimization problem (6.1.4) in $\mathbb{B}(\mathbf{u}; \epsilon)$. Since $\|\mathbf{u}_n\|_2 = 1$ for all n, $\mathbf{x}_n$ are nonzero. The neural network in (6.2.1) is locally asymptotically stable in the sense of Lyapunov stability theory at a local maximizer from $\mathbb{S}$. We obtain the following theorem.

Theorem 6.3.1 *Suppose that* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$ *is a local maximizer of the optimization problem* (6.1.4) *with* $\mathbf{u}_n \in \mathbb{R}^{I_n}$. *If the initial unit values* $\mathbf{x}_n(0)$ *of the neural network in* (6.2.1) *belong to* $\mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$, *given in* (6.3.1), *with*

$$\delta(\mathbf{u})\mathbf{u} = (\delta(\mathbf{u}_1)\mathbf{u}_1, \delta(\mathbf{u}_2)\mathbf{u}, \ldots, \delta(\mathbf{u}_N)\mathbf{u}_N),$$

then the solution of (6.2.1) *is locally asymptotically stable in the sense of Lyapunov stability theory at* $\{\delta(\mathbf{u}_1)\mathbf{u}_1, \delta(\mathbf{u}_2)\mathbf{u}, \ldots, \delta(\mathbf{u}_N)\mathbf{u}_N\}$, *where* $\prod_{n=1}^{N} \delta(\mathbf{u}_n) = 1$ *with* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} \in \mathbb{S}_+$; *or* $\prod_{n=1}^{N} \delta(\mathbf{u}_n) = -1$ *with* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} \in \mathbb{S}_-$.

Proof For illustration, without loss of generality, we assume that $N = 3$ and $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \in \mathbb{S}_+$. We can see that $F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) > 0$ for all $\mathbf{x} \in \mathbb{B}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3; \epsilon)$ with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. For (6.2.1), we define the Lyapunov function as

$$V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \sigma_* - \frac{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{\|\mathbf{x}_1\|_2 \|\mathbf{x}_2\|_2 \|\mathbf{x}_3\|_2}.$$

From the proof of Theorem 6.2.1, we have $\|\mathbf{x}_n(t)\|_2 = 1$ for $t \geq 0$. Together with the definition of $\mathbb{B}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3; \epsilon)$, it is clear that $V(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) > 0$, where $\mathbf{x} \in \mathbb{B}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3; \epsilon)$ and $\mathbf{x}_n \neq \alpha_n \mathbf{u}_n$ with $\prod_{n=1}^{3} \delta(\alpha_n) = 1$. Let

$$G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \frac{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{\|\mathbf{x}_1\|_2 \|\mathbf{x}_2\|_2 \|\mathbf{x}_3\|_2},$$

we have

$$\frac{dV}{dt} = -\left(\frac{d\mathbf{x}_1}{dt}\right)^\top \left[\nabla_{\mathbf{x}_1} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right] - \left(\frac{d\mathbf{x}_2}{dt}\right)^\top \left[\nabla_{\mathbf{x}_2} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right] - \left(\frac{d\mathbf{x}_3}{dt}\right)^\top \left[\nabla_{\mathbf{x}_3} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right]. \quad (6.3.2)$$

Computing the partial derivative $\nabla_{\mathbf{x}_1} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, we obtain

$$\begin{aligned}\nabla_{\mathbf{x}_1} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) =& \frac{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2 F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1}}{(\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2)^2} - \frac{\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2 F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{(\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2)^2} \cdot \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|_2} \\ =& \frac{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1}}{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2} - \frac{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{\|\mathbf{x}_1\|_2^3\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2}\mathbf{x}_1 \\ =& \frac{1}{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2}\left(F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1} - \frac{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{\|\mathbf{x}_1\|_2^2}\mathbf{x}_1\right).\end{aligned}$$

By routine manipulations, we have

$$\begin{aligned}&\left(\frac{d\mathbf{x}_1}{dt}\right)^\top \left[\nabla_{\mathbf{x}_1} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right] \\ &= \frac{(F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1} - \mathbf{x}_1)^\top}{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)} \nabla_{\mathbf{x}_1} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \\ &= \frac{1}{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2\|\mathbf{x}_3\|_2}\left[\frac{(F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1})^\top F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1}}{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)} - \frac{F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{\|\mathbf{x}_1\|_2^2}\right].\end{aligned}$$

By the Cauchy-Schwartz inequality, we obtain

$$F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)^2 \le \left(F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1}\right)^\top F(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{-1}\|\mathbf{x}_1\|_2^2.$$

Hence

$$\left(\frac{d\mathbf{x}_1}{dt}\right)^\top \left[\nabla_{\mathbf{x}_1} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right] \ge 0,$$

and the equality holds if and only if $\mathbf{x}_n = \alpha_n \mathbf{u}_n$ with $\prod_{n=1}^3 \delta(\alpha_n)\alpha_n = 1$. Similarly, we have

$$\left(\frac{d\mathbf{x}_2}{dt}\right)^\top \left[\nabla_{\mathbf{x}_2} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right] \ge 0, \quad \left(\frac{d\mathbf{x}_3}{dt}\right)^\top \left[\nabla_{\mathbf{x}_3} G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\right] \ge 0.$$

According to (6.3.2), we have

$$\frac{dV(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{dt} < 0$$

for $t \geq 0$, where $\mathbf{x} \in \mathbb{B}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3; \epsilon)$ and $\mathbf{x}_n \neq \alpha_n \mathbf{u}_n$ with $\prod_{n=1}^{3} \delta(\alpha_n) = 1$. The neural network in (6.2.1) is locally asymptotically stable in the sense of Lyapunov stability theory at $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \in \mathbb{S}_+$.

Moreover, we can prove that the solution of (6.2.1) is locally asymptotically stable in the sense of Lyapunov stability theory at $(\delta(\mathbf{u}_1)\mathbf{u}_1, \delta(\mathbf{u}_2)\mathbf{u}_2, \delta(\mathbf{u}_3)\mathbf{u}_3)$ with $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \in \mathbb{S}_+$ and $\prod_{n=1}^{3} \delta(\mathbf{u}_n) = 1$.

The case of local maximizers $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \in \mathbb{S}_-$ is similar to that in $\mathbb{S}_+$. Moreover, we can prove that the solution of (6.2.1) is locally asymptotically stable in the sense of Lyapunov stability theory at

$$(\delta(\mathbf{u}_1)\mathbf{u}_1, \delta(\mathbf{u}_2)\mathbf{u}_2, \delta(\mathbf{u}_3)\mathbf{u}_3)$$

with $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \in S_-$ and $\prod_{n=1}^{3} \delta(\mathbf{u}_n) = -1$.

In general, we prove that the solution of (6.2.1) is *locally asymptotically stable* in the sense of Lyapunov stability theory at

$$(\delta(\mathbf{u}_1)\mathbf{u}_1, \delta(\mathbf{u}_2)\mathbf{u}, \ldots, \delta(\mathbf{u}_N)\mathbf{u}_N),$$

where

$$\prod_{n=1}^{N} \delta(\mathbf{u}_n) = \begin{cases} 1, & \text{with} \quad (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) \in \mathbb{S}_+; \\ -1, & \text{with} \quad (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) \in \mathbb{S}_-. \end{cases}$$

The theorem is proved. □

Remark 6.3.1 Theorem 6.3.1 indicates, independent of the nonzero initial conditions, that the neural network in (6.2.1) generally has a sub-linear speed of convergence. The neural networks in (6.2.3) and (6.5.3) have similar results.

6.3.2 Symmetric Tensor Rank-One Approximation

By Lemma 6.2.2, σ_* must be nonzero for the neural network in (6.2.3). Suppose that $\mathbf{u}$ is a local maximizer of the optimization problem (6.1.7), we define a neighborhood of $\mathbf{u}$ as

$$\mathbb{B}(\mathbf{u}; \epsilon) = \{\mathbf{x} : \|\mathbf{u} - \mathbf{x}\|_2 \leq \hat{\epsilon}\}, \tag{6.3.3}$$

where $0 < \hat{\epsilon} \leq \epsilon_0 := \min \|\mathbf{u} - \mathbf{v}\|_2$ and $\mathbf{v} \neq \mathbf{u}$ is another local maximizer of (6.1.7).

Since $f(\mathbf{x})$ is continuous and differentiable, $\mathbf{u}$ is a stationary point of $f(\mathbf{x})$, there exists $\epsilon \leq \hat{\epsilon}$ such that $\delta(f(\mathbf{x}))$ is the same as $\delta(f(\mathbf{u}))$ and $|f(\mathbf{x})| \leq |f(\mathbf{u})|$ for all $\mathbf{x} \in \mathbb{B}(\mathbf{u}; \epsilon)$.

Theorem 6.3.1 states that the neural network in (6.2.1) is locally asymptotically stable in the sense of Lyapunov stability theory at a local maximizer of (6.1.4), belonging to $\mathbb{S}$. For the neural network in (6.2.3), we have a similar result, summarized in following theorem.

Theorem 6.3.2 *Suppose that* $\mathbf{u} \in \mathbb{R}^I$ *is a local maximizer of the problem* (6.1.7). *For* $\mathbf{x}(0) \in \mathbb{B}(\delta(\mathbf{u}), \epsilon)$, *defined by* (6.3.3), *then the solution of* (6.2.3) *is locally asymptotically stable in the sense of Lyapunov stability theory at* $\delta(\mathbf{u})\mathbf{u}$.

Proof Suppose $\mathbf{u} \in \mathbb{R}^I$ is a local maximizer of (6.1.7). According to the sign of $f(\mathbf{u})$ and the value of the positive integer N, the proof procedure is divided into four cases: (i) N is even and $f(\mathbf{u}) < 0$, (ii) N is even and $f(\mathbf{u}) > 0$, (iii) N is odd and $f(\mathbf{u}) < 0$, and (iv) N is odd and $f(\mathbf{u}) > 0$.

For (i), we have $0 > f(\mathbf{x}) \geq f(\mathbf{u})$ with $\mathbf{x} \in \mathbb{B}(\mathbf{u}, \epsilon)$. We cannot change the sign of $\mathbf{u}$ to ensure $f(\delta(\mathbf{u})\mathbf{u}) > 0$, because if $(\sigma; \mathbf{v})$ is a Z-eigenpair of the symmetric tensor $\mathcal{A}$, then $(\sigma; -\mathbf{v})$ is also a Z-eigenpair [23]. Hence, by replacing the symmetric tensor $\mathcal{A}$, we rewrite the neural network in (6.2.3) as

$$\frac{d\mathbf{x}(t)}{dt} = \frac{(-\mathcal{A})\mathbf{x}(t)^{N-1}}{(-\mathcal{A})\mathbf{x}(t)^N} - \mathbf{x}(t) = \frac{\mathcal{A}\mathbf{x}(t)^{N-1}}{\mathcal{A}\mathbf{x}(t)^N} - \mathbf{x}(t),$$

and the associated Lyapunov function is

$$V(\mathbf{x}) = -\sigma_* - \frac{(-\mathcal{A})\mathbf{x}^N}{\|\mathbf{x}\|_2^N} = -\sigma_* + \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_2^N}, \quad \sigma_* = f(\mathbf{u}) < 0.$$

By Theorem 6.2.2, we have $\|\mathbf{x}(t)\|_2 = 1$ for all $t \geq 0$. According to the definition of $\mathbb{B}(\mathbf{u}, \epsilon)$ and the property of $f(\mathbf{x})$ with $\mathbf{x} \in \mathbb{B}(\mathbf{u}, \epsilon)$, we have $V(\mathbf{x}) > 0$ where $\mathbf{x} \in \mathbb{B}(\mathbf{u}, \epsilon)$ and $\mathbf{x} \neq \alpha\mathbf{u}$ with a nonzero α.

We then have

$$\nabla_{\mathbf{x}} G(\mathbf{x}) = \frac{N}{\|\mathbf{x}\|_2^N}\left(\mathcal{A}\mathbf{x}^{N-1} - \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_2^2}\mathbf{x}\right), \quad G(\mathbf{x}) = \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_2^N}.$$

After some manipulation, we obtain

$$\begin{aligned}
\frac{dV}{dt} &= \left(\frac{d\mathbf{x}(t)}{dt}\right)^\top [\nabla_{\mathbf{x}} G(\mathbf{x})] = \frac{N}{\|\mathbf{x}\|_2^N}\left(\frac{\mathcal{A}\mathbf{x}^{N-1}}{\mathcal{A}\mathbf{x}^N} - \mathbf{x}\right)^\top \left(\mathcal{A}\mathbf{x}^{N-1} - \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_2^2}\mathbf{x}\right) \\
&= \frac{N}{\|\mathbf{x}\|_2^N}\left(\frac{(\mathcal{A}\mathbf{x}^{N-1})^\top \mathcal{A}\mathbf{x}^{N-1}}{\mathcal{A}\mathbf{x}^N} - \frac{\mathcal{A}\mathbf{x}^N}{\|\mathbf{x}\|_2^2} - \mathcal{A}\mathbf{x}^N + \mathcal{A}\mathbf{x}^N\right) \\
&= \frac{N}{\|\mathbf{x}\|_2^{N+2}\mathcal{A}\mathbf{x}^N}\left(\|\mathbf{x}\|_2^2((\mathcal{A}\mathbf{x}^{N-1})^\top \mathcal{A}\mathbf{x}^{N-1}) - (\mathcal{A}\mathbf{x}^N)^2\right) \\
&= -\frac{N}{\|\mathbf{x}\|_2^{N+2}(-\mathcal{A})\mathbf{x}^N}\left(\|\mathbf{x}\|_2^2(((-\mathcal{A})\mathbf{x}^{N-1})^\top (-\mathcal{A})\mathbf{x}^{N-1}) - ((-\mathcal{A})\mathbf{x}^N)^2\right) \\
&\leq 0
\end{aligned}$$

by the Cauchy-Schwartz inequality and the equality holds if and only if $\mathbf{x} = \alpha\mathbf{u}$ with a nonzero α. Note that $f(\mathbf{u}) = f(\delta(\mathbf{u})\mathbf{u})$ with $\delta(\mathbf{u}) = \pm 1$. The case of $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}, \epsilon)$ is similar to that of $\mathbf{x} \in \mathbb{B}(\mathbf{u}, \epsilon)$. Case (i) is proved.

Since the last three cases are similar to Case (i), we leave the proof as an exercise. □

6.4 Generalized Models: TCCA

In Sect. 6.1.1, when we define the tensor singular values and tensor singular vectors, we restrict $\|\mathbf{x}_n\|_2 = 1$ for all n. In this section, we give a more general form of the constraints. For N given symmetric positive definite matrices $\mathbf{B}_n \in \mathbb{R}^{I_n \times I_n}$ and $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, let $\mathbf{x}_n \in \mathbb{R}^{I_n}$ be nonzero and $\mathbf{x}_n^\top \mathbf{B}_n \mathbf{x}_n = 1$ for all n. If the $(N+1)$-tuple $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is a solution of the nonlinear equations [30, Definition 2.7]:

$$F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)_{-n} = \sigma \mathbf{B}_n \mathbf{x}_n, \tag{6.4.1}$$

then $\sigma \in \mathbb{R}$ and the unit vectors $\mathbf{x}_n$ are called the *restricted singular value* of $\mathcal{A}$ and the corresponding *mode-n restricted singular vector* for all n.

When $N = 2$, the restricted singular values of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2}$ are reduced to the $(\mathbf{B}_1, \mathbf{B}_2)$-singular values of the matrix $\mathcal{A}$ [31, Definition 3], where $\mathbf{B}_1 \in \mathbb{R}^{I_1 \times I_1}$ and $\mathbf{B}_2 \in \mathbb{R}^{I_2 \times I_2}$ are symmetric positive definite matrices.

Luo et al. [32] develop the tensor canonical correlation analysis (TCCA), which straightforwardly yet naturally generalizes the canonical correlation analysis (CCA), to analyze the data of an arbitrary number of views through the covariance tensor. For N given views $\{\mathbf{X}_n\}_{n=1}^N$ of M instances, and $\mathbf{X}_n = [\mathbf{x}_{n1}, \mathbf{x}_{n2}, \ldots, \mathbf{x}_{nM}] \in \mathbb{R}^{I_n \times M}$, the main problem of the TCCA is equivalent to solving the following constraint optimization problem:

$$\max C \times_1 \mathbf{h}_1^\top \times_2 \mathbf{h}_2^\top \cdots \times_N \mathbf{h}_N^\top \quad \text{s.t.} \quad \mathbf{h}_n^\top \mathbf{C}_{nn} \mathbf{h}_n = 1, \text{ for all } n, \tag{6.4.2}$$

where

$$C = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_{1m} \otimes \mathbf{x}_{2m} \otimes \cdots \otimes \mathbf{x}_{Nm}, \quad \mathbf{C}_{nn} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_{nm} \mathbf{x}_{nm}^\top.$$

Since the matrices $\mathbf{C}_{nn}$ may be symmetric positive semi-definite, we usually add a regularization term in the constraints of the optimization problem (6.4.2) to control the model complexity. Thus, the constraints of the optimization problem (6.4.2) become

$$\mathbf{h}_n^\top (\mathbf{C}_{nn} + \varepsilon \mathbf{I}_{I_n}) \mathbf{h}_n = 1, \text{ for all } n,$$

where ε is a nonnegative tradeoff parameter to ensure that $\mathbf{C}_{nn} + \varepsilon \mathbf{I}_{I_n}$ is symmetric positive definite.

The optimization problem (6.4.2) is a special case of

$$\max \mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top \quad \text{s.t.} \quad \mathbf{x}_n^\top \mathbf{B}_n \mathbf{x}_n = 1, \text{ for all } n, \tag{6.4.3}$$

where $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\mathbf{B}_n \in \mathbb{R}^{I_n \times I_n}$ is symmetric positive definite.

The first-order necessary condition for (6.4.3) is the system of nonlinear equations (6.4.1) with $\mathbf{x}_n^\top \mathbf{B}_n \mathbf{x}_n = 1$. Since our methods for solving (6.4.3) is based on (6.4.1), the approximate $\widetilde{\mathbf{x}}_n$ may satisfy $\mathcal{A} \times_1 \widetilde{\mathbf{x}}_1^\top \times_2 \widetilde{\mathbf{x}}_2^\top \cdots \times_N \widetilde{\mathbf{x}}_N^\top < 0$. However, we can change the sign of $\widetilde{\mathbf{x}}_n$ to ensure that $\mathcal{A} \times_1 \mathbf{y}_1^\top \times_2 \mathbf{y}_2^\top \cdots \times_N \mathbf{y}_N^\top > 0$, where $\mathbf{y}_n = \delta(\widetilde{\mathbf{x}}_n)\widetilde{\mathbf{x}}_n$ with $\prod_{n=1}^N \delta(\widetilde{\mathbf{x}}_n) = -1$.

We next design some strategies to solve (6.4.3), solving (6.1.4) numerically.

We use the matrix factorizations of $\mathbf{B}_n$ to transfer (6.4.3) to (6.1.4). Consider the Cholesky factorization $\mathbf{B}_n = \mathbf{G}_n \mathbf{G}_n^\top$, where $\mathbf{G}_n \in \mathbb{R}^{I_n \times I_n}$ is lower triangular with positive diagonal entries. Let $\mathbf{y}_n = \mathbf{G}_n^\top \mathbf{x}_n$, then the optimization problem (6.4.3) is simplified as

$$\max \widetilde{\mathcal{A}} \times_1 \mathbf{y}_1^\top \times_2 \mathbf{y}_2^\top \cdots \times_N \mathbf{y}_N^\top \quad \text{s.t.} \quad \mathbf{y}_n^\top \mathbf{y}_n = 1, \text{ for all } n, \tag{6.4.4}$$

where $\widetilde{\mathcal{A}} = \mathcal{A} \times_1 \mathbf{G}_1^{-1} \times_2 \mathbf{G}_2^{-1} \cdots \times_N \mathbf{G}_N^{-1}$.

We can use HOMP [15, 16] or the neural network in (6.2.1) to solve the optimization problem. Luo et al. [32] also apply the method for (6.4.2).

Next, we will generalize the neural network in (6.2.1) to directly solve (6.4.3). For the nonzero restricted singular values of $\mathcal{A}$, the dynamics of the neural network model is

$$\mathbf{B}_n \frac{d\mathbf{x}_n(t)}{dt} = \frac{F\big(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\big)_{-n}}{F\big(\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_N(t)\big)} - \mathbf{B}_n \mathbf{x}_n(t), \tag{6.4.5}$$

for $t \geq 0$, where $\mathbf{x}_n(t) \in \mathbb{R}^{I_n}$ represents the state.

Note that (6.4.3) can be rewritten as

$$\max \frac{\mathcal{A} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top}{\sqrt{\mathbf{x}_1^\top \mathbf{B}_1 \mathbf{x}_1} \sqrt{\mathbf{x}_2^\top \mathbf{B}_2 \mathbf{x}_2} \ldots \sqrt{\mathbf{x}_N^\top \mathbf{B}_N \mathbf{x}_N}} \quad \text{s.t.} \quad \mathbf{x}_n^\top \mathbf{x}_n = 1, \text{ for all } n. \tag{6.4.6}$$

Some manipulation shows that the optimization problems (6.4.3) and (6.4.6) are equivalent. We list the results of the neural network in (6.4.5), such as the locally asymptotic stability of its solutions. Since the proof is similar to that of (6.2.1), we omit it.

Theorem 6.4.1 *Given the neural network in* (6.4.5) *and the arbitrary initial value* $\mathbf{x}_n(0)$ *for all* n, $\mathbf{x}_n(t)^\top \mathbf{B}_n \mathbf{x}_n(t)$ *converges exponentially to* 1 *as* $t \to +\infty$, *and the convergence is independent of* $\mathcal{A}$.

For an arbitrary nonzero initial value $\mathbf{x}_n(0) \in \mathbb{R}^{I_n}$, a discrete-time iterative algorithm corresponding to the neural network in (6.4.5) is

$$\mathbf{B}_n(\mathbf{x}_n(k+1)-\mathbf{x}_n(k)) = \alpha\left(\frac{F\big(\mathbf{x}_1(k),\mathbf{x}_2(k),\ldots,\mathbf{x}_N(k)\big)_{-n}}{F\big(\mathbf{x}_1(k),\mathbf{x}_2(k),\ldots,\mathbf{x}_N(k)\big)} - \mathbf{x}_n(k)\right), \tag{6.4.7}$$

for all n, where $\alpha > 0$ is the *learning rate*.

According to Theorem 6.4.1, the value of α can be selected so that $0 < \alpha < 0.5$. Since $\mathbf{B}_n$ is symmetric positive definite, then for $\mathbf{b} \in \mathbb{R}^{I_n}$, we can use the preconditioning conjugate gradient method [33] for solving $\mathbf{B}_n\mathbf{x} = \mathbf{b}$.

From Theorem 6.4.1, we deduce the following results.

Corollary 6.4.1 *For bounded initial values $\mathbf{x}_n(0)$ for all n, the state vectors of the neural network in (6.4.5) are bounded.*

Using $G_n(\mathbf{x}_1(t),\mathbf{x}_2(t),\ldots,\mathbf{x}_N(t))$ to denote the right-hand side of the neural network in (6.4.5), we have the following lemma.

Lemma 6.4.1 *Suppose that there exists N unit vectors $\mathbf{x}_n^* \in \mathbb{R}^{I_n}$ such that $G_n(\mathbf{x}_1^*,\mathbf{x}_2^*,\ldots,\mathbf{x}_N^*) = \mathbf{0}_I$ for all n. For the neural network in (6.4.5), let $\mathbf{x}_n(0) \in \mathbb{R}^{I_n}$ with $\mathbf{x}_n(0)^\top\mathbf{B}_n\mathbf{x}_n(0) = 1$. If $\mathbf{x}_n(t) \to \mathbf{x}_n^*$ as $t \to +\infty$, then $\sigma^* = F(\mathbf{x}_1^*,\mathbf{x}_2^*,\ldots,\mathbf{x}_N^*)$ and $\mathbf{x}_n^*$ are the restricted singular value and the corresponding mode-n restricted singular vectors of $\mathcal{A}$, respectively.*

Theorem 6.4.2 *Assume that $\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_N\}$ is a local maximizer of the optimization problem (6.4.3) with $\mathbf{u}_n \in \mathbb{R}^{I_n}$. If $\mathbf{x}_n(0) \in \widetilde{\mathbb{B}}(\delta(\mathbf{u})\mathbf{u};\epsilon)$ satisfies $\mathbf{x}_n(0)^\top\mathbf{B}_n\mathbf{x}_n(0) = 1$, with $\delta(\mathbf{u})\mathbf{u} = (\delta(\mathbf{u}_1)\mathbf{u}_1,\delta(\mathbf{u}_2)\mathbf{u},\ldots,\delta(\mathbf{u}_N)\mathbf{u}_N)$, then the solution of (6.4.5) is locally asymptotically stable in the sense of Lyapunov stability theory at $\{\delta(\mathbf{u}_1)\mathbf{u}_1,\delta(\mathbf{u}_2)\mathbf{u},\ldots,\delta(\mathbf{u}_N)\mathbf{u}_N\}$, where $\prod_{n=1}^N \delta(\mathbf{u}_n) = 1$ with $F(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_N) > 0$; or $\prod_{n=1}^N \delta(\mathbf{u}_n) = -1$ with $F(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_N) < 0$.*

It should be noted that the set $\widetilde{\mathbb{B}}(\delta(\mathbf{u})\mathbf{u};\epsilon)$ in the above theorem is similar to the set $\mathbb{B}(\delta(\mathbf{u})\mathbf{u};\epsilon)$ in (6.3.1).

6.5 Neural Networks for Generalized Tensor Eigenvalues

In this section, we assume that N is even and all tensors are real symmetric tensors. A tensor $\mathcal{C}$ is *positive definite* [21], if $\mathcal{C}\mathbf{x}^N > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^I$. A *symmetric-definite tensor pair* $\{\mathcal{A},\mathcal{B}\}$ is defined as: $\mathcal{A}$ and $\mathcal{B}$ are symmetric and $\mathcal{B}$ is positive definite (see [33, Section 8.7.1] for the matrix cases). Chang et al. [34] prove the existence of the H-eigenvalues for symmetric-definite tensor pairs. Assume that both $\mathcal{A}\mathbf{x}^{N-1}$ and $\mathcal{B}\mathbf{x}^{N-1}$ are not identically zero for all nonzero $\mathbf{x} \in \mathbb{R}^I$. The pair $(\lambda;\mathbf{x})$ is a generalized eigenpair (with generalized eigenvalue λ and the corresponding

generalized eigenvector $\mathbf{x}$) of the tensor pair $\{\mathcal{A}, \mathcal{B}\}$, if

$$(\mathcal{A} - \lambda\mathcal{B})\mathbf{x}^{N-1} = \mathbf{0}_I. \tag{6.5.1}$$

If $\mathcal{B}$ is the identity tensor, then the pair $(\lambda; \mathbf{x})$ is called as an H-*eigenpair* of $\mathcal{A}$ [13, 21].

There is a clear interpretation of the maximal and minimal H-eigenvalues of a symmetric-definite tensor pair. Ding and Wei [35] derive the following theorem as a generalization of the Rayleigh-Ritz theorem for a Hermitian matrix [33, Section 8.1.1]. If $\mathcal{B}$ is the identity tensor, then this theorem simplifies into the single tensor case [21, Theorem 5].

Theorem 6.5.1 ([35, Theorem 3.5]) *Let $\{\mathcal{A}, \mathcal{B}\}$ be a symmetric-definite tensor pair. Denote the maximal generalized eigenvalue of $\{\mathcal{A}, \mathcal{B}\}$ as*

$$\lambda_{\max} = \max\left\{\lambda \in \mathbb{R} : \mathcal{A}\mathbf{x}^{N-1} = \lambda\mathcal{B}\mathbf{x}^{N-1}, \mathbf{x} \in \mathbb{R}^I \backslash \{\mathbf{0}_I\}\right\},$$

and the minimal generalized eigenvalue $\lambda_{\min}$ of $\{\mathcal{A}, \mathcal{B}\}$ similarly. Then

$$\lambda_{\max} = \max_{\mathbf{x}\in\mathbb{R}^I\backslash\{\mathbf{0}_I\}} \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N} = \max_{\substack{\mathbf{x}\in\mathbb{R}^I \\ \|\mathbf{x}\|_2=1}} \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}, \qquad \lambda_{\min} = \min_{\mathbf{x}\in\mathbb{R}^I\backslash\{\mathbf{0}_I\}} \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N} = \min_{\substack{\mathbf{x}\in\mathbb{R}^I \\ \|\mathbf{x}\|_2=1}} \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}.$$

For a given symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, we can formulate the following nonlinear programming problem

$$\max_{\mathbf{x}\in\mathbb{R}^I} \quad \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\|\mathbf{x}\|_2^N \quad \text{s.t.} \quad \|\mathbf{x}\|_2 = 1. \tag{6.5.2}$$

A pair $(\lambda; \mathbf{x})$ is called a local maximal (resp. minimal) generalized H-eigenpair of the symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, if $\mathbf{x}$ is the local maximizer (resp. minimizer) of (6.5.2). We compute the generalized H-eigenvectors of a symmetric-definite tensor pair by (6.5.2), rather than the local maximal (or the local minimal) generalized H-eigenvalues, because of its close relationship with the generalized tensor eigenvalue problem (6.5.1).

Theorem 6.5.2 ([36, Theorem 3.2]) *Any pair $(\lambda; \mathbf{x})$ is a solution to (6.5.1) if and only if the scaled version with $\|\mathbf{x}\|_2 = 1$ is a KKT point of (6.5.2), where λ is the Lagrange multiplier.*

Analogous to the neural network in (6.2.5), we design a neural network for computing the local extremal generalized H-eigenvalue of the symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$. The dynamics of the new neural network model is

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{B}\mathbf{x}(t)^N\mathcal{A}\mathbf{x}(t)^{N-1} - \mathcal{A}\mathbf{x}(t)^N\mathcal{B}\mathbf{x}(t)^{N-1}, \tag{6.5.3}$$

for all $t \geq 0$, where $\mathbf{x}(t) \in \mathbb{R}^I$ represents the state. The solution of the neural network in (6.5.3) has the following important property.

Theorem 6.5.3 *If $\mathbf{x}(t)$ is a solution of the neural network in (6.5.3) for $t \geq 0$, then $\|\mathbf{x}(t)\|_2^2$ is invariant for $t \geq 0$, i.e., $\|\mathbf{x}(t)\|_2^2 = \|\mathbf{x}(0)\|_2^2$.*

Proof For $t \geq 0$, let $\mathbf{x}(t)$ be a solution of (6.5.3) starting from any nonzero initial value $\mathbf{x}(0)$. It follows that

$$\frac{d\|\mathbf{x}(t)\|_2^2}{dt} = 2\mathbf{x}(t)^\top \frac{d\mathbf{x}(t)}{dt} = 2\mathbf{x}^\top \left[\mathcal{B}\mathbf{x}(t)^N \mathcal{A}\mathbf{x}(t)^{N-1} - \mathcal{A}\mathbf{x}(t)^N \mathcal{B}\mathbf{x}(t)^{N-1}\right] = 0.$$

Hence $\|\mathbf{x}(t)\|_2^2$ is invariant for $t \geq 0$. □

For an arbitrary nonzero $\mathbf{x}(0) \in \mathbb{R}^I$, a discrete-time iterative algorithm corresponding to (6.5.3) is

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \alpha \left(\mathcal{B}\mathbf{x}(k)^N \mathcal{A}\mathbf{x}(k)^{N-1} - \mathcal{A}\mathbf{x}(k)^N \mathcal{B}\mathbf{x}(k)^{N-1}\right), \tag{6.5.4}$$

where $\alpha > 0$ is the *learning rate*.

With $G(\mathbf{x}(t))$, denoting the right-hand side of (6.5.3), we have the following lemma.

Lemma 6.5.1 *Suppose that there exists a unit vector $\mathbf{x}^* \in \mathbb{R}^I$ such that $G(\mathbf{x}^*) = \mathbf{0}_I$. Consider the neural network in (6.5.3)with $\mathbf{x}(0) \in \mathbb{R}^I$ and $\|\mathbf{x}(0)\|_2 = 1$. If $\mathbf{x}(t) \to \mathbf{x}^*$ as $t \to +\infty$, then $\lambda^* = \mathcal{A}\mathbf{x}_*^N / \mathcal{B}\mathbf{x}_*^N := g(\mathbf{x}_*)$ and $\mathbf{x}^*$ are the generalized eigenvalue and the corresponding generalized eigenvector of the symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, respectively.*

Given a symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, denote the set of all generalized eigenvectors corresponding to the local maximal generalized eigenvalues $\lambda_{\max}^{\text{Local}}$ and the local minimal generalized eigenvalues $\lambda_{\min}^{\text{Local}}$ by $\mathbb{S}_{\max}$ and $\mathbb{S}_{\min}$, respectively. It is obvious that the maximal (resp. minimal) generalized eigenvalue is one of the local maximal (resp. minimal) generalized eigenvalues. A neighbourhood of $\mathbf{u} \in \mathbb{S}_{\max}$ (resp. $\mathbf{u} \in \mathbb{S}_{\min}$) is defined by

$$\mathbb{B}(\mathbf{u}; \epsilon) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{u}\|_2 \leq \epsilon\} \tag{6.5.5}$$

where $0 < \epsilon \leq \epsilon_0 := \min \|\mathbf{u} - \mathbf{v}\|_2$ and $\mathbf{v} \in \mathbb{S}_{\max}$ (resp. $\mathbf{u} \in \mathbb{S}_{\min}$) with $\mathbf{u} \neq \mathbf{v}$. Since $\mathbf{u}$ is nonzero, all vectors in $\mathbb{B}(\mathbf{u}; \epsilon)$ are nonzero for a small enough ϵ.

We show that the (6.5.3) is locally asymptotically stable in the sense of Lyapunov stability theory at a generalized eigenvector $\mathbf{u}$, in $\mathbb{S}_{\max}$ or $\mathbb{S}_{\min}$.

Theorem 6.5.4 *Suppose that the unit vector $\mathbf{u} \in \mathbb{R}^I$ is a generalized eigenvector of a symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$ corresponding to a local maximal (resp. minimal) generalized eigenvalue. If $\mathbf{x}(0) \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$ defined by (6.5.5), then the solution of (6.5.3) is locally asymptotically stable in the sense of Lyapunov stability theory at $\delta(\mathbf{u})\mathbf{u}$.*

Proof If $(\lambda; \mathbf{x})$ is a generalized eigenpair of the symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, then so is $(\lambda; -\mathbf{x})$. We show that the neural network in (6.5.3) is locally asymptotically stable in the sense of Lyapunov stability theory at a generalized eigenvector $\pm\mathbf{u}$, with $\mathbf{u} \in \mathbb{S}_{\max}$ (resp. $\mathbf{u} \in \mathbb{S}_{\min}$).

Let $\mathbf{u} \in \mathbb{S}_{\max}$ and $g(\mathbf{u}) > 0$. It is easy to see $g(-\mathbf{u}) > 0$ because of the even N. Then, we have $0 < g(\mathbf{x}) \leq g(\delta(\mathbf{u})\mathbf{u})$ for all $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$.

For (6.5.3), we define the Lyapunov function

$$V(\mathbf{x}) = \lambda_{\max}^{\text{Local}} - \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}, \qquad \lambda_{\max}^{\text{Local}} = g(\delta(\mathbf{u})\mathbf{u}) > 0.$$

We have $V(\mathbf{x}) > 0$ for $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$ but not parallel to $\delta(\mathbf{u})\mathbf{u}$. Computing the partial derivative of $V(\mathbf{x})$ with respect to $\mathbf{x}$, we have

$$\frac{dV}{dt} = -\left(\frac{d\mathbf{x}}{dt}\right)^\top \left[\nabla_{\mathbf{x}} \left(\frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\right)\right].$$

We also have

$$\nabla_{\mathbf{x}} \left(\frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\right) = \frac{N\mathcal{A}\mathbf{x}^{N-1}}{\mathcal{B}\mathbf{x}^N} - \frac{N\mathcal{A}\mathbf{x}^N \mathcal{B}\mathbf{x}^{N-1}}{(\mathcal{B}\mathbf{x}^N)^2}.$$

Hence

$$\begin{aligned}
\left(\frac{d\mathbf{x}}{dt}\right)^\top \left[\nabla_{\mathbf{x}} \left(\frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\right)\right] &= (\mathcal{B}\mathbf{x}^N \mathcal{A}\mathbf{x}^{N-1} - \mathcal{A}\mathbf{x}^N \mathcal{B}\mathbf{x}^{N-1})^\top \left(\frac{m\mathcal{A}\mathbf{x}^{N-1}}{\mathcal{B}\mathbf{x}^N} - \frac{N\mathcal{A}\mathbf{x}^N \mathcal{B}\mathbf{x}^{N-1}}{(\mathcal{B}\mathbf{x}^N)^2}\right) \\
&= N\left((\mathcal{A}\mathbf{x}^{N-1})^\top \mathcal{A}\mathbf{x}^{N-1} - 2\frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}(\mathcal{A}\mathbf{x}^{N-1})^\top \mathcal{B}\mathbf{x}^{N-1}\right. \\
&\quad \left. + \left(\frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\right)^2 (\mathcal{B}\mathbf{x}^{N-1})^\top \mathcal{B}\mathbf{x}^{N-1}\right) \\
&\geq N\left(\mathcal{A}\mathbf{x}^{N-1} - \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\mathcal{B}\mathbf{x}^{N-1}\right)^\top \left(\mathcal{A}\mathbf{x}^{N-1} - \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}\mathcal{B}\mathbf{x}^{N-1}\right) \\
&\geq 0,
\end{aligned}$$

and the equality holds if and only if $\mathbf{x}$ is parallel to $\mathbf{u}$.

Thus we have

$$\frac{dV}{dt} < 0$$

for $t \geq 0$, where $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$ is not parallel to $\mathbf{u}$.

Assume that $\mathbf{u} \in \mathbb{S}_{\max}$ and $g(\mathbf{u}) < 0$. It is easy to see that $g(-\mathbf{u}) < 0$ because of the even N. Then, we have $g(\mathbf{x}) \leq g(\delta(\mathbf{u})\mathbf{u}) < 0$ for all $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$. Define the Lyapunov function

$$V(\mathbf{x}) = \lambda_{\max}^{\text{Local}} - \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}, \quad \text{with} \quad \lambda_{\max}^{\text{Local}} = g(\delta(\mathbf{u})\mathbf{u}) < 0.$$

Since $g(\mathbf{x}) \leq g(\delta(\mathbf{u})\mathbf{u}) < 0$ for $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$, then $0 < -g(\delta(\mathbf{u})\mathbf{u}) \leq -g(\mathbf{x})$ for $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$. Hence, we have $V(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{B}(\delta(\mathbf{u})\mathbf{u}; \epsilon)$ but not parallel to $\mathbf{u}$. The proof is similar to that of $\mathbf{u} \in \mathbb{S}_{\max}$ and $g(\mathbf{u}) > 0$.

Since $\lambda_{\min}^{\text{Local}}$ is a local minimal generalized eigenvalue of the symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, then $-\lambda_{\min}^{\text{Local}}$ is the local maximal generalized eigenvalue of the symmetric-definite tensor pair $\{-\mathcal{A}, \mathcal{B}\}$ with its generalized eigenvector $\mathbf{u} \in \mathbb{S}_{\min}$.

For the symmetric-definite tensor pair $\{-\mathcal{A}, \mathcal{B}\}$, the neural network in (6.5.3) can be represented by

$$\frac{d\mathbf{x}(t)}{dt} = -\mathcal{B}\mathbf{x}(t)^N \mathcal{A}\mathbf{x}(t)^{N-1} + \mathcal{A}\mathbf{x}(t)^N \mathcal{B}\mathbf{x}(t)^{N-1}, \tag{6.5.6}$$

for all $t \geq 0$. The associated Lyapunov function is defined as

$$V(\mathbf{x}) = -\lambda_{\min}^{\text{Local}} - \frac{(-\mathcal{A})\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N} = -\lambda_{\min}^{\text{Local}} + \frac{\mathcal{A}\mathbf{x}^N}{\mathcal{B}\mathbf{x}^N}.$$

The proof is similar for the local maximal generalized eigenvalues of the symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$. The theorem is proved. □

6.6 Numerical Examples

In this section, we present numerical examples of neural networks for finding locally optimal rank-one approximations. We terminate the corresponding iterations using the criterion

$$\left|F\big(\mathbf{x}_1(k+1), \mathbf{x}_2(k+1), \ldots, \mathbf{x}_N(k+1)\big) - F\big(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)\big)\right| < 10^{-8},$$

for *nonsymmetric* tensors and

$$\left|f\big(\mathbf{x}(k+1)\big) - f\big(\mathbf{x}(k)\big)\right| < 10^{-8},$$

for *symmetric* tensors, or when the algorithm runs over 1000 iterations.

Note that $\mathbf{x}_n(k) \in \mathbb{R}^{I_n}$ and $\mathbf{x}(k) \in \mathbb{R}^{I}$ are computed from the first-order difference approximation (6.2.2) and HOPM, and the first-order difference approximation (6.2.4) and SHOPM, respectively.

For a local maximizer $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ of the optimization problem (6.1.4), we have

$$\mathbb{S}_\tau = \left\{ (\delta(\tau)\tau) \cdot \left(\delta(\mathbf{v}_1)\mathbf{v}_1 \otimes \delta(\mathbf{v}_2)\mathbf{v}_2 \otimes \cdots \otimes \delta(\mathbf{v}_N)\mathbf{v}_N\right) : \delta(\tau) \prod_{n=1}^{N} \delta(\mathbf{v}_n) = 1 \right\}$$

being the set of all locally optimal rank-one approximations corresponding to the locally optimal rank-one approximation $\tau \cdot (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_N)$ of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, where $\tau = F(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$. It is obvious that the best rank-one approximation of $\mathcal{A}$ belongs to $\mathbb{S}_\tau$. The related results are shown in tables and the running time is measured by the Matlab "tic-toc" pairs in seconds. We only list one locally optimal rank-one approximation of $\mathcal{A}$ for each $\mathbb{S}_\tau$ in all examples. We follow the same practice for symmetric tensors.

In Sect. 6.6.3, we present three numerical methods for computing the restricted singular values and associated restricted singular vectors.

Finally, we present numerical examples to illustrate that we can utilize the neural network in (6.5.3) to compute the generalized eigenvector of a symmetric-definite tensor pair (cf. Sect. 6.5), corresponding to the maximal or the minimal generalized eigenvalue. We terminate the iterations when

$$\left\| \mathcal{A}\mathbf{x}(k)^{N-1} - \lambda(k)\mathcal{B}\mathbf{x}(k)^{N-1} \right\|_2 \leq 10^{-8},$$

or the number of the iterations exceeds 5000, with $\mathbf{x}(k)$ from the first-order difference approximation (6.5.4) and $\lambda(k) = \frac{\mathcal{A}\mathbf{x}(k)^N}{\mathcal{B}\mathbf{x}(k)^N}$.

We only record the Matlab "tic-toc" estimates of the execution times by the first-order difference approximations (6.2.2), (6.2.4) and (6.5.4) for the local maximizers of the optimization problems (6.1.4), (6.1.7) and (6.5.2), respectively.

6.6.1 Nonsymmetric Tensor Examples

We present numerical results for nonsymmetric tensors. The locally optimal rank-one approximation *ratio* of a real tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is investigated by Qi [19, 37]:

$$\rho(\mathcal{A}) := \max \frac{|\langle \mathcal{A}, \mathcal{X} \rangle|}{\|\mathcal{A}\|_F \|\mathcal{X}\|_F},$$

where $\mathcal{X} = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_N$ with nonzero $\mathbf{x}_n \in \mathbb{R}^{I_n}$.

Clearly, if $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N)$ is a locally optimal rank-one approximation of $\mathcal{A}$, with each $\|\mathbf{u}_n\|_2 = 1$ and $\sigma_* = F(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$, then $\rho(\mathcal{A}) = |\sigma_*|/\|\mathcal{A}\|_F$.

Suppose that $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N)$ is returned by the first-order difference approximation (6.2.2), then we measure the quality of the rank-one approximation by the residual:

$$\text{RES}_{\text{NET}}(\mathcal{A}) = \|\mathcal{A} - \sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N)\|_F.$$

There are well known connections exist between the singular value decomposition [33, Section 2.4] of a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ and the Schur decomposition [33, Section 7.4.1] of its symmetric embedding

$$\mathbf{sym}(\mathbf{A}) = \begin{pmatrix} \mathbf{0}_{I \times I} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0}_{J \times J} \end{pmatrix}.$$

Ragnarsson and Van Loan [38] develop similar connections for tensors by building on Lim's variational approach. Through the symmetric embedding $\mathbf{sym}(\mathcal{A})$, if $(\sigma; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ is a solution of (6.1.1) with $\sigma \neq 0$, then $(\sigma_\alpha; \mathbf{x}_\alpha)$ is an Z-eigenpair of $\mathbf{sym}(\mathcal{A})$, where

$$\begin{cases} \sigma_\alpha &= \alpha_1 \alpha_2 \ldots \alpha_N \frac{N!}{\sqrt{N^N}} \sigma, \\ \mathbf{x}_\alpha &= \frac{1}{\sqrt{N}} \left(\alpha_1 \mathbf{x}_1^\top, \alpha_2 \mathbf{x}_2^\top, \ldots, \alpha_N \mathbf{x}_N^\top\right)^\top, \\ \alpha &= [\alpha_1, \alpha_2, \ldots, \alpha_N] = [1, \underbrace{\pm 1, \ldots, \pm 1}_{N-1}]. \end{cases}$$

If $(\sigma; \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ can be computed by the neural network in (6.2.1), then we can define

$$\mathbf{C}(\sigma_\alpha, \mathbf{u}_\alpha) = \mathbf{U}^\top \left[(N-1)\mathbf{sym}(\mathcal{A})\mathbf{u}_\alpha^{N-1} - \sigma \mathbf{I}_I \right] \mathbf{U} \in \mathbb{R}^{(I-1)\times(I-1)},$$

where the columns of $\mathbf{U} \in \mathbb{R}^{I \times (I-1)}$ form an orthonormal basis for $\mathbf{u}_\alpha^\perp$ with $I = I_1 + I_2 + \ldots + I_N$.

Suppose that $(\sigma; \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$ is a singular pair of the tensor $\mathcal{A}$, then

$$\begin{cases} \mathbf{C}(\sigma_\alpha; \mathbf{u}_\alpha) \text{ positive definite} & \Rightarrow \mathbf{u}_n \text{ is a local maximum of (6.1.4)}, \\ \mathbf{C}(\sigma_\alpha; \mathbf{u}_\alpha) \text{ negative definite} & \Rightarrow \mathbf{u}_n \text{ is a local maximum of (6.1.4)}, \\ \mathbf{C}(\sigma_\alpha; \mathbf{u}_\alpha) \text{ indefinite} & \Rightarrow \mathbf{u}_n \text{ is a saddle point of (6.1.4)}. \end{cases}$$

Example 6.6.1 ([37, Example 1]) Consider the positive tensor $\mathcal{A} \in \mathbb{R}^{3\times 3\times 3}$ with

$$\begin{cases} a_{111} = 0.4333,\ a_{121} = 0.4278,\ a_{131} = 0.4140,\ a_{211} = 0.8154, \\ a_{221} = 0.0199,\ a_{231} = 0.5598,\ a_{311} = 0.0643,\ a_{321} = 0.3815, \\ a_{331} = 0.8834,\ a_{112} = 0.4866,\ a_{122} = 0.8087,\ a_{132} = 0.2073, \\ a_{212} = 0.7641,\ a_{222} = 0.9924,\ a_{232} = 0.8752,\ a_{312} = 0.6708, \\ a_{322} = 0.8296,\ a_{332} = 0.1325,\ a_{113} = 0.3871,\ a_{123} = 0.0769, \\ a_{133} = 0.3151,\ a_{213} = 0.1355,\ a_{223} = 0.7727,\ a_{233} = 0.4089, \\ a_{313} = 0.9715,\ a_{323} = 0.7726,\ a_{333} = 0.5526. \end{cases}$$

If we employ the first-order difference approximation (6.2.2), then we get a locally optimal rank-one tensor $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \mathbf{u}_3)$ with

$$\begin{cases} \sigma_* = -2.8167,\ \rho(\mathcal{A}) = 0.9017,\ \mathrm{RES}_{\mathrm{NET}} = 1.3510. \\ \mathbf{u}_1 = (0.4281, 0.6557, 0.6219)^\top, \\ \mathbf{u}_2 = (-0.5706, -0.6467, -0.5062)^\top, \\ \mathbf{u}_3 = (0.4501, 0.7094, 0.5423)^\top. \end{cases}$$

It takes 0.0625 s. Figure 6.1 displays the convergence of the global maximizer estimation for the optimization problem (6.1.4).

Comparing to [8, Example 3.12], we see that $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \mathbf{u}_3)$ is the best rank-one approximation.

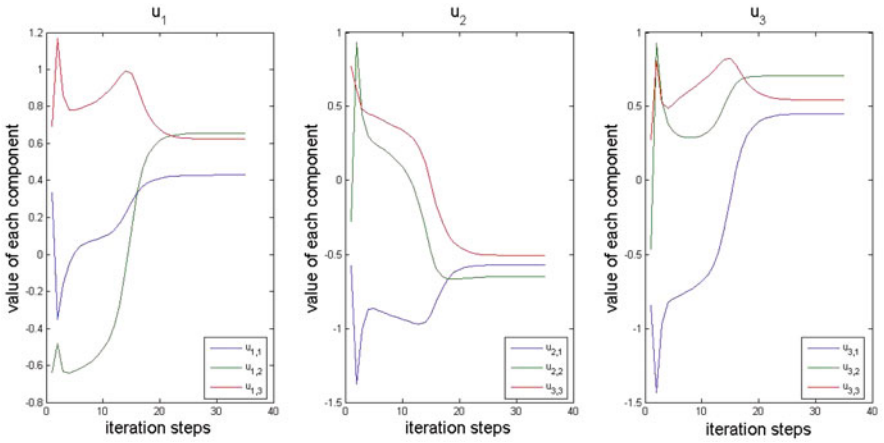

Fig. 6.1 Estimation of a global maximizer of (6.1.4) with Example 6.6.1

Example 6.6.2 ([8, Example 3.14]) Consider the tensor $\mathcal{A} \in \mathbb{R}^{I\times I\times I}$ given as

$$a_{i_1i_2i_3} = \cos(i_1 + 2i_2 + 3i_3).$$

For $I = 5$, by applying the first-order difference approximation (6.2.2), we obtain a locally optimal rank-one tensor $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \mathbf{u}_3)$ with

$$\begin{cases} \sigma_* = -6.0996, \quad \rho(\mathcal{A}) = 0.7728, \quad \mathrm{RES}_{\mathrm{NET}} = 5.0093, \\ \mathbf{u}_1 = (\ 0.4295,\ \ 0.5611,\ \ 0.1768, -0.3700, -0.5767)^\top, \\ \mathbf{u}_2 = (\ 0.6210, -0.2957, -0.3749,\ \ 0.6077, -0.1309)^\top, \\ \mathbf{u}_3 = (-0.4528,\ \ 0.4590, -0.4561,\ \ 0.4441, -0.4231)^\top. \end{cases}$$

It takes 0.2188 s. Figure 6.2 presents the convergence of the global maximizer estimation for the optimization problem (6.1.4).

Comparing to [8, Example 3.14], we know that $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \mathbf{u}_3)$ is the best rank-one approximation.

For I from 5 to 100, we apply the first-order difference approximation (6.2.2) to find the best rank-one approximations. The computational results are listed in Table 6.1.

Example 6.6.3 ([8, Example 3.15]) Consider $\mathcal{A} \in RT_{4,I}$ given as

$$a_{i_1i_2i_3i_4} = \begin{cases} \displaystyle\sum_{j=1}^{4} \arcsin\left((-1)^{i_j}\frac{j}{i_j}\right), & \text{if all } i_j \geq j, \\ 0, & \text{otherwise.} \end{cases}$$

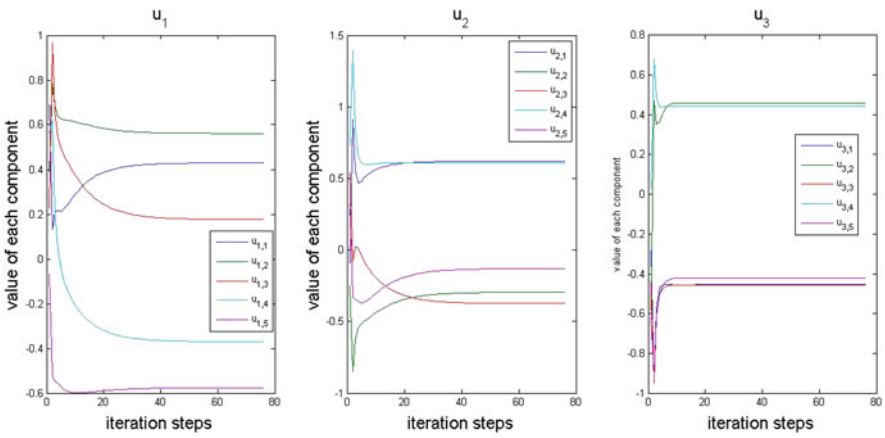

Fig. 6.2 Estimation of a global maximizer of (6.1.4) with Example 6.6.2

Table 6.1 Related results for Example 6.6.2

I	σ_*	Time	$\rho(\mathcal{A})$	$\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$
5	−6.0996	0.2969	0.7728	5.0093
10	−14.7902	1.9063	0.6629	16.7026
15	25.4829	0.5156	0.6203	32.2270
20	33.7020	3.8125	0.5329	53.5095
25	46.7997	7.6875	0.5295	74.9817
30	64.9106	3.0781	0.5587	96.3619
35	80.7697	7.0938	0.5517	122.1175
40	−95.0878	5.1719	0.5316	151.5162
95	−338.5683	65.3750	0.5171	560.4087
100	368.6851	67.1719	0.5214	603.3848

For $I = 5$, if we use the first-order difference approximation (6.2.2), then we obtain two locally optimal rank-one tensors $\sigma_*^i \cdot (\mathbf{u}_1^i \otimes \mathbf{u}_2^i \otimes \mathbf{u}_3^i \otimes \mathbf{u}_4^i)$ with

$$\begin{cases} \sigma_*^1 = -15.3155, \quad \rho(\mathcal{A}) = 0.7076, \quad \mathrm{RES}_{\mathrm{NET}} = 15.2957, \\ \mathbf{u}_1^1 = (-0.6711, -0.2776, -0.4398, -0.3285, -0.4138)^\top, \\ \mathbf{u}_2^1 = (-0.0000, -0.1709, -0.6708, -0.3985, -0.6017)^\top, \\ \mathbf{u}_3^1 = (-0.0000, -0.0000, -0.8048, -0.1804, -0.5654)^\top, \\ \mathbf{u}_4^1 = (0.0000, 0.0000, 0.0000, -0.0073, -1.0000)^\top, \end{cases}$$

and

$$\begin{cases} \sigma_*^2 = -13.4801, \quad \rho(\mathcal{A}) = 0.6228, \quad \mathrm{RES}_{\mathrm{NET}} = 16.9354, \\ \mathbf{u}_1^2 = (-0.2089, -0.5643, -0.4178, -0.5183, -0.4413)^\top, \\ \mathbf{u}_2^2 = (-0.0000, 0.7280, 0.2800, 0.5241, 0.3420)^\top, \\ \mathbf{u}_3^2 = (-0.0000, -0.0000, -0.2742, -0.8293, -0.4870)^\top, \\ \mathbf{u}_4^2 = (0.0000, -0.0000, -0.0000, -0.9777, 0.2102)^\top. \end{cases}$$

It takes 0.1719 s and 0.4063 s, respectively. We conclude that $\sigma_*^1 \cdot (\mathbf{u}_1^1 \otimes \mathbf{u}_2^1 \otimes \mathbf{u}_3^1 \otimes \mathbf{u}_4^1)$ is the best rank-one approximation and the results are consistent with those in [8, Example 3.15].

For I from 5 to 12, we apply the first-order difference approximation (6.2.2) to obtain the best rank-one approximation. The computational results are listed in Table 6.2.

Remark 6.6.1 For $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, suppose that $\tau_* \cdot (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_N)$ is the best rank-one approximation returned by the semi-definite relaxation methods [8], we measure the quality of the rank-one approximation by the residual:

$$\mathrm{RES}_{\mathrm{SPD}}(\mathcal{A}) = \|\mathcal{A} - \tau_* \cdot (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_N)\|_F.$$

Table 6.2 Related results for Example 6.6.3

I	σ_*	Time	$\rho(\mathcal{A})$	$\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$
5	−15.3155	0.1719	0.7076	15.2957
6	−25.3934	0.1406	0.7478	22.5473
7	33.0389	0.1719	0.6874	34.9113
8	−44.6644	0.1406	0.7052	44.9052
9	−54.3981	0.4063	0.6782	58.9426
10	67.3464	0.3125	0.6872	71.2040
11	−78.9268	0.4063	0.6731	86.7219
12	−93.0160	0.4688	0.6775	100.9830

Let $\sigma_* \cdot (\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_N)$ be the best rank-one approximation returned by the first-order difference approximation (6.2.2). When the iterations are terminated, $\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$ is not significantly bigger than $\mathrm{RES}_{\mathrm{SPD}}(\mathcal{A})$ and the CPU time for (6.2.2) is smaller than the CPU time for the semi-definite relaxation methods, as shown in [8].

Example 6.6.4 ([8, Example 3.19]) We compare the first-order difference approximation (6.2.2) with HOPM in [15, Algorithm 3.2], which is widely used for computing rank-one approximation for nonsymmetric tensors. Meanwhile, we can also use the semi-definite relaxation methods to check the convergence of HOPM. The tensor $\mathcal{A} \in T_{3,I}$ is given by

$$a_{i_1 i_2 i_3} = \tan\left(i_1 - \frac{i_2}{2} + \frac{i_3}{3}\right).$$

Suppose that $\lambda \cdot (\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_N)$ is a rank-one approximation returned by HOPM, then we measure its quality by the residual:

$$\mathrm{RES}_{\mathrm{HOPM}}(\mathcal{A}) = \|\mathcal{A} - \lambda \cdot (\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_N)\|_F.$$

Suppose $I = 20$. When we implement (6.2.2) and HOPM 100 times on the testing tensor $\mathcal{A}$, the absolute values of σ_* and λ are chosen from the set

$$\{382.6558, 385.8545, 386.0731, 399.3615, 400.6770, 508.8158, 508.8218\}.$$

The computational results are presented in Table 6.3.

According to Table 6.3, we can use (6.2.2) and HOPM to compute all locally optimal rank-one approximations of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. However, we do not know how to choose an optimal initial value to ensure that the sequence, which converges to the global minimizer.

Table 6.3 Related results for Example 6.6.4

Values	Occ.	Time	$\text{RES}_{\text{NET}}(\mathcal{A})$	Occ.	Time	$\text{RES}_{\text{HOPM}}(\mathcal{A})$
382.6558	2	69.3516	1.8119e+3	2	63.7344	1.8119e+3
385.8545	15	9.0479	1.8112e+3	18	21.0113	1.8112e+3
386.0731	9	9.3576	1.8112e+3	9	21.4833	1.8112e+3
399.3615	11	3.9105	1.8083e+3	15	8.4948	1.8083e+3
400.6770	16	4.0439	1.8080e+3	10	8.1656	1.8080e+3
508.8158	21	0.7649	1.7806e+3	24	1.6680	1.7806e+3
508.8218	26	0.8305	1.7806e+3	22	1.6747	1.7806e+3

6.6.2 *Symmetric Tensor Examples*

We report some numerical results for symmetric tensors here. Qi [37] defines the locally optimal symmetric rank-one approximation *ratio* of a symmetric tensor $\mathcal{A} \in RT_{N,I}$ as

$$\rho(\mathcal{A}) := \max \frac{|\langle \mathcal{A}, \mathcal{X} \rangle|}{\|\mathcal{A}\|_F \|\mathcal{X}\|_F},$$

where $\mathcal{X} = \mathbf{x}^{\otimes N}$ with a nonzero vector $\mathbf{x} \in \mathbb{R}^I$.

If $\sigma_* \cdot \mathbf{u}^{\otimes N}$ is a locally optimal symmetric rank-one approximation of $\mathcal{A}$, with $\|\mathbf{u}\|_2 = 1$ and $\sigma_* = f(\mathbf{u})$, then $\rho(\mathcal{A}) = |\sigma_*|/\|\mathcal{A}\|_F$. Estimates for $\rho(\mathcal{A})$ are given by Qi in [37]. Let $\sigma_* \cdot \mathbf{u}^{\otimes N}$ be the locally optimal symmetric rank-one approximation returned by the first-order difference approximation (6.2.4). We measure the quantity of the approximation by the residual:

$$\text{RES}_{\text{NET}}(\mathcal{A}) = \left\| \mathcal{A} - \sigma_* \cdot \mathbf{u}^{\otimes N} \right\|_F.$$

As discussed in [15, 22, 23], a Z-eigenpair may be associated with a local maximum, a local minimum, or a saddle point. For a given Z-eigenpair $(\sigma; \mathbf{x})$ normalized so that $\|\mathbf{x}\|_2 = 1$, we can categorize it by the projected Hessian of the Lagrangian, i.e.,

$$\mathbf{C}(\sigma; \mathbf{x}) = \mathbf{U}^\top \left[(N-1)\mathcal{A}\mathbf{x}^{N-1} - \sigma \mathbf{I}_I \right] \mathbf{U} \in \mathbb{R}^{(I-1)\times(I-1)},$$

where the columns of $\mathbf{U} \in \mathbb{R}^{I\times(I-1)}$ form an orthonormal basis for $\mathbf{x}^\perp$. We have three situations

$$\begin{cases} \mathbf{C}(\sigma; \mathbf{x}) \text{ positive definite} & \Rightarrow \mathbf{x} \text{ is a local maximum of (6.1.7)}, \\ \mathbf{C}(\sigma; \mathbf{x}) \text{ negative definite} & \Rightarrow \mathbf{x} \text{ is a local maximum of (6.1.7), or} \\ \mathbf{C}(\sigma; \mathbf{x}) \text{ indefinite} & \Rightarrow \mathbf{x} \text{ is a saddle point of (6.1.7).} \end{cases}$$

For the illustrative examples in this subsection, the solutions of the neural network in (6.2.3) are the local maxima of (6.1.7).

Example 6.6.5 ([23, Example 3.6], [19, Example 4.2]) Consider the symmetric tensor $\mathcal{A} \in RT_{3,3}$ with

$$a_{111} = -0.1281,\ a_{112} = 0.0516,\ a_{113} = -0.0954,\ a_{122} = -0.1958,\ a_{123} = -0.1790;$$
$$a_{133} = -0.2676,\ a_{222} = 0.3251,\ a_{223} = 0.2513,\ a_{233} = 0.1773,\ a_{333} = 0.0338.$$

Applying the first-order difference approximation (6.2.4), we obtain two locally optimal symmetric rank-one tensors $\sigma_*^i \cdot (\mathbf{u}^i)^{\otimes 3}$ $(i = 1, 2)$ with

$$\begin{cases} \sigma_*^1 = -0.8730, \quad \rho(\mathcal{A}) = 0.8890, \quad \mathrm{RES}_{\mathrm{NET}} = 0.4498, \\ \mathbf{u}^1 = (0.3922, -0.7249, -0.5664)^\top; \\ \sigma_*^2 = -0.4306, \quad \rho(\mathcal{A}) = 0.4385, \quad \mathrm{RES}_{\mathrm{NET}} = 0.8826, \\ \mathbf{u}^2 = (0.7187, \ 0.1245, \ 0.6840)^\top. \end{cases}$$

It takes 0.0313 s and 0.0469 s, respectively. We conclude that $\sigma_*^1 \cdot (\mathbf{u}^1)^{\otimes 3}$ is the best symmetric rank-one approximation, which is consistent with [8, Example 3.3].

Example 6.6.6 ([37, Example 2]) Consider the positive symmetric tensor $\mathcal{A} \in RT_{3,3}$ with

$$a_{111} = 0.0517,\ a_{112} = 0.3579,\ a_{113} = 0.5298,\ a_{122} = 0.7544,\ a_{123} = 0.2156,$$
$$a_{133} = 0.3612,\ a_{222} = 0.3943,\ a_{223} = 0.0146,\ a_{233} = 0.6718,\ a_{333} = 0.9723.$$

Employing the first-order difference approximation (6.2.4), we obtain a locally optimal symmetric rank-one tensor $\sigma_* \cdot \mathbf{u}^{\otimes 3}$ with

$$\begin{cases} \mathbf{u} = (-0.5204, -0.5113, -0.6839)^\top, \\ \sigma_* = -2.1110, \quad \rho(\mathcal{A}) = 0.8574, \quad \mathrm{RES}_{\mathrm{NET}} = 1.2672. \end{cases}$$

It takes 0.0781 s. Figure 6.3 shows the convergence of the global maximizer estimation for the optimization problem (6.1.7).

Comparing to [8, Example 3.3], it is obvious that $\sigma_* \cdot \mathbf{u}^{\otimes 3}$ is the best symmetric rank-one approximation and $|\sigma_*|$ is the Z-spectral radius of $\mathcal{A}$.

Example 6.6.7 ([8, Example 3.5]) Consider the symmetric tensor $\mathcal{A} \in RT_{3,I}$ with

$$a_{ijk} = \frac{(-1)^i}{i} + \frac{(-1)^j}{j} + \frac{(-1)^k}{k}.$$

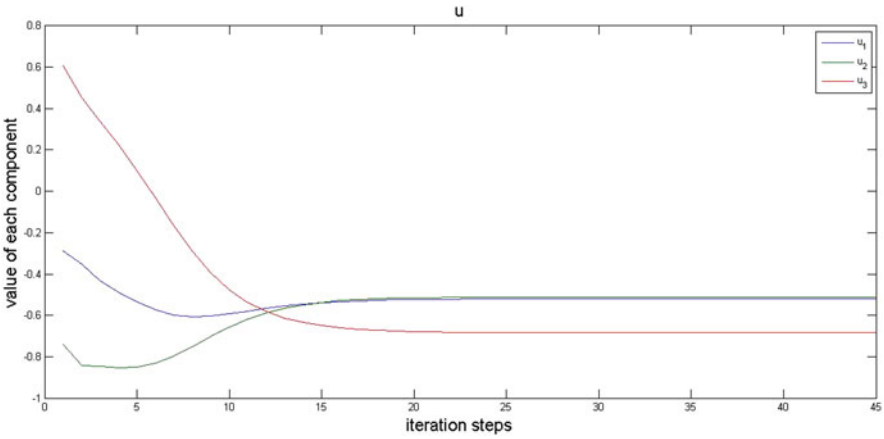

Fig. 6.3 Estimation of a global maximizer of (6.1.7) with Example 6.6.6

For $I = 5$, when the first-order difference approximation (6.2.4) is applied, we get two locally optimal symmetric rank-one tensors $\sigma_*^i \cdot (\mathbf{u}^i)^{\otimes 3}$ $(i = 1, 2)$ with

$$\begin{cases} \sigma_*^1 = 9.9779,\ \rho(\mathcal{A}) = 0.8813,\ \mathrm{RES_{NET}} = 5.3498, \\ \mathbf{u}^1 = (-0.7313, -0.1375, -0.4674, -0.2365, -0.4146)^\top; \\ \sigma_*^2 = 4.2876,\ \rho(\mathcal{A}) = 0.3787,\ \mathrm{RES_{NET}} = 10.4783, \\ \mathbf{u}^2 = (-0.1859,\ 0.7158,\ 0.2149,\ 0.5655,\ 0.2950)^\top. \end{cases}$$

It takes 0.0313 s and 0.0625 s, respectively. We see that $\sigma_*^1 \cdot (\mathbf{u}^1)^{\otimes 3}$ is the best symmetric rank-one approximation, consistent with [8, Example 3.7].

For I from 10 to 105, we apply the first-order difference approximation (6.2.4) to find the best symmetric rank-one approximation. The computational results are presented in Table 6.4.

Example 6.6.8 ([8, Example 3.6]) Consider the symmetric tensor $\mathcal{A} \in RT_{4,I}$ with

$$a_{i_1 i_2 i_3 i_4} = \arctan\left((-1)^{i_1}\frac{i_1}{I}\right) + \arctan\left((-1)^{i_2}\frac{i_2}{I}\right) + \arctan\left((-1)^{i_3}\frac{i_3}{I}\right) + \arctan\left((-1)^{i_4}\frac{i_4}{I}\right).$$

For $I = 5$, making use of the first-order difference approximation (6.2.4), we obtain two locally optimal symmetric rank-one tensors $\sigma_*^i \cdot (\mathbf{u}^i)^{\otimes 4}$ $(i = 1, 2)$ with

$$\begin{cases} \sigma_*^1 = -23.5741,\ \rho(\mathcal{A}) = 0.8135,\ \mathrm{RES_{NET}} = 16.8501, \\ \mathbf{u}^1 = (0.4403,\ 0.2382,\ 0.5602,\ 0.1354,\ 0.6459)^\top; \\ \sigma_*^2 = \ \ 13.0779,\ \rho(\mathcal{A}) = 0.4513,\ \mathrm{RES_{NET}} = 25.8579, \\ \mathbf{u}^2 = (0.3174,\ 0.5881,\ 0.1566,\ 0.7260,\ 0.0418)^\top. \end{cases}$$

Table 6.4 Related results for Example 6.6.7

I	σ_*	Time	$\rho(\mathcal{A})$	$\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$
10	−17.8002	0.0156	0.8042	13.1565
15	−26.4770	0.0156	0.7932	20.3254
20	24.1589	0.0313	0.7698	28.3238
25	42.5109	0.0625	0.7651	35.7736
30	−50.1376	0.0625	0.7528	43.8345
35	−58.3303	0.0938	0.7501	1.4221
40	−65.9255	0.0938	0.7423	59.5166
45	74.0191	0.0625	0.7404	67.1900
50	81.5934	0.0625	0.7349	75.3064
55	−89.4178	0.0625	0.7335	83.0399
100	158.9902	0.0938	0.7158	155.0992
105	−166.8380	0.1250	0.7153	162.9868

Table 6.5 Related results for Example 6.6.8

I	σ_*	Time	$\rho(\mathcal{A})$	$\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$
5	−23.5741	0.0781	0.8135	16.8501
10	77.0689	0.0625	0.7268	72.8350
15	−165.0965	0.0938	0.7081	164.6400
20	282.9708	0.0938	0.6911	295.9706
25	−435.3152	0.1250	0.6849	463.0760
30	617.5361	0.1563	0.6779	669.7284
35	−834.2093	0.1563	0.6749	912.1421
40	1.0808e+3	0.1875	0.6710	1.1941e+3
45	−1.3618e+3	0.2813	0.6693	1.5118e+3
50	1.6727e+3	0.5000	0.6669	1.8691e+3
55	−2.0180e+3	0.6875	0.6657	2.2622e+3
60	2.3933e+3	0.6563	0.6640	2.6948e+3

It takes 0.0469 s and 0.0625 s, respectively. Comparing to [8, Example 3.6], $\sigma_*^1 \cdot (\mathbf{u}^1)^{\otimes 4}$ is the best symmetric rank-one approximation.

For I from 5 to 60, we apply the first-order difference approximation (6.2.4) for the best symmetric rank-one approximation. The computational results are in Table 6.5.

Example 6.6.9 ([8, Example 3.7]) Consider the symmetric tensor $\mathcal{A} \in RT_{5,I}$ with

$$a_{i_1 i_2 i_3 i_4 i_5} = (-1)^{i_1} \ln(i_1) + (-1)^{i_2} \ln(i_2) + (-1)^{i_3} \ln(i_3) + (-1)^{i_4} \ln(i_4) + (-1)^{i_5} \ln(i_5).$$

For $I = 5$, taking advantage of the first-order difference approximation (6.2.4), we get two locally optimal symmetric rank-one tensors $\sigma_*^i \cdot (\mathbf{u}^i)^{\otimes 5}$ $(i = 1, 2)$ with

$$\begin{cases} \sigma_*^1 = \ \ 110.0083,\ \rho(\mathcal{A}) = 0.7709,\ \mathrm{RES}_{\mathrm{NET}} = 90.8819, \\ \mathbf{u}^1 = (-0.3906,\ -0.2782,\ -0.5666,\ -0.1669,\ -0.6490)^\top; \\ \sigma_*^2 = -69.8573,\ \rho(\mathcal{A}) = 0.4598,\ \mathrm{RES}_{\mathrm{NET}} = 124.4248, \\ \mathbf{u}^2 = (-0.4132,\ -0.5505,\ -0.1902,\ -0.6943,\ -0.0894)^\top. \end{cases}$$

It takes 0.0313 s and 0.0156 s, respectively, and $\sigma_*^1 \cdot (\mathbf{u}^1)^{\otimes 5}$ is the best symmetric rank-one approximation, comparable to [8, Example 3.7].

For I from 5 to 20, we apply the first-order difference approximation (6.2.4) to find the best symmetric rank-one approximations, as summarized in Table 6.6.

Remark 6.6.2 Suppose that $\tau_* \cdot \mathbf{v}^{\otimes N}$ is the best symmetric rank-one approximation returned by the semi-definite relaxation method [8], then we measure its quality by the residual:

$$\mathrm{RES}_{\mathrm{SPD}}(\mathcal{A}) = \left\| \mathcal{A} - \tau_* \cdot \mathbf{v}^{\otimes N} \right\|_F.$$

If $\sigma_* \cdot \mathbf{u}^{\otimes N}$ is the best rank-one approximation by the first-order difference approximation (6.2.4), its residual $\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$ is not significantly bigger than $\mathrm{RES}_{\mathrm{SPD}}(\mathcal{A})$ and the CPU time for (6.2.4) is smaller than the CPU time for the semi-definite relaxation methods.

Table 6.6 Computational results for Example 6.6.9

I	σ_*	Time	$\rho(\mathcal{A})$	$\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$
5	−110.0083	0.0313	0.7709	90.8818
6	−209.6261	0.0625	0.8109	151.2544
7	297.4530	0.0625	0.7376	272.3047
8	−470.6419	0.0781	0.7675	393.1087
9	−619.1962	0.0313	0.7174	601.3126
10	883.2849	0.0625	0.7408	801.0223
11	−1.1067e+3	0.0625	0.7040	1.1165e+3
12	−1.4780e+3	0.0625	0.7229	1.4126e+3
13	1.7897e+3	0.0625	0.6945	1.8544e+3
14	−2.3836e+3	0.1250	0.7103	2.3633e+3
15	−2.6966e+3	0.1094	0.6974	2.8494e+3
16	3.3277e+3	0.2656	0.7008	3.3870e+3
17	−3.8543e+3	0.1094	0.6819	4.1347e+3
18	−4.6369e+3	0.0938	0.6936	4.8160e+3
19	−5.2891e+3	0.1563	0.6775	5.7416e+3
20	6.2367e+3	0.2344	0.6878	6.5816e+3

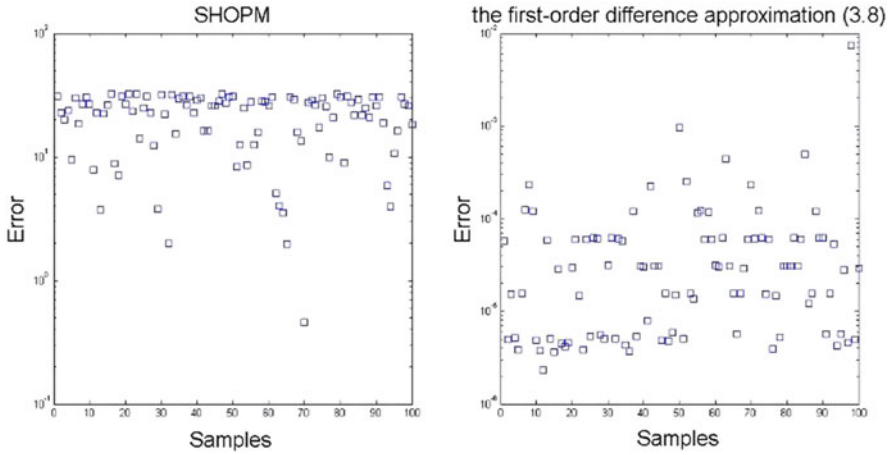

Fig. 6.4 Comparison (6.2.4) and SHOPM in Example 6.6.10

Table 6.7 Related results for Example 6.6.10

Values	Occ.	Time	$\mathrm{RES}_{\mathrm{NET}}(\mathcal{A})$
29.0085	36	0.1905	56.1536
32.7139	35	0.1496	54.1218
32.9758	29	0.1487	53.9626

Example 6.6.10 ([8, Example 3.10]) We will compare the first-order difference approximation (6.2.4) with S-HOPM in [22, Algorithm 2], which is widely used for computing rank-one approximatiosn for symmetric tensors. We can also use the semi-definite relaxation methods to check the convergence of S-HOPM. The symmetric tensor $\mathcal{A} \in RT_{3,I}$ is given by $a_{i_1 i_2 i_3} = \sin(i_1 + i_2 + i_3)$.

Suppose $I = 20$. If the unit vector $\mathbf{u} \in \mathbb{R}^I$ is derived from the first-order difference approximation (6.2.4) or S-HOPM, then we define Error $:= \|\mathcal{A}\mathbf{u}^2 - \sigma\mathbf{u}\|_2$, where $\sigma = \mathcal{A}\mathbf{u}^3$. When we test (6.2.4) and S-HOPM on the tensor $\mathcal{A}$ 100 times, the values of "Error" are shown in Fig. 6.4.

According to Fig. 6.4, we cannot use S-HOPM to compute a rank-one approximation of $\mathcal{A}$. By the first-order difference approximation (6.2.4), we find $\sigma_* \in \{29.0885, 32.7139, 32.9758\}$. The computational results are presented in Table 6.7.

6.6.3 Restricted Singular Values

In Sect. 6.4, we introduce three methods for finding the local maximizers of the optimization problem (6.4.3): Cholesky factorization+HOPM (Chol+HOPM), Cholesky factorization+the first-order difference approximation (6.2.2) (Chol+NN I) and the first-order difference approximation (6.4.7) (NN II). We compare these methods via an example.

We implement three methods: Chol+HOPM, Chol+NN I and NN II, and terminate the iterations when

$$\max_{n=1,2,\ldots,N} \|F(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k))_{-n} - \sigma(k)\mathbf{B}_n\mathbf{x}_n(k)\|_2 < \text{tol}$$

for NN II, or

$$\max_{n=1,2,\ldots,N} \left\|\widetilde{F}(\mathbf{y}_1(k), \mathbf{y}_2(k), \ldots, \mathbf{y}_N(k))_{-n} - \widetilde{\sigma}(k)\mathbf{y}_n(k)\right\|_2 < \text{tol}$$

for Chol+HOPM and Chol+NN, or $k > 5000$.

Here we have

$$\begin{cases} \sigma(k) = F(\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)), & \mathbf{x}_n(0)^\top \mathbf{B}_n \mathbf{x}_n(0) = 1; \\ \widetilde{\sigma}(k) = \widetilde{F}(\mathbf{y}_1(k), \mathbf{y}_2(k), \ldots, \mathbf{y}_N(k)), & \mathbf{y}_n(0)^\top \mathbf{y}_n(0) = 1; \end{cases}$$

and the symbol $\widetilde{F}$ is related to the tensor $\widetilde{\mathcal{A}}$, as given in (6.4.4) for F.

Example 6.6.11 Consider $\mathcal{A} \in \mathbb{R}^{I\times I\times I}$ given in Example 6.6.2. Three symmetric positive definite matrices $\mathbf{B}_i \in \mathbb{R}^{I\times I}$ are chosen from [39] with $i = 1, 2, 3$ as follows:

$$\mathbf{B}_1 = \mathbf{U}\mathbf{\Sigma}_1\mathbf{U}^\top, \quad \mathbf{B}_2 = \mathbf{U}\mathbf{\Sigma}_2\mathbf{U}^\top, \quad \mathbf{B}_3 = \mathbf{U}\mathbf{\Sigma}_3\mathbf{U}^\top,$$

in which $\mathbf{U}$ is a randomly chosen orthogonal matrix (obtained from a QR factorization of a matrix whose entries are drawn from a uniform distribution over $(-1, 1)$); the entries of

$$\mathbf{\Sigma}_i = \text{diag}(\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{iI}), \quad i = 1, 2, 3,$$

are: (a) $\sigma_{ij} = (i + j) + \eta_{ij}$; (b) $\sigma_{ij} = (i + j) + \eta_{ij}$ and $\sigma_{1j} = (1 + \eta_{1j}) * 10^{\frac{2}{7}j}$; (c) $\sigma_{ij} = (i + j) + \eta_{ij}$ and $\sigma_{2j} = (2 + \eta_{2j}) * 10^{\frac{3}{7}j}$; (d) $\sigma_{ij} = (i + j) + \eta_{ij}$ and $\sigma_{3j} = (3 + \eta_{3j}) * 10^{\frac{4}{7}j}$; (e) $\sigma_{1j} = (1 + j) + \eta_{ij}$, $\sigma_{2j} = (2 + \eta_{2j}) * 10^{\frac{3}{7}j}$ and $\sigma_{3j} = (3 + \eta_{3j}) * 10^{\frac{4}{7}j}$; and (f) $\sigma_{1j} = (1 + \eta_{1j}) * 10^{\frac{2}{7}j}$, $\sigma_{2j} = (2 + \eta_{2j}) * 10^{\frac{3}{7}j}$ and $\sigma_{3j} = (3 + \eta_{3j}) * 10^{\frac{4}{7}j}$ with $i = 1, 2, 3$ and $j = 1, 2, \ldots, I$, where η is drawn from a uniform distribution over $[0, 1)$.

We note that all $\mathbf{B}_n$ are well-conditioned or have a moderate condition number.

In this example, we have $I = 10$ and $\text{tol} = 1e - 13$. The computational results are in Table 6.8.

Table 6.8 Related results of Chol+HOPM, Chol+NN I and NN II for Example 6.6.11

	Chol+HOPM		Chol+NN I		NN II	
Cases	σ	Time	σ	Time	σ	Time
(a)	2.7236	10.8594	2.7236	39.0313	2.7236	43.8594
(b)	5.3084	9.1094	−5.3084	24.5156	5.3084	33.9531
(c)	2.8049	12.3438	2.8049	40.0313	−2.8049	486.9063
(d)	2.5504	8.2188	−2.5504	30.5469	2.5504	484.3438
(e)	3.5350	4.0625	−3.5350	15.1250	3.5350	482.5000
(f)	5.3357	10.6719	5.3357	41.8906	5.3357	486.4063

6.6.4 *Symmetric-Definite Tensor Pairs*

In this subsection, all testing tensors are quoted from Kolda and Mayo [36]. If the pair $(\lambda; \mathbf{x})$ is a generalized eigenpair of a symmetric-definite tensor pair $\{\mathcal{A}, \mathcal{B}\}$, we define

$$\mathbf{C}(\lambda; \mathbf{x}) = \mathbf{U}^\top [(N-1)\mathbf{H}(\mathbf{x}) - \lambda N \mathbf{I}_I] \mathbf{U} \in \mathbb{R}^{(I-1)\times(I-1)},$$

where the columns of $\mathbf{U} \in \mathbb{R}^{I\times(I-1)}$ form an orthonormal basis for $\mathbf{x}^\perp$ and $\mathbf{H}(\mathbf{x})$ is given in [36, Theorem 3.1]. One hundred tests of the first-order difference approximation (6.5.4) are performed, each with a different random initial vector. Selecting the initial vectors from either a uniform or normal distribution had no effect on the finial result.

Example 6.6.12 ([36, Example 5.2]) We consider a special case of the generalized tensor eigenvalue problem with the identity tensor $\mathcal{B}$. All entries of the symmetric tensor $\mathcal{A}$ are given in Table 6.9. We compute all the local maxima and the local minima of the optimization problem (6.5.2).

For the local maxima of the optimization problem (6.5.2), a summary of the results is provided in Table 6.10. We elaborate on the last column in Table 6.10. For example, 14 times out of 100 times, $\lambda(k)$ converges to 4.8422, then it takes (2.0045×14) s for the 14 tests of (6.5.3) (Table 6.11).

We use (6.5.4) for computing all local minima of the optimization problem (6.5.2) by replacing $\mathcal{A}$ in the (6.5.4) with $-\mathcal{A}$, as summarized in Table 6.12. For example, 36 times out of 100 times, $\lambda(k)$ converges to -10.7440, then it takes (1.1385×36) s for the 36 tests of (6.5.4).

Example 6.6.13 ([36, Example 5.5]) For a general symmetric-definite tensor pair, all the entries of $\mathcal{A} \in RT_{6,4}$ and positive definite $\mathcal{B} \in RT_{6,4}$ are given in Tables 6.9 and 6.11. We compute maxima and all local minima of the optimization problem (6.5.2) by (6.5.4).

For all local maxima of the optimization problem (6.5.2), a summary of the results is provided in Table 6.13. For example, in the last column in Table 6.13,

Table 6.9 $\mathcal{A}$ from Examples 6.6.12 and 6.6.13

$a_{111111} = 0.2888$	$a_{111112} = -0.0013$	$a_{111113} = -0.1422$	$a_{111114} = -0.0323$
$a_{111122} = -0.1070$	$a_{111123} = -0.0899$	$a_{111124} = -0.2487$	$a_{111133} = 0.0231$
$a_{111134} = -0.0106$	$a_{111144} = 0.0740$	$a_{111222} = 0.1490$	$a_{111223} = 0.0527$
$a_{111224} = -0.0710$	$a_{111233} = -0.1039$	$a_{111234} = -0.0250$	$a_{111244} = 0.0169$
$a_{111333} = 0.2208$	$a_{111334} = 0.0662$	$a_{111344} = 0.0046$	$a_{111444} = 0.0943$
$a_{112222} = -0.1144$	$a_{112223} = -0.1295$	$a_{112224} = -0.0484$	$a_{112233} = 0.0238$
$a_{112234} = -0.0237$	$a_{112244} = 0.0308$	$a_{112333} = 0.0142$	$a_{112334} = 0.0006$
$a_{112344} = -0.0044$	$a_{112444} = 0.0353$	$a_{113333} = 0.0947$	$a_{113334} = -0.0610$
$a_{113344} = -0.0293$	$a_{113444} = 0.0638$	$a_{114444} = 0.2326$	$a_{122222} = -0.2574$
$a_{122223} = 0.1018$	$a_{122224} = 0.0044$	$a_{122233} = 0.0248$	$a_{122234} = 0.0562$
$a_{122244} = 0.0221$	$a_{122333} = 0.0612$	$a_{122334} = 0.0184$	$a_{122344} = 0.0226$
$a_{122444} = 0.0247$	$a_{123333} = 0.0847$	$a_{123334} = -0.0209$	$a_{123344} = -0.0795$
$a_{123444} = -0.0323$	$a_{124444} = -0.0819$	$a_{133333} = 0.5486$	$a_{133334} = -0.0311$
$a_{133344} = -0.0592$	$a_{133444} = 0.0386$	$a_{134444} = -0.0138$	$a_{144444} = 0.0246$
$a_{222222} = 0.9207$	$a_{222223} = -0.0908$	$a_{222224} = 0.0633$	$a_{222233} = 0.1116$
$a_{222234} = -0.0318$	$a_{222244} = 0.1629$	$a_{222333} = 0.1797$	$a_{222334} = -0.0348$
$a_{222344} = -0.0058$	$a_{222444} = 0.1359$	$a_{223333} = 0.0584$	$a_{223334} = -0.0299$
$a_{223344} = -0.0110$	$a_{223444} = 0.1375$	$a_{224444} = -0.1405$	$a_{233333} = 0.2613$
$a_{233334} = 0.0809$	$a_{233344} = 0.0205$	$a_{233444} = 0.0196$	$a_{234444} = 0.0226$
$a_{244444} = -0.2487$	$a_{333333} = 0.6007$	$a_{333334} = -0.0272$	$a_{333344} = -0.1343$
$a_{333444} = -0.0233$	$a_{334444} = -0.0227$	$a_{344444} = -0.3355$	$a_{444444} = -0.5937$

Table 6.10 Local maximal H-eigenpairs for Example 6.6.12

Occ.	λ_*	$\mathbf{u}_*^\top$	Time
14	4.8422	(0.5895, −0.2640, −0.4728, 0.5994)	2.0045
15	5.8493	(0.6528, 0.5607, −0.0627, −0.5055)	0.9271
34	8.7371	(0.4837, 0.5502, 0.6671, −0.1354)	1.2757
15	9.6386	(0.5364, −0.5625, 0.5490, −0.3210)	2.3302
22	14.6941	(0.5426, −0.4853, 0.4760, 0.4936)	0.7131

62 times out of 100 times, $\lambda(k)$ converges to 11.3476, then it takes (1.7467×64) s for the 64 tests of (6.5.4).

We use (6.5.4) for computing all local minima of the optimization problem (6.5.2) by replacing $\mathcal{A}$ with $-\mathcal{A}$ while keeping $\mathcal{B}$ unchange. A summary of the results is provided in Table 6.14. For example, in the last column in Table 6.14, 48 times out of 100 times, $\lambda(k)$ converges to -6.3985, then it takes (3.4261×48) s for the 64 tests of (6.5.4).

Table 6.11 $\mathcal{B}$ from Example 6.6.13

$b_{111111} = 0.2678$	$b_{111112} = -0.0044$	$b_{111113} = -0.0326$	$b_{111114} = -0.0081$
$b_{111122} = 0.0591$	$b_{111123} = -0.0009$	$b_{111124} = -0.0045$	$b_{111133} = 0.0533$
$b_{111134} = -0.0059$	$b_{111144} = 0.0511$	$b_{111222} = -0.0020$	$b_{111223} = -0.0072$
$b_{111224} = -0.0016$	$b_{111233} = -0.0005$	$b_{111234} = 0.0007$	$b_{111244} = -0.0006$
$b_{111333} = -0.0185$	$b_{111334} = 0.0001$	$b_{111344} = -0.0058$	$b_{111444} = -0.0046$
$b_{112222} = 0.0651$	$b_{112223} = -0.0013$	$b_{112224} = -0.0050$	$b_{112233} = 0.0190$
$b_{112234} = -0.0023$	$b_{112244} = 0.0190$	$b_{112333} = -0.0011$	$b_{112334} = -0.0014$
$b_{112344} = 0.0000$	$b_{112444} = -0.0043$	$b_{113333} = 0.0498$	$b_{113334} = -0.0061$
$b_{113344} = 0.0169$	$b_{113444} = -0.0060$	$b_{114444} = 0.0486$	$b_{122222} = -0.0054$
$b_{122223} = -0.0078$	$b_{122224} = -0.0016$	$b_{122233} = -0.0006$	$b_{122234} = 0.0008$
$b_{122244} = -0.0006$	$b_{122333} = -0.0067$	$b_{122334} = 0.0001$	$b_{122344} = -0.0022$
$b_{122444} = -0.0016$	$b_{123333} = -0.0002$	$b_{123334} = 0.0006$	$b_{123344} = -0.0002$
$b_{123444} = 0.0006$	$b_{124444} = -0.0003$	$b_{133333} = -0.0286$	$b_{133334} = 0.0017$
$b_{133344} = -0.0056$	$b_{133444} = 0.0001$	$b_{134444} = -0.0051$	$b_{144444} = -0.0073$
$b_{222222} = 0.3585$	$b_{222223} = -0.0082$	$b_{222224} = -0.0279$	$b_{222233} = 0.0610$
$b_{222234} = -0.0076$	$b_{222244} = 0.0636$	$b_{222333} = -0.0042$	$b_{222334} = -0.0044$
$b_{222344} = -0.0002$	$b_{222444} = -0.0145$	$b_{223333} = 0.0518$	$b_{223334} = -0.0067$
$b_{223344} = 0.0184$	$b_{223444} = -0.0069$	$b_{224444} = 0.0549$	$b_{233333} = -0.0059$
$b_{233334} = -0.0034$	$b_{233344} = -0.0002$	$b_{233444} = -0.0039$	$b_{234444} = 0.0010$
$b_{244444} = -0.0208$	$b_{333333} = 0.2192$	$b_{333334} = -0.0294$	$b_{333344} = 0.0477$
$b_{333444} = -0.0181$	$b_{334444} = 0.0485$	$b_{344444} = -0.0304$	$b_{444444} = 0.2305$

Table 6.12 Local minimal H-eigenpairs for Example 6.6.12

Occ.	λ_*	$\mathbf{u}_*^\top$	Time
36	−10.7440	(0.4664, 0.4153, −0.5880, −0.5140)	1.1385
27	−8.3200	(0.5970, −0.5816, −0.4740, −0.2842)	1.1765
13	−4.1781	(0.4397, 0.5139, −0.5444, 0.4962)	7.3834
16	−3.7179	(0.6843, 0.5519, 0.3136, 0.3589)	1.2734
8	−2.9314	(0.3161, 0.5173, 0.4528, −0.6537)	2.3867

Table 6.13 Local maximal generalized eigenpairs for Example 6.6.13

Occ.	λ_*	$\mathbf{u}_*^\top$	Time
20	2.9979	(0.8224, 0.4083, −0.0174, −0.3958)	0.6398
18	3.7394	(0.2185, −0.9142, 0.2197, −0.2613)	2.2335
62	11.3476	(0.8224, 0.4083, −0.0174, −0.3958)	1.7467

Table 6.14 Local minimal generalized eigenpairs for Example 6.6.13

Occ.	λ_*	$\mathbf{u}_*^\top$	Time
48	−6.3985	(0.0733, 0.1345, 0.3877, 0.9090)	3.4261
16	−3.5998	(0.7899, 0.4554, 0.2814, 0.2991)	1.2686
19	−3.2777	(0.6888, −0.6272, −0.2914, −0.2174)	0.8808
17	−1.1507	(0.1935, 0.5444, 0.2991, −0.7594)	3.2960

References

1. A. Cichocki, R. Unbehauen, *Neural Networks for Optimization and Signal Processing* (Wiley, New York, 1993)
2. L. Liao, H. Qi, L. Qi, Neurodynamical optimization. J. Glob. Optim. **28**(2), 175–195 (2004)
3. A. Rodriguezvazquez, R. Dominguezcastro, A. Rueda, J. Huertas, E. Sanchezsinencio, Nonlinear switched capacitor 'neural' networks for optimization problems. IEEE Trans. Circuits Syst. **37**(3), 384–398 (1990)
4. A. Cichocki, Neural network for singular value decomposition. Electron. Lett. **28**(8), 784–786 (1992)
5. K. Diamantaras, S. Kung, Cross-correlation neural network models. IEEE Trans. Signal Process. **42**(11), 3218–3223 (1994)
6. D. Feng, Z. Bao, X. Zhang, A cross-associative neural network for SVD of non-squared data matrix in signal processing. IEEE Trans. Neural Netw. **12**(5), 1215–1221 (2001)
7. S. Fiori, Singular value decomposition learning on double stiefel manifold. Int. J. Neural Syst. **13**(3), 155–170 (2003)
8. J. Nie, L. Wang, Semidefinite relaxations for best rank-1 tensor approximations. SIAM J. Matrix Anal. Appl. **35**(3), 1155–1179 (2014)
9. L. Liu, W. Wu, Dynamical system for computing largest generalized eigenvalue, in *Lecture Notes in Computer Science*, vol. 3971 (Springer, Berlin, 2006), pp. 399–404
10. L. Liu, H. Shao, D. Nan, Recurrent neural network model for computing largest and smallest generalized eigenvalue. Neurocomputing **71**(16), 3589–3594 (2008)
11. Y. Tang, J. Li, Notes on "Recurrent neural network model for computing largest and smallest generalized eigenvalue". Neurocomputing **73**(4), 1006–1012 (2010)
12. S. Friedland, G. Ottaviani, The number of singular vector tuples and uniqueness of best rank-one approximation of tensors. Found. Comput. Math. **14**(6), 1209–1242 (2014)
13. L. Lim, Singular values and eigenvalues of tensors: a variational approach, in *IEEE CAMSAP 2005: First International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (IEEE, Piscataway, 2005), pp. 129–132
14. L. Wang, M. Chu, B. Yu, Orthogonal low rank tensor approximation: alternating least squares method and its global convergence. SIAM J. Matrix Anal. Appl. **36**(1), 1–19 (2015)
15. L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(r_1, r_2, ..., r_n) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000)
16. T. Zhang, G. Golub, Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. **23**(2), 534–550 (2001)
17. S. Friedland, L. Lim, Nuclear norm of higher-order tensors. Math. Comput. **87**, 1255–1281 (2018)
18. J. Nie, Symmetric tensor nuclear norms. SIAM J. Appl. Algebr. Geom. **1–1**, 599–625 (2017)
19. X. Zhang, C. Ling, L. Qi, The best rank-1 approximation of a symmetric tensor and related spherical optimization problems. SIAM J. Matrix Anal. Appl. **33**(3), 806–821 (2012)
20. K. Chang, K. Pearson, T. Zhang, Some variational principles for Z-eigenvalues of nonnegative tensors. Linear Algebra Appl. **438**(11), 4166–4182 (2013)
21. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
22. E. Kofidis, P. Regalia, On the best rank-1 approximation of higher-order supersymmetric tensors. SIAM J. Matrix Anal. Appl. **23**(3), 863–884 (2002)
23. T. Kolda, J. Mayo, Shifted power method for computing tensor eigenpairs. SIAM J. Matrix Anal. Appl. **32**(4), 1095–1124 (2011)
24. L. Ljung, Analysis of recursive stochastic algorithms. IEEE Trans. Autom. Control **22**(4), 551–575 (1977)
25. M. Che, A. Cichocki, Y. Wei, Neural networks for computing best rank-one approximations of tensors and its applications. Neurocomputing **267**(3), 114–133 (2017)
26. N. Samardzija, R. Waterland, A neural network for computing eigenvectors and eigenvalues. Biol. Cybern. **65**(4), 211–214 (1991)

27. J. Vegas, P. Zufiria, Generalized neural networks for spectral analysis: dynamics and liapunov function. Neural Netw. **17**(16), 233–245 (2004)
28. M. Che, X. Wang, Y. Wei, H. Yan, A collective neurodynamic optimization approach to the low multilinear rank approximation of tensors. Southwestern Univ. Financ. Econ. (2019). Preprint
29. X. Wang, M. Che, Y. Wei, Neural network for the computation of the best rank-one approximation of the fourth-order partially symmetric tensors. Numer. Math. Theory, Methods Appl. **11**, 673–700 (2018)
30. M. Che, L. Qi, Y. Wei, Perturbation bounds of tensor eigenvalue and singular value problems with even order. Linear Multilinear Algebra **64**(4), 622–652 (2016)
31. C. Van Loan, Generalizing the singular value decomposition. SIAM J. Numer. Anal. **13**(1), 76–83 (1976)
32. Y. Luo, D. Tao, K. Ramamonhanarao, C. Xu, Y. Wen, Tensor canonical correlation analysis for multi-view dimension reduction. IEEE Trans. Knowl. Data Eng. **27**(11), 3111–3124 (2015)
33. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)
34. K. Chang, K. Pearson, T. Zhang, On eigenvalue problems of real symmetric tensors. J. Math. Anal. Appl. **350**(1), 416–422 (2009)
35. W. Ding, Y. Wei, Generalized tensor eigenvalue problems. SIAM J. Matrix Anal. Appl. **36**(3), 1073–1099 (2015)
36. T. Kolda, J. Mayo, An adaptive shifted power method for computing generalized tensor eigenpairs. SIAM J. Matrix Anal. Appl. **35**(4), 1563–1581 (2014)
37. L. Qi, The best rank-one approximation ratio of a tensor space. SIAM J. Matrix Anal. Appl. **32**(2), 430–442 (2011)
38. S. Ragnarsson, C. Van Loan, Block tensors and symmetric embeddings. Linear Algebra Appl. **438**(2), 853–874 (2013)
39. D. Chu, L. De Lathauwer, B. De Moor, On the computation of the restricted singular value decomposition via the cosine-sine decomposition. SIAM J. Matrix Anal. Appl. **22**(2), 580–601 (2000)

Chapter 7
US- and U-Eigenpairs of Complex Tensors

In this chapter we discuss the computation of US-eigenpairs of complex symmetric tensors and U-eigenpairs of complex tensors. We derive an iterative algorithm for computing US-eigenpairs of complex symmetric tensors which is based on the Takagi factorization of complex symmetric matrices that are denoted as the QRCST Algorithm. We observe that multiple US-eigenpairs can be found from a local permutation heuristic, which is effectively a tensor similarity transformation, resulting in the permuted version of QRCST. We also present a generalization of their techniques to general complex tensors and derive a higher-order power type method for computing a US- or a U-eigenpair, similar to the higher-order power method for computing Z-eigenpairs of real symmetric tensors or a best rank-one approximation of real tensors. We illustrate the algorithms via numerical examples.

7.1 Preliminaries

7.1.1 US- and U-Eigenpairs

For a given symmetric tensor $\mathcal{A} \in CT_{N,I}$, a nonnegative number $\sigma \in \mathbb{R}$ is called a unitary symmetric eigenvalue (US-eigenvalue [1]) of $\mathcal{A}$, if there exist a unit vector $\mathbf{v} \in \mathbb{C}^I$ and $\sigma \in \mathbb{R}$ such that

$$\mathcal{A}\overline{\mathbf{v}}^{N-1} = \sigma \mathbf{v}, \quad \overline{\mathcal{A}}\mathbf{v}^{N-1} = \sigma \overline{\mathbf{v}}. \tag{7.1.1}$$

In this case, $\mathbf{v}$ is called a unitary symmetric eigenvector (US-eigenvector) of $\mathcal{A}$ associated with the US-eigenvalue σ and $(\sigma; \mathbf{v})$ is called a US-eigenpair of $\mathcal{A}$. The concept of the US-eigenvalues arises in quantum entanglement. The absolute largest σ is the entanglement eigenvalue and the corresponding complex symmetric rank-one tensor $\mathbf{v}^{\otimes N}$ is the closest symmetric separable state [1, 2].

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_7

When $N = 2$, the US-eigenvalues are the same as the Takagi values of $\mathcal{A}$. The tensor $\mathcal{A}$ is reduced to a complex symmetric matrix.

In Sect. 1.3.3, we present the definition of the U-eigenpairs of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. In this section, we assume that $I_n = I$ for all n. For a given $\mathcal{A} \in CT_{N,I}$, let $\mathbf{v}_n \in \mathbb{C}^I$ be nonzero vectors with $\|\mathbf{v}_n\|_2 = 1$ for all n, σ be nonnegative and $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ be a solution of the following nonlinear equations,

$$F(\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_N)_{-n} = \sigma \mathbf{v}_n, \quad \overline{F}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)_{-n} = \sigma \overline{\mathbf{v}}_n, \tag{7.1.2}$$

where

$$\begin{cases} F(\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_N)_{-n} = \mathcal{A} \times_1 \overline{\mathbf{v}}_1^\top \cdots \times_{n-1} \overline{\mathbf{v}}_{n-1}^\top \times_{n+1} \overline{\mathbf{v}}_{n+1}^\top \cdots \times_N \overline{\mathbf{v}}_N^\top \\ \qquad\qquad\qquad\qquad = \mathcal{A} \times_1 \mathbf{v}_1^* \cdots \times_{n-1} \mathbf{v}_{n-1}^* \times_{n+1} \mathbf{v}_{n+1}^* \cdots \times_N \mathbf{v}_N^*, \\ \overline{F}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)_{-n} = \overline{\mathcal{A}} \times_1 \mathbf{v}_1^\top \cdots \times_{n-1} \mathbf{v}_{n-1}^\top \times_{n+1} \mathbf{v}_{n+1}^\top \cdots \times_N \mathbf{v}_N^\top, \end{cases}$$

then $\sigma \in \mathbb{R}$ and $\mathbf{v}_n$ are called the unitary eigenvalue (U-eigenvalue) and the corresponding mode-n unitary eigenvector (the mode-n U-eigenvector) $\mathcal{A}$, respectively.

We call $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ a U-eigenpair of $\mathcal{A}$. Ragnarsson et al. [3] derive a method for obtaining a symmetric embedding $\mathbf{sym}(\mathcal{A})$ from $\mathcal{A} \in RT_{N,I}$. Similarly, for a general $\mathcal{A} \in CT_{N,I}$, we derive a complex symmetric embedding $\mathbf{csym}(\mathcal{A}) \in CT_{N,NI}$. We can establish a relationship between the US- and U-eigenpairs.

If $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ is a U-eigenpair of $\mathcal{A} \in CT_{N,I}$, then $(\sigma_\alpha; \mathbf{v}_\alpha)$ is a US-eigenpair of the symmetric tensor $\mathbf{csym}(\mathcal{A}) \in CT_{N,NI}$, where $\sigma_\alpha = \alpha_1 \alpha_2 \ldots \alpha_N \frac{N!}{\sqrt{N^N}} \sigma$ and

$$\mathbf{v}_\alpha = \frac{1}{\sqrt{N}} \left(\alpha_1 \mathbf{v}_1^\top, \alpha_2 \mathbf{v}_2^\top, \ldots, \alpha_N \mathbf{v}_N^\top \right)^\top, \quad \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N] = [1, \underbrace{\pm 1, \ldots, \pm 1}_{N-1}].$$

Furthermore, if θ is the argument of σ_α, then the pair $(\sigma, e^{\iota\theta/N}\mathbf{v})$ is a US-eigenpair of $\mathbf{csym}(\mathcal{A})$.

It is obvious when $N = 2$ that, i.e., $\mathcal{A}$ is a complex matrix, U-eigenvalues are the same as singular values. The Takagi factorization (another form for the singular value decomposition) of $A \in \mathbb{C}^{I \times I}$ can be written as

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top, \tag{7.1.3}$$

where $\mathbf{U}, \mathbf{V} \in \mathbb{C}^{I \times I}$ are unitary matrices and $\Sigma \in \mathbb{R}^{I \times I}$ is diagonal and positive semi-definite. Denote the main diagonal entry of Σ by σ_i.

The ith columns of $\mathbf{U}$ and $\mathbf{V}$ are respectively the left and the right Takagi vectors of $\mathbf{A}$, corresponding to the Takagi value σ_i. We call $(\sigma; \mathbf{u}, \mathbf{v})$ as a Takagi tuple of $\mathbf{A}$,

which satisfies

$$\begin{cases} \mathbf{A}\overline{\mathbf{v}} = \sigma\mathbf{u}, \quad \overline{\mathbf{A}}\mathbf{v} = \sigma\overline{\mathbf{u}}, \\ \mathbf{A}^\top\overline{\mathbf{u}} = \sigma\mathbf{v}, \quad \mathbf{A}^*\mathbf{u} = \sigma\overline{\mathbf{v}}, \\ \mathbf{u}^*\mathbf{u} = 1, \quad \mathbf{v}^*\mathbf{v} = 1, \quad \sigma \in \mathbb{R}. \end{cases}$$

In particular, if $\mathbf{U} = \mathbf{V}$, $\mathbf{A}$ is a complex symmetric matrix and $\mathbf{\Sigma} \in \mathbb{R}^{I\times I}$ is diagonal, (7.1.3) is the Takagi factorization of $\mathbf{A}$ [4–6].

7.2 Takagi Factorization of Complex Matrices

In order to analyze Algorithm 7.2.3 for the Takagi factorization, it is convenient to define the *left and right Takagi invariant subspaces* of a complex matrix, respectively generalized from the left and the right Takagi vectors.

Definition 7.2.1 Let $\mathbf{A} \in \mathbb{C}^{I\times I}$. A Takagi invariant subspace pair of $\mathbf{A}$ is made up of two subspace $\mathscr{X}$ and $\mathscr{Y}$ of $\mathbb{C}^I$, with the property that $\mathbf{x} \in \mathscr{X}$ and $\mathbf{y} \in \mathscr{Y}$ imply that $\mathbf{A}\overline{\mathbf{x}} \in \mathscr{Y}$ and $\mathbf{A}^\top\overline{\mathbf{y}} \in \mathscr{X}$. In other words, we have $\mathbf{A}\overline{\mathscr{X}} \subseteq \mathscr{Y}$ and $\mathbf{A}^\top\overline{\mathscr{Y}} \subseteq \mathscr{X}$, where $\overline{\mathscr{X}} = \{\overline{\mathbf{x}} : \mathbf{x} \in \mathscr{X}\}$ and $\overline{\mathscr{Y}} = \{\overline{\mathbf{y}} : \mathbf{y} \in \mathscr{Y}\}$. The dimensions of $\mathscr{X}$ and $\mathscr{Y}$ are the same.

Let the Takagi values of $\mathbf{A} \in \mathbb{C}^{I\times I}$ satisfy $\sigma_1 \geq \cdots \geq \sigma_K > \sigma_{K+1} \geq \cdots \geq \sigma_I \geq 0$ and $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and the right Takagi vectors corresponding to σ_i with $i = 1, 2, \ldots, K$. From Definition 7.2.1, the subspaces $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K\}$ and $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K\}$ are respectively called the left and the right dominant Takagi invariant subspaces of dimension K of $\mathbf{A}$.

Algorithm 7.2.1 Orthogonal iteration (simultaneous iteration or left and right Takagi iteration) for complex matrices

Input: Given $\mathbf{A} \in \mathbb{C}^{I\times I}$ and an integer $1 < P < I$ with $\sigma_1 \geq \cdots \geq \sigma_K > \sigma_{K+1} \geq \cdots \geq \sigma_I \geq 0$

Output: $\mathbf{\Sigma}_K \in \mathbb{R}^{K\times K}$ is diagonal positive definite, and $\mathbf{U}_K, \mathbf{V}_K \in \mathbb{C}^{I\times K}$ are unitary such that $\mathbf{A}\overline{\mathbf{V}}_K = \mathbf{U}_K\mathbf{\Sigma}_K$ and $\mathbf{A}^\top\overline{\mathbf{U}}_K = \mathbf{V}_K\mathbf{\Sigma}_K$

Given $\mathbf{X}_0, \mathbf{Y}_0 \in \mathbb{C}^{I\times K}$ such that $\mathbf{X}_0^*\mathbf{X}_0 = \mathbf{Y}_0^*\mathbf{Y}_0 = \mathbf{I}_K$

for $k = 0, 1, 2, \ldots$ **do**

Compute $\widehat{\mathbf{Y}}_{k+1} = \mathbf{A}\overline{\mathbf{X}}_k$ and $\widehat{\mathbf{X}}_{k+1} = \mathbf{A}^\top\overline{\mathbf{Y}}_k$

Factor $\widehat{\mathbf{X}}_{k+1} = \mathbf{X}_{k+1}\mathbf{R}_{1,k+1}$ and $\widehat{\mathbf{Y}}_{k+1} = \mathbf{Y}_{k+1}\mathbf{R}_{2,k+1}$ (thin QR decompositions)

end for

if convergence **then**

Compute $\mathbf{\Lambda}_\infty = \mathbf{Y}_\infty^*\mathbf{A}\overline{\mathbf{X}}_\infty$ and $\mathbf{\Sigma}_\infty = |\mathbf{\Lambda}_\infty|$

Choose $\widetilde{\mathbf{U}}_\infty = \mathbf{Y}_\infty\mathbf{D}_\infty$ and $\widetilde{\mathbf{V}}_\infty = \mathbf{X}_{k+1}\mathbf{D}_{k+1}$, where $\mathbf{D}_\infty \in \mathbb{C}^{K\times K}$ is diagonal with entires $\mathbf{D}_\infty(j, j) = \exp\left(\iota\frac{\arg(\mathbf{y}_{\infty,j}^*\mathbf{A}\overline{\mathbf{x}}_{\infty,j})}{2}\right)$ and $j = 1, 2, \ldots, K$

end if

From this definition, we present the result for Takagi values.

Lemma 7.2.1 *For $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I\times I}$, if there exist unitary matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{C}^{I\times I}$ such that $\mathbf{B} = \mathbf{P}^*\mathbf{A}\overline{\mathbf{Q}}$, then the Takagi values of $\mathbf{A}$ and $\mathbf{B}$ are the same.*

7.2.1 Orthogonal Type Iteration for Complex Matrices

Let the Takagi values of $\mathbf{A} \in \mathbb{C}^{I\times I}$ satisfy $\sigma_1 \geq \cdots \geq \sigma_K > \sigma_{K+1} \geq \cdots \geq \sigma_I \geq 0$. We compute the largest K Takagi values and the associated left and right Takagi vectors of $\mathbf{A}$, similar to the orthogonal iteration for symmetric matrix eigenvalue problems. This idea is summarized in Algorithm 7.2.1.

In particular, when $K = 1$, Algorithm 7.2.1 is reduced to Algorithm 7.2.2, which can be used to compute the largest Takagi value and the associated left and right Takagi vectors of $\mathbf{A}$ with $\sigma_1 > \sigma_2 \geq \cdots \geq \sigma_I \geq 0$. When the algorithms in this chapter terminate, the symbol "∞" stands for the maximal iteration step.

Algorithm 7.2.2 Power method iteration for complex matrices

Input: Given $\mathbf{A} \in \mathbb{C}^{I\times I}$ with $\sigma_1 > \sigma_2 \geq \cdots \geq \sigma_I \geq 0$
Output: Takagi value σ_1 and associated left and right Takagi vectors $\mathbf{u}_1$ and $\mathbf{v}_1$
Choose $\mathbf{x}_0, \mathbf{y}_0 \in \mathbb{C}^I$ with $\|\mathbf{x}_0\|_2 = \|\mathbf{y}_0\|_2 = 1$
for $k = 0, 1, 2, \ldots$ **do**
  Compute $\widehat{\mathbf{y}}_{k+1} = \mathbf{A}\overline{\mathbf{x}}_k$ and $\widehat{\mathbf{x}}_{k+1} = \mathbf{A}^\top\overline{\mathbf{y}}_k$
  Normalize $\mathbf{x}_{k+1} = \widehat{\mathbf{x}}_{k+1}/\|\widehat{\mathbf{x}}_{k+1}\|_2$ and $\mathbf{y}_{k+1} = \widehat{\mathbf{y}}_{k+1}/\|\widehat{\mathbf{y}}_{k+1}\|_2$
end for
if convergence **then**
  Compute $\lambda_\infty = \mathbf{y}_\infty^*\mathbf{A}\overline{\mathbf{x}}_\infty$ and $\widetilde{\sigma}_\infty = |\lambda_\infty|$
  Choose $\widetilde{\mathbf{u}}_\infty = \exp\left(\iota\frac{\arg(\lambda_\infty)}{2}\right)\mathbf{y}_\infty$ and $\widetilde{\mathbf{v}}_\infty = \exp\left(\iota\frac{\arg(\lambda_\infty)}{2}\right)\mathbf{x}_\infty$
end if

We analyze the convergence of Algorithm 7.2.1, or that of the subspaces $\mathrm{span}\{\mathbf{X}_k\}$ and $\mathrm{span}\{\mathbf{Y}_k\}$ of $\mathbb{C}^I$. Assume that $\sigma_K > \sigma_{K+1}$, we have

$$\begin{cases} \mathrm{span}\{\mathbf{X}_{k+1}\} = \mathrm{span}\{\widehat{\mathbf{X}}_{k+1}\} = \mathrm{span}\{\mathbf{A}^\top\overline{\mathbf{Y}}_k\}, \\ \mathrm{span}\{\mathbf{Y}_{k+1}\} = \mathrm{span}\{\widehat{\mathbf{Y}}_{k+1}\} = \mathrm{span}\{\mathbf{A}\overline{\mathbf{X}}_k\}. \end{cases}$$

According to the Takagi factorization of $\mathbf{A}$, we have

$$\begin{cases} \mathrm{span}\{\mathbf{Y}_{2k}\} = \mathrm{span}\{\mathbf{U}\Sigma^{2k}\mathbf{U}^*\mathbf{Y}_0\}, \quad \mathrm{span}\{\mathbf{Y}_{2k+1}\} = \mathrm{span}\{\mathbf{U}\Sigma^{2k+1}\mathbf{V}^\top\mathbf{X}_0\}, \\ \mathrm{span}\{\mathbf{X}_{2k}\} = \mathrm{span}\{\mathbf{V}\Sigma^{2k}\mathbf{V}^*\mathbf{X}_0\}, \quad \mathrm{span}\{\mathbf{X}_{2k+1}\} = \mathrm{span}\{\mathbf{V}\Sigma^{2k+1}\mathbf{U}^\top\mathbf{Y}_0\}. \end{cases}$$

For the convergence of span$\{\mathbf{Y}_k\}$, let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$. For $k = 0, 1, 2, \ldots$, we deduce that

$$\begin{cases} (\mathbf{A}\mathbf{A}^*)^k & = \underbrace{(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\overline{\mathbf{V}}\boldsymbol{\Sigma}^\top\mathbf{U}^*)(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\overline{\mathbf{V}}\boldsymbol{\Sigma}^\top\mathbf{U}^*)\ldots(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\overline{\mathbf{V}}\boldsymbol{\Sigma}^\top\mathbf{U}^*)}_{2k \text{ times}} = \mathbf{U}\boldsymbol{\Sigma}^{2k}\mathbf{U}^*, \\ (\mathbf{A}\mathbf{A}^*)^k\mathbf{A} & = \underbrace{(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\overline{\mathbf{V}}\boldsymbol{\Sigma}^\top\mathbf{U}^*)(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\overline{\mathbf{V}}\boldsymbol{\Sigma}^\top\mathbf{U}^*)\ldots(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\overline{\mathbf{V}}\boldsymbol{\Sigma}^\top\mathbf{U}^*)}_{2k \text{ times}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \\ & = \mathbf{U}\boldsymbol{\Sigma}^{2k+1}\mathbf{V}^\top, \\ (\mathbf{A}^\top\overline{\mathbf{A}})^k & = \underbrace{(\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\overline{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^*)(\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\overline{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^*)\ldots(\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\overline{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^*)}_{2k \text{ times}} = \mathbf{V}\boldsymbol{\Sigma}^{2k}\mathbf{V}^*, \\ (\mathbf{A}^\top\overline{\mathbf{A}})^k\mathbf{A}^\top & = \underbrace{(\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\overline{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^*)(\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\overline{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^*)\ldots(\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\overline{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^*)}_{2k \text{ times}}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top \\ & = \mathbf{V}\boldsymbol{\Sigma}^{2k+1}\mathbf{U}^\top, \end{cases}$$

since $\mathbf{V}^\top\overline{\mathbf{V}} = \mathbf{U}^*\mathbf{U} = \mathbf{U}^\top\overline{\mathbf{U}} = \mathbf{V}^*\mathbf{V} = \mathbf{I}_I$. Then, we have

$$\begin{aligned} \mathbf{U}\boldsymbol{\Sigma}^{2k}\mathbf{U}^*\mathbf{Y}_0 &= \mathbf{U}\mathrm{diag}(\sigma_1^{2k}, \sigma_2^{2k}, \ldots, \sigma_N^{2k})\mathbf{U}^*\mathbf{Y}_0 \\ &= \sigma_K^{2k}\mathbf{U}\mathrm{diag}((\sigma_1/\sigma_K)^{2k}, \ldots, 1, \ldots, (\sigma_N/\sigma_K)^{2k})\mathbf{U}^*\mathbf{Y}_0, \\ \mathbf{U}\boldsymbol{\Sigma}^{2k+1}\mathbf{V}^\top\mathbf{X}_0 &= \mathbf{U}\mathrm{diag}(\sigma_1^{2k+1}, \sigma_2^{2k+1}, \ldots, \sigma_N^{2k+1})\mathbf{V}^\top\mathbf{X}_0 \\ &= \sigma_K^{2k+1}\mathbf{U}\mathrm{diag}((\sigma_1/\sigma_K)^{2k+1}, \ldots, 1, \ldots, (\sigma_N/\sigma_K)^{2k+1})\mathbf{V}^\top\mathbf{X}_0. \end{aligned}$$

Since $\sigma_i/\sigma_K \geq 1$ if $i \leq K$, and $\sigma_i/\sigma_K < 1$ if $i > K$, we get

$$\begin{cases} \mathrm{diag}((\sigma_1/\sigma_K)^{2k}, \ldots, 1, \ldots, (\sigma_N/\sigma_K)^{2k})\mathbf{U}^*\mathbf{Y}_0 = \begin{pmatrix} \mathbf{P}_{2k} \\ \mathbf{Q}_{2k} \end{pmatrix}, \\ \mathrm{diag}((\sigma_1/\sigma_K)^{2k+1}, \ldots, 1, \ldots, (\sigma_N/\sigma_K)^{2k+1})\mathbf{V}^\top\mathbf{X}_0 = \begin{pmatrix} \mathbf{P}_{2k+1} \\ \mathbf{Q}_{2k+1} \end{pmatrix}. \end{cases}$$

Since $(\sigma_{K+1}/\sigma_K)^k$ converges to 0 as $k \to \infty$, $\mathbf{Q}_k \in \mathbb{C}^{(N-K)\times K}$ approaches the zero matrix and $\mathbf{P}_k \in \mathbb{C}^{K\times K}$ does not. Indeed, if $\mathbf{P}_0$ has full rank, so does $\mathbf{P}_k$. Let $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) = (\mathbf{U}_K, \widehat{\mathbf{U}}_K)$, with $\mathbf{U}_K = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K)$. Then

$$\begin{cases} \mathbf{U}\boldsymbol{\Sigma}^{2k}\mathbf{U}^*\mathbf{Y}_0 = \sigma_K^{2k}\mathbf{U}\begin{pmatrix} \mathbf{P}_{2k} \\ \mathbf{Q}_{2k} \end{pmatrix} = \sigma_K^{2k}(\mathbf{U}_K\mathbf{P}_{2k} + \widehat{\mathbf{U}}_K\mathbf{Q}_{2k}), \\ \mathbf{U}\boldsymbol{\Sigma}^{2k+1}\mathbf{V}^\top\mathbf{X}_0 = \sigma_K^{2k+1}\mathbf{U}\begin{pmatrix} \mathbf{P}_{2k+1} \\ \mathbf{Q}_{2k+1} \end{pmatrix} = \sigma_K^{2k+1}(\mathbf{U}_K\mathbf{P}_{2k+1} + \widehat{\mathbf{U}}_K\mathbf{Q}_{2k+1}). \end{cases}$$

Consequently span$\{\mathbf{Y}_k\}$ converges to span span$\{\mathbf{U}_K\}$,

$$\begin{cases} \text{span}\{\mathbf{Y}_{2k}\} = \text{span}\{(\mathbf{A}\mathbf{A}^*)^k\mathbf{Y}_0\} = \text{span}\{\mathbf{U}_K\mathbf{P}_{2k} + \widehat{\mathbf{U}}_K\mathbf{Q}_{2k}\} \\ \text{span}\{\mathbf{Y}_{2k+1}\} = \text{span}\{(\mathbf{A}\mathbf{A}^*)^k\mathbf{A}\mathbf{X}_0\} = \text{span}\{\mathbf{U}_K\mathbf{P}_{2k+1} + \widehat{\mathbf{U}}_K\mathbf{Q}_{2k+1}\}, \end{cases}$$

The left Takagi invariant subspace is spanned by the first P left Takagi vectors. The convergence of the right Takagi invariant subspace span$\{\mathbf{X}_k\}$ is similar.

In particular, we can set $|\mathbf{X}_0| = |\mathbf{Y}_0| = \mathbf{I}_I$ and $K = I$ in Algorithm 7.2.1. The next theorem shows under mild assumptions that we can use Algorithm 7.2.1 for computing the Takagi factorization of complex matrices.

Theorem 7.2.1 *Let $\mathbf{A} \in \mathbb{C}^{I\times I}$. Consider running the orthogonal iteration (Algorithm 7.2.1) on matrix $\mathbf{A}$ with $|\mathbf{X}_0| = |\mathbf{Y}_0| = \mathbf{I}_I$ and $K = I$. If all the Takagi values of $\mathbf{A}$ have distinct values and if all the principal submatrices $\mathbf{U}(1:j, 1:j)$ and $\mathbf{V}(1:j, 1:j)$ are of full column ranks, then as $k \to \infty$, $\mathbf{A}_k = \mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k$ converges to $\mathbf{D}\Sigma$, where $\mathbf{D} \in \mathbb{C}^{I\times I}$ satisfies $|\mathbf{D}| = \mathbf{I}_I$.*

Equivalently, as $k \to \infty$, $\widetilde{\mathbf{A}}_i = \widetilde{\mathbf{U}}_i^*\mathbf{A}\overline{\widetilde{\mathbf{V}}}_i$ converges to Σ. The Takagi values will appear in decreasing order.

Proof Since $\mathbf{X}_0$ and $\mathbf{Y}_0$ are unitary matrices, then $\mathbf{X}_k$ and $\mathbf{Y}_k$ are also unitary. The Takagi values of $\mathbf{A}$ and $\mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k$ are the same. Write $\mathbf{X}_k = [\mathbf{X}_{1k}, \mathbf{X}_{2k}]$, $\mathbf{Y}_k = [\mathbf{Y}_{1k}, \mathbf{Y}_{2k}]$, where $\mathbf{X}_{1k}$ and $\mathbf{Y}_{1k}$ have K columns, and $\mathbf{X}_{2k}$ and $\mathbf{Y}_{2k}$ have $I - K$ columns, then

$$\mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k = \begin{pmatrix} \mathbf{Y}_{1k}^*\mathbf{A}\overline{\mathbf{X}}_{1k} & \mathbf{Y}_{1k}^*\mathbf{A}\overline{\mathbf{X}}_{2k} \\ \mathbf{Y}_{2k}^*\mathbf{A}\overline{\mathbf{X}}_{1k} & \mathbf{Y}_{2k}^*\mathbf{A}\overline{\mathbf{X}}_{2k} \end{pmatrix}. \tag{7.2.1}$$

Since span$\{\mathbf{X}_{1k}\}$ and span$\{\mathbf{Y}_{1k}\}$ converge to a right and a left Takagi subspaces of $\mathbf{A}$, respectively, span$\{\mathbf{A}^\top\overline{\mathbf{Y}}_{1k}\}$ and span$\{\mathbf{A}\overline{\mathbf{X}}_{1k}\}$ converge to the same subspaces. All off-diagonal submatrices on the right-hand side of (7.2.1) converge to zero. Since this is true for all $P < N$, every off diagonal entry of $\mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k$ converges to zero, so $\mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k$ converges to a diagonal matrix. □

7.2.2 QR Type Algorithm for Complex Matrices

Algorithm 7.2.3 summarizes the main process of the QR algorithm for the Takagi factorization of complex matrices, originated from the QR algorithm for matrices.

When Algorithm 7.2.3 terminates, we do not restrict the entries of the final diagonal matrix to be real. For the Takagi factorization of $\mathbf{A} = \mathbf{U}\Sigma\mathbf{Q}^\top$, then for arbitrary diagonal matrix $\mathbf{D} \in \mathbb{C}^{I\times I}$, whose diagonal entries satisfy $d_{ii} = \exp(\iota\varphi_i)$, we have

$$\begin{aligned} \mathbf{A} &= (\mathbf{U}\mathbf{D}^{-\alpha})\mathbf{D}^\alpha\Sigma\mathbf{D}^\beta(\mathbf{V}\mathbf{D}^{-\beta})^\top = (\mathbf{U}\mathbf{D}^{-\alpha})(\mathbf{D}^{\alpha+\beta}\Sigma)(\mathbf{V}\mathbf{D}^{-\beta})^\top \\ &= (\mathbf{U}\mathbf{D}^{-\alpha})(\Sigma\mathbf{D}^{\alpha+\beta})(\mathbf{V}\mathbf{D}^{-\beta})^\top, \end{aligned}$$

Algorithm 7.2.3 QR algorithm for Takagi factorization of complex matrices

Input: Given $\mathbf{A} \in \mathbb{C}^{I\times I}$
Output: diagonal matrix $\Lambda \in \mathbb{C}^{I\times I}$ and unitary $\mathbf{U}, \mathbf{V} \in \mathbb{C}^{I\times I}$ such that $\mathbf{A}\overline{\mathbf{V}} = \mathbf{U}\Lambda$
Set $\mathbf{A}_0 := \mathbf{A}$ and $\mathbf{U}_0, \mathbf{V}_0 \leftarrow \mathbf{I}_I$
for $k = 0, 1, 2, \ldots$ **do**
  Factor $\mathbf{A}_k = \mathbf{P}_k\mathbf{R}_k$ and $\mathbf{A}_k^\top = \mathbf{Q}_k\mathbf{T}_k$ (QR decompositions)
  Compute $\mathbf{A}_{k+1} = \mathbf{P}_k^*\mathbf{A}_k\overline{\mathbf{Q}}_k$, $\mathbf{U}_{k+1} = \mathbf{U}_k\mathbf{P}_k$ and $\mathbf{V}_{k+1} = \mathbf{V}_k\mathbf{Q}_k$
end for

where $\varphi_i \in (-\pi, \pi]$ with all i and $\alpha, \beta \in \mathbb{R}$ are nonzero. The absolute values of the diagonal entries of $\mathbf{D}^{\alpha+\beta}\mathbf{\Sigma}$, $\mathbf{\Sigma}$ and $\mathbf{\Sigma}\mathbf{D}^{\alpha+\beta}$ are the same.

Since $\mathbf{A}_{k+1} = \mathbf{R}_k\overline{\mathbf{Q}}_k = \mathbf{P}_k^*\mathbf{T}_k^\top = \mathbf{P}_k^*\mathbf{A}_k\overline{\mathbf{Q}}_k$, then by Lemma 7.2.1, $\mathbf{A}_{k+1}$ and $\mathbf{A}_k$ have the same Takagi values. We claim that $\mathbf{A}_k$ computed by the QR iteration is identical to $\mathbf{P}_k^*\mathbf{A}\overline{\mathbf{Q}}_k$ implicitly computed by the orthogonal iteration.

Lemma 7.2.2 *Let $\mathbf{A} \in \mathbb{C}^{I\times I}$ and $\mathbf{A}_k = \mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k$, where $\mathbf{Y}_k$ and $\mathbf{X}_k$ are computed from Algorithm 7.2.1 with $\mathbf{Y}_0^*\mathbf{Y}_0 = \mathbf{X}_0^*\mathbf{X}_0 = \mathbf{I}_I$. As $k \to \infty$, $\mathbf{A}_k$ converges to $\mathbf{D}\Sigma$ if all the Takagi values are distinct, where $\mathbf{D} \in \mathbb{C}^{I\times I}$ is diagonal with $|\mathbf{D}| = \mathbf{I}_I$. The choice of $\mathbf{D}$ depends on $\mathbf{P}_k$, $\mathbf{Q}_k$ and $\mathbf{A}$.*

Proof We prove by the induction. From Algorithm 7.2.3, we deduce that

$$\mathbf{A}\overline{\mathbf{X}}_k = \mathbf{Y}_{k+1}\mathbf{R}_{2,k+1}, \quad \mathbf{A}^\top\overline{\mathbf{Y}}_k = \mathbf{X}_{k+1}\mathbf{R}_{1,k+1},$$

where $\mathbf{X}_{k+1}$ and $\mathbf{Y}_{k+1}$ are unitary, and $\mathbf{R}_{1,k+1}$ and $\mathbf{R}_{2,k+1}$ are upper triangular. Then $\mathbf{Y}_k^*\mathbf{A}\overline{\mathbf{X}}_k = \mathbf{Y}_k^*(\mathbf{Y}_{k+1}\mathbf{R}_{2,k+1})$ is the product of the unitary matrix $\mathbf{Q}_2 = \mathbf{Y}_k^*\mathbf{Y}_{k+1}$ and the upper triangular $\mathbf{R}_2 = \mathbf{R}_{2,k+1} = \mathbf{Y}_{k+1}^*\mathbf{A}\overline{\mathbf{X}}_k$. Similarly, $\mathbf{X}_k^*(\mathbf{A}^\top\overline{\mathbf{Y}}_k) = \mathbf{X}_k^*(\mathbf{X}_{k+1}\mathbf{R}_{1,k+1})$ is the product of the unitary matrix $\mathbf{Q}_1 = \mathbf{X}_k^*\mathbf{X}_{k+1}$ and the upper triangular matrix $\mathbf{R}_1 = \mathbf{X}_{k+1}^*\mathbf{A}^\top\overline{\mathbf{Y}}_k$. These must be the unique QR decompositions $\mathbf{A}_k = \mathbf{Q}_2\mathbf{R}_2$ and $\mathbf{A}_k^\top = \mathbf{Q}_1\mathbf{R}_1$ (except for the scaling of each column of $\mathbf{Q}$ and the row of $\mathbf{R}$ by -1). Then

$$\begin{cases} \mathbf{Y}_{k+1}^*\mathbf{A}\overline{\mathbf{X}}_{k+1} = (\mathbf{Y}_{k+1}^*\mathbf{A}\overline{\mathbf{X}}_k)(\mathbf{X}_k^\top\overline{\mathbf{X}}_{k+1}) = \mathbf{R}_{2,k+1}(\mathbf{X}_k^\top\overline{\mathbf{X}}_{k+1}) = \mathbf{R}_2\overline{\mathbf{Q}}_2, \\ \mathbf{X}_{k+1}^*\mathbf{A}^\top\overline{\mathbf{Y}}_{k+1} = (\mathbf{X}_{k+1}^*\mathbf{A}^\top\overline{\mathbf{Y}}_k)(\mathbf{Y}_k^\top\overline{\mathbf{Y}}_{k+1}) = \mathbf{R}_{1,k+1}(\mathbf{Y}_k^\top\overline{\mathbf{Y}}_{k+1}) = \mathbf{R}_1\overline{\mathbf{Q}}_1. \end{cases}$$

This is precisely how the QR iteration maps $\mathbf{A}_k$ to $\mathbf{A}_{k+1}$, so $\mathbf{Y}_{k+1}^*\mathbf{A}\bar{\mathbf{X}}_{k+1} = \mathbf{A}_{k+1}$ as desired. □

7.3 Iterative Algorithm for US-Eigenpairs

We suppose that $\mathcal{A} \in CT_{N,I}$, is symmetric and the imaginary part of $a_{i_1 i_2 \ldots i_N}$ is not zero for all i_n and n. We deduce some properties of US-eigenpairs of $\mathcal{A}$. We also present an iterative algorithm for computing the US-eigenpairs and analyze its convergence.

7.3.1 Properties of US-Eigenpairs

We present an obvious lemma on a property of US-eigenpairs of complex symmetric tensors without the proof.

Lemma 7.3.1 *Suppose that $\mathcal{A} \in CT_{N,I}$ is symmetric and $\theta \in (-\pi, \pi]$. If $(\sigma; \mathbf{v})$ is a* US*-eigenpair of $\mathcal{A}$, then there exists a scalar $\varphi \in (-\pi, \pi]$ such that $(e^{\iota\theta}\sigma; e^{-\iota\varphi}\mathbf{v})$ and $(e^{-\iota\theta}\sigma; e^{\iota\varphi}\mathbf{v})$ are respectively solutions of the two equations in (7.1.1) with $N\varphi = \theta$.*

Two complex symmetric tensors $\mathcal{A}, \mathcal{B} \in CT_{N,I}$ are *similar* [7], if for a nonsingular $\mathbf{P} \in \mathbb{C}^{I\times I}$, we have $\mathcal{B} := \mathcal{A}\mathbf{P}^N = \mathcal{A} \times_1 \mathbf{P} \times_2 \mathbf{P} \cdots \times_N \mathbf{P}$. The inverse transformation of $\mathcal{B}$ by $\mathbf{P}^{-1}$ will retrieve $\mathcal{A}$:

$$\mathcal{B}(\mathbf{P}^{-1})^N = (\mathcal{A}\mathbf{P}^N)(\mathbf{P}^{-1})^N = \mathcal{A} \times_1 (\mathbf{P}^{-1}\mathbf{P}) \times_2 (\mathbf{P}^{-1}\mathbf{P}) \cdots \times_N (\mathbf{P}^{-1}\mathbf{P}) = \mathcal{A}.$$

For a symmetric $\mathcal{A} \in CT_{N,I}$ and a nonsingular $\mathbf{P} \in \mathbb{C}^{I\times I}$, it is easy to prove that $\mathcal{B} = \mathcal{A}\mathbf{P}^N$ is symmetric (see [7, Proposition 10] for the real symmetric tensor case).

In particular, assume that $\widetilde{\mathcal{A}} \in CT_{N,I}$ is similar to $\mathcal{A}$ through a similarity transformation by a unitary matrix $\mathbf{Q}^* \in \mathbb{C}^{I\times I}$. Let $\widetilde{\mathcal{A}} = \mathcal{A}(\mathbf{Q}^*)^N$ and $\theta \in (-\pi, \pi]$ be the argument of $\widetilde{\mathcal{A}}_{11\ldots1} = \widetilde{\mathcal{A}}\mathbf{e}_1^N$, where $\mathbf{e}_1$ is the first column of $\mathbf{I}_I$. It follows that if $\theta \neq 0$ and $(\sigma; e^{\iota\theta/N}\mathbf{e}_1)$ is a US-eigenpair of $\widetilde{\mathcal{A}}$, then $(\sigma; \mathbf{q}_1)$ is a US-eigenpair of $\mathcal{A}$, where $\mathbf{q}_1$ denotes the first column of $\mathbf{Q}$. Then we have

$$\begin{aligned} e^{-\iota\theta(N-1)/N}\widetilde{\mathcal{A}}\mathbf{e}_1^{N-1} = e^{\iota\theta/N}\sigma\mathbf{e}_1 &\Leftrightarrow \mathcal{A} \times_1 \mathbf{Q}^* \times_2 (\mathbf{e}_1^\top\mathbf{Q}^*) \cdots \times_N (\mathbf{e}_1^\top\mathbf{Q}^*) = \sigma\mathbf{e}_1 \\ &\Leftrightarrow \mathcal{A} \times_1 \mathbf{Q}^* \times_2 \mathbf{q}_1^* \cdots \times_N \mathbf{q}_1^* = \sigma\mathbf{e}_1 \\ &\Leftrightarrow \mathbf{Q}^*\mathcal{A}\overline{\mathbf{q}}_1^{N-1} = \sigma\mathbf{e}_1 \\ &\Leftrightarrow \mathcal{A}\overline{\mathbf{q}}_1^{N-1} = \sigma\mathbf{Q}\mathbf{e}_1 = \sigma\mathbf{q}_1. \end{aligned} \tag{7.3.1}$$

Also, we have

$$e^{\iota\theta(N-1)/N}\overline{\widetilde{\mathcal{A}}}\mathbf{e}_1^{N-1} = e^{-\iota\theta/N}\sigma\mathbf{e}_1 \Leftrightarrow \overline{\mathcal{A}}\mathbf{q}_1^{N-1} = \sigma\overline{\mathbf{q}}_1.$$

We note that $\sigma = |\widetilde{\mathcal{A}}\mathbf{e}_1^N|$. In fact, we do not need to apply $\mathbf{e}_1$ and the above result easily generalizes to any $\mathbf{e}_i$, the ith column of $\mathbf{I}_I$ for all i, as in Theorem 7.3.1.

Theorem 7.3.1 *For a symmetric tensor $\mathcal{A} \in CT_{N,I}$, if $(\sigma; e^{\iota\theta/N}\mathbf{e}_i)$ is a US-eigenpair of $\mathcal{A}(\mathbf{Q}^*)^N$, then $(\sigma; \mathbf{q}_i)$ is a* US*-eigenpair of $\mathcal{A}$, where $\mathbf{q}_i$ is the ith column of the unitary matrix $\mathbf{Q} \in \mathbb{C}^{I\times I}$ and $\theta \in (-\pi, \pi]$ is the argument of $(\mathcal{A}(\mathbf{Q}^*)^N)\mathbf{e}_i^N$.*

Proof The proof follows from (7.1.1) and (7.3.1) with $\mathbf{e}_1$ replaced by $\mathbf{e}_i$. □

If two complex symmetric tensors are similar through a similarity transformation by a unitary matrix, then the corresponding US-eigenpairs satisfy the following theorem.

Theorem 7.3.2 *Suppose that $\mathcal{A}, \mathcal{B} \in CT_{N,I}$ are complex symmetric tensors. If there exists a unitary matrix $\mathbf{Q} \in \mathbb{C}^{I\times I}$ such that $\mathcal{B} = \mathcal{A}(\mathbf{Q}^*)^N$ and $(\sigma; \mathbf{v})$ is a* US*-eigenpair of $\mathcal{B}$, then $(\sigma; \mathbf{Qv})$ is a* US*-eigenpair of $\mathcal{A}$.*

Proof We leave the proof as an exercise. □

7.3.2 QR Algorithms for Complex Symmetric Tensors

In Sect. 7.2, we propose a QR-type algorithm for computing the Takagi factorization of complex matrices. The algorithm is efficient for complex symmetric matrices. The QR algorithm for a complex symmetric matrix starts from an initial complex symmetric matrix $\mathbf{A}_0 \in \mathbb{C}^{I\times I}$, and computes its QR factorization $\mathbf{Q}_1\mathbf{R}_1 = \mathbf{A}_0$. Next, the QR-type product is reversed to give $\mathbf{A}_1 = \mathbf{R}_1\overline{\mathbf{Q}}_1 = \mathbf{Q}_1^*\mathbf{A}_0\overline{\mathbf{Q}}_1$. In subsequent iterations, we get $\mathbf{Q}_k\mathbf{R}_k = \mathbf{A}_{k-1}$ and

$$\mathbf{A}_k = \mathbf{Q}_k^*\mathbf{A}_{k-1}\overline{\mathbf{Q}}_k = (\mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_k)^*\mathbf{A}_0\overline{(\mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_k)} = \underline{\mathbf{Q}}_k^*\mathbf{A}_0\overline{\underline{\mathbf{Q}}}_k,$$

where $\underline{\mathbf{Q}}_k = \mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_k$.

Under mild hypothesis, $\mathbf{A}_k$ converges to a diagonal form, where its absolute value contains the Takagi values of $\mathbf{A}_0$. In this section, we first present the QRCST algorithm to compute US-eigenpairs of complex symmetric tensors. Moreover, with a permutation strategy to scramble tensor entries, a more efficient version of the QRCST is proposed to produce possibly more distinct US-eigenpairs.

The QRCST is summarized in Algorithm 7.3.1. The core of QRCST is a two-step iteration, indicated as Steps 1 and 2, for the ith mode-$(1, 2)$ slice with all i.

Some remarks about Algorithm 7.3.1 are presented in the following:

(a) Since the core computation of the QRCST lies in the QR factorization of $O(I^3)$ complexity for complex matrices and the corresponding similarity transform of $O(NI^{N+1})$ complexity, the overall complexity is therefore dominated by the better.
(b) $\mathcal{A}_0$ is a finite sum of R rank-one tensors, the complexity of the QRCST will be reduced to $O(NRI^2)$.

For structured tensors, we can improve the computation of the similarity transformation by unitary matrices in Step 2 of Algorithm 7.3.1. For example, for Hankel tensors [8, 9], we can apply the fast Hankel tensor-vector product technique [10].

For a given symmetric $\mathcal{A} \in CT_{N,I}$, the QRCST algorithm produces at most I US-eigenpairs in one pass, as some US-eigenairs may be identical. According to

Algorithm 7.3.1 QRCST

Input: Given a symmetric $\mathcal{A} \in CT_{N,I}$, tolerance τ, maximum number of iteration $K_{\max}$
Output: US-eigenpairs $(\sigma_i; \mathbf{q}_i)$
for $i = 1, 2, \ldots, I$ **do**
  Set $\mathcal{A}_0 := \mathcal{A}$ and $\mathbf{Q} \leftarrow \mathbf{I}_I$
  Compute $\epsilon \leftarrow \|\mathcal{A}_0 \mathbf{e}_i^{N-1} - \mathbf{e}_i\|_2 / \|\mathcal{A}_0 \mathbf{e}_i^{N-2}\|_2$
  iteration count $k \leftarrow 1$
  while $\epsilon \leftarrow \tau$ or $k \leq K_{\max}$ **do**
    Step 1: factorize $\mathbf{Q}_k \mathbf{R}_k = \mathcal{A}_{k-1} \mathbf{e}_i^{N-2}$ and Compute $\mathbf{Q} \leftarrow \mathbf{Q}\mathbf{Q}_k$
    Step 2: compute $\mathcal{A}_k = \mathcal{A}_{k-1}(\mathbf{Q}_k^*)^N$
    Compute $\epsilon \leftarrow \|\mathcal{A}_k \mathbf{e}_i^{N-1} - \mathbf{e}_i\|_2 / \|\mathcal{A}_k \mathbf{e}_i^{N-2}\|_2$
  end while
end for
if converged **then**
  Let $\sigma_i \leftarrow |\mathcal{A}_k \mathbf{e}_i^N|$ and θ be the argument of $\mathcal{A}_k \mathbf{e}_i^N$
  Compute $\mathbf{q}_i \leftarrow e^{\iota\theta/N} \mathbf{Q}(:, i)$, where $\mathbf{Q}(:, i)$ is the ith column of $\mathbf{Q}$
end if

Theorem 7.3.1, we may improve the efficiency of Algorithm 7.3.2 by scrambling the entries in $\mathcal{A}_0$.

Let $\mathfrak{S}$ be the set of all permutation matrices and $I!$ be its cardinality. We can apply QRCST on a complex symmetric tensor with different permutations. The resulting Permuted QRCST algorithm for complex symmetric tensors is summarized in Algorithm 7.3.2.

Algorithm 7.3.2 Permuted QRCST

Input: Given a symmetric $\mathcal{A}_0 \in CT_{N,I}$
Output: US-eigenpairs $(\sigma_i; \mathbf{q}_i)$
for $p = 1, 2, \ldots, I!$ **do**
  Choose $\mathbf{P}_p \in \mathfrak{S}$ and compute $\mathcal{A}_p \leftarrow \mathcal{A}_0 \mathbf{P}_p^N$
  Apply Algorithm 7.3.1 with $\mathcal{A}_p$ and collect distinct convergent US-eigenpairs
end for

As the cardinality of $\mathfrak{S}$ is $I!$, Permuted QRCST quickly becomes impractical for large I.

7.3.3 Convergence

We next analyze the convergence of Algorithm 7.3.1. We set $\underline{\mathbf{Q}}_0 = \mathbf{I}_I$ and $\underline{\mathbf{Q}}_k = \mathbf{Q}_1 \mathbf{Q}_2 \ldots \mathbf{Q}_k$. Also, $\underline{\mathbf{q}}_{i,k}$ is the ith column of $\underline{\mathbf{Q}}_k$ and $r_{ij,k}$ is the (i, j)-entry of $\mathbf{R}_k$. A

few iterations of the QRCST can be better visualized as follows:

$$\begin{cases} \mathbf{Q}_1\mathbf{R}_1 = \mathcal{A}_0\mathbf{e}_i^{N-2}, \quad \mathcal{A}_1 = \mathcal{A}_0(\mathbf{Q}_1^*)^N = \mathcal{A}_0(\underline{\mathbf{Q}}_1^*)^N, \\ \mathbf{Q}_2\mathbf{R}_2 = \mathcal{A}_1\mathbf{e}_i^{m-2} = \mathcal{A}_0 \times_1 \underline{\mathbf{Q}}_1^* \times_2 \underline{\mathbf{Q}}_1^* \times_3 \underline{\mathbf{q}}_{i,1}^* \cdots \times_N \underline{\mathbf{q}}_{i,1}^* = \underline{\mathbf{Q}}_1^*(\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{i,1}^{N-2})\overline{\underline{\mathbf{Q}}}_1, \\ \quad \mathcal{A}_2 = \mathcal{A}_1(\mathbf{Q}_2^*)^N = \mathcal{A}_0(\underline{\mathbf{Q}}_2^*)^N, \\ \mathbf{Q}_3\mathbf{R}_3 = \mathcal{A}_2\mathbf{e}_i^{N-2} = \mathcal{A}_0 \times_1 \underline{\mathbf{Q}}_2^* \times_2 \underline{\mathbf{Q}}_2^* \times_3 \underline{\mathbf{q}}_{i,2}^* \cdots \times_N \underline{\mathbf{q}}_{i,2}^* = \underline{\mathbf{Q}}_2^*(\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{i,2}^{N-2})\overline{\underline{\mathbf{Q}}}_2. \end{cases} \tag{7.3.2}$$

From (7.3.2), we have for $k = 1, 2, \ldots,$

$$\mathcal{A}_{k-1} = \mathcal{A}_0(\underline{\mathbf{Q}}_{k-1}^*)^N, \quad \mathbf{Q}_k\mathbf{R}_k = \mathcal{A}_{k-1}\mathbf{e}_i^{N-2} = \underline{\mathbf{Q}}_{k-1}^*(\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{i,k-1}^{m-2})\overline{\underline{\mathbf{Q}}}_{k-1}. \tag{7.3.3}$$

Multiplying $\underline{\mathbf{Q}}_{k-1}$ to both sides of (7.3.3), then we get

$$\begin{aligned} \underline{\mathbf{Q}}_k\mathbf{R}_k &= (\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{i,k-1}^{N-2})\overline{\underline{\mathbf{Q}}}_{k-1} \\ &\Leftrightarrow \left(\underline{\mathbf{q}}_{1,k}\ \underline{\mathbf{q}}_{2,k}\ \cdots\ \underline{\mathbf{q}}_{n,k}\right)\begin{pmatrix} r_{11,k} & r_{12,k} & \ldots & r_{1I,k} \\ & r_{22,k} & \ldots & r_{2I,k} \\ & & \ddots & \vdots \\ & & & r_{II,k} \end{pmatrix} \\ &= \left[\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{i,k-1}^{N-2}\right]\left(\overline{\underline{\mathbf{q}}}_{1,k-1}\ \overline{\underline{\mathbf{q}}}_{2,k-1}\ \cdots\ \overline{\underline{\mathbf{q}}}_{I,k-1}\right). \end{aligned} \tag{7.3.4}$$

It is obvious that (7.3.4) is the tensor generalization of the matrix orthogonal iteration (or called the simultaneous or the Takagi subspace iteration) for finding dominant invariant Takagi subspace of complex symmetric matrices (Algorithm 7.2.1).

If we set $i = 1$ and extract the first columns on both sides of (7.3.4), then we get

$$\underline{\mathbf{q}}_{1,k}r_{11,k} = (\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{1,k-1}^{N-2})\overline{\underline{\mathbf{q}}}_{1,k-1} = \mathcal{A}_0\overline{\underline{\mathbf{q}}}_{1,k-1}^{N-1} \quad \text{or} \quad \underline{\mathbf{q}}_{1,k} = \frac{\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{1,k-1}^{N-1}}{\left\|\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{1,k-1}^{N-1}\right\|_2}. \tag{7.3.5}$$

According to (7.3.5), we derive a higher order power method for computing the US-eigenpairs of complex symmetric tensors. This process is summarized in Algorithm 7.3.3. The convergence analysis of Algorithm 7.3.3 will be presented in the next subsection.

If we start Algorithm 7.3.1 with $i = 2$, then the first two columns of (7.3.4) proceed as

$$\underline{\mathbf{q}}_{1,k}r_{11,k} = (\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{2,k-1}^{N-2})\overline{\underline{\mathbf{q}}}_{1,k-1}, \quad \underline{\mathbf{q}}_{1,k}r_{12,k} + \underline{\mathbf{q}}_{2,k}r_{22,k} = (\mathcal{A}_0\overline{\underline{\mathbf{q}}}_{2,k-1}^{N-2})\overline{\underline{\mathbf{q}}}_{2,k-1}.$$

Algorithm 7.3.3 Higher order power method for complex symmetric tensors

Input: Given a symmetric $\mathcal{A} \in CT_{N,I}$
Output: US-eigenpairs $(\sigma; \mathbf{v})$
Choose an initial vector $\mathbf{x}_0 \in \mathbb{C}^n$ with $\|\mathbf{x}_0\|_2 = 1$
for $k = 0, 1, 2, \ldots$ **do**
  Compute $\mathbf{y}_{k+1} = \mathcal{A}\overline{\mathbf{x}}_k^{N-1}$ and normalize $\mathbf{x}_{k+1} = \mathbf{y}_{k+1}/\|\mathbf{y}_{k+1}\|_2$
end for
Compute $\lambda_\infty = \mathcal{A}\overline{\mathbf{x}}_\infty^N$
Set $\sigma_\infty = |\lambda_\infty|$ and $\mathbf{v}_\infty = \exp\left(\iota\frac{\theta}{N}\right)\mathbf{x}_\infty$, where $\theta \in (-\pi, \pi]$ is the argument of λ_∞

With convergence at $k = \infty$, we have

$$\underline{\mathbf{q}}_{1,\infty} r_{11,\infty} = (\mathcal{A}_0 \overline{\underline{\mathbf{q}}}_{2,\infty}^{N-2})\overline{\underline{\mathbf{q}}}_{1,\infty}, \quad \underline{\mathbf{q}}_{1,\infty} r_{12,\infty} + \underline{\mathbf{q}}_{2,\infty} r_{22,\infty} = (\mathcal{A}_0 \overline{\underline{\mathbf{q}}}_{2,\infty}^{N-2})\overline{\underline{\mathbf{q}}}_{2,\infty}, \tag{7.3.6}$$

from which it is easy to check that $r_{12,\infty} = \overline{\underline{\mathbf{q}}}_{1,\infty}^\top \mathcal{A}_0 \overline{\underline{\mathbf{q}}}_{2,\infty}^{N-1} = r_{11,\infty}\underline{\mathbf{q}}_{1,\infty}^* \underline{\mathbf{q}}_{2,\infty} = 0$. In fact, for any i, if $\underline{\mathbf{Q}}_k$ and $\mathbf{R}_k$ converge to $\underline{\mathbf{Q}}_\infty$ and $\mathbf{R}_\infty$ respectively, then (7.3.4) implies

$$\mathbf{R}_\infty = \begin{pmatrix} r_{11,\infty} & 0 & \ldots & 0 \\ 0 & r_{22,\infty} & \ldots & 0 \\ \vdots & \ldots & \ddots & \vdots \\ 0 & \ldots & 0 & r_{II,\infty} \end{pmatrix} = \underline{\mathbf{Q}}_\infty^* (\mathcal{A}_0 \overline{\underline{\mathbf{q}}}_{i,\infty}^{N-2})\overline{\underline{\mathbf{Q}}}_\infty, \tag{7.3.7}$$

where $\mathbf{R}_\infty$ is a diagonal matrix due to the symmetry on the right. We observe that if $r_{ii,\infty}$ is not nonnegative, then we can derive a strategy to transform $\mathbf{R}_\infty$ to a positive semi-definite diagonal form. Any symmetric $\mathbf{A} \in \mathbb{C}^{I\times I}$ can be represented as $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, where $\mathbf{Q} \in \mathbb{C}^{I\times I}$ is unitary and $\mathbf{\Lambda} \in \mathbb{C}^{I\times I}$ is diagonal. If $\mathbf{\Lambda}$ is not positive semi-definite, then the Takagi factorization of $\mathbf{A}$ can be represented as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top$, where $\mathbf{\Sigma} = \mathbf{D}^{-1}\mathbf{\Lambda}\mathbf{D}^{-1}$ and $\mathbf{U} = \mathbf{P}\mathbf{D}$. The choice of the diagonal matrix $\mathbf{D} \in \mathbb{C}^{I\times I}$ is to ensure that the diagonal matrix $\mathbf{\Sigma}$ is positive semi-definite.

Suppose that $\mathbf{R}_\infty$ is a positive semi-definite diagonal matrix. If we pre-multiply $\underline{\mathbf{Q}}_\infty$ onto (7.3.7) and post-multiply with $\mathbf{e}_j$, then we have

$$\underline{\mathbf{q}}_{j,\infty} r_{jj,\infty} = (\mathcal{A}_0 \overline{\underline{\mathbf{q}}}_{i,\infty}^{N-2})\overline{\underline{\mathbf{q}}}_{j,\infty}$$

which leads to the US-eigenpair $(r_{ii,\infty}; \underline{\mathbf{q}}_{i,\infty})$ with $i = j$.

7.3.4 Convergence of Algorithm 7.3.3

Let $N = 2L$. For a given symmetric $\mathcal{A} \in CT_{N,I}$, we define

$$g(\mathbf{z}) = \mathcal{A}\overline{\mathbf{z}}^N = (\underbrace{\mathbf{z} \otimes \mathbf{z} \otimes \cdots \otimes \mathbf{z}}_{K})^* \mathbf{A} \overline{(\underbrace{\mathbf{z} \otimes \mathbf{z} \otimes \cdots \otimes \mathbf{z}}_{K})},$$

where a_{ij} in $\mathbf{A}$ are given by $a_{ij} = a_{i_1 i_2 \ldots i_l j_1 j_2 \ldots j_l}$, with

$$\begin{cases} i = I^{K-1}(i_1 - 1) + \cdots + I(i_{K-1} - 1) + i_K, & 1 \le i_1, i_2, \ldots, i_K \le I, \\ j = I^{K-1}(j_1 - 1) + \cdots + I(j_{K-1} - 1) + j_K, & 1 \le j_1, j_2, \ldots, j_K \le I. \end{cases}$$

Here $\mathbf{z} \in \mathbb{C}^I$ satisfies $\|\mathbf{z}\|_2 = 1$.

If there exists a unitary vector $\mathbf{z}_0 \in \mathbb{C}^I$ such that $g(\mathbf{z}_0)$ is complex, where the nonzero $\theta_0 \in (-\pi, \pi]$ is its argument, let $\widetilde{\mathbf{z}}_0 = e^{\iota\theta_0/N}\mathbf{z}_0$. Then $g(\widetilde{\mathbf{z}}_0)$ is nonnegative. Hence we can assume that $g(\mathbf{z})$ is nonnegative for all unit vectors $\mathbf{z} \in \mathbb{C}^I$.

Let $\mathbf{z} = \mathbf{x} + \iota\mathbf{y}$ where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$. Then $g(\mathbf{z})$ can be viewed as a real bivariate function of its real and imaginary parts [11, 12]. An auxiliary function of g is defined as follows: $f(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ where $f(\mathbf{x}, \mathbf{y}) = g(\mathbf{z})$ and $\mathbf{z} = \mathbf{x} + \iota\mathbf{y}$. Furthermore, a useful mapping $\mathbf{w} = \varphi(\mathbf{z})$ is given by $(\Re(\mathbf{z}), \Im(\mathbf{z})) \in \mathbb{R}^{2I}$, where $\Re(\mathbf{z})$ and $\Im(\mathbf{z})$ return the real and imaginary parts of $\mathbf{z}$, respectively. The mapping $\varphi : \mathbb{C}^I \to \mathbb{R}^{2I}$ is invertible. Substituting $\mathbf{z} = \varphi^{-1}(\mathbf{w})$ into $g(\mathbf{z})$, we have

$$f(\mathbf{w}) = g\left(\varphi^{-1}(\mathbf{w})\right) : \mathbb{R}^{2I} \to \mathbb{R}.$$

Let $g(\mathbf{z})$ be a function whose values are real or $\pm\infty$ and whose domain is a convex subset $\mathcal{S}$ of $\mathbb{C}^I$, and the set $\mathcal{S}_1 \in \mathbb{R}^{2I}$ is the range of φ with domain $\mathcal{S}$. It is obvious that $\mathcal{S}$ is a convex subset in $\mathbb{C}^I$ if and only if $\mathcal{S}_1$ is a convex subset in $\mathbb{R}^{2I}$. Then $f(\mathbf{w})$ is said to be convex on $\mathcal{S}_1$, if the epigraph of $f(\mathbf{w})$:

$$\mathrm{epi}(f) = \{(\mathbf{u}, v) : \mathbf{u} \in \mathcal{S}_1, v \in \mathbb{R}, v \ge f(\mathbf{u})\}$$

is a convex subset of $\mathbb{R}^{2I+1}$.

Theorem 7.3.3 (Convergence of Algorithm 7.3.3) *For any symmetric $\mathcal{A} \in CT_{N,I}$ with an even N, if the associate function $g(\mathbf{z})$ is convex (concave) on $\mathbb{C}^I$, then Algorithm 7.3.3 converges to a local maximum (minimum) of the restriction of g to the unit sphere $\Sigma = \{\mathbf{z} \in \mathbb{C}^I : \|\mathbf{z}\|_2 = 1\}$, for any initialization, except for saddle points or crest lines leading to such saddle points.*

Proof First consider the case that $g(\mathbf{z})$ is convex on the convex set $\mathcal{S} \in \mathbb{C}^I$. According to [13, Proposition 1], $f(\mathbf{w})$ is convex on $\mathcal{S}_1$. This implies that $\mathrm{epi}(f)$ is a convex subset of $\mathbb{R}^{2I+1}$. Hence a tangent hyperplane at any point $(\mathbf{w}, f(\mathbf{w}))$ is

a supporting hyperplane of epi(f). By the proof of [14, Theorem 4], we have

$$f(\varphi(\mathbf{z}_{k+1})) \geq f(\varphi(\mathbf{z}_k)), \quad \mathbf{z}_k, \mathbf{z}_{k+1} \in \Sigma,$$

where $\mathbf{z}_k$ and $\mathbf{z}_{k+1}$ are derived from Algorithm 7.3.3 with some initial vectors $\mathbf{z}_0 \in \Sigma$. Hence we have $g(\mathbf{z}_k) \geq g(\mathbf{z}_{k+1})$. The convergence follows from the fact that the restriction of g to Σ is bounded from above, namely, $\frac{g(\mathbf{z})}{\|\mathbf{z}\|_2^N} = \frac{(\mathbf{z}\otimes\mathbf{z}\otimes\cdots\otimes\mathbf{z})^* \mathbf{A} \overline{(\mathbf{z}\otimes\mathbf{z}\otimes\cdots\otimes\mathbf{z})}}{\|\mathbf{z}\|_2^N} \leq \sigma_1$, where σ_1 is the largest Takagi value of $\mathbf{A}$. The case of g being concave can be treated similarly, by replacing g with $-g$. □

7.4 Iterative Algorithms for U-Eigenpairs

Suppose that $\mathcal{A} \in CT_{N,I}$ and the imaginary part of $a_{i_1 i_2 \ldots i_N}$ is not zero for all i_n and n. We investigate the properties of U-eigenpairs of $\mathcal{A}$, design an iterative algorithm for their computation and analyze its convergence.

7.4.1 Properties of U-Eigenpairs

A basic property of a U-eigenpair of $\mathcal{A} \in CT_{N,I}$ is given in the following lemma.

Lemma 7.4.1 *Let $\mathcal{A} \in CT_{N,I}$ and $\theta \in (-\pi, \pi]$. If $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ is a* U*-eigenpair of $\mathcal{A}$, then there exists N scalars $\varphi_n \in (-\pi, \pi]$ such that*

$$(e^{\iota\theta}\sigma; e^{-\iota\varphi_1}\mathbf{v}_1, \ldots, e^{-\iota\varphi_N}\mathbf{v}_N), \quad (e^{-\iota\theta}\sigma; e^{\iota\varphi_1}\mathbf{v}_1, e^{\iota\varphi_2}\mathbf{v}_2, \ldots, e^{\iota\varphi_N}\mathbf{v}_N)$$

are solutions of (7.1.2) with $\varphi_1 + \varphi_2 + \cdots + \varphi_N = \theta$. In particular, for all n, we can have all φ_n identical with $\varphi_n = \theta/N$.

Proof From (7.1.2), the lemma is proved. □

For $\mathcal{A} \in CT_{N,I}$, let $\mathbf{U}_n \in \mathbb{C}^{I\times I}$ be complex matrices for all n. Define $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_N \mathbf{U}_N \in CT_{N,I}$ with entries $b_{j_1 j_2 \ldots j_N} = \sum_{i_1=1}^{I} \sum_{i_2=1}^{I} \cdots \sum_{i_N=1}^{I} a_{i_1 i_2 \ldots i_N} u_{1,j_1 i_1} u_{2,j_2 i_2} \ldots u_{N,j_N i_N}$, where $u_{n,i_n j_n}$ is the (i_n, j_n)-entry of $\mathbf{U}_n$. By some tedious manipulations, we have the following theorem.

Theorem 7.4.1 *Assume that $\mathbf{U}_n \in \mathbb{C}^{I\times I}$ are nonsingular for all n and $\mathbf{V}_n = \mathbf{U}_n^{-1}$. If $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_N \mathbf{U}_N$, then $\mathcal{A} = \mathcal{B} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \cdots \times_N \mathbf{V}_N$.*

In particular, if $\mathbf{U}_n$ is unitary and $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}_1^* \times_2 \mathbf{U}_2^* \cdots \times_N \mathbf{U}_N^*$, then we say that $\mathcal{A}$ is similar to $\mathcal{B}$ through similarity transformations by N unitary matrices $\mathbf{U}_n$. The following theorem states the relationship between the U-eigenpairs of $\mathcal{A}$ and $\mathcal{B}$.

Theorem 7.4.2 *Let $\mathcal{A}, \mathcal{B} \in CT_{N,I}$. If $\mathcal{A}$ is similar to $\mathcal{B}$ through similarity transformations by N unitary matrices $\mathbf{U}_n \in \mathbb{C}^{I\times I}$, that is, $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}_1^* \times_2 \mathbf{U}_2^* \cdots \times_N \mathbf{U}_N^*$, then they have the same* U*-eigenvalues.*

Moreover, if $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ is a U*-eigenpair of $\mathcal{B}$ with $\|\mathbf{v}_n\|_2 = 1$ and $\sigma \geq 0$, then $(\sigma; \mathbf{U}_1\mathbf{v}_1, \mathbf{U}_2\mathbf{v}_2, \ldots, \mathbf{U}_N\mathbf{v}_N)$ is also a* U*-eigenpair of $\mathcal{A}$.*

Proof Since $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ is a U-eigenpair of $\mathcal{B}$, then for all n, we have

$$\begin{cases} \mathcal{B} \times_1 \mathbf{v}_1^* \cdots \times_{n-1} \mathbf{v}_{n-1}^* \times_{n+1} \mathbf{v}_{n+1}^* \cdots \times_N \mathbf{v}_N^* = \sigma \mathbf{v}_n, \\ \overline{\mathcal{B}} \times_1 \mathbf{v}_1^\top \cdots \times_{n-1} \mathbf{v}_{n-1}^\top \times_{n+1} \mathbf{v}_{n+1}^\top \cdots \times_N \mathbf{v}_N^\top = \sigma \overline{\mathbf{v}}_n. \end{cases}$$

According to the assumptions, we have

$$\begin{aligned} &\mathcal{B} \times_1 \mathbf{v}_1^* \cdots \times_{n-1} \mathbf{v}_{n-1}^* \times_{n+1} \mathbf{v}_{n+1}^* \cdots \times_N \mathbf{v}_N^* \\ &\quad = \mathcal{B} \times_1 \overline{\mathbf{v}}_1^\top \cdots \times_{n-1} \overline{\mathbf{v}}_{n-1}^\top \times_{n+1} \overline{\mathbf{v}}_{n+1}^\top \cdots \times_N \overline{\mathbf{v}}_N^\top \\ &\quad = \mathbf{U}_n^* \left(\mathcal{A} \times_1 (\overline{\mathbf{U}_1\mathbf{v}_1})^\top \cdots \times_{n-1} (\overline{\mathbf{U}_{n-1}\mathbf{v}_{n-1}})^\top \times_{n+1} (\overline{\mathbf{U}_{n+1}\mathbf{v}_{n+1}})^\top \cdots \times_N (\overline{\mathbf{U}_N\mathbf{v}_N})^\top \right). \end{aligned}$$

Since $\mathbf{U}_n \in \mathbb{C}^{I\times I}$ are unitary and $\|\mathbf{v}_n\|_2 = 1$ for all n, then $(\mathbf{U}_n\mathbf{v}_n)^*(\mathbf{U}_n\mathbf{v}_n) = \mathbf{v}_n^*(\mathbf{U}_n^*\mathbf{U}_n)\mathbf{v}_n = 1$. Moreover, we have

$$\begin{cases} \mathcal{A} \times_1 (\overline{\mathbf{U}_1\mathbf{v}_1})^\top \cdots \times_{n-1} (\overline{\mathbf{U}_{n-1}\mathbf{v}_{n-1}})^\top \times_{n+1} (\overline{\mathbf{U}_{n+1}\mathbf{v}_{n+1}})^\top \cdots \times_N (\overline{\mathbf{U}_N\mathbf{v}_N})^\top \\ \quad = \sigma(\mathbf{U}_n\mathbf{v}_n), \\ \overline{\mathcal{A}} \times_1 (\mathbf{U}_1\mathbf{v}_1)^\top \cdots \times_{n-1} (\mathbf{U}_{n-1}\mathbf{v}_{n-1})^\top \times_{n+1} (\mathbf{U}_{n+1}\mathbf{v}_{n+1})^\top \cdots \times_N (\mathbf{U}_N\mathbf{v}_N)^\top \\ \quad = \sigma(\overline{\mathbf{U}_n\mathbf{v}_n}). \end{cases}$$

Hence $(\sigma; \mathbf{U}_1\mathbf{v}_1, \mathbf{U}_2\mathbf{v}_2, \ldots, \mathbf{U}_N\mathbf{v}_N)$ is a U-eigenpair of $\mathcal{A}$. □

Let θ be the argument of $\widetilde{\mathcal{A}} \times_1 \mathbf{e}_1^\top \times_2 \mathbf{e}_1^\top \cdots \times_N \mathbf{e}_1^\top$, where $\mathbf{e}_1$ is the first column of $\mathbf{I}_I$. If σ is a U-eigenvalue of $\mathcal{B}$ and $e^{\iota\theta/N}\mathbf{e}_1$ is the mode-n U-eigenvector of $\mathcal{B}$ associated to σ, then $(\sigma, \mathbf{u}_{1,1}, \mathbf{u}_{2,1}, \ldots, \mathbf{u}_{N,1})$ is a U-eigenpair of $\mathcal{A}$, where $\mathbf{u}_{n,1}$ is the first column of $\mathbf{U}_n$. More generally, we have the following corollary from Theorem 7.4.2.

Corollary 7.4.1 *For $\mathcal{A} \in CT_{N,I}$. If $(\sigma, e^{\iota\theta/N}\mathbf{e}_i, e^{\iota\theta/N}\mathbf{e}_i, \ldots, e^{\iota\theta/N}\mathbf{e}_i)$ is a* U*-eigenpair of $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}_1^* \times_2 \mathbf{U}_2^* \cdots \times_N \mathbf{U}_N^*$, where $\mathbf{U}_n \in \mathbb{C}^{I\times I}$ are unitary, then $(\sigma; \mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \ldots, \mathbf{u}_{N,i})$ is a* U*-eigenpair of $\mathcal{A}$, where $\mathbf{u}_{n,i}$ and $\mathbf{e}_i$ is the i*th *column of $\mathbf{U}_n$ and $\mathbf{I}_I$, respectively, and $\theta \in (-\pi, \pi]$ is the argument of $\mathcal{B}\times_1 \mathbf{e}_i^\top \times_2 \mathbf{e}_i^\top \cdots \times_N \mathbf{e}_i^\top$.*

Algorithm 7.4.1 QRCT

Input: Given a $\mathcal{A} \in CT_{N,I}$, tolerance τ, maximum number of iteration $K_{\max}$
Output: U-eigenpairs $(\sigma_i; \mathbf{p}_{1,i}, \mathbf{p}_{2,i}, \ldots, \mathbf{p}_{N,i})$
for $i = 1, 2, \ldots, I$ **do**
Set $\mathcal{A}_0 := \mathcal{A}$ and compute $\epsilon_n \leftarrow \|F_0(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n}$
$- \mathbf{e}_{n,i}\|_2 / \|F_0(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n,-(n+1)}\|_2$
Set $\mathbf{Q}_n \leftarrow \mathbf{I}_I$ and $\epsilon \leftarrow \max_{1 \le n \le N} \epsilon_n$
iteration count $\zeta \leftarrow 1$
while $\epsilon > \tau$ or $\zeta \le K_{\max}$ **do**
Step 1: factorize $\mathbf{Q}_{n,k}\mathbf{R}_{n,k} = F_{k-1}(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n,-(n+1)}$
Step 2: compute

$$\mathcal{A}_k = \mathcal{A}_{k-1} \times_1 \mathbf{Q}_{1,k}^* \times_2 \mathbf{Q}_{2,k}^* \cdots \times_N \mathbf{Q}_{N,k}^*$$

$$\epsilon_n \leftarrow \|F_k(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n} - \mathbf{e}_i\|_2 / \|F_k(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n,-(n+1)}\|_2$$

Compute $\mathbf{Q}_n \leftarrow \mathbf{Q}_n\mathbf{Q}_{n,k}$ and $\epsilon \leftarrow \max_{1 \le n \le N} \epsilon_n$
end while
end for
if converged **then**
$\sigma_i \leftarrow |F_\zeta(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)|$, where θ is the argument of $F_\zeta(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)$
$\mathbf{q}_i \leftarrow e^{\iota\theta/N}\mathbf{Q}_n(:, i)$, where $\mathbf{Q}_n(:, i)$ is the ith column of $\mathbf{Q}_n$
end if

7.4.2 QR Algorithms for Complex Tensors

In Sect. 7.3, we propose an iterative algorithm for the US-eigenpairs of complex symmetric tensors based on the Takagi factorization and Theorem 7.3.1. In this section, we present an iterative algorithm for computing the U-eigenpairs of general complex tensors based on the Takagi factorization and Corollary 7.4.1.

We present the QR-type algorithm for computing the Takagi factorization of general complex matrices in Sect. 7.2. The QR-type algorithm for general complex matrices starts from an initial complex matrix $\mathbf{A}_0 \in \mathbb{C}^{I \times I}$, and computes two QR factorizations $\mathbf{U}_1\mathbf{R}_1 = \mathbf{A}_0$ and $\mathbf{V}_1\mathbf{T}_1 = \mathbf{A}_0^\top$. The QR-type product is then reversed to give $\mathbf{A}_1 = \mathbf{U}_1^*\mathbf{A}_0\overline{\mathbf{V}}_1$. In subsequent iterations, we get $\mathbf{U}_k\mathbf{R}_k = \mathbf{A}_{k-1}$, $\mathbf{V}_k\mathbf{T}_k = \mathbf{A}_{k-1}^\top$ and

$$\mathbf{A}_k = \mathbf{U}_k^*\mathbf{A}_{k-1}\overline{\mathbf{V}}_k = (\mathbf{U}_1\mathbf{U}_2 \ldots \mathbf{U}_k)^*\mathbf{A}_0\overline{(\mathbf{V}_1\mathbf{V}_2 \ldots \mathbf{V}_k)} = \underline{\mathbf{U}}_k^*\mathbf{A}_0\overline{\underline{\mathbf{V}}}_k,$$

where $\underline{\mathbf{U}}_k = \mathbf{U}_1\mathbf{U}_2 \ldots \mathbf{U}_k$ and $\underline{\mathbf{V}}_k = \mathbf{V}_1\mathbf{V}_2 \ldots \mathbf{V}_k$.

In the following, we present a QRCT algorithm for U-eigenpairs of general complex tensors. Moreover, with a permutation strategy to scramble tensor entries, a more efficient version of the QRCT is proposed to compute more U-eigenpairs. The core of the QRCT consists of Steps 1 and 2, for all ith mode-(m, n) slices with all i and $m < n$.

Remark 7.4.1 In Algorithm 7.4.1, we define

$$\begin{cases} F_k(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i) = \mathcal{A}_k \times_1 \mathbf{e}_i^\top \times_2 \mathbf{e}_i^\top \cdots \times_N \mathbf{e}_i^\top, \\ F_k(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n} = \mathcal{A}_k \times_1 \mathbf{e}_i^\top \cdots \times_{n-1} \mathbf{e}_i^\top \times_{n+1} \mathbf{e}_i^\top \cdots \times_N \mathbf{e}_i^\top, \\ F_k(\mathbf{e}_i, \mathbf{e}_i, \ldots, \mathbf{e}_i)_{-n,-(n+1)} = \mathcal{A}_k \times_1 \mathbf{e}_i^\top \cdots \times_{n-1} \mathbf{e}_i^\top \times_{n+2} \mathbf{e}_i^\top \cdots \times_N \mathbf{e}_i^\top, \end{cases}$$

and when $k = N$, we replace $k + 1$ by 1.

For $\mathcal{A} \in CT_{N,I}$, the QRCT produces at most I U-eigenpairs in one pass. According to Theorem 7.4.2, a more efficient version of the QRCT can be devised, by scrambling the entries in $\mathcal{A}_0$, as summarized below.

Algorithm 7.4.2 Permuted QRCT

Input: Without loss of generality, let $\mathcal{A}_0 \in CT_{N,I}$ with $N = 3$.
Output: U-eigenpairs $(\sigma_i; \mathbf{q}_{1,i}, \mathbf{q}_{2,i}, \mathbf{q}_{3,i})$
for $p_1 = 1, 2, \ldots, \bar{P}_1$ **do**
  for $p_2 = 1, 2, \ldots, \bar{P}_2$ **do**
    for $p_3 = 1, 2, \ldots, \bar{P}_3$ **do**
      Choose $\mathbf{P}_{p_i} \in \mathfrak{S}$ and compute $\mathcal{A}_{p_1,p_2,p_3} \leftarrow \mathcal{A}_0 \times_1 \mathbf{P}_1^\top \times_2 \mathbf{P}_2^\top \times_3 \mathbf{P}_3^\top$
      Apply Algorithm 7.4.1 with $\mathcal{A}_{p_1,p_2,p_3}$ and collect distinct converged U-eigenpairs
    end for
  end for
end for

Again, the Permuted QRCT quickly becomes impractical for large I and N, since we need to preform QRCST $(I!)^N$ times in one pass.

7.4.3 Convergence

We consider the convergence of Algorithm 7.4.1. We generalize Algorithm 7.3.3, which can be used to compute the US-eigenpairs of complex symmetric tensors and the U-eigenpairs of complex tensors.

Without loss of generality, let $\mathcal{A} \in CT_{3,I}$. For $n = 1, 2, 3$, we define $\underline{\mathbf{Q}}_{n,k} = \mathbf{Q}_{n,1}\mathbf{Q}_{n,2}\ldots\mathbf{Q}_{n,k}$, with $\underline{\mathbf{Q}}_{n,0} = \mathbf{I}_I$ and $\underline{\mathbf{Q}}_{n,1} = \mathbf{Q}_{n,1}$. We also use $\underline{\mathbf{q}}_{k,i,\zeta}$ to denote the ith column in $\underline{\mathbf{Q}}_{n,k}$ and $r_{n,ij,k}$ for the (i, j)-entry of $\mathbf{R}_{n,k}$.

A few iterations of Algorithm 7.4.1 can be visualized as follows:

$$\begin{cases} \mathbf{Q}_{1,1}\mathbf{R}_{1,1} = \mathcal{A}_0 \times_3 \mathbf{e}_i^\top, \quad \mathbf{Q}_{2,1}\mathbf{R}_{2,1} = \mathcal{A}_0 \times_1 \mathbf{e}_i^\top, \quad \mathbf{Q}_{3,1}\mathbf{R}_{3,1} = (\mathcal{A}_0 \times_2 \mathbf{e}_i^\top)^\top, \\ \quad \mathcal{A}_1 = \mathcal{A}_0 \times_1 \mathbf{Q}_{1,1}^* \times_2 \mathbf{Q}_{2,1}^* \times_3 \mathbf{Q}_{3,1}^* = \mathcal{A}_0 \times_1 \underline{\mathbf{Q}}_{1,1}^* \times_2 \underline{\mathbf{Q}}_{2,1}^* \times_3 \underline{\mathbf{Q}}_{3,1}^*, \\ \mathbf{Q}_{1,2}\mathbf{R}_{1,2} = \mathcal{A}_1 \times_3 \mathbf{e}_i^\top = \underline{\mathbf{Q}}_{1,1}^* \mathcal{A}_0 \times_3 \underline{\mathbf{q}}_{3,i,1}^* \overline{\underline{\mathbf{Q}}}_{2,1} = \underline{\mathbf{Q}}_{1,1}^* \mathcal{A}_0 \times_3 \overline{\underline{\mathbf{q}}}_{3,i,1}^\top \overline{\underline{\mathbf{Q}}}_{2,1}, \\ \mathbf{Q}_{2,2}\mathbf{R}_{2,2} = \mathcal{A}_1 \times_1 \mathbf{e}_i^\top = \underline{\mathbf{Q}}_{2,1}^* \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,i,1}^* \overline{\underline{\mathbf{Q}}}_{3,1} = \underline{\mathbf{Q}}_{2,1}^* \mathcal{A}_0 \times_1 \overline{\underline{\mathbf{q}}}_{1,i,1}^\top \overline{\underline{\mathbf{Q}}}_{3,1}, \\ \mathbf{Q}_{3,2}\mathbf{R}_{3,2} = (\mathcal{A}_1 \times_2 \mathbf{e}_i^\top)^\top = (\underline{\mathbf{Q}}_{1,1}^* (\mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,i,1}^*) \underline{\mathbf{Q}}_{3,1})^\top = (\underline{\mathbf{Q}}_{1,1}^* \mathcal{A}_0 \times_2 \overline{\underline{\mathbf{q}}}_{2,i,1}^\top \overline{\underline{\mathbf{Q}}}_{3,1})^\top, \\ \quad \mathcal{A}_2 = \mathcal{A}_1 \times_1 \mathbf{Q}_{1,2}^* \times_2 \mathbf{Q}_{2,2}^* \times_3 \mathbf{Q}_{3,2}^* = \mathcal{A}_0 \times_1 \underline{\mathbf{Q}}_{1,2}^* \times_2 \underline{\mathbf{Q}}_{2,2}^* \times_3 \underline{\mathbf{Q}}_{3,2}^* \\ \mathbf{Q}_{1,3}\mathbf{R}_{1,3} = \mathcal{A}_2 \times_3 \mathbf{e}_i^\top = \underline{\mathbf{Q}}_{1,2}^* \mathcal{A}_0 \times_3 \underline{\mathbf{q}}_{3,i,2}^* \overline{\underline{\mathbf{Q}}}_{2,2} = \underline{\mathbf{Q}}_{1,2}^* \mathcal{A}_0 \times_3 \overline{\underline{\mathbf{q}}}_{3,i,2}^\top \overline{\underline{\mathbf{Q}}}_{2,2}, \\ \mathbf{Q}_{2,3}\mathbf{R}_{2,3} = \mathcal{A}_2 \times_1 \mathbf{e}_i^\top = \underline{\mathbf{Q}}_{2,2}^* \mathcal{A}_0 \bar{\times}_1 \underline{\mathbf{q}}_{1,i,2}^* \overline{\underline{\mathbf{Q}}}_{3,2} = \underline{\mathbf{Q}}_{2,2}^* \mathcal{A}_0 \bar{\times}_1 \overline{\underline{\mathbf{q}}}_{1,i,2}^\top \overline{\underline{\mathbf{Q}}}_{3,2}, \\ \mathbf{Q}_{3,3}\mathbf{R}_{3,3} = (\mathcal{A}_2 \times_2 \mathbf{e}_i^\top)^\top = (\underline{\mathbf{Q}}_{1,2}^* (\mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,i,2}^*) \underline{\mathbf{Q}}_{3,2})^\top = (\underline{\mathbf{Q}}_{1,2}^* \mathcal{A}_0 \times_2 \overline{\underline{\mathbf{q}}}_{2,i,2}^\top \overline{\underline{\mathbf{Q}}}_{3,2})^\top. \end{cases} \tag{7.4.1}$$

From (7.4.1), we see that in general for $k = 1, 2, \ldots,$

$$\begin{cases} \quad \mathcal{A}_{k-1} = \mathcal{A}_0 \times_1 \underline{\mathbf{Q}}_{1,k-1}^* \times_2 \underline{\mathbf{Q}}_{2,it-1}^* \times_3 \underline{\mathbf{Q}}_{3,k-1}^*, \\ \mathbf{Q}_{1,k}\mathbf{R}_{1,k} = \underline{\mathbf{Q}}_{1,k-1}^* \mathcal{A}_0 \times_3 \underline{\mathbf{q}}_{3,i,k-1}^* \overline{\underline{\mathbf{Q}}}_{2,k-1}, \\ \mathbf{Q}_{2,k}\mathbf{R}_{2,k} = \underline{\mathbf{Q}}_{2,k-1}^* \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,i,k-1}^* \overline{\underline{\mathbf{Q}}}_{3,k-1}, \\ \mathbf{Q}_{3,k}\mathbf{R}_{3,k} = \left(\underline{\mathbf{Q}}_{1,k-1}^* \mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,i,k-1}^* \overline{\underline{\mathbf{Q}}}_{3,k-1}\right)^\top. \end{cases} \tag{7.4.2}$$

Multiplying $\underline{\mathbf{Q}}_{1,k-1}$, $\underline{\mathbf{Q}}_{2,k-1}$, and $\underline{\mathbf{Q}}_{3,k-1}$ to the both sides of (7.4.2), we obtain

$$\begin{cases} \underline{\mathbf{Q}}_{1,k} R_{1,k} = \mathcal{A}_0 \times_3 \underline{\mathbf{q}}_{3,i,k-1}^* \overline{\underline{\mathbf{Q}}}_{2,k-1}, \quad \underline{\mathbf{Q}}_{2,k} R_{2,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,i,k-1}^* \overline{\underline{\mathbf{Q}}}_{3,it-1}, \\ \underline{\mathbf{Q}}_{3,k} R_{3,k} = (\mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,i,k-1}^*)^\top \overline{\underline{\mathbf{Q}}}_{1,k-1}. \end{cases} \tag{7.4.3}$$

It is obvious that (7.4.3) describes a tensor generalization of the matrix orthogonal iteration for finding the dominant invariant left and right Takagi subspaces of matrices.

If we set $i = 1$ and extract the first columns of (7.4.3), then we have

$$\begin{cases} \underline{\mathbf{q}}_{1,1,k} r_{1,11,k} = (\mathcal{A}_0 \times_3 \underline{\mathbf{q}}_{3,1,k-1}^*) \overline{\underline{\mathbf{q}}}_{2,1,k-1} = \mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,1,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^*, \\ \underline{\mathbf{q}}_{2,1,k} r_{2,11,k} = (\mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,1,k-1}^*) \overline{\underline{\mathbf{q}}}_{3,1,k-1} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,i,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^*, \\ \underline{\mathbf{q}}_{3,1,k} r_{3,11,k} = (\mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,1,k-1}^*)^\top \overline{\underline{\mathbf{q}}}_{1,1,k-1} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,1,k-1}^* \times_2 \underline{\mathbf{q}}_{2,1,k-1}^*, \end{cases}$$

or,

$$\begin{cases} \underline{\mathbf{q}}_{1,1,k} r_{1,11,k} = \mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,1,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^* / \left\| \mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,1,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^* \right\|_2, \\ \underline{\mathbf{q}}_{2,1,k} r_{2,11,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,i,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^* / \left\| \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,i,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^* \right\|_2, \\ \underline{\mathbf{q}}_{3,1,k} r_{3,11,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,1,k-1}^* \times_2 \underline{\mathbf{q}}_{2,1,k-1}^* / \left\| \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,1,k-1}^* \times_2 \underline{\mathbf{q}}_{2,1,k-1}^* \right\|_2, \end{cases}$$

which is just the higher order power method for computing the U-eigenpairs of general complex tensors. This algorithm is summarized in Algorithm 7.4.3. With the complex symmetric embedding and Theorem 7.3.3, we shall analyze its convergence.

Algorithm 7.4.3 Higher order power method for complex tensors

Input: Given a $\mathcal{A} \in CT_{N,I}$
Output: a U-eigenpair $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$
Choose initial vectors $\mathbf{x}_{n,0} \in \mathbb{C}^I$ with $\|\mathbf{x}_{n,0}\|_2 = 1$
for $k = 0, 1, 2, \ldots$ **do**
  Compute $\mathbf{y}_{n,k+1} = F(\overline{\mathbf{x}_{1,k}}, \overline{\mathbf{x}_{2,k}}, \ldots, \overline{\mathbf{x}_{N,k}})_{-n}$
  Normalize $\mathbf{x}_{n,k+1} = \mathbf{y}_{n,k+1} / \|\mathbf{y}_{n,k+1}\|_2$
end for
Compute $\lambda_\infty = F(\overline{\mathbf{x}_{1,\infty}}, \overline{\mathbf{x}_{2,\infty}}, \ldots, \overline{\mathbf{x}_{N,\infty}})$
Set $\sigma_\infty = |\lambda_\infty|$ and $\mathbf{v}_{n,\infty} = \exp\left(\iota \frac{\theta}{N}\right) \mathbf{x}_{n,\infty}$, where $\theta \in (\pi, \pi]$ is the argument of λ_∞

If we start Algorithm 7.4.1 with $i = 2$, then the first two columns of (7.4.2) proceed as

$$\begin{cases} \underline{\mathbf{q}}_{1,1,k} r_{1,11,k} = \mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,1,k-1}^* \times_3 \underline{\mathbf{q}}_{3,2,k-1}^*, \\ \underline{\mathbf{q}}_{1,1,k} r_{1,12,k} + \underline{\mathbf{q}}_{1,2,k} r_{1,22,k} = \mathcal{A}_0 \times_2 \underline{\mathbf{q}}_{2,2,k-1}^* \times_3 \underline{\mathbf{q}}_{3,2,k-1}^*, \\ \underline{\mathbf{q}}_{2,1,k} r_{2,11,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,2,k-1}^* \times_3 \underline{\mathbf{q}}_{3,1,k-1}^*, \\ \underline{\mathbf{q}}_{2,1,k} r_{2,12,k} + \underline{\mathbf{q}}_{2,2,k} r_{2,22,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,2,k-1}^* \times_3 \underline{\mathbf{q}}_{3,2,k-1}^*, \\ \underline{\mathbf{q}}_{3,1,k} r_{3,11,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,1,k-1} \times_2 \underline{\mathbf{q}}_{2,2,k-1}^*, \\ \underline{\mathbf{q}}_{3,1,k} r_{3,12,k} + \underline{\mathbf{q}}_{3,2,k} r_{3,22,k} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,2,k-1}^* \times_2 \underline{\mathbf{q}}_{2,2,k-1}^*. \end{cases}$$

Assuming convergence at $k = \infty$, we have

$$\begin{cases} \underline{\mathbf{q}}_{1,1,\infty} r_{1,11,\infty} = \mathcal{A}_0 \times_2\underline{\mathbf{q}}^*_{2,1,\infty} \times_3 \underline{\mathbf{q}}^*_{3,2,\infty}, \\ \underline{\mathbf{q}}_{1,1,\infty} r_{1,12,\infty} + \underline{\mathbf{q}}_{1,2,\infty} r_{1,22,\infty} = \mathcal{A}_0 \times_2 \underline{\mathbf{q}}^*_{2,2,\infty} \times_3 \underline{\mathbf{q}}^*_{3,2,\infty}, \\ \underline{\mathbf{q}}_{2,1,\infty} r_{2,11,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,2,\infty} \times_3 \underline{\mathbf{q}}^*_{3,1,\infty}, \\ \underline{\mathbf{q}}_{2,1,\infty} r_{2,12,\infty} + \underline{\mathbf{q}}_{2,2,\infty} r_{2,22,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,2,\infty} \times_3 \underline{\mathbf{q}}^*_{3,2,\infty}, \\ \underline{\mathbf{q}}_{3,1,\infty} r_{3,11,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}_{1,1,\infty} \times_2 \underline{\mathbf{q}}^*_{2,2,\infty}, \\ \underline{\mathbf{q}}_{3,1,\infty} r_{3,12,\infty} + \underline{\mathbf{q}}_{3,2,\infty} r_{3,22,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,2,\infty} \times_2 \underline{\mathbf{q}}^*_{2,2,\infty}. \end{cases} \tag{7.4.4}$$

Due to (7.4.4), it is easy to check that

$$\begin{cases} r_{1,12,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,1,\infty} \times_2 \underline{\mathbf{q}}^*_{2,2,\infty} \times_3 \underline{\mathbf{q}}^*_{3,2,\infty} = 0, \\ r_{2,12,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,2,\infty} \times_2 \underline{\mathbf{q}}^*_{2,1,\infty} \times_3 \underline{\mathbf{q}}^*_{3,2,\infty} = 0, \\ r_{3,12,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,2,\infty} \times_2 \underline{\mathbf{q}}^*_{2,2,\infty} \times_3 \underline{\mathbf{q}}^*_{3,1,\infty} = 0, \end{cases} \tag{7.4.5}$$

and

$$r_{k,22,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,2,\infty} \times_2 \underline{\mathbf{q}}^*_{2,2,\infty} \times_3 \underline{\mathbf{q}}^*_{3,2,\infty}, \quad n = 1, 2, 3.$$

For all i, if $\underline{\mathbf{Q}}_{n,k}$ and $\mathbf{R}_{n,k}$ converge to $\underline{\mathbf{Q}}_{n,\infty}$ and $\mathbf{R}_{n,\infty}$, respectively, then from (7.4.2), we have

$$\begin{aligned} \mathbf{R}_{1,\infty} &= \begin{pmatrix} r_{1,11,\infty} & 0 & \dots & 0 \\ 0 & r_{1,22,\infty} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_{1,II,\infty} \end{pmatrix} = \underline{\mathbf{Q}}^*_{1,\infty} \left(\mathcal{A}_0 \times_3 \underline{\mathbf{q}}^*_{3,i,\infty} \right) \overline{\underline{\mathbf{Q}}}_{2,\infty}, \\ \mathbf{R}_{2,\infty} &= \begin{pmatrix} r_{2,11,\infty} & 0 & \dots & 0 \\ 0 & r_{2,22,\infty} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_{2,II,\infty} \end{pmatrix} = \underline{\mathbf{Q}}^*_{2,\infty} \left(\mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,i,\infty} \right) \overline{\underline{\mathbf{Q}}}_{3,\infty}, \\ \mathbf{R}_{3,\infty} &= \begin{pmatrix} r_{3,11,\infty} & 0 & \dots & 0 \\ 0 & r_{3,22,\infty} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_{3,II,\infty} \end{pmatrix} = \underline{\mathbf{Q}}^*_{3,\infty} \left(\mathcal{A}_0 \times_2 \underline{\mathbf{q}}^*_{2,i,\infty} \right)^\top \overline{\underline{\mathbf{Q}}}_{1,\infty}. \end{aligned} \tag{7.4.6}$$

where $\mathbf{R}_{n,\infty}$ is diagonal due to the Takagi factorization of a complex matrix. We do not know whether all diagonal entries in $\mathbf{R}_{n,\infty}$ are nonnegative. However, we can convert all $\mathbf{R}_{n,\infty}$ to diagonal forms through the following strategy: let $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}^\top \in \mathbb{C}^{I\times I}$, where $\mathbf{P}, \mathbf{Q} \in \mathbb{C}^{I\times I}$ are unitary matrices and $\Lambda \in \mathbb{C}^{I\times I}$ is diagonal. If Λ is not positive semi-definite, then the Takagi factorization is $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where

$$\mathbf{\Sigma} = \mathbf{D}^{-1}\mathbf{\Lambda}\mathbf{D}^{-1}, \quad \mathbf{U} = \mathbf{P}\mathbf{D}, \quad \mathbf{V} = \mathbf{Q}\mathbf{D},$$

and the diagonal $\mathbf{D} \in \mathbb{C}^{I\times I}$ is unitary with a positive semi-definite diagonal matrix Σ.

According to (7.4.5), we see that all $\mathbf{R}_{n,\infty}$ are identical, denoted by $\mathbf{R}_\infty$. For each n, if we pre-multiply (7.4.6) with $\underline{\mathbf{Q}}_{n,\infty}$ and post-multiply it with $\mathbf{e}_j$, then we have

$$\begin{cases} \underline{\mathbf{q}}_{1,j,\infty} r_{jj,\infty} = \mathcal{A}_0 \times_2 \underline{\mathbf{q}}^*_{2,j,\infty} \times_3 \underline{\mathbf{q}}^*_{3,i,\infty}, \\ \underline{\mathbf{q}}_{2,j,\infty} r_{jj,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,i,\infty} \times_3 \underline{\mathbf{q}}^*_{3,j,\infty}, \\ \underline{\mathbf{q}}_{3,j,\infty} r_{jj,\infty} = \mathcal{A}_0 \times_1 \underline{\mathbf{q}}^*_{1,j,\infty} \times_2 \underline{\mathbf{q}}^*_{2,i,\infty}. \end{cases}$$

which leads to the U-eigenpair $(r_{ii,\infty}; \underline{\mathbf{q}}_{1,i,\infty}, \underline{\mathbf{q}}_{2,i,\infty}, \underline{\mathbf{q}}_{3,i,\infty})$ when $i = j$. For a more general $\mathcal{A} \in CT_{N,I}$, we can prove the convergence of Algorithm 7.4.1 similarly.

We can also compute the U-eigenpairs of $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$ by a slightly modified version of Algorithm 7.4.1. The process is similar, so we omit its details. Note that by modifying Algorithm 7.4.1, we can produce at most $\min\{I_1, I_2, \ldots, I_N\}$ U-eigenpairs in one pass.

7.5 Special Case: Real Symmetric Tensors

In the previous two sections, we propose the QRCST algorithm for the US-eigenpairs and the QRCT algorithm for the U-eigenpairs, if the imaginary parts of the entries of the target tensors are nonzero. We consider how to apply Algorithm 7.3.1 for computing the US- or U-eigenpairs of $\mathcal{A} \in RT_{N,I}$.

Let $(\sigma; \mathbf{v})$ be a US-eigenpair of a symmetric $\mathcal{A} \in RT_{N,I}$, then

$$\mathcal{A}\overline{\mathbf{v}}^{N-1} = \sigma\mathbf{v}, \quad \overline{\mathcal{A}}\mathbf{v}^{N-1} = \sigma\overline{\mathbf{v}}, \quad \|\mathbf{v}\|_2 = 1.$$

Since the argument of $\iota = \sqrt{-1}$ is $\pi/4$, then $(\sigma; e^{-\iota\pi/4}\mathbf{v})$ is also a US-eigenpair of $\iota\mathcal{A}$, where the imaginary part of the entries of $\iota\mathcal{A}$ is nonzero. For the US-eigenpairs of $\mathcal{A}$, we apply Algorithm 7.3.1 or 7.3.2 to the complex symmetric tensor $\iota\mathcal{A}$.

Similarly, if $(\sigma; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)$ is a U-eigenpair of $\mathcal{A} \in RT_{N,I}$, that is,

$$\begin{cases} F(\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_N)_{-n} = \sigma \mathbf{v}_n, \\ \overline{F}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)_{-n} = \sigma \overline{\mathbf{v}}_n, \quad \|\mathbf{v}_n\|_2 = 1, \end{cases}$$

then $(\sigma; e^{-\iota\theta_1}\mathbf{v}_1, e^{-\iota\theta_2}\mathbf{v}_2, \ldots, e^{-\iota\theta_N}\mathbf{v}_N)$ is also a U-eigenpair of $\iota\mathcal{A}$ with $\sum_{n=1}^{N} \theta_n = \pi/2$. For the US-eigenpairs of $\mathcal{A}$, we need to apply Algorithm 7.4.1 or 7.4.2 to $\iota\mathcal{A}$.

7.6 Numerical Examples

We choose the accuracy tolerance $\tau = 10^{-10}$ and include examples for US- and U-eigenpairs. Note that $\|\mathbf{x}\|_2 = \|\overline{\mathbf{x}}\|_2$ for any vector $\mathbf{x} \in \mathbb{C}^I$. If $\mathcal{A} \in CT_{N,I}$ is symmetric, then

$$\mathrm{err}_k := \left\| \mathcal{A}\overline{\mathbf{x}}_k^{N-1} - \lambda_k \mathbf{x}_k \right\|_2 < \tau$$

or

$$\mathrm{err}_k := \left\| \mathcal{A}\overline{\mathbf{v}}_k^{N-1} - \sigma_k \mathbf{v}_k \right\|_2 < \tau$$

as the convergence criterion in Algorithm 7.3.3. Meanwhile, when $\mathcal{A} \in CT_{N,I}$, for all n, we define

$$\mathrm{err}_{n,k} := \left\| F(\overline{\mathbf{x}}_{1,k}, \overline{\mathbf{x}}_{2,k}, \ldots, \overline{\mathbf{x}}_{N,k})_{-n} - \lambda_n \mathbf{x}_{n,k} \right\|_2,$$

or

$$\mathrm{err}_{n,k} := \left\| F(\overline{\mathbf{v}}_{1,k}, \overline{\mathbf{v}}_{2,k}, \ldots, \overline{\mathbf{v}}_{N,k})_{-n} - \sigma_k \mathbf{v}_{n,k} \right\|_2,$$

Then we select

$$\mathrm{err}_k := \max_{n=1,2,\ldots,N} \mathrm{err}_{n,k} < \tau$$

as the convergence criterion of Algorithm 7.4.3.

Next, we give an example for computing some US-eigenvalues of two given complex symmetric tensors.

Example 7.6.1 The symmetric $\mathcal{A}, \mathcal{B} \in CT_{3,2}$ are selected from [1] with

$$\begin{cases} a_{111} = 2,\ a_{112} = \ \ 1,\ a_{122} = -1,\ a_{222} = 1; \\ b_{111} = 2,\ b_{112} = -1,\ b_{122} = -2,\ b_{222} = 1. \end{cases}$$

For $\mathcal{A}$, when we apply Algorithm 7.3.3 for 100 different initial starting points, we obtain two US-eigenvalues $\sigma_1 = 2.1745$ and $\sigma_2 = 2.3547$. The US-eigenvectors corresponding to σ_1 are:

$$\begin{cases} (-0.4863 - 0.8423\iota, -0.1163 - 0.2014\iota)^\top, \\ (-0.4863 + 0.8423\iota, -0.1163 + 0.2014\iota)^\top, \\ (0.9726, 0.2326)^\top, \end{cases}$$

and that corresponding to σ_2 are:

$$\begin{cases} (0.2536 + 0.4392\iota, -0.4309 - 0.7464\iota)^\top, \\ (0.2536 - 0.4392\iota, -0.4309 + 0.7464\iota)^\top, \\ (-0.5071, 0.8617)^\top. \end{cases}$$

For $\mathcal{B}$, after applying Algorithm 7.3.3 100 times, we obtain one US-eigenvalue $\sigma = 3.1623$. The US-eigenvector associated with σ are:

$$\begin{cases} (-0.4436 + 0.5506\iota, -0.5506 - 0.4436\iota)^\top, \\ (0.6987 - 0.1088\iota, -0.1088 - 0.6987\iota)^\top, \\ (-0.2551 + 0.6595\iota, 0.6595 + 0.2551\iota)^\top, \\ (-0.4436 - 0.5506\iota, -0.5506 + 0.4436\iota)^\top, \\ (0.6987 + 0.1088\iota, -0.1088 + 0.6987\iota)^\top, \\ (-0.2551 - 0.6595\iota, 0.6595 - 0.2551\iota)^\top. \end{cases}$$

After applying Algorithm 7.3.2 to compute the US-eigenpairs of $\mathcal{A}$ and $\mathcal{B}$, we need to replace $\mathcal{A}$ and $\mathcal{B}$ by $\iota\mathcal{A}$ and $\iota\mathcal{B}$, respectively. For $\mathcal{A}$, by Algorithm 7.3.2, then we obtain two US-eigenpairs:

$$\begin{cases} \sigma = 2.3547, \quad \mathbf{v} = (-0.4863 - 0.8423\iota, -0.1163 - 0.2014\iota)^\top, \\ \sigma = 2.1745, \quad \mathbf{v} = (0.2536 - 0.4392\iota, -0.4309 + 0.7464\iota)^\top. \end{cases}$$

As for $\mathcal{B}$, we obtain one US-eigenvalue $\sigma = 3.1623$ by Algorithm 7.3.3 and two associated US-eigenvectors:

$$\begin{cases} \mathbf{v} = (-0.2551 + 0.6595\iota, 0.6595 + 0.2551\iota)^\top, \\ \mathbf{v} = (0.6987 + 0.1088\iota, -0.1088 + 0.6987\iota)^\top. \end{cases}$$

Next, two tensors in following example are chosen from [15, Example 3] and [16, section 4.1].

Example 7.6.2 Consider the nonnegative tensor $\mathcal{A} \in CT_{4,2}$ with nonzero entries

$$\mathcal{A}_{1111} = 25.1, \quad \mathcal{A}_{1212} = 25.6, \quad \mathcal{A}_{2121} = 24.8, \quad \mathcal{A}_{2222} = 23,$$

and the entries of $\mathcal{B} \in CT_{3,3}$ are given by

$$\begin{cases} \mathcal{B}_{111} = 0.0072, \ \mathcal{B}_{121} = -0.4413, \ \mathcal{B}_{131} = 0.1941, \mathcal{B}_{211} = -0.4413, \ \mathcal{B}_{221} = 0.0940, \\ \mathcal{B}_{231} = 0.5901, \ \mathcal{B}_{311} = 0.1941, \ \mathcal{B}_{321} = -0.4099, \mathcal{B}_{331} = -0.1012, \ \mathcal{B}_{112} = -0.4413, \\ \mathcal{B}_{122} = 0.0940, \ \mathcal{B}_{132} = -0.4099, \ \mathcal{B}_{212} = 0.0940, \mathcal{B}_{222} = 0.2183, \ \mathcal{B}_{232} = 0.2950, \\ \mathcal{B}_{312} = 0.5901, \ \mathcal{B}_{322} = 0.2950, \ \mathcal{B}_{332} = 0.2229, \mathcal{B}_{113} = 0.1941, \ \mathcal{B}_{123} = 0.5901, \\ \mathcal{B}_{133} = -0.1012, \ \mathcal{B}_{213} = -0.4099, \ \mathcal{B}_{223} = 0.2950, \mathcal{B}_{233} = 0.2229, \ \mathcal{B}_{313} = -0.1012, \\ \mathcal{B}_{323} = 0.2229, \ \mathcal{B}_{333} = -0.4891. \end{cases}$$

We apply Algorithm 7.4.3 to compute some U-eigenpairs of the symmetric tensors $\mathcal{A}$ and $\mathcal{B}$. For $\mathcal{A}$, according to [17, Example 3.11], we know that 25.6 is a singular value of $\mathcal{A}$ and the associated mode-n ($n = 1, 2, 3, 4$) singular vectors are

$$\mathbf{v}_1 = (-1, 0)^\top, \quad \mathbf{v}_2 = (0, -1)^\top, \quad \mathbf{v}_3 = (1, 0)^\top, \quad \mathbf{v}_4 = (0, -1)^\top,$$

and $25.6 \cdot (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 \otimes \mathbf{v}_4)$ is the real best rank-one approximation of $\mathcal{A}$.

When we apply Algorithm 7.4.3 100 times, we obtain a U-eigenvalue $\sigma = 25.6$ and the associated mode-n unitary eigenvectors

$$\mathbf{u}_n = e^{\iota\theta_n}\mathbf{v}_n, \quad n = 1, 2, 3, 4,$$

where $e^{\iota(\theta_1+\theta_2+\theta_3+\theta_4)} = 1$ with $\theta_n \in (-\pi, \pi]$.

Meanwhile, for $\mathcal{B}$, according to [17, Example 3.11], we know that 1 is a singular value and associated mode-n ($n = 1, 2, 3$) singular vectors are

$$\begin{cases} \mathbf{v}_1 = (0.7955, 0.2491, 0.5524)^\top, \\ \mathbf{v}_2 = (-0.0050, 0.9142, -0.4051)^\top, \\ \mathbf{v}_3 = (-0.6060, 0.3195, 0.7285)^\top, \end{cases}$$

and $1 \cdot (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3)$ is the real best rank-one approximation of $\mathcal{B}$. When we apply Algorithm 7.4.3 100 times, we obtain a U-eigenvalue $\sigma = 1$ and the associated mode-n unitary eigenvectors are

$$\mathbf{u}_n = e^{\iota\theta_n}\mathbf{v}_n, \quad n = 1, 2, 3,$$

where $e^{\iota(\theta_1+\theta_2+\theta_3)} = 1$ with $\theta_k \in (-\pi, \pi]$.

We illustrate the feasibility of Algorithm 7.3.3 where the tensors are chosen from [9, 10].

Example 7.6.3 Consider symmetric $\mathcal{A} \in CT_{3,n}$ and $\mathcal{B} \in CT_{4,n}$ with

$$\begin{aligned} a_{i_1i_2i_3} =& \exp((-0.01 + 2\pi\iota 0.20)(i_1 + i_2 + i_3 - 2)) \\ &+ \exp((-0.02 + 2\pi\iota 0.22)(i_1 + i_2 + i_3 - 2)) + e_{i_1+i_2+i_3-2}, \\ b_{i_1i_2i_3i_4} =& \exp((-0.01 + 2\pi\iota 0.20)(i_1 + i_2 + i_3 + i_4 - 3)) \\ &+ \exp((-0.02 + 2\pi\iota 0.22)(i_1 + i_2 + i_3 + i_4 - 3)) + e_{i_1+i_2+i_3+i_4-3}, \end{aligned}$$

where $e_{i_1+i_2+i_3-2}$ and $e_{i_1+i_2+i_3+i_4-3}$ are two different complex white Gaussian noise.

We apply Algorithm 7.3.3 to compute some US-eigenpairs of the symmetric tensors $\mathcal{A}$ and $\mathcal{B}$ with $I = 5$. For $\mathcal{A}$, when we apply Algorithm 7.3.3 100 times, we obtain a US-eigenvalue $\sigma = 18.0492$ and the associated US-eigenvector is

$$\begin{aligned} \mathbf{u} =& \alpha(-0.4019 + 0.1710\iota, -0.2677 - 0.3921\iota, \\ & 0.2203 - 0.3541\iota, 0.4520 + 0.1410\iota, -0.1232 + 0.4131\iota)^\top, \end{aligned}$$

where the nonzero $\alpha \in \mathbb{C}$ satisfies $|\alpha| = 1$.

For $\mathcal{B}$, when we apply Algorithm 7.3.3 100 times, we obtain a US-eigenvalue $\sigma = 37.9441$ and the associated US-eigenvector is

$$\begin{aligned} \mathbf{u} =& \alpha(-0.1701 + 0.4419\iota, -0.5187 - 0.0502\iota, \\ & - 0.0797 - 0.4268\iota, 0.3549 - 0.2051\iota, 0.2564 + 0.2862\iota)^\top, \end{aligned}$$

where the nonzero $\alpha \in \mathbb{C}$ satisfies $|\alpha| = 1$.

7.7 Conclusions and Further Considerations

In the above sections, we present and analyze Algorithms 7.3.1 and 7.4.1 for computing the US-eigenpairs of complex symmetric tensors and the U-eigenpairs of complex tensors, respectively. We generalize the higher order power method to compute the US- or U-eigenpairs of complex tensors, summarized in Algorithm 7.4.3.

In the following, we introduce two types of the Takagi factorizations of complex tensors [18].

Definition 7.7.1 If $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ can be represented as

$$\mathcal{A} = \sum_{r=1}^{R} \sigma_r \cdot (\mathbf{U}_1(:, r) \otimes \mathbf{U}_2(:, r) \otimes \cdots \otimes \mathbf{U}_N(:, r)), \tag{7.7.1}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_R > 0$ and $\mathbf{U}_n \in \mathbb{C}^{I_n \times R}$ satisfy $\mathbf{U}_n^* \mathbf{U}_n = \mathbf{I}_R$, then (7.7.1) is called the Takagi factorization of $\mathcal{A}$, and $(\sigma_r; \mathbf{U}_1(:, r), \mathbf{U}_2(:, r), \ldots, \mathbf{U}_N(:, r))$ is a Takagi pair of $\mathcal{A}$ with $r = 1, 2, \ldots, R$.

Definition 7.7.2 Suppose that $\mathcal{A} \in CT_{N,I}$ is symmetric. If

$$\mathcal{A} = \sum_{r=1}^{R} \sigma_r \cdot \left(\underbrace{\mathbf{U}(:, r) \otimes \mathbf{U}(:, r) \otimes \cdots \otimes \mathbf{U}(:, r)}_{N} \right), \tag{7.7.2}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_R > 0$ and $\mathbf{U} \in \mathbb{C}^{I \times R}$ is unitary, then (7.7.2) is the Takagi factorization of the complex symmetric tensor $\mathcal{A}$. We call $(\sigma_r; \mathbf{U}_n(:, r))$ a Takagi pair of $\mathcal{A}$ with $r = 1, 2, \ldots, R$.

We also introduce another form of Takagi pairs. The Takagi values of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, denoted by $\sigma_k(\mathcal{A})$, where $k = 1, 2, \ldots, K$ and

$$K = \min\{\operatorname{rank}(\mathbf{A}_{(1)}), \operatorname{rank}(\mathbf{A}_{(2)}), \ldots, \operatorname{rank}(\mathbf{A}_{(N)})\},$$

are defined as follows.

Let $\mathbb{S}_n^{(1)} := \{\mathbf{x} \in \mathbb{C}^{I_n} : \|\mathbf{x}\|_2 = 1\}$. Define the first Takagi value of $\mathcal{A}$ by

$$\begin{aligned} \sigma_1(\mathcal{A}) &:= \sup_{\mathbf{x}_n \in \mathbb{S}_n^{(1)},\ n=1,2,\ldots,N} \left| \mathcal{A} \times_1 \mathbf{x}_1^* \times_2 \mathbf{x}_2^* \cdots \times_N \mathbf{x}_N^* \right| \\ &= \sup_{\mathbf{x}_n \in \mathbb{S}_n^{(1)},\ n=1,2,\ldots,N} \left| \overline{\mathcal{A}} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top \right|. \end{aligned} \tag{7.7.3}$$

Since the Cartesian product $\mathbb{S}^{(1)} := \mathbf{S}_1^{(1)} \times \mathbf{S}_2^{(1)} \times \cdots \times \mathbf{S}_N^{(1)}$ of unit spheres is a compact set, an extremal solution of (7.7.3) exists (i.e., the supremum in (7.7.3) is a maximum) and is attained by an N-tuple

$$\left(\mathbf{u}_1^{(1)}, \ldots, \mathbf{u}_N^{(1)} \right) \in \mathbb{S}^{(1)}.$$

Subsequent Takagi values of $\mathcal{A}$ are defined in an inductive manner by setting

$$\mathbb{S}_n^{(k)} := \left\{ \mathbf{x} \in \mathbb{C}^{I_n} : \|\mathbf{x}\|_2 = 1, \mathbf{x}^* \mathbf{u}_n^{(j)} = 0,\ j = 1, 2, \ldots, k-1 \right\}$$

for all n and $k = 1, 2, \ldots, K$, and

$$\begin{aligned} \sigma_k(\mathcal{A}) &:= \sup_{\mathbf{x}_n \in \mathbb{S}_n^{(k)},\ n=1,2,\ldots,N} \left| \mathcal{A} \times_1 \mathbf{x}_1^* \times_2 \mathbf{x}_2^* \cdots \times_N \mathbf{x}_N^* \right| \\ &= \sup_{\mathbf{x}_n \in \mathbb{S}_n^{(k)},\ n=1,2,\ldots,N} \left| \overline{\mathcal{A}} \times_1 \mathbf{x}_1^\top \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top \right|. \end{aligned} \tag{7.7.4}$$

Since the Cartesian product $\mathbb{S}^{(k)} := \mathbf{S}_1^{(k)} \times \mathbf{S}_2^{(k)} \times \cdots \times \mathbf{S}_N^{(k)}$ of unit spheres is a compact set, an extremal solution of (7.7.4) exists (i.e., the supremum in (7.7.4) is a maximum) and is attained by an N-tuple

$$\left(\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_N^{(k)}\right) \in \mathbb{S}^{(k)}.$$

It follows that $\mathbf{u}_n^{(1)}, \ldots, \mathbf{u}_n^{(K)}$ are unit in $\mathbb{C}^{I_n}$. If $K < I_n$, for any n, then we extend the collection of orthogonal elements $\mathbf{u}_n^{(1)}, \ldots, \mathbf{u}_n^{(K)}$ to a complete orthonormal basis of $\mathbb{C}^{I_n}$. This construction leads to a collection of orthonormal bases

$$\{\mathbf{u}_1^{(i_1)}, i_1 = 1, 2, \ldots, I_1\}, \ldots, \{\mathbf{u}_N^{(i_N)}, i_N = 1, 2, \ldots, I_N\} \tag{7.7.5}$$

for the vector spaces $\mathbb{C}^{I_1}, \ldots, \mathbb{C}^{I_N}$, respectively.

Definition 7.7.3 The Takagi values of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ are the numbers $\sigma_1, \ldots, \sigma_K$ with $K = \min\{\text{rank}(\mathbf{A}_{(1)}), \text{rank}(\mathbf{A}_{(2)}), \ldots, \text{rank}(\mathbf{A}_{(N)})\}$ defined by (7.7.3) and (7.7.4). The Takagi vectors of order k are the extremal solutions $(\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_N^{(k)})$ in $\mathbb{S}^{(k)}$ that attain the maximum in (7.7.4).

One of our future plan is to consider the numerical computation of the Takagi factorization of complex tensors. In order to study multivariate polynomial functions in complex variables and their corresponding symmetric tensor representations, Jiang et al. [19] introduce two other types of eigenvalues of complex tensors: C-eigenvalues and G-eigenvalues. Another future research topic is to consider the numerical computation of C- and G-eigenvalues.

References

1. G. Ni, L. Qi, M. Bai, Geometric measure of entanglement and U-eigenvalues of tensors. SIAM J. Matrix Anal. Appl. **35**(1), 73–87 (2014)
2. T. Wei, P. Goldbart, Geometric measure of entanglement and applications to bipartite and multipartite quantum states. Phys. Rev. A **68**(4), 042307 (2003)
3. S. Ragnarsson, C. Van Loan, Block tensors and symmetric embeddings. Linear Algebra Appl. **438**(2), 853–874 (2013)
4. L. Autonne, Sur les matrices hypo-hermitiennes et sur les matrices unitaires. Ann. Univ. Lyon Nouv. Sér. I FASC **38**, 1–77 (1915)
5. R. Horn, C. Johnson, *Matrix Analysis*, 2nd edn. (Cambridge University, Cambridge, 2014)
6. L. Hua, On the theory of automorphic functions of a matrix variable, II-the classification of hypercircles under the symplectic group. Am. J. Math. **66**(4), 531–563 (1944)
7. L. Qi, Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**(6), 1302–1324 (2005)
8. L. Qi, Hankel tensors: Associated Hankel matrices and Vandermonde decomposition. Commun. Math. Sci. **13**(1), 113–125 (2015)
9. J. Papy, L. De Lathauwer, S. Van Huffel, Exponential data fitting using multilinear algebra: the single-channel and multi-channel case. Numer. Linear Algebra Appl. **12**(8), 809–826 (2005)

10. W. Ding, L. Qi, Y. Wei, Fast Hankel tensor-vector products and application to exponential data fitting. Numer. Linear Algebra Appl. **22**(5), 814–832 (2015)
11. L. Sorber, M. Van Barel, L. De Lathauwer, Unconstrained optimization of real functions in complex variables. SIAM J. Optim. **22**(3), 879–898 (2012)
12. S. Zhang, Y. Xia, J. Wang, A complex-valued projection neural network for constrained optimization of real functions in complex variables. IEEE Trans. Neural Netw. **26**(12), 3227–3238 (2015)
13. S. Zhang, Y. Xia, W. X. Zheng, A complex-valued neural dynamical optimization approach and its stability analysis. Neural Netw. **61**, 59–67 (2015)
14. E. Kofidis, P. Regalia, On the best rank-1 approximation of higher-order supersymmetric tensors. SIAM J. Matrix Anal. Appl. **23**(3), 863–884 (2002)
15. L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(r_1, r_2, ..., r_n) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000)
16. M. Ishteva, P. Absil, P. Van Dooren, Jacobi algorithm for the best low multilinear rank approximation of symmetric tensors. SIAM J. Matrix Anal. Appl. **34**(2), 651–672 (2013)
17. J. Nie, L. Wang, Semidefinite relaxations for best rank-1 tensor approximations. SIAM J. Matrix Anal. Appl. **35**(3), 1155–1179 (2014)
18. A. Boralevi, J. Draisma, E. Horobet, E. Robeva, *Orthogonal and Unitary Tensor Decomposition from an Algebraic Perspective* (2014). ArXiv preprint:1512.08031v1
19. B. Jiang, Z. Li, S. Zhang, Characterizing real-valued multivariate complex polynomials and their symmetric tensor representations. SIAM J. Matrix Anal. Appl. **37**(1), 381–408 (2016)

Chapter 8
Randomized Algorithms

Randomized algorithms provide a useful tool for scientific computing. Compared with standard deterministic algorithms, randomized algorithms are often faster and robust. This chapter is devoted to an adaptive randomized algorithms for computing the approximate tensor decompositions. We present an adaptive randomized algorithm for the low multilinear rank approximation with unknown multilinear rank and give its probabilistic error bound. Finally, we develop an adaptive randomized algorithm for the tensor train approximations. We illustrate our algorithms via several numerical examples.

8.1 Preliminaries

We present adaptive randomized algorithms for low multilinear rank and tensor train approximations, summarized in Algorithms 8.2.2 and 8.4.2, generalizing Algorithm 4.2 in [1] to the case $N > 2$. We list these two problems as follows.

Problem 8.1.1 Suppose that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. Find N columnwise orthogonal matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times \mu_n}$ with $\mu_n \leq I_n$, such that

$$a_{i_1 i_2 \ldots i_N} \approx \sum_{j_1, \ldots, j_N = 1}^{I_1, \ldots, I_N} a_{j_1 j_2 \ldots j_N} p^{(1)}_{i_1 j_1} p^{(2)}_{i_2 j_2} \cdots p^{(N)}_{i_N j_N},$$

where $\mathbf{P}^{(n)} = \mathbf{Q}^{(n)} (\mathbf{Q}^{(n)})^{\top} \in \mathbb{R}^{I_n \times I_n}$ is a projection matrix.

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4_8

Problem 8.1.2 Suppose that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. Find N order-3 tensors $\mathcal{U}^{(n)} \in \mathbb{R}^{\mu_{n-1} \times I_n \times \mu_n}$ with $\mu_0 = \mu_N = 1$, such that

$$a_{i_1 i_2 \ldots i_N} \approx \sum_{r_1=1}^{\mu_1} \sum_{r_2=1}^{\mu_2} \cdots \sum_{r_{N-1}=1}^{\mu_{N-1}} u^{(1)}_{i_1, r_1} u^{(2)}_{r_1, i_2, r_2} \ldots u^{(N)}_{r_{N-1}, i_N},$$

where $u^{(n)}_{r_{n-1}, i_n, r_n}$ is the (r_{n-1}, i_n, r_n)-element of $\mathcal{U}^{(n)}$, with $r_0 = r_N = 1$.

Many researchers design the numerical algorithms for Problem 8.1.1, such as the higher-order orthogonal iteration [2], the Newton-Grassmann method [3], the Riemannian trust-region method [4], the Quasi-Newton method [5], semi-definite programming (SDP) [6], and Lanczos-type iterations [7, 8]. Readers may refer to the nice surveys [9, 10] on this topic. Solving Problem 8.1.1 without columnwise orthogonal constraints has been investigated. For example, if the columns of each $\mathbf{Q}_n$ are extracted from the mode-n unfolding matrix $\mathbf{A}_{(n)}$, then the solution of Problem 8.1.1 is called as the *CUR*-type decomposition of $\mathcal{A}$, and can be obtained using different versions of the cross approximation method [11–16]. For Problem 8.1.1, we restrict the entries of $\mathcal{A}$ and $\mathbf{Q}_n$ to be nonnegative with the latter not columnwise orthogonal, the solution of Problem 8.1.1 is sometimes called a nonnegative Tucker decomposition [17–20],

The solution for Problem 8.1.2 is called the *tensor train* approximation. Oseledets [21] provides a quasi-best approximation in a tensor train decomposition for a given TT-rank from the SVDs of the matrices obtained from the corresponding unfoldings by successive projections. Holtz et al. [22] generalize ALS to obtain a modified approach (MALS) for the tensor train approximations. In [21, 22], the solutions $\mathcal{V}^{(n)}_* \in \mathbb{R}^{\mu_{n-1} \times I_n \times \mu_n}$ of Problem 8.1.2 are restricted so that reshape$(\mathcal{V}^{(n)}_*, [\mu_{n-1} I_n, \mu_n])$ is columnwise orthogonal with $n = 1, 2, \ldots, N-1$, which is not true in general.

For the large-scale symmetric eigenvalue problems [23, 24], the singular value decompositions [25] and the linear equations [26], we can use the TT format of vectors and matrices to overcome the curse of dimensionality and to make storage and computational costs feasible; see [27] and the references therein.

Low-rank matrix approximations, such as the truncated singular value decomposition [28, page 291] and the rank-revealing QR decomposition [29], play a fundamental role in data analysis and scientific computing. Halko et al. [1] present a modular framework to construct randomized algorithms for computing partial matrix decompositions. The topic for decomposing a tensor by the strategy of randomized algorithms has been developed [11, 30–32].

An important advantage of the adaptive randomized algorithms is that nearly optimal approximations are possible, even with very noisy data.

The same subroutine (Algorithm 8.2.1) is needed in both Algorithms 8.2.2 and 8.4.1, that is, under certain constraints, Problems 8.1.1 and 8.1.2 have the same subproblem (Problem 8.2.1). We analyze the convergence of Algorithms 8.2.2 and 8.4.1 via the convergence of Algorithm 8.2.1.

We need to use the Frobenius norm of matrices to analyze Algorithm 8.2.1. The core problem is to estimate the singular values of matrices. Reynolds et al. [33] consider the system of linear equations with Khatri-Rao structures, when they introduce a randomized variation of ALS for rank reduction of canonical tensor formats.

To better understand Algorithms 8.2.2 and 8.4.1 the approximate tensor decompositions, we compare these two algorithms with other known algorithms on several test tensors.

There are two subroutines in Algorithm 8.2.1, distinguished by a Boolean flag "take_max". We have not presented the theoretical analysis for these two subroutines. We only compare these two subroutines on several test matrices, from Hansen Tools [34].

8.1.1 Basic Operations

For $\mathbf{x}_n \in \mathbb{R}^{I_n}$ with all n and $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the *tensor-vector* multiplication [7] (abbreviation *tenvec*) is defined as

$$\mathcal{A} \times_2 \mathbf{x}_2^\top \cdots \times_N \mathbf{x}_N^\top, \quad \mathcal{A} \times_1 \mathbf{x}_1^\top \times_3 \mathbf{x}_3^\top \cdots \times_N \mathbf{x}_N^\top, \quad \ldots, \quad \mathcal{A} \times_1 \mathbf{x}_1^\top \cdots \times_{N-1} \mathbf{x}_{N-1}^\top.$$

We recall the mode-(n, m) product [27] (called *tensor-tensor product*) of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_M}$ with common modes $I_n = J_m$ that produces an order $(M + N - 2)$ tensor $\mathcal{C} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N \times J_1 \times \cdots \times J_{m-1} \times J_{m+1} \times \cdots \times J_M}$:

$$\mathcal{C} = \mathcal{A} \times_n^m \mathcal{B},$$

where its entries are

$$c_{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N j_1 \ldots j_{m-1} i_{m+1} \ldots j_N} = \sum_{i_n=1}^{I_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} b_{j_1 \ldots j_{m-1} i_n j_{m+1} \ldots \ldots j_M}.$$

The mode-n unfolding matrix of an order N tensor can be understood as the process of the construction of a matrix containing all the mode-n vectors of the tensor. The order of the columns is not unique and chosen in accordance with [10]. The mode-n unfolding matrix of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, denoted by $\mathbf{A}_{(n)}$, arranges the mode-n fibers into columns of a matrix. More specifically, a tensor element $(i_1, i_2, \ldots, i_N)$ maps to a matrix element (i_n, j), where

$$\begin{aligned} j =& i_1 + (i_2 - 1) I_1 + \cdots + (i_{n-1} - 1) I_1 \ldots I_{n-2} + (i_{n+1} - 1) I_1 \ldots I_{n-1} \\ &+ \cdots + (i_N - 1) I_1 \ldots I_{n-1} I_{n+1} \ldots I_{N-1}. \end{aligned}$$

We briefly review another form for unfolding a tensor into a matrix. We denote the unfolding matrix of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ by $\mathbf{A}_{([n])}$ [21]. In detail, the (i, j)-element of $\mathbf{A}_{([n])}$ is defined as $\mathbf{A}_{([n])}(i, j) = a_{i_1 i_2 \ldots i_N}$, where

$$\begin{cases} i = i_1 + (i_2 - 1)I_1 + \cdots + (i_n - 1)I_1 \ldots I_{n-1}, \\ j = i_{n+1} + (i_{n+2} - 1)I_{n+1} + \cdots + (i_N - 1)I_{n+1} \ldots I_{N-1}. \end{cases}$$

8.1.2 Tucker Decomposition

A *Tucker decomposition* [35] of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as

$$\mathcal{A} \approx \mathcal{G} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}, \tag{8.1.1}$$

where $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ are called the *mode-n factor matrices* and $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ the *core tensor* of the decomposition with the set $\{R_1, R_2, \ldots, R_N\}$.

Comparing with Problem 8.1.1, $\mathbf{U}_n$ is not restricted to be columnwise orthogonal in (8.1.1). We refer to the survey [10] for more details. We summarize the basic techniques, which are also needed to motivate the tensor train decomposition [21, 36].

The Tucker decomposition is closely related to the mode-n unfolding matrix $\mathbf{A}_{(n)}$ for all n. In particular, the relation (8.1.1) implies

$$\mathbf{A}_{(n)} \approx \mathbf{U}^{(n)} \mathbf{G}_{(n)} (\mathbf{U}^{(N)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \cdots \otimes \mathbf{U}^{(1)})^\top.$$

It follows that the rank of $\mathbf{A}_{(n)}$ is less than or equal to R_n, as the mode-n factor $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ at most has rank R_n. This motivates us to define the multilinear rank of $\mathcal{A}$ as the tuple

$$\{R_1, R_2, \ldots, R_N\}, \quad \text{where the rank of } \mathbf{A}_{(n)} \text{ is equal to } R_n.$$

Algorithm 8.1.1 Higher-order singular value decomposition (HOSVD) [10]

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and the multilinear rank $\{R_1, R_2, \ldots, R_N\}$.
Output: N columnwise orthogonal matrix $\mathbf{U} \in \mathbb{R}^{I_n \times R_n}$ and the core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$.

1: **for** $n = 1, 2, \ldots, N$ **do**
2: $\quad \mathbf{U}_n \leftarrow R_n$ leading left singular vectors of $\mathbf{A}_{(n)}$.
3: **end for**
4: $\mathcal{G} \leftarrow \mathcal{A} \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \cdots \times_N \mathbf{U}_N^\top$.
5: Return $\mathcal{G}, \mathbf{U}_1, \ldots, \mathbf{U}_N$.

By applying SVD to $\mathbf{A}_{(n)}$ for all n, we obtain a special form of the Tucker decomposition of a given tensor, which is referred to as the higher-order singular value decomposition (HOSVD) [37], as shown in Algorithm 8.1.1. The HOSVD has these properties: all factor matrices are columnwise orthogonal and the core tensor has the properties of all-orthogonality and ordering. In detail, we have

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N},$$

in which $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ is orthogonal and $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ satisfies

(i) for all i_n and j_n with $i_n \neq j_n$, $\langle \mathcal{S}(:, \ldots, :, i_n, :, \ldots, :), \mathcal{S}(:, \ldots, :, j_n, :, \ldots, :) \rangle = 0$;
(ii) for all n,

$$\begin{aligned} \|\mathcal{S}(:, \ldots, :, 1, :, \ldots, :)\|_F &\geq \|\mathcal{S}(:, \ldots, :, 2, :, \ldots, :)\|_F \\ &\geq \cdots \\ &\geq \|\mathcal{S}(:, \ldots, :, I_n, :, \ldots, :)\|_F \\ &\geq 0. \end{aligned}$$

The Frobenius norms $\|\mathcal{S}(:, \ldots, :, i_n, :, \ldots, :)\|_F$, denoted by $\sigma_{i_n}^{(n)}$, are the mode-n singular values of $\mathcal{A}$ and the vector $\mathbf{U}^{(n)}(:, i_n)$ is an i_nth mode-n singular vector. The mode-n singular values of $\mathcal{A}$ are different from its singular values, defined in [38].

When $R_n < \operatorname{rank}(\mathbf{A}_{(n)})$ for one or more n, the decomposition is called the *truncated HOSVD*. The truncated HOSVD is not optimal in terms of giving the best fit as measured by the norm of the difference, but it is a good starting point for an iterative ALS algorithm. With respect to the Frobenius norm of tensors, Problem 8.1.1 can be rewritten as the optimization problem

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(N)}} \quad & \left\| \mathcal{A} - \mathcal{G} \times_1 \mathbf{Q}^{(1)} \cdots \times_N \mathbf{Q}^{(N)} \right\|_F^2, \\ \text{subject to} \quad & \mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}, \\ & \mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n} \text{ is columnwise orthogonal;} \end{aligned}$$

which is equivalent to the following maximization problem

$$\begin{aligned} \max_{\mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(N)}} \quad & \left\| \mathcal{A} \times_1 \mathbf{Q}^{(1)\top} \cdots \times_N \mathbf{Q}^{(N)\top} \right\|_F^2, \\ \text{subject to} \quad & \mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n} \text{ is columnwise orthogonal.} \end{aligned}$$

If $\mathbf{Q}_*^{(n)}$ is a solution of the above maximization problem, then we call $\mathcal{A} \times_1 \mathbf{P}^{(1)} \cdots \times_N \mathbf{P}^{(N)}$ a *low multilinear rank approximation* of $\mathcal{A}$, where $\mathbf{P}^{(n)} = \mathbf{Q}_*^{(n)}(\mathbf{Q}_*^{(n)})^\top$.

The higher-order orthogonal iteration (HOOI) for the low multilinear rank approximation of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is summarized in Algorithm 8.1.2.

Algorithm 8.1.2 Higher-order orthogonal iteration (HOOI) [2, Algorithm 4.2]

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the multilinear rank $\{R_1, R_2, \ldots, R_N\}$ and maximum number of iterations $K_{\max}$.
Output: N columnwise orthogonal matrix $\mathbf{U} \in \mathbb{R}^{I_n \times R_n}$ and the core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$.

1: initialize $\mathbf{U}_n \in \mathbb{R}^{I_n \times R_n}$ with all n using HOSVD.
2: **for** $k = 1, 2, \ldots, K_{\max}$ **do**
3: **for** $n = 1, 2, \ldots, N$ **do**
4: Compute $\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}_1^\top \cdots \times_{n-1} \mathbf{U}_{n-1}^\top \times_{n+1} \mathbf{U}_{n+1}^\top \cdots \times_N \mathbf{U}_N^\top$.
5: $\mathbf{U}_n \leftarrow R_n$ leading left singular vectors of $\mathbf{B}_{(n)}$.
6: **end for**
7: **end for**
8: $\mathcal{G} \leftarrow \mathcal{A} \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \cdots \times_N \mathbf{U}_N^\top$.
9: Return $\mathcal{G}, \mathbf{U}_1, \ldots, \mathbf{U}_N$.

8.1.3 Tensor Train Decomposition

The storage of the core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ renders the Tucker decomposition increasingly unattractive as N gets larger. In order to overcome this disadvantage, a simple non-recursive form of the tensor decomposition is presented, called the *tensor train decomposition* [21, 36, 39], a special case of the *Hierarchical Tucker format* [40–42].

The tensor train decomposition of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ can be described in the standard scalar form

$$a_{i_1 i_2 \ldots i_N} \approx \sum_{j_1=1}^{R_1} \sum_{j_2=1}^{R_2} \cdots \sum_{j_{N-1}=1}^{R_{N-1}} q_{i_1, j_1}^{(1)} q_{j_1, i_2, j_2}^{(2)} \cdots q_{j_{N-1}, i_N}^{(N)},$$

or equivalently by using slice representations:

$$a_{i_1 i_2 \ldots i_N} \approx \mathbf{Q}^{(1)}(i_1)\mathbf{Q}^{(2)}(i_2) \ldots \mathbf{Q}^{(N)}(i_N), \tag{8.1.2}$$

where slice matrices are defined as

$$\mathbf{Q}^{(n)}(i_n) = \mathcal{Q}^{(n)}(:, i_n, :) \in \mathbb{R}^{R_{n-1} \times R_n},$$

i.e., $\mathbf{Q}^{(n)}(i_n)$ is an i_nth lateral slice of the core $\mathcal{Q}^{(n)} \in \mathbb{R}^{R_{n-1}\times I_n \times R_n}$ for all n with $R_0 = R_N = 1$. By the tensor-tensor product, (8.1.2) can be also represented as

$$\mathcal{A} \approx \mathcal{Q}^{(1)} \times_3^1 \mathcal{Q}^{(2)} \times_3^1 \cdots \times_3^1 \mathcal{Q}^{(N)}.$$

The tensor train decomposition is closely related with the unfolding matrices $\mathbf{A}_{([n])}$ of $\mathcal{A}$. Equation (8.1.2) implies that $\text{rank}(\mathbf{A}_{([n])}) \leq R_n$ for $n = 1, 2, \ldots, N-1$ [9, 27]. The tuple containing the ranks of these matricizations is called the TT-rank of $\mathcal{A}$.

8.2 Low Multilinear Rank Approximations

Based on the key idea for probabilistic algorithms to approximate matrix decompositions [1], we develop adaptive randomized algorithms for the low multilinear rank approximation of tensors in $\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$. We shall reveal the relationship between our randomized algorithms and the Randomized Tucker decomposition [32, Algorithm 2].

8.2.1 Randomized Algorithms for Low Multilinear Rank Approximations

If we add the constraints on $\mathbf{Q}_n \in \mathbb{R}^{I_n\times\mu_n}$ in Problem 8.1.1, that is, $\mathbf{Q}_n^\top\mathbf{Q}_n = \mathbf{I}_{\mu_n}$, then we can rewrite Problem 8.1.1 as follows.

Problem 8.2.1 Suppose that $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ and ϵ is a prescribed tolerance. Find N columnwise orthogonal matrices $\mathbf{Q}_n \in \mathbb{R}^{I_n\times\mu_n}$ with $\mu_n \leq I_n$, such that

$$\left\|\mathcal{A} - \mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\times_2(\mathbf{Q}_2\mathbf{Q}_2^\top)\times_3\cdots\times_N(\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F \leq \sqrt{N}\epsilon.$$

If the N-tuple $\{\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_N\}$ is a solution of Problem 8.2.1, then we have

$$\begin{aligned}
&\mathcal{A} - \mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\times_2(\mathbf{Q}_2\mathbf{Q}_2^\top)\cdots\times_N(\mathbf{Q}_N\mathbf{Q}_N^\top)\\
&= (\mathcal{A} - \mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)) + (\mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top) - \mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\times_2(\mathbf{Q}_2\mathbf{Q}_2^\top)) + \ldots\\
&\quad +(\mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\cdots\times_{N-1}(\mathbf{Q}_{N-1}\mathbf{Q}_{N-1}^\top)\\
&\quad -\mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\cdots\times_{N-1}(\mathbf{Q}_{N-1}\mathbf{Q}_{N-1}^\top)\times_N(\mathbf{Q}_N\mathbf{Q}_N^\top)).
\end{aligned} \tag{8.2.1}$$

The equality (8.2.1) can be found in [43, 44] for the low multilinear rank approximation of $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$.

It follows from (8.2.1) that

$$\left\|\mathcal{A}-\mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\times_2(\mathbf{Q}_2\mathbf{Q}_2^\top)\times_3\cdots\times_N(\mathbf{Q}_N\mathbf{Q}_N^\top)\right\|_F^2\leq\sum_{n=1}^{N}\left\|\mathcal{A}-\mathcal{A}\times_n(\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F^2. \tag{8.2.2}$$

The result relies on the orthogonality of the projector in the Frobenius norm [44], i.e., for all n, we have

$$\|\mathcal{A}\|_F^2=\left\|\mathcal{A}\times_n(\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F^2+\left\|\mathcal{A}\times_n(\mathbf{I}_{I_n}-\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F^2,$$

and the fact that $\|\mathbf{AP}\|_F\leq\|\mathbf{A}\|_F$ with $\mathbf{A}\in\mathbb{R}^{I\times J}$, where the orthogonal projection $\mathbf{P}$ satisfies [28]

$$\mathbf{P}^2=\mathbf{P},\quad \mathbf{P}^\top=\mathbf{P},\quad \mathbf{P}\in\mathbb{R}^{J\times J}.$$

In order to obtain a solution $\{\mathbf{Q}_1,\mathbf{Q}_2,\ldots,\mathbf{Q}_N\}$ for Problem 8.2.1, we need to consider the following subproblem for all n.

Subproblem 8.2.1 *Suppose that $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ and ϵ is a prescribed tolerance. For a given n, find an columnwise orthogonal matrix $\mathbf{Q}_n\in\mathbb{R}^{I_n\times\mu_n}$ with $\mu_n\leq I_n$, such that*

$$\left\|\mathcal{A}-\mathcal{A}\times_n(\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F=\left\|\mathcal{A}\times_n(\mathbf{I}_{I_n}-\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F\leq\epsilon.$$

For each n, if the columnwise orthogonal matrix $\mathbf{Q}_n$ satisfies Problem 8.2.1 with the same ϵ, we have

$$\begin{aligned}&\left\|\mathcal{A}-\mathcal{A}\times_1(\mathbf{Q}_1\mathbf{Q}_1^\top)\times_2(\mathbf{Q}_2\mathbf{Q}_2^\top)\times_3\cdots\times_N(\mathbf{Q}_N\mathbf{Q}_N^\top)\right\|_F^2\\&\leq\sum_{n=1}^{N}\left\|\mathcal{A}-\mathcal{A}\times_n(\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F^2=N\epsilon^2.\end{aligned}$$

By this hypothesis, we obtain a solution for Problem 8.2.1 by solving Subproblem 8.2.1, for n from 1 to N. For such a given n, we present a randomized algorithm for seeking an columnwise orthogonal matrix $\mathbf{Q}_n\in\mathbb{R}^{I_n\times\mu_n}$ such that

$$\left\|\mathcal{A}-\mathcal{A}\times_n(\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F=\left\|\mathcal{A}\times_n(\mathbf{I}_{I_n}-\mathbf{Q}_n\mathbf{Q}_n^\top)\right\|_F\leq\epsilon,$$

where ϵ is a computational tolerance.

Starting with an initial empty basis matrix $\mathbf{Q}_0$, we generate a columnwise orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{I_n \times \mu_n}$ such that the range of $\mathbf{Q}$ captures the action of the mode-n unfolding matrix of the tensor $\mathcal{A}$ as follows:

for $k = 0, 1, \ldots,$ until convergence **do**
  Draw $(N-1)$ standard Gaussian vectors $\mathbf{x}_m \in \mathbb{R}^{I_m}$ for all m $(m \neq n)$.
  Compute

$$\mathbf{y}_k = \mathcal{A} \bar{\times}_1 \mathbf{x}_1 \ldots \bar{\times}_{n-1} \mathbf{x}_{n-1} \bar{\times}_{n+1} \mathbf{x}_{n+1} \ldots \bar{\times}_N \mathbf{x}_N.$$

  Compute $\widetilde{\mathbf{y}}_k = (\mathbf{I}_{I_n} - \mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^\top)\mathbf{y}_k$ if $k > 1$.
  Normalize $\mathbf{y}_k = \widetilde{\mathbf{y}}_k / \|\widetilde{\mathbf{y}}_k\|_2$, and form $\mathbf{Q}_k = [\mathbf{Q}_{k-1}, \mathbf{y}_k]$.
end for
Write $\mathbf{Q}$ as $\mathbf{Q}_k$.

The above process requires μ_n tenvecs and $O(I\mu_n^2)$ additional operations, where $\mu_n \leq \mu_{\max}$ is the number of the columns of $\mathbf{Q}$. To terminate the computation, we use one or both of the following stopping criteria: (1) fix maximum number of iterations, i.e., the desired size of basis $\mathbf{Q}$ by $\mu_{\max}$; and (2) find a basis $\mathbf{Q}$ that allows approximation of $\mathbf{A}_{(n)}$ with a relative accuracy ε.

Unfortunately, the matrix $\mathbf{Q}$ generated by the above process is not columnwise orthogonal and the Frobenius norm of the error matrix $\mathbf{I}_{\mu_n} - \mathbf{Q}^\top\mathbf{Q}$ is not sufficiently small. The reason is that the vector $\widetilde{\mathbf{y}}_k$ becomes small as the basis starts to capture the mode-n unfolding matrix of the tensor $\mathcal{A}$. In finite-precision arithmetics, their directions are unreliable.

To overcome this difficulty, we re-project the normalized vector $\mathbf{y}_k$ onto range$(\mathbf{Q}_{k-1})^\perp$ in Step 13 to Step 17 of Algorithm 8.2.1, where range$(\mathbf{A})$ is the subspace generated by the columns of $\mathbf{A} \in \mathbb{R}^{I \times J}$ with $I \geq J$, and $\mathbb{K}^\perp$ is the orthogonal complement in $\mathbb{R}^n$ of a subspace $\mathbb{K}$.

Algorithm 8.2.1 requires $(R + \mu_n)$ tenvecs and $O(I\mu_n^2)$ additional operations, where $\mu_n \leq \mu_{\max}$ is the number of the columns of $\mathbf{Q}$, derived by Algorithm 8.2.1. The calculations in Algorithm 8.2.1 can be organized so that each iteration processes a block of samples simultaneously. This revision leads to dramatic improvements in speed, exploiting higher-level linear algebra subroutines (e.g., BLAS3) or parallel processors.

The main part of Algorithm 8.2.1 consists of Steps 4–18, which generates $\mathbf{q}_k$, from $\mathbf{y}_k$. There are two ways to generate $\mathbf{q}_k$, distinguished by the Boolean flag "take_max".

When $N = 2$ and the value of "take_max" is "False", Algorithm 8.2.1 is reduced to Algorithm 4.2 in [1] for constructing a subspace that captures most of the action of the matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$. In Sect. 8.5.3, we illustrate Algorithm 8.2.1 with different values of the Boolean flag "take_max" via certain testing matrices. The failure probability stated for Algorithm 8.2.1 is pessimistic, since it is derived from a simple uniform bound argument. In practice, the error is reliable in a range of circumstances, if we take $R = 10$.

Algorithm 8.2.1 Adaptive randomized algorithm for Subproblem 8.2.1

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, a mode n, an integer R, the prescribed tolerance ϵ, a Boolean flag "take_max" and maximum number of iterations $\mu_{\max}$.
Output: An columnwise orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{I_n \times \mu_n}$.

1: Draw $(N-1)$ independent families $\{\mathbf{x}_m^{(r)} \in \mathbb{R}^{I_m} : r = 1, 2, \ldots, R\}$ of standard Gaussian vectors for all m $(m \neq n)$.
2: Compute $\mathbf{y}_r = \mathcal{A} \bar{\times}_1 \mathbf{x}_1^{(r)} \ldots \bar{\times}_{n-1} \mathbf{x}_{n-1}^{(r)} \bar{\times}_{n+1} \mathbf{x}_{n+1}^{(r)} \ldots \bar{\times}_N \mathbf{x}_N^{(r)}$ with $r = 1, 2, \ldots, R$.
3: Start with an empty basis matrix $\mathbf{Q}_0$ and set $k = 0$.
4: **while** $\max\{\|\mathbf{y}_{k+1}\|_2, \|\mathbf{y}_{k+2}\|_2, \ldots, \|\mathbf{y}_{k+R}\|_2\} > \epsilon$ or $k < \mu_{\max}$ **do**
5: Set $k = k + 1$.
6: **if** the value of "take_max" is "True" **then**
7: Choose $k_0 \in \{k+1, \ldots, k+R\}$ such that $\|\mathbf{y}_{k_0}\|_2 = \max\{\|\mathbf{y}_{k+1}\|_2, \|\mathbf{y}_{k+2}\|_2, \ldots, \|\mathbf{y}_{k+R}\|_2\}$.
8: Overwrite $\mathbf{y}_k$ by $(\mathbf{I}_{I_n} - \mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^\top)\mathbf{y}_{k_0}$ if $k > 1$.
9: **else**
10: Overwrite $\mathbf{y}_k$ by $(\mathbf{I}_{I_n} - \mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^\top)\mathbf{y}_k$ if $k > 1$.
11: **end if**
12: Compute $\mathbf{q}_k = \mathbf{y}_k / \|\mathbf{y}_k\|_2$ and form $\mathbf{Q}_k = [\mathbf{Q}_{k-1}, \mathbf{q}_k]$.
13: Draw $(N-1)$ standard Gaussian vectors $\mathbf{x}_m \in \mathbb{R}^{I_m}$ for all $m \neq n$.
14: Compute $\mathbf{y}_{k+R} = (\mathbf{I}_{I_n} - \mathbf{Q}_k\mathbf{Q}_k^\top)(\mathcal{A} \bar{\times}_1 \mathbf{x}_1 \ldots \bar{\times}_{n-1} \mathbf{x}_{n-1} \bar{\times}_{n+1} \mathbf{x}_{n+1} \ldots \bar{\times}_N \mathbf{x}_N)$.
15: **for** $i = k+1, k+2, \ldots, k+R-1$ **do**
16: Overwrite $\mathbf{y}_i$ by $\mathbf{y}_i - (\mathbf{q}_k^\top \mathbf{y}_i)\mathbf{q}_k$.
17: **end for**
18: **end while**
19: Set $\mathbf{Q} = \mathbf{Q}_k$ and μ_k as the number of all columns of $\mathbf{Q}$.

By the relationship between Problem 8.2.1 and Subproblem 8.2.1, we construct an adaptive randomized algorithm for Problem 8.2.1, summarized in Algorithm 8.2.2.

Algorithm 8.2.2 Adaptive randomized algorithm for Problem 8.2.1

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, an integer R, the prescribed tolerance ϵ, a Boolean flag "take_max" and maximum number of iterations $\mu_{\max}$.
Output: N columnwise orthogonal matrices $\mathbf{Q}_n \in \mathbb{R}^{I_n \times \mu_n}$ for all n.

1: Set the temporary tensor: $\mathcal{C} = \mathcal{A}$.
2: **for** $n = 1, 2, \ldots, N$ **do**
3: Implement Algorithm 8.2.1 with $\mathcal{C}$ to generate the columnwise orthogonal matrix $\mathbf{Q}_{n,k} \in \mathbb{R}^{I_n \times \mu_n}$.
4: Set $\mathbf{Q}_n = \mathbf{Q}_{n,k}$, μ_n as the number of all columns of $\mathbf{Q}$,
5: Compute $\mathcal{C} = \mathcal{C} \times_n \mathbf{Q}_n^\top$ and let $I_n = \mu_n$.
6: **end for**

If we take $R = 10$ in Algorithm 8.2.2, then the error estimation is reliable under certain assumptions. In Step 5 of Algorithm 8.2.2, the temporary tensor $\mathcal{C}$ is updated for each n. The reason is (8.2.1) and the fact $\|\mathbf{AQ}\|_F \leq \|\mathbf{A}\|_F$ with $A \in \mathbb{R}^{I \times J}$ and any columnwise orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{J \times K}$ $(K \leq J)$. For clarity, we assume $I_1 \sim I_2 \sim \cdots \sim I_N \sim I$ and $\mu_1 \sim \mu_2 \sim \cdots \sim \mu_N \sim \mu$ in complexity estimates

[7, Page A2], where μ_n is the number of the columns of $\mathbf{Q}_n$. If the size of C is not updated for each n, Algorithm 8.2.2 requires $N(R+\mu)$ tenvecs and $O(NI\mu^2)$ additional operations. Similar to Algorithm 3.2 in [44], a more general formwork of Algorithm 8.2.2 is in the following:

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, an integer R, the prescribed tolerance ϵ, a Boolean flag "take_max", maximum number of iterations $\mu_{\max}$, and a processing order $\mathbf{p} \in \mathbb{S}_N$.

Output: N columnwise orthogonal matrices $\mathbf{Q}_n \in \mathbb{R}^{I_n \times \mu_n}$.

1: Set the temporary tensor: $C = \mathcal{A}$.
2: **for** $n = p_1, p_2, \ldots, p_N$ **do**
3: Implement Algorithm 8.2.1 with C to generate the columnwise orthogonal matrix $\mathbf{Q}_{n,k} \in \mathbb{R}^{I_n \times \mu_n}$.
4: Set $\mathbf{Q}_n = \mathbf{Q}_{n,k}$, μ_n as the number of all columns of $\mathbf{Q}$.
5: Compute $C = C \times_n \mathbf{Q}_n^\top$ and let $I_n = \mu_n$.
6: **end for**

If the value of "take_max" is "False" and the multlilinear rank of $\mathcal{A}$ is $\{R_1, R_2, \ldots, R_N\}$, then Algorithm 8.2.2 can be represented as follows.

Algorithm 8.2.3 Randomized algorithm for Problem 8.2.1 with known multilinear rank $\{R_1, R_2, \ldots, R_N\}$

1: Set the temporary tensor: $C = \mathcal{A}$.
2: **for** $n = 1, 2, \ldots, N$ **do**
3: Compute $\mathbf{Z}_{(n)} = \mathbf{A}_{(n)}\mathbf{\Omega}_{(n)}$, where $\mathbf{\Omega}_{(n)} = \mathbf{\Omega}'_1 \odot \cdots \odot \mathbf{\Omega}'_{n-1} \odot \mathbf{\Omega}'_{n+1} \odot \cdots \odot \mathbf{\Omega}'_N$ and $\mathbf{\Omega}'_m \in \mathbb{R}^{I_m \times R_m + 10}$ is a standard Gaussian matrix with $m \neq n$ and $m = 1, 2, \ldots, N$.
4: Compute $\mathbf{Q}_n$ as a columnwise orthogonal basis of $\mathbf{Z}_{(n)}$ by using the QR decomposition and let $\mathbf{Q}_n = \mathbf{Q}_n(:, 1 : R_n)$..
5: Set $C = C \times \mathbf{Q}_n^\top$ and let $I_n = R_n$.
6: **end for**

When the multlilinear rank of $\mathcal{A}$ is $\{R_1, R_2, \ldots, R_N\}$, the Randomized Tucker decomposition can be estimated as follows [32, Algorithm 2].

Algorithm 8.2.4 Randomized Tucker decomposition for solving Problem 8.2.1 with known multilinear rank $\{R_1, R_2, \ldots, R_N\}$

1: Set the temporary tensor: $C = \mathcal{A}$.
2: **for** $n = 1, 2, \ldots, N$ **do**
3: Compute $\mathbf{Z}_{(n)} = \mathbf{A}_{(n)}\mathbf{\Omega}_{(n)}$, where $\mathbf{\Omega}_{(n)}$ is an $(\prod_{k \neq n} I_k)$-by-$R_n + 10$ standard Gaussian matrix.
4: Compute $\mathbf{Q}_n$ as a columnwise orthogonal basis of $\mathbf{Z}_{(n)}$ by using the QR decomposition and let $\mathbf{Q}_n = \mathbf{Q}_n(:, 1 : R_n)$.
5: Set $C = C \times \mathbf{Q}_n^\top$ and let $I_n = R_n$.
6: **end for**

For each n, choosing the random matrix $\mathbf{\Omega}_{(n)}$ in the Randomized Tucker decomposition requires $O(I^{N-1}K)$ additional space; drawing the random matrix $\mathbf{\Omega}_{(n)}$ in Algorithm 8.2.2 requires $O((N-1)IK)$ additional space, where we assume that $R_1 \sim R_2 \sim \cdots \sim R_N \sim K$ for clarity. Comparing with the Randomized Tucker decomposition, when we implement Algorithm 8.2.2 to obtain a solution for Problem 8.2.1, the multilinear rank of the target tensor may be unknown. Before implementing Algorithm 8.2.2 to compute the low multilinear rank approximation of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, we choose N random Gaussian matrices $\widehat{\mathbf{U}}_n$ and precondition $\mathcal{A}$ as $\mathcal{A} := \mathcal{A} \times_1 \widehat{\mathbf{U}}_1 \times_2 \widehat{\mathbf{U}}_2 \cdots \times_N \widehat{\mathbf{U}}_N$. This strategy is firstly considered in [32].

According to probabilistic error bounds for Algorithm 4.1 in [1], the derivation bounds for the Frobenius norm and average Frobenius error for the Randomized Tucker decomposition is easy to obtain. However, we can not use the derivation bounds for the Frobenius norm for Algorithm 4.1 [1] directly to analyze Algorithm 8.2.2, which will be considered in the next section.

8.2.2 More Considerations

As shown in Algorithm 8.2.3, our analysis is specialized to the case where the test matrices $\mathbf{\Omega}'_m$ are standard Gaussian matrices. But there are potential benefits from implementing the proposed algorithms using test matrices drawn from another distribution. The choice of distribution leads to some tradeoffs in the range of permissible parameters; the costs of randomness, arithmetic, and communication to generate the test matrices; the storage costs for the test matrices and the sketch; the arithmetic costs for sketching and updates; the numerical stability of matrix approximation algorithms; and the quality of a priori error bounds.

Let us list some of the contending distributions along with background references. We have ranked these in decreasing order of reliability.

(a) For each n, Step 3 in Algorithm 8.2.3 generates matrices $\mathbf{\Omega}'_m$ with orthonormal columns that span uniformly random subspaces of dimension $R_m + 10$.

(b) This chapter focuses on test matrices with the standard Gaussian matrices. Benefits include excellent practical performance and accurate a priori error bounds.

When we consider the low multilinear rank approximation of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, for each n, the test matrices $\mathbf{\Omega}'_m$ have the form $\mathbf{\Omega}'_m = \mathbf{\Omega}'_{1m} + \iota \mathbf{\Omega}'_{2m}$, where $\mathbf{\Omega}'_{1m} \in \mathbb{R}^{I_m \times (R_m+10)}$ and $\mathbf{\Omega}'_{2m} \in \mathbb{R}^{I_m \times (R_m+10)}$ are standard Gaussian matrices.

(c) These test matrices have independent Rademacher[1] entries. Their behavior is similar to Gaussian test matrices, but there are minor improvements in the cost of storage and arithmetic, as well as the amount of randomness required.

[1] A Rademacher random variable takes the values ± 1 with equal probability.

(d) For each n, the test matrices $\mathbf{\Omega}'_m$ take the form

$$\mathbf{\Omega}'_m = \mathbf{DHP},$$

where $\mathbf{D} \in \mathbb{R}^{I_m \times I_m}$ are diagonal with independent Rademacher variables, $\mathbf{H} \in \mathbb{R}^{I_m \times I_m}$ is orthonormal and $\mathbf{P} \in \mathbb{R}^{I_m \times (R_m+10)}$ are restrictions onto $R_n + 10$ coordinates, chosen uniformly at random. In practice, one can choose $\mathbf{H}$ with efficient fast computationable algorithms, such as the discrete cosine transform matrix, the Walsh-Hadamard transform matrix, the fast Fourier transform matrix or the discrete wavelet transform matrix.

(e) Let s be a sparsity parameter. For each n, each column of $\mathbf{\Omega}'_m$ is placed independent Rademacher random variables in s uniformly random locations; the remaining entries of the test matrices are zero. This type of test matrices is sometime called a sparse embedding matrix.

The interested readers can refer to [1, 45, 46] on the computation of low-rank approximations via randomized algorithms based on the above test matrices. In the future, based on these test matrices, we will consider how to improve our proposed algorithms.

8.3 Theoretical Analysis

We prove Theorem 8.3.1, which is the essential basis for analyzing Algorithm 8.2.2. We introduce a particular type of random matrices for this: the row are independently distributed random vectors, but the columns are not (instead of the standard case where all entries are independently and identically distributed (i.i.d)). Such matrices are studied extensively by Vershynin [47] and we rely heavily on this research for our estimates.

8.3.1 Probabilistic Error Bounds

For each n and a given positive integer $R > 0$, let $\mathbf{\Omega}_{n,m} \in \mathbb{R}^{I_m \times (\mu_n + R)}$ with $\mu_n + R \leq I_n$ be standard Gaussian matrix with all m and $m \neq n$. Define $\mathbf{\Omega}_{(n)} = \mathbf{\Omega}_1 \odot \cdots \odot \mathbf{\Omega}_{n-1} \odot \mathbf{\Omega}_{n+1} \odot \cdots \odot \mathbf{\Omega}_N$. Suppose that $\mathbf{A}_{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^\top$, where $\mathbf{U}_n \in \mathbb{R}^{I_n \times I_n}$ is orthogonal, $\mathbf{\Sigma}_n \in \mathbb{R}^{I_n \times I_n}$ is diagonal whose entries are nonnegative and $\mathbf{V}_n \in \mathbb{R}^{I_1 \ldots I_{n-1} I_{n+1} \ldots I_N \times I_n}$ is columnwise orthogonal. Let $\mathbf{U}_n = (\mathbf{Q}_n, \mathbf{Q}_n^\perp)$ with $\mathbf{Q}_n \in \mathbb{R}^{I_n \times \mu_n}$ and $\mathbf{Q}_n^\perp \in \mathbb{R}^{I_n \times (I_n - \mu_n)}$, $\widetilde{\mathbf{\Sigma}}_n \in \mathbb{R}^{\mu_n \times \mu_n}$ and $\widetilde{\mathbf{\Sigma}}_n \in \mathbb{R}^{(I_n - \mu_n) \times (I_n - \mu_n)}$ be the upper-left and lower-right blocks of $\mathbf{\Sigma}_n$, respectively, and $\mathbf{V}_n = (\mathbf{P}_n, \mathbf{P}_n^\perp)$ with $\mathbf{P}_n \in \mathbb{R}^{I_1 \ldots I_{n-1} I_{n+1} \ldots I_N \times \mu_n}$ and $\mathbf{P}_n^\perp \in \mathbb{R}^{I_1 \ldots I_{n-1} I_{n+1} \ldots I_N \times (I_1 \ldots I_{n-1} I_{n+1} \ldots I_N - \mu_n)}$. Let $\mathbf{\Omega}_{n,1} = \mathbf{P}_n^\top \mathbf{\Omega}_{(n)}$ and $\mathbf{\Omega}_{n,2} = (\mathbf{P}_n^\perp)^\top \mathbf{\Omega}_{(n)}$.

Lemma 8.3.1 *For each n, $\mathbf{\Omega}_{n,1}^{\top} \in \mathbb{R}^{(\mu_n+R)\times\mu_n}$ is a random matrix with isotropic rows and $\mathbf{\Omega}_{n,2} \in \mathbb{R}^{I_1\ldots I_{n-1}I_{n+1}\ldots I_N-\mu_n\times(\mu_n+R)}$ is a random matrix with isotropic columns. Moreover, the columns of $\mathbf{\Omega}_{n,2}$ satisfy $\|\mathbf{\Omega}_{n,2}(:,r_n)\|_2 = \sqrt{I_1\ldots I_{n-1}I_{n+1}\ldots I_N-\mu_n}$ almost surely for $r_n = 1,2,\ldots,\mu_n+R$.*

This lemma is a general form of Lemma 8.3.3 with a similar proof. In order to present the probabilistic error for Algorithm 8.2.2, we introduce the defintion.

Definition 8.3.1 ([47, Definition 5.7]) Let X be a random variable that satisfies one of the three following equivalent properties:

1. $\mathscr{P}\{|X| > t\} \le \exp(1-t^2/K_1^2)$ for all $t \ge 0$,
2. $(\mathscr{E}\{|X|^p\})^{1/p} \le K_2\sqrt{p}$ for all $p \ge 1$,
3. $\mathscr{E}\{\exp(X^2/K_3^2)\} \le 1$,

where the constants K_i $(i = 1,2,3)$ differ from each other by at most an absolute constant factor (see [47, Lemma 5.5] for a proof of the equivalence of these properties). Then X is called a sub-Gaussian random variable. The sub-Gaussian norm of X is defined as the smallest K_2 in property 2, i.e.,

$$\|X\|_{\psi_2} = \sup_{p\ge 1}\frac{(\mathscr{E}\{|X|^p\})^{1/p}}{\sqrt{p}}.$$

We establish the following theorem to analyze the probabilistic error bound for Algorithm 8.2.2. In Sect. 8.3.3, we give a rigorous proof for this theorem.

Theorem 8.3.1 *Suppose that $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ where $I_n \le I_1\ldots I_{n-1}I_{n+1}\ldots I_N$ with all n. Let the value of "take_max" be "False" in Algorithm 8.2.2. For given positive integers $\mu_n > 0$ and $R > 0$ with $\mu_n + R \le I_n$, assume that $\mathbf{Q}_n \in \mathbb{R}^{I_n\times\mu_n}$ are derived by Algorithm 8.2.2. If for all n, $\mathbf{\Omega}_{n,1}$ have full row rank, then*

$$\left\|\mathcal{A}-\mathcal{A}\times_1\left(\mathbf{Q}_1\mathbf{Q}_1^{\top}\right)\times_2\left(\mathbf{Q}_2\mathbf{Q}_2^{\top}\right)\cdots\times_N\left(\mathbf{Q}_N\mathbf{Q}_N^{\top}\right)\right\|_F$$
$$\le\sum_{n=1}^{N}\sqrt{1+\frac{\sqrt{I_1\ldots I_{n-1}I_{n+1}\ldots I_N-\mu_n}+C_n\sqrt{\mu_n+R}+t}{\sqrt{\mu_n+R}-C_n'\sqrt{\mu_n}-t}}\,\Delta_{\mu_n+1}(\mathbf{A}_{(n)})$$

with probability at least $1-2\sum_{n=1}^{N}\left(\exp(-c_nt^2)+\exp(-c_n't^2)\right)$, where

$$\Delta_{\mu_n+1}(\mathbf{A}_{(n)}) = \left(\sum_{i_n=\mu_n+1}^{I_n}\sigma_{i_n}(\mathbf{A}_{(n)})^2\right)^{1/2},$$

and $\sigma_{i_n}(\mathbf{A}_{(n)})$ is the i_nth singular value of $\mathbf{A}_{(n)}$. Here $C_n = C_{K_n}$ and $c_n = c_{K_n} \ge 0$ depend only on the sub-Gaussian norm $K_n = \max_j\|\mathbf{\Omega}_{n,2}(:,j)\|_{\psi_2}$, and $C_n' = C_{K_n}$ and $c_n' = c_{K_n} \ge 0$ depend only on the sub-Gaussian norm $K_n = \max_j\|\mathbf{\Omega}_{n,1}(:,j)\|_{\psi_2}$ for all n.

Remark 8.3.1 In the above theorem, we need to assume that there exists t (> 0) such that

$$\sqrt{\mu_n + R} - C'_n\sqrt{\mu_n} - t > 0, \quad n = 1, 2, \ldots, N.$$

Remark 8.3.2 In practice, when we use Algorithm 8.2.2 to compute a low multilinear rank approximation of tensors in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, we restrict the approximation multilinear rank $\{\mu_1, \mu_2, \ldots, \mu_N\}$, that is, for each mode n in Algorithm 8.2.2, we shall give an upper bound for μ_n with all n.

8.3.2 *Some Results for Sub-Gaussian Matrices*

We introduce the definition of sub-Gaussian random variables and sub-Gaussian norm in Sect. 8.2, as shown in Definition 8.3.1. Examples of sub-Gaussian random variables include Gaussian and Bernoulli random variables. We present definitions for sub-Gaussian random vectors and their norm.

Definition 8.3.2 ([47, Definition 5.7]) A random vector $\mathbf{x} \in \mathbb{R}^I$ is called a sub-Gaussian random vector if $\langle \mathbf{x}, \mathbf{y} \rangle$ is a sub-Gaussian random variable for all $\mathbf{y} \in \mathbb{R}^I$. The sub-Gaussian norm of $\mathbf{x}$ is subsequently defined by

$$\|\mathbf{x}\|_{\psi_2} = \sup_{\mathbf{y} \in S_{I-1}} \|\langle \mathbf{x}, \mathbf{y} \rangle\|_{\psi_2}.$$

Definition 8.3.3 ([47, Definition 5.19]) A random vector $\mathbf{x} \in \mathbb{R}^I$ is called *isotropic* if its second moment matrix $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{x}) = \mathscr{E}\{\mathbf{x}\mathbf{x}^\top\}$ is equal to the identity matrix. This definition is equivalent to

$$\mathscr{E}\{\langle \mathbf{x}, \mathbf{y} \rangle^2\} = \|\mathbf{y}\|_2$$

for all $\mathbf{y} \in \mathbb{R}^I$.

The following theorem from [47] provides bounds on the condition numbers of matrices whose rows are independent sub-Gaussian isotropic random variables.

Theorem 8.3.2 ([47, Theorem 5.38]) *Let* $\mathbf{A}$ *be an* $I \times J$ *matrix whose rows* $\mathbf{A}(i, :)$ *are independent sub-Gaussian isotropic random vectors in* $\mathbb{R}^J$. *Then for every* $t \geq 0$, *with probability at least* $1 - 2\exp(-ct^2)$, *one has*

$$\sqrt{I} - C\sqrt{J} - t \leq \sigma_{\min}(\mathbf{A}) \leq \sigma_{\max}(\mathbf{A}) \leq \sqrt{I} + C\sqrt{J} + t.$$

Here $C = C_K$ *and* $c = c_K \geq 0$ *depend only on the sub-Gaussian norm* $K = \max_i \|\mathbf{A}(i, :)\|_{\psi_2}$.

Similarly, the following theorem from [47] provides bounds on the condition numbers of matrices whose columns are independent sub-Gaussian isotropic random variables.

Theorem 8.3.3 ([47, Theorem 5.58]) *Let* $\mathbf{A}$ *be an* $I \times J$ $(I \geq J)$ *matrix whose columns* $\mathbf{A}(:, j)$ *are independent sub-Gaussian isotropic random vectors in* $\mathbb{R}^I$ *with* $\|\mathbf{A}(:, j)\|_2 = \sqrt{I}$ *almost surely (a.s.). Then for every* $t \geq 0$*, with probability at least* $1 - 2\exp(-ct^2)$*, one has*

$$\sqrt{I} - C\sqrt{J} - t \leq \sigma_{\min}(\mathbf{A}) \leq \sigma_{\max}(\mathbf{A}) \leq \sqrt{I} + C\sqrt{J} + t.$$

Here $C = C_K$ *and* $c = c_K \geq 0$ *depend only on the sub-Gaussian norm* $K = \max_j \|\mathbf{A}(:, j)\|_{\psi_2}$.

8.3.3 *Proof of Theorem 8.3.1*

For clarity, suppose that $n = 1$ and R is a positive integer. Let $\mathbf{\Omega}_m \in \mathbb{R}^{I_m \times (\mu_1 + R)}$ $(\mu_1 + R \leq I_1)$ be a standard Gaussian matrix with $m = 2, 3, \ldots, N$. Define $\mathbf{\Omega}_{(1)} = \mathbf{\Omega}_2 \odot \mathbf{\Omega}_3 \odot \cdots \odot \mathbf{\Omega}_N$. As seen in [33], we see that $\mathbf{\Omega}_{(1)}$ is a random matrix whose columns are independent from one another but whose rows are not. In the following lemma (similar to [33, Lemma 11] without the proof), we show that each column of $\mathbf{\Omega}_{(1)}$ is isotropic.

Lemma 8.3.2 *Let* $\mathbf{\Omega}'_m \in \mathbb{R}^{I_m \times (\mu_1 + R)}$ $(\mu_1 + R \leq I_1)$ *be standard Gaussian matrix with* $m = 2, 3, \ldots, N$*, and* $\mathbf{Q} \in \mathbb{R}^{I_2 \ldots I_N \times R_Q}$ *be columnwise orthogonal with* $R_Q < \mu_1 + R < I_2 \ldots I_N$*. Define* $\mathbf{\Omega}_{(1)} = \mathbf{\Omega}'_2 \odot \mathbf{\Omega}'_3 \odot \cdots \odot \mathbf{\Omega}'_N$*. Then* $\mathbf{Q}^\top \mathbf{\Omega}_{(1)}$ *is a random matrix with isotropic columns.*

Remark 8.3.3 According to Lemma 8.3.2, it is easy to see that $\|\mathbf{Q}^\top \mathbf{\Omega}_{(1)}(:, r_1)\|_2 = \sqrt{R_Q}$ almost surely with $r_1 = 1, 2, \ldots, \mu_1 + R$.

Suppose that $\mathbf{A}_{(1)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{I_1 \times I_1}$ is orthogonal, $\mathbf{\Sigma} \in \mathbb{R}^{I_1 \times I_1}$ is diagonal whose entries are nonnegative and $\mathbf{V} \in \mathbb{R}^{I_2 \ldots I_N \times I_1}$ is columnwise orthogonal. Let $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_1^\perp)$ with $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times \mu_1}$ and $\mathbf{U}_1^\perp \in \mathbb{R}^{I_1 \times (I_1 - \mu_1)}$, $\mathbf{\Sigma}_1 \in \mathbb{R}^{\mu_1 \times \mu_1}$ and $\mathbf{\Sigma}_2 \in \mathbb{R}^{(I_1 - \mu_1) \times (I_1 - \mu_1)}$ be the upper-left and lower-right blocks of $\mathbf{\Sigma}$, respectively, and $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_1^\perp)$ with $\mathbf{V}_1 \in \mathbb{R}^{I_2 \ldots I_N \times \mu_1}$ and $\mathbf{V}_1^\perp \in \mathbb{R}^{I_2 \ldots I_N \times (I_2 \ldots I_N - \mu_1)}$. Let $\mathbf{\Omega}_1 = \mathbf{V}_1^\top \mathbf{\Omega}_{(1)}$ and $\mathbf{\Omega}_2 = (\mathbf{V}_1^\perp)^\top \mathbf{\Omega}_{(1)}$. From Lemma 8.3.2, it is easily to derive the following lemma.

Lemma 8.3.3 $\mathbf{\Omega}_1^\top \in \mathbb{R}^{(\mu_1 + R) \times \mu_1}$ *is a random matrix with isotropic rows and* $\mathbf{\Omega}_2 \in \mathbb{R}^{(I_2 \ldots I_N - \mu_1) \times (\mu_1 + R)}$ *is a random matrix with isotropic columns. Moreover, the columns of* $\mathbf{\Omega}_2$ *satisfy* $\|\mathbf{\Omega}_2(:, r_1)\|_2 = \sqrt{I_2 \ldots I_N - \mu_1}$ *almost surely for* $r_1 = 1, 2, \ldots, \mu_1 + R$.

The main result is presented in the following theorem.

Theorem 8.3.4 *Suppose that* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *with* $I_1 \leq I_2 \ldots I_N$. *Let the value of "take_max" be "False" in Algorithm 8.2.1. For given positive integers* $\mu_1 > 0$ *and* $R > 0$ *with* $\mu_1 + R \leq I_1$, *suppose that* $\mathbf{Q}_1 \in \mathbb{R}^{I_1 \times \mu_1}$ *is derived by Algorithm 8.2.1. If* $\mathbf{\Omega}_1$ *is of full row rank, then*

$$\left\| \mathcal{A} - \mathcal{A} \times_1 \left(\mathbf{Q}_1 \mathbf{Q}_1^\top \right) \right\|_F \leq \sqrt{1 + \frac{\sqrt{I_2 \ldots I_N - \mu_1} + C\sqrt{\mu_1 + R} + t}{\sqrt{\mu_1 + R} - C'\sqrt{\mu_1} - t}} \Delta_{\mu_1+1}(\mathbf{A}_{(1)})$$

with probability at least $1 - 2\exp(-ct^2) - 2\exp(-c't^2)$, *where*

$$\Delta_{\mu_1+1}(\mathbf{A}_{(1)}) = \left(\sum_{i_1=\mu_1+1}^{I_1} \sigma_{i_1}(\mathbf{A}_{(1)})^2 \right)^{1/2}.$$

Here $C = C_K$ *and* $c = c_K \geq 0$ *depend only on the sub-Gaussian norm* $K = \max_i \|\mathbf{\Omega}_2(:, i)\|_{\psi_2}$; $C' = C_K$ *and* $c' = c_K \geq 0$ *depend only on the sub-Gaussian norm* $K = \max_i \|\mathbf{\Omega}_1(:, i)\|_{\psi_2}$.

Proof To obtain the matrix $\mathbf{Q}_1$, Algorithm 8.2.1 is similar to the following process: (a) choose $\mathbf{\Omega}_{(1)} = \mathbf{\Omega}_2 \odot \mathbf{\Omega}_3 \odot \cdots \odot \mathbf{\Omega}_N$, where $\mathbf{\Omega}_m \in \mathbb{R}^{I_m \times (\mu_1 + R)}$ with $\mu_1 + R \leq I_1$ are standard Gaussian matrix with $m = 2, 3, \ldots, N$; (b) compute $\mathbf{Y} = \mathbf{A}_{(1)}\mathbf{\Omega}_{(1)}$; (c) factor $\mathbf{QR} = \mathbf{Y}$ by the QR decomposition; (d) form $\mathbf{Q}_1 = \mathbf{Q}(:, 1 : \mu_1)$.

According to Theorem 9.1 in [1], we have

$$\left\| \mathcal{A} - \mathcal{A} \times_1 (\mathbf{Q}_1 \mathbf{Q}^\top) \right\|_F = \left\| (\mathbf{I}_{I_1} - \mathbf{Q}_1 \mathbf{Q}^\top) \mathbf{A}_{(1)} \right\|_F \leq \sqrt{1 + \|\mathbf{\Omega}_2\|_2 \left\| \mathbf{\Omega}_1^\dagger \right\|_2} \|\mathbf{\Sigma}_2\|_F.$$

Note that $\|\mathbf{\Sigma}_2\|_F = (\sum_{i_1=\mu_1+1}^{I_1} \sigma_{i_1}(\mathbf{A}_{(1)})^2)^{1/2}$. Now we estimate $\|\mathbf{\Omega}_2\|_2$ and $\|\mathbf{\Omega}_1^\dagger\|_2$. According to Theorem 8.3.3 and Lemma 8.3.3, with probability at least $1 - 2\exp(-ct^2)$, we have

$$\|\mathbf{\Omega}_2\|_2 \leq \sqrt{I_2 \ldots I_N - \mu_1} + C\sqrt{\mu_1 + R} + t,$$

for $t > 0$, where $C = C_K$ and $c = c_K \geq 0$ depend only on the sub-Gaussian norm $K = \max_j \|\mathbf{\Omega}_2(:, j)\|_{\psi_2}$.

According to Theorem 8.3.2 and Lemma 8.3.3, with probability at least $1 - 2\exp(-c't^2)$, we have

$$\left\| \mathbf{\Omega}_1^\dagger \right\|_2 = \left\| (\mathbf{\Omega}_1^\dagger)^\top \right\|_2 = \left\| (\mathbf{\Omega}_1^\top)^\dagger \right\|_2 \leq \frac{1}{\sqrt{\mu_1 + R} - C'\sqrt{\mu_1} - t},$$

for $t > 0$, where $C' = C_K$ and $c' = c_K \geq 0$ depend only on the sub-Gaussian norm $K = \max_i \|(\mathbf{\Omega}_1)^\top(i, :)\|_{\psi_2} = \max_i \|\mathbf{\Omega}_1(:, i)\|_{\psi_2}$. □

The rigorous proof for Theorem 8.3.1 comes from Theorem 8.3.4 and (8.2.2).

8.4 Randomized Tensor Train Approximation

In Sect. 8.2, we present an adaptive randomized algorithm for a low multilinear rank approximation with unknown multiliear rank. In Sect. 8.3, we make the theoretical analysis for the algorithm. We next develop an adaptive randomized algorithm for approximate tensor train decompositions of tensors with unknown TT-rank.

We briefly introduce the idea of TT-SVD for computing the approximate tensor train decompositions. Suppose that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and ϵ is the prescribed tolerance, the process [21, Algorithm 1] for computing an approximate tensor train decomposition of the tensor $\mathcal{A}$ is as follows:

Compute truncation parameter $\delta = \frac{\epsilon}{\sqrt{N-1}} \|\mathcal{A}\|_F$.
Set the temporary matrix: $\mathbf{C} = \mathbf{A}_{(1)}$ and $\mu_0 = 1$.
for $n = 1, 2, \ldots, N-1$ **do**
  Form $\mathbf{C} = \text{reshape}(\mathbf{C}, [\mu_{n-1} I_n, I_{n+1} \ldots I_N])$.
  Compute the δ-truncated SVD: $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top + \mathbf{E}$ with $\|\mathbf{E}\|_F \leq \delta$ and $\mu_n = \text{rank}_\delta(\mathbf{C})$.
  Form $\mathcal{Q}^{(n)} = \text{reshape}(\mathbf{U}, [\mu_{n-1}, I_n, \mu_n])$ and $\mathbf{C} = \mathbf{\Sigma}\mathbf{V}^\top$.
end for
Form $\mathcal{Q}^{(N)} = \text{reshape}(\mathbf{C}, [\mu_{N-1}, I_N, \mu_N])$ with $\mu_N = 1$.
Return $\mathcal{B}$ in TT-format with cores $\mathcal{Q}^{(1)}, \mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(N)}$.

When we implement the above process, the approximate tensor train decomposition satisfies

$$\|\mathcal{A} - \mathcal{B}\|_F \leq \epsilon \|\mathcal{A}\|_F.$$

Similar to the Randomized Tucker decomposition [32], we consider the randomized range finder algorithm (Algorithm 4.1 in [1]) to estimate the basis of the temporary matrix $\mathbf{C}$ for each n in the above process, summarized as Algorithm 8.4.1.

Algorithm 8.4.1 Randomized algorithm for tensor train approximation

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and the TT-rank $\{\mu_1, \mu_2, \ldots, \mu_{N-1}\}$.
output: Cores $\mathcal{Q}^{(1)}, \mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(N)}$ of the TT-approximation $\mathcal{B}$ of the tensor $\mathcal{A}$
Set the temporary tensor: $\mathcal{C} = \mathcal{A}$ and $\mu_0 = 1$.
for $n = 1, 2, \ldots, N-1$ **do**
  Form $\mathbf{C} = \text{reshape}(\mathcal{C}, [\mu_{n-1} I_n, I_{n+1} \ldots I_N])$.
  Compute $\mathbf{Z} = \mathbf{C}\mathbf{\Omega}$, where $\mathbf{\Omega} \in \mathbb{R}^{I_{n+1} \ldots I_N \times (\mu_n + R)}$ is a random Gaussian matrix.
  Compute $\mathbf{U}$ as a columnwise orthogonal basis of $\mathbf{Z}$ by using the QR decomposition.
  Let $\mathbf{U} = \mathbf{U}(:, 1 : \mu_n)$.
  Form $\mathcal{Q}^{(n)} = \text{reshape}(\mathbf{U}, [\mu_{n-1}, I_n, \mu_n])$.
  Compute $\mathcal{C} = \mathcal{C} \times_1 \mathbf{U}^\top$.
end for
Form $\mathbf{C} = \text{reshape}(\mathcal{C}, [\mu_{N-1}, I_N, \mu_N])$ and $\mathcal{Q}^{(N)} = \text{reshape}(\mathbf{C}, [\mu_{N-1}, I_N, \mu_N])$ with $\mu_N = 1$.
Return $\mathcal{B}$ in TT-format with cores $\mathcal{Q}^{(1)}, \mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(N)}$.

In Algorithm 8.4.1, R is referred to as the oversampling parameter. In general, we set $R = 5$ or 10. Furthermore, the probabilistic error bound of Algorithm 8.4.1 is easy to obtain, due to the theoretical results for Algorithm 4.1 in [1]. The disadvantage of Algorithm 8.4.1 is that we need to know the TT-rank of the target tensors. In the remainder of this section, we design an adaptive randomized algorithm for an approximate tensor train decomposition of a tensor in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with unknown TT-rank. The probabilistic error bound of this approximate tensor train decomposition can be derived from Theorem 8.3.3.

If we add the orthogonal constraints to Problem 8.1.2 as $\mathbf{Q}_n^\top \mathbf{Q}_n = \mathbf{I}_{\mu_n}$, where $\mathbf{Q}_n = \text{reshape}(\mathcal{Q}^{(n)}, [\mu_{n-1} I_n, \mu_n])$ with $n = 1, 2, \ldots, N-1$, then we rewrite Problem 8.1.2 as follows.

Problem 8.4.1 Suppose that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and ϵ is a prescribed tolerance. Find N core tensors $\mathcal{Q}^{(n)} \in \mathbb{R}^{\mu_{n-1} \times I_n \times \mu_n}$ such that

$$\|\mathcal{A} - \mathcal{B}\|_F := \left\| \mathcal{A} - \mathcal{Q}^{(1)} \times_3^1 \mathcal{Q}^{(2)} \cdots \times_3^1 \mathcal{Q}^{(N)} \right\|_F \le \sqrt{N-1}\epsilon, \tag{8.4.1}$$

with $\mu_n \le \min\{I_1 \ldots I_n, I_{n+1} \ldots I_N\}$ and $\mu_0 = \mu_N = 1$, where the core tensors $\mathcal{Q}^{(n)} \in \mathbb{R}^{\mu_{n-1} \times I_n \times \mu_n}$ satisfy

$$\mathbf{Q}_n^\top \mathbf{Q}_n = \mathbf{I}_{\mu_n}, \quad \mathbf{Q}_n = \text{reshape}(\mathcal{Q}^{(n)}, [\mu_{n-1} I_n, \mu_n]), \quad n = 1, 2, \ldots, N-1.$$

When the TT-rank of $\mathcal{A}$ is unknown, Algorithm 8.4.2 is summarized as the adaptive randomized algorithm for solving Problem 8.4.1.

Algorithm 8.4.2 Adaptive randomized algorithm for TT-approximation

Input: A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, a positive integer R and the prescribed tolerance ϵ.
output: Cores $\mathcal{Q}^{(1)}, \mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(N)}$ of the TT-approximation $\mathcal{B}$ of the tensor $\mathcal{A}$ with approximate TT-rank $\{\mu_1, \mu_2, \ldots, \mu_{N-1}\}$.

1: Set the temporary tensor $\mathcal{C} = \mathcal{A}$ and $\mu_0 = 1$.
2: **for** $n = 1, 2, \ldots, N-1$ **do**
3: Implement Algorithm 8.2.1 with the tensor $\mathcal{C}$ to generate the columnwise orthogonal matrix $\mathbf{Q}_k \in \mathbb{R}^{I_n \mu_{n-1} \times k}$.
4: Set $\mu_n = k$ and form $\mathcal{Q}^{(n)} = \text{reshape}(\mathbf{Q}_k, [\mu_{n-1}, I_n, \mu_n])$.
5: Compute $\mathcal{C} = \mathcal{C} \times_1 \mathbf{U}^\top$.
6: **end for**
7: Form $\mathcal{Q}^{(N)} = \text{reshape}(\mathcal{C}, [\mu_{N-1}, I_N, \mu_N])$ with $\mu_N = 1$.
8: Return $\mathcal{B}$ in TT-format with cores $\mathcal{Q}^{(1)}, \mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(N)}$.

For clarity, we assume $\mu_1 \sim \mu_2 \sim \cdots \sim \mu_{N-1} \sim \mu$ for Problem 8.1.2 in complexity estimates. Algorithm 8.4.2 requires $(N-1)(R+\mu)$ tenvecs and $O((N-1)I\mu^2)$ additional operations. For a given tolerance ϵ, if the columnwise orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{\mu_{n-1} I_n \times \mu_n}$ satisfies

$$\left\| \mathcal{C}_n - \mathcal{C}_n \times_1 (\mathbf{Q}\mathbf{Q}^\top) \right\|_F^2 \le \epsilon^2,$$

where $C_n \in \mathbb{R}^{\mu_{n-1} I_n \times I_{n+1} \times \cdots \times I_N}$ is generated from Steps 2 to 6 in Algorithm 8.4.2, for $n = 1, 2, \ldots, N-1$, then the TT-approximation $\mathcal{B}$ with $\mathcal{Q}^{(n)} \in \mathbb{R}^{\mu_{n-1} \times I_n \times \mu_n}$ satisfies

$$\|\mathcal{A} - \mathcal{B}\|_F := \left\| \mathcal{A} - \mathcal{Q}^{(1)} \times_3^1 \mathcal{Q}^{(2)} \times_3^1 \cdots \times_3^1 \mathcal{Q}^{(N)} \right\|_F \leq \sqrt{N-1}\epsilon,$$

with $\mu_0 = \mu_N = 1$.

Remark 8.4.1 In practice, when we apply Algorithm 8.4.2, we restrict the approximation TT-rank $\{\mu_1, \mu_2, \ldots, \mu_{N-1}\}$, that is, for each n in Algorithm 8.4.2, we shall give an upper bound for μ_n with $n = 1, 2, \ldots, N-1$.

8.5 Numerical Examples

In order to test the algorithms in this chapter, we choose $R = 10$. We use these three functions "ttv", "ttm" and "ttt" in [48] to implement the tensor-vector, the tensor-matrix and the tensor-tensor products, respectively.

8.5.1 Low Multilinear Rank Approximations

We compare Algorithm 8.2.2 with several known algorithms for computing low multilinear rank approximations of a tensor via three examples. These algorithms are:

- Tucker-ALS: higher-order orthogonal iteration [48];
- mlsvd: truncated multilinear singular value decomposition [44];
- lmlra_aca: low multilinear rank approximation by adaptive cross-approximation [11, 49] and ;
- mlsvd_rsi: truncated multilinear singular value decomposition [44] by a randomized SVD algorithm based on randomized subspace iteration [1].

The first two examples are chosen from Caiafa and Cichocki [11]. The form of the third example is similar to that of [32].

Example 8.5.1 We apply Algorithm 8.2.2 to the Tucker tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *(*given in Eq. (8.1.1)*)*, where their multilinear rank is $\{R, R, \ldots, R\}$, by randomly generating matrices $\mathbf{U}_n \in \mathbb{R}^{I_n \times R}$ and the core tensors $\mathcal{G} \in \mathbb{R}^{R \times R \times \cdots \times R}$ for different R.

The entries are generated by using Gaussian independent identically distributed with zero mean and the resulting tensor $\mathcal{A}$ is normalized, i.e., $\mathcal{A} := \mathcal{A}/\|\mathcal{A}\|_F$.

Table 8.1 The relative errors derived by Algorithm 8.2.2 for the exact representation case of the random Tucker tensors with a given multilinear rank $\{R, R, R\}$ for Example 8.5.1

Rank of $\mathcal{A}$	Rank of $\widehat{\mathcal{A}}$	Relative error	CPU times
10	10	4.4e−15	0.19 s
20	20	4.2e−15	0.25 s
30	30	1.8e−14	0.36 s
40	40	7.8e−15	0.52 s
50	50	8.3e−15	0.65 s
60	60	9.4e−15	0.86 s
70	70	1.4e−14	1.1 s
80	80	1.1e−14	1.3 s

The relative error is defined as

$$e = \left\| \mathcal{A} - \widehat{\mathcal{A}} \right\|_F / \|\mathcal{A}\|_F, \tag{8.5.1}$$

where $\widehat{\mathcal{A}} = \mathcal{A} \times_1 (\mathbf{Q}_1\mathbf{Q}_1^\top) \times_2 (\mathbf{Q}_2\mathbf{Q}_2^\top) \times_3 \cdots \times_N (\mathbf{Q}_N\mathbf{Q}_N^\top)$ and the columnwise orthogonal matrices $\mathbf{Q}_n$ are derived form our Algorithm 8.2.2, Tucker-ALS, mlsvd, or lmlra_aca.

For $N = 3$ and $\{I_1, I_2, I_3\} = \{100, 100, 100\}$, suppose that the value of R is chosen from the set $\{10, 20, \ldots, 80\}$. When we apply Algorithm 8.2.2 to compute the low multilinear rank approximation of the tensor $\mathcal{A}$ with different R, the relative errors are shown in Table 8.1. The relative errors were under 10^{-14}, relative to the maximal attainable machine precision 10^{-16}.

We compare Algorithm 8.2.2 with other algorithms (Tucker-ALS [48], mlsvd and lmlra_aca [49]) to compute the approximate tensor $\mathcal{B}$ of $\mathcal{A}$, where the multilinear rank $\{P, P, P\}$ of $\mathcal{B}$ satisfies $1 \le P \le R = 30$. For a given positive integer P, we find three columnwise orthogonal matrices $\mathbf{Q}_n \in \mathbb{R}^{I_n \times P}$ by Algorithm 8.2.2 with $n = 1, 2, 3$.

The relative errors given in (8.5.1), and the CPU time of Algorithm 8.2.2, Tucker-ALS, mlsvd, and lmlra_aca, applied to a multilinear rank $\{30, 30, 30\}$ random Tucker tensor, are shown in Fig. 8.1 for $P = 1, 2, \ldots, 30$.

Example 8.5.2 The entries of the testing tensors are generated by sampling some smooth functions. We analyze the case of applying our Algorithm 8.2.2, *Tucker-ALS*, *mlsvd* and *lmlra_aca* to find the low multilinear rank approximations for these testing tensors.

We note that the approximate multilinear rank $\{P, P, \ldots, P\}$ is selected from a set of integers and the multilinear rank of the testing tensors is unknown.

We set $N = 3$ and $\{I_1, I_2, I_3\} = \{200, 200, 200\}$. We consider two tensors generated by sampling two families of smooth functions, respectively, as follows:

$$a_{ijk} = \frac{1}{i + j + k}, \quad b_{ijk} = \frac{1}{\ln(i + 2j + 3k)}.$$

The tensor $\mathcal{A}$ is chosen from [11].

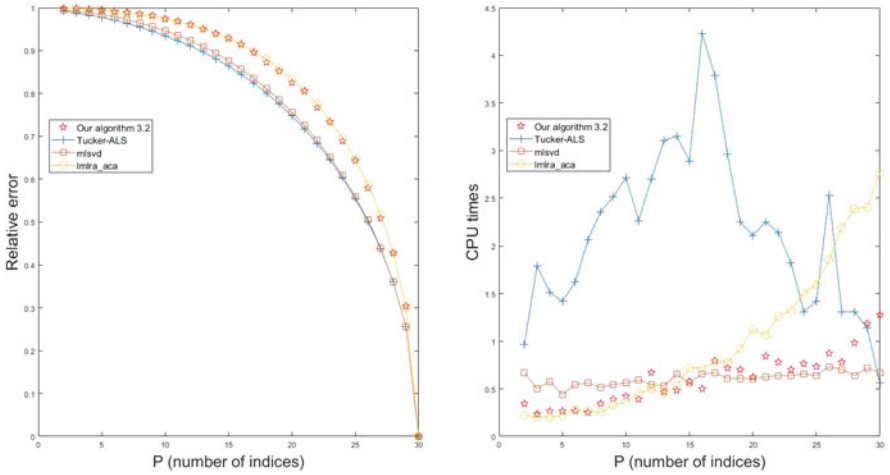

Fig. 8.1 Example 8.5.1 applying Algorithm 8.2.2 to the order-3 tensors ($I_1 = I_2 = I_3 = 100$): a randomly generated multilinear rank {30, 30, 30} Tucker tensor. The comparisons with the relative error and CPU times achieved by Tucker-ALS, mlsvd and lmlra_aca are shown

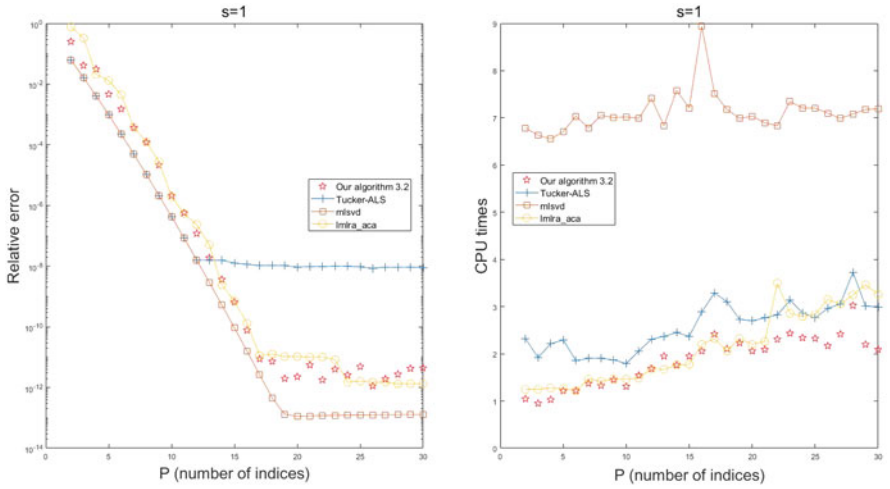

Fig. 8.2 Example 8.5.2 applying Algorithm 8.2.2, Tucker-ALS, mlsvd and lmlra_aca to the tensor $\mathcal{A}$

In Figs. 8.2 and 8.3 respectively, the relative errors given in (8.5.1), and the CPU times of Algorithm 8.2.2, Tucker-ALS, mlsvd and lmlra_aca, applied to $\mathcal{A}$ and $\mathcal{B}$, are shown for $P = 1, 2, \ldots, 30$.

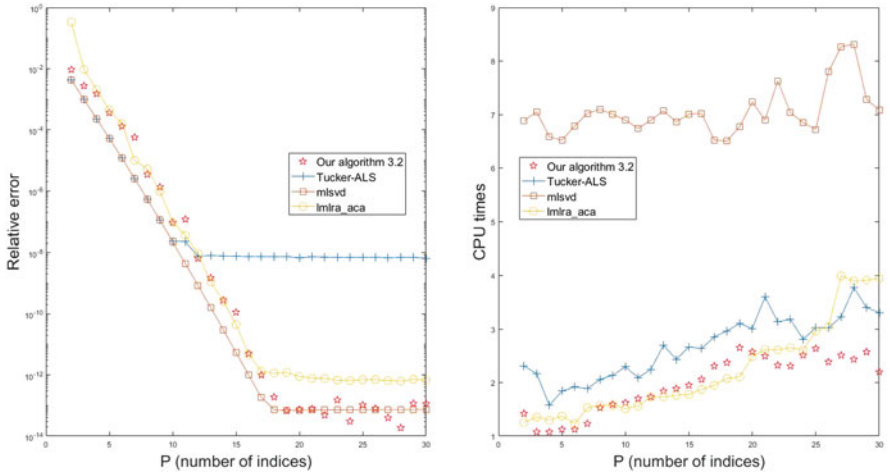

Fig. 8.3 Example 8.5.2 applying Algorithm 8.2.2, Tucker-ALS, mlsvd and lmlra_aca to the tensor $\mathcal{B}$

Example 8.5.3 We present the results of Algorithm 8.2.2 to $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The form of $\mathcal{A}$ is given as $\mathcal{A} = \mathcal{B} + \beta \mathcal{N}$, where $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\mathcal{B} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_N \mathbf{U}_N,$$

with known multilinear rank $\{R_1, R_2, \ldots, R_N\}$, and $\mathcal{N}$ is an unstructured perturbation tensor and β controls the noise level. The entries of the core tensor $\mathcal{G}$, mode-n factor matrix $\mathbf{U}_n$ and the perturbation tensor $\mathcal{N}$ are randomly drawn from a normal distribution with mean zero and variance unit.

The resulting tensor $\mathcal{A}$ is normalized, i.e., $\mathcal{A} := \mathcal{A}/\|\mathcal{A}\|_F$. The following signal-to-noise ratio (SNR) measure will be used:

$$\text{SNR [dB]} = 10 \log \left(\frac{\|\mathcal{B}\|_F^2}{\|\beta \mathcal{N}\|_F^2} \right).$$

Let $N = 3$, $\{I_1, I_2, I_3\} = \{200, 200, 200\}$ and $\{R_1, R_2, R_3\} = \{30, 30, 30\}$. When we use Algorithm 8.2.2 to find three columnwise orthogonal matrices $\mathbf{Q}_n \in \mathbb{R}^{I_n \times \mu_n}$, we have $\mu_n \geq R_n$. We assume that $\mu_n = R_n + 10$. The Fit value for approximating the tensor $\mathcal{A}$ is defined by

$$\text{Fit} = 1 - \left\| \mathcal{A} - \widehat{\mathcal{A}} \right\|_F / \|\mathcal{A}\|_F,$$

where $\widehat{\mathcal{A}} = \mathcal{A} \times_1 (\mathbf{Q}_1 \mathbf{Q}_1^\top) \times_2 (\mathbf{Q}_2 \mathbf{Q}_2^\top) \times_3 (\mathbf{Q}_3 \mathbf{Q}_3^\top)$ is an approximation to the tensor $\mathcal{A}$ and the columnwise orthogonal matrices $\mathbf{Q}_n$ are derived form Algorithm 8.2.2, Tucker-ALS, mlsvd, or lmlra_aca. The results of Algorithm 8.2.2, applied to the tensor $\mathcal{A}$ with different noise level, are shown in Fig. 8.4.

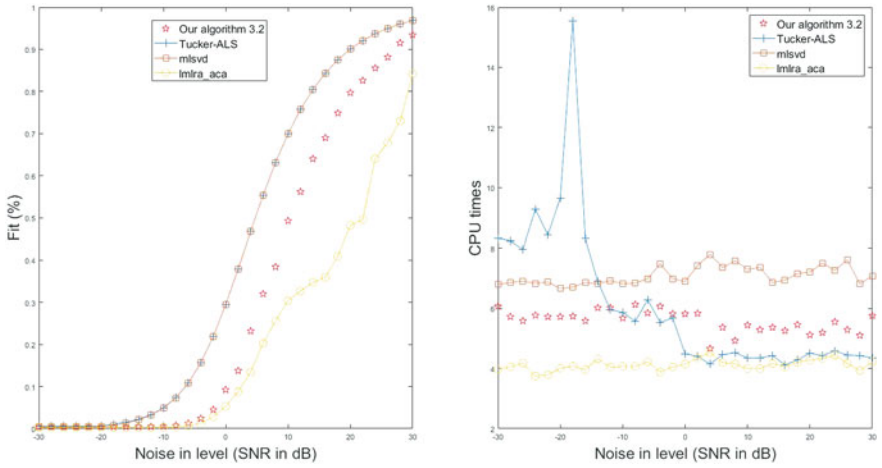

Fig. 8.4 Comparison between Algorithm 8.2.2, Tucker-ALS, mlsvd and lmlra_aca, in term of mean of Fit and the standard derivation for Example 8.5.3

Algorithm 8.2.2 is better than the existing algorithms. Figure 8.1 shows that the proposed algorithm is less accurate than and with similar speed to mlsvd. Figures 8.2, 8.3 and 8.4 show that the proposed algorithm is competitive with but not convincingly better than existing algorithms. When implementing Algorithm 8.2.2, we have not utilized higher-level linear algebra subroutines (e.g., BLAS3) or parallel processors.

Example 8.5.4 The entries of the test tensors are generated by sampling some smooth functions. We apply Algorithm 8.2.2, the Randomized Tucker decomposition [32, Algorithm 2] and mlsvd_rsi [49] to find the low multilinear rank approximations for these tensors. The multilinear rank of the tensors is unknown.

Set $N = 3$ and $\{I_1, I_2, I_3\} = \{200, 200, 200\}$. The test tensor $\mathcal{A}$ is defined in Example 8.5.2. In Fig. 8.5, the relative errors given in (8.5.1), and the CPU times of our Algorithm 8.2.2, the Randomized Tucker decomposition and mlsvd_rsi, applied to the tensor $\mathcal{A}$ with $s = 1$, are shown for $P = 1, 2, \ldots, 30$.

8.5.2 *Tensor Train Approximation*

We have illustrated that it is effective to use Algorithm 8.2.2 to compute the low multilinear rank approximations of tensors in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ via the some numerical examples. We give two examples to illustrate the efficiency of Algorithm 8.4.2 for the approximation tensor train decompositions.

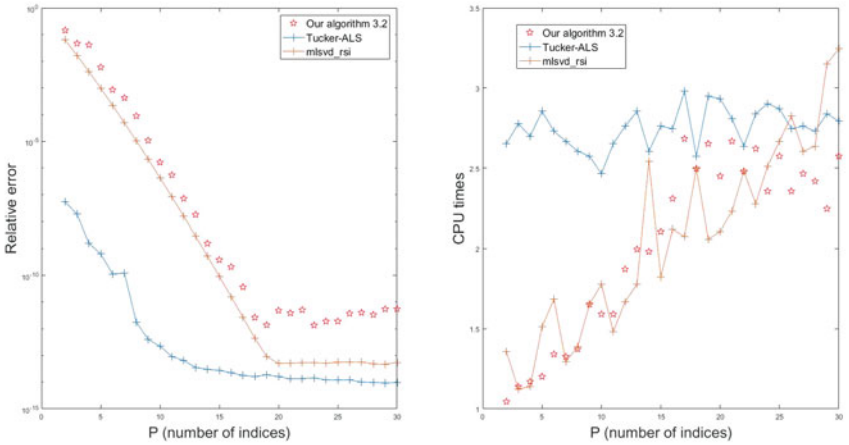

Fig. 8.5 Example 8.5.4 applying Algorithm 8.2.2, the Randomized Tucker decomposition and mlsvd_rsi to the tensor $\mathcal{A}$ with $s = 1$

Table 8.2 Example 8.5.5, CPU times with the same relative error level from Algorithm 8.4.2 and tt_tensor for tensors with generated randomly with TT-rank $\{R, R, R\}$

TT-rank	2	4	6	8	10	12
Relative error level	1e−15	1e−15	1e−14	1e−15	1e−13	1e−15
CPU times (Algorithm 3.2)	6.08 s	7.27 s	7.47 s	8.23 s	8.89 s	9.89 s
CPU times (tt_tensor)	53.38 s	61.48 s	53.09 s	82.73 s	61.16 s	106.52 s

Example 8.5.5 We apply Algorithm 8.4.2 to $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with the TT-format, given in (8.1.2), where their TT-rank is

$$\{R_1, R_2, \ldots, R_{N-1}\}$$

by the random tensors $\mathcal{Q}^{(n)} \in \mathbb{R}^{R_{n-1} \times I_n \times R_n}$ with $R_0 = R_N = 1$.

The entries are Gaussian independent identically distributed with zero mean and the resulting tensor $\mathcal{A}$ was normalized, i.e., $\mathcal{A} := \mathcal{A}/\|\mathcal{A}\|_F$.

Let $N = 4$ and $\{I_1, I_2, I_3, I_4\} = \{100, 100, 100, 100\}$. We assume the TT-rank $\{R_1, R_2, R_3\} = \{R, R, R\}$. When we apply Algorithm 8.4.2 and tt_tensor [49] for the approximate tensor train decomposition of the tensor $\mathcal{A}$, CPU times are shown in Table 8.2 with the same relative error level.

Furthermore, for a given order-4 tensor $\mathcal{A}$ with TT-rank $\{15, 15, 15\}$, by applying Algorithm 8.4.2 and tt_tensor to compute its approximate tensor train approximation with TT-rank $\{P, P, P\}$ $(P \leq R)$, CPU times with the same relative error level are shown in Table 8.3.

Table 8.3 Example 8.5.5 for $\mathcal{A}$ with TT-rank $\{R, R, R\}$, CPU times for the same relative error level from Algorithm 8.4.2 and tt_tensor for $P = 3, 4, \ldots, 15$

P	Relative error level	CPU time (Algorithm 3.2)	CPU time (tt_tensor)
3	1e−1	5.47 s	56.06 s
5	1e−1	6.22 s	55.03 s
7	1e−1	6.95 s	58.20 s
9	1e−1	7.86 s	56.77 s
11	1e−1	8.61 s	59.67 s
13	1e−1	12.23 s	65.56 s
15	1e−14	13.44 s	111.63 s

We compare Algorithm 8.4.2 with some existed algorithms for computing tensor train approximations via three examples. These algorithms are:

- tt_tensor: the tensor train approximation the SVDs of the matrices obtained from the corresponding unfoldings by a successive of projections [21, 50];
- dmrg_cross: DMRG-cross method for the tensor train approximation [50, 51] and;
- greedy2_cross: the tensor train approximation by the greedy cross interpolation scheme [50, 52].

Example 8.5.6 The entries of the test tensors are generated by sampling some smooth functions. For each test tensor, we use the function *"tt_tensor"* with accuracy 10^{-15} in [49] to derive its approximate TT-rank. Based on the approximate TT-rank, we apply Algorithm 8.4.2 to find the tensor train approximation. We also compare Algorithm 8.4.2 with *"tt_tensor"*, *"dmrg_cross"* and *"greedy2_cross"* for finding the tensor train approximation of the test tensors, in terms of CPU times.

The relative error for a tensor train approximation of a tensor in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as

$$e = \left\| \mathcal{A} - \widehat{\mathcal{A}} \right\|_F / \|\mathcal{A}\|_F, \tag{8.5.2}$$

where $\widehat{\mathcal{A}} = \mathcal{U}^{(1)} \times_3^1 \mathcal{U}^{(2)} \times_3^1 \cdots \times_3^1 \mathcal{U}^{(N)}$ and the the core tensors $\mathcal{U}^{(n)}$ are derived from Algorithm 8.4.2 or tt_tensor.

We assume that the test tensors are chosen from the set $\mathbb{R}^{I \times I \times I \times I \times I}$ with $I = 3, 4, \ldots, 30$. The entries of three test tensors $\mathcal{A}$ and $\mathcal{B}$ are

$$\begin{cases} a_{i_1 i_2 i_3 i_4 i_5} = (-1)^{i_1} \ln(i_1) + (-1)^{i_2} \ln(i_2) + (-1)^{i_3} \ln(i_3) + (-1)^{i_4} \ln(i_4) + (-1)^{i_5} \ln(i_5), \\ b_{i_1 i_2 i_3 i_4 i_5} = \sin(i_1 + i_2 + i_3 + i_4 + i_5), \end{cases}$$

where $i_n = 1, 2, \ldots, I$ and $n = 1, 2, \ldots, 5$. The test tensors are chosen from [53] and [54]. For each I, we can derive the approximate TT-ranks of $\mathcal{A}$ and $\mathcal{B}$ via the function "tt_tensor".

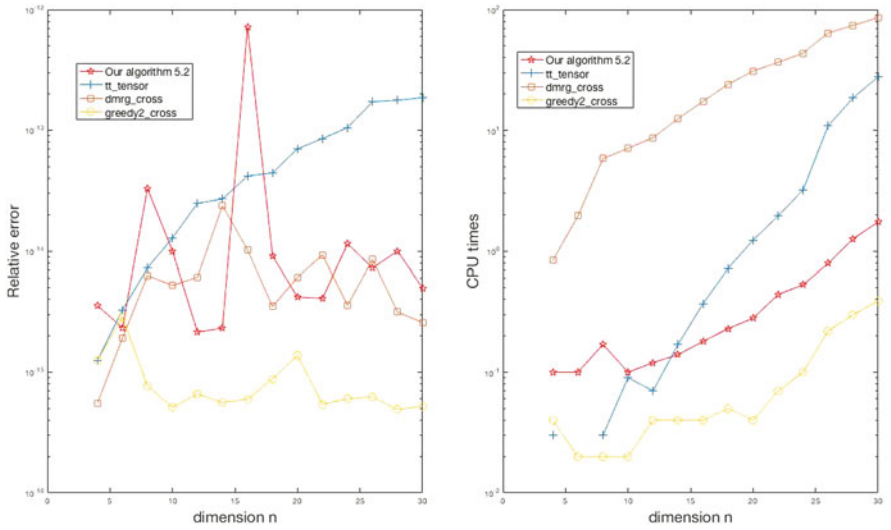

Fig. 8.6 Applying Algorithm 8.4.2, tt_tensor, "dmrg_cross" and "greedy2_cross" to the tensor $\mathcal{A}$ with unknown TT-rank for Example 8.5.6

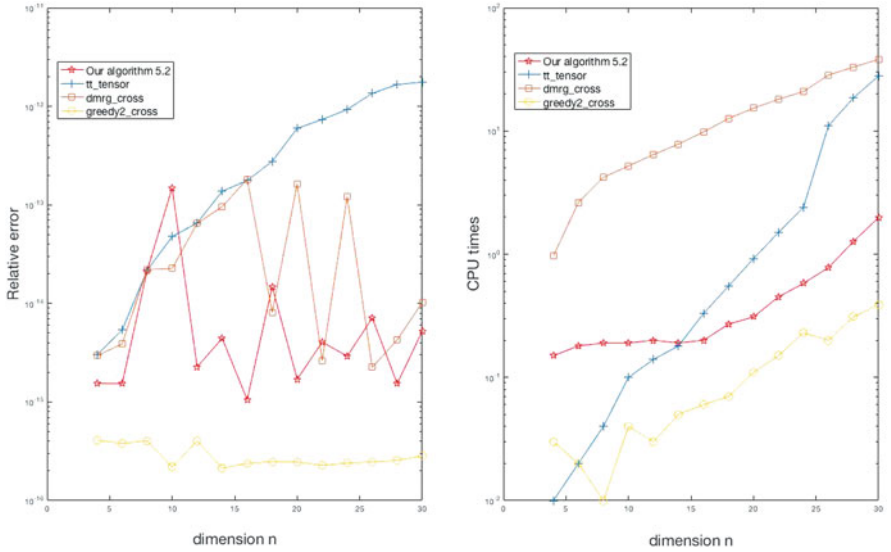

Fig. 8.7 Applying Algorithm 8.4.2, tt_tensor, "dmrg_cross" and "greedy2_cross" to the tensor $\mathcal{B}$ with unknown TT-rank for Example 8.5.6

In Figs. 8.6 and 8.7, the relative errors (8.5.2), and the CPU times of Algorithm 8.4.2, "tt_tensor", "dmrg_cross" and "greedy2_cross", applied to the tensors $\mathcal{A}$ and $\mathcal{B}$, are shown for $I = 3, 4, \ldots, 30$.

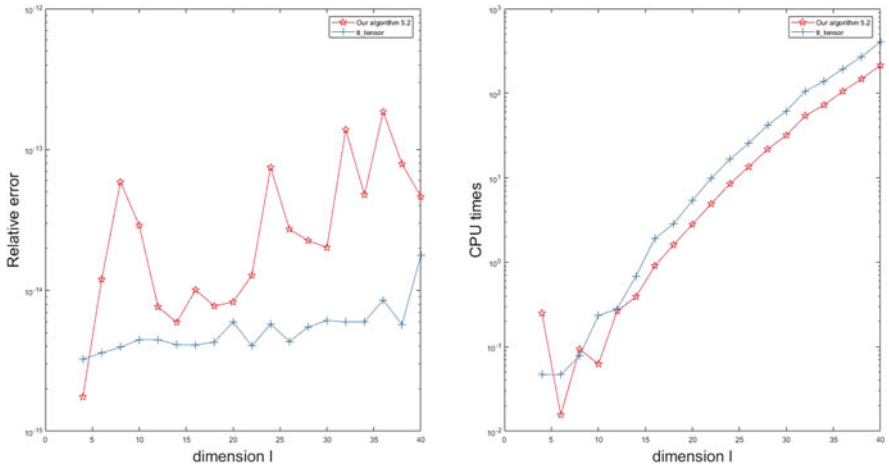

Fig. 8.8 Applying Algorithm 8.4.2 and tt_tensor to the sparse tensor $\mathcal{A}$ with 20 I nonzero entries, for $I = 5, 10, \ldots, 40$

Table 8.4 Example 8.5.7, CPU times and relative errors with the same approximate TT-rank from Algorithm 8.4.2 and tt_tensor for the sparse tensors $\mathcal{A}$ with $40K$ nonzero entries, for $K = 1, 10, 100, 1000, 1000$

The values of K	1	10	100	1000	10,000
CPU times (Algorithm 8.4.2)	234.89 s	217.43 s	222.86 s	219.30 s	229.74 s
Relative error (Algorithm 8.4.2)	1.32e−15	0.46e−15	8.65e−14	1.23e−10	4.58e−13
CPU times (tt_tensor)	408.16 s	394.15 s	416.48 s	385.17 s	393.56 s
Relative error (tt_tensor)	1.62e−15	0.65e−15	5.72e−15	1.10e−14	1.17e−14

Example 8.5.7 In the previous two examples, the test tensors are dense. In this example, we apply our Algorithm 8.4.2 to compute the tensor train decomposition of some sparse tensors with unknown TT-rank.

We set $I_n = I$ with $n = 1, 2, \ldots, 5$ and use the function "sptenrand" in the MATLAB Tensor Toolbox [48] to generate the test sparse tensors, with

$$\mathcal{A} = \text{sptenrand}([I, I, I, I, I], 40\,K)$$

creating a random sparse tensor $\mathcal{A}$ with approximately $40\,K$ nonzero entries, where K is a positive integer. The relative errors (8.5.2), and the CPU times of Algorithm 8.4.2 and tt_tensor, applied to the sparse tensors $\mathcal{A} \in \mathbb{R}^{I\times I\times I\times I\times I}$ with different I, where the number of nonzero entries in $\mathcal{A}$ is linear with respect to I, are shown in Fig. 8.8.

We assume that $I = 40$. For different positive integer K, when applying Algorithm 8.4.2 and tt_tensor to the sparse tensor $\mathcal{A} \in \mathbb{R}^{40\times 40\times 40\times 40\times 40}$ with $40\,K$ nonzero entries, CPU times and relative errors with the same approximate TT-rank are shown in Table 8.4.

Table 8.5 Example 8.5.8. For **deriv2** and **phillips**, $\varepsilon = 10^{-4}$; the another parameter in **deriv2** is 1

Matrix type	Num_c	CPU_t	Err	ANum_c	ACPU_t	AErr
baart	9	0.43 s	1.06e−11	9	0.34 s	1.05e−11
deriv2	207	8.20 s	9.00e−5	122	4.53 s	8.63e−5
foxgood	38	1.44 s	1.81e−11	24	0.97 s	2.67e−11
heat	507	23.44 s	2.76e−12	431	18.62 s	8.33e−11
phillips	360	17.02 s	9.46e−5	214	8.90 s	8.82e−5
shaw	27	0.99 s	1.27e−11	17	0.64 s	1.07e−11

As shown in Figs. 8.6 and 8.7, Algorithm 8.4.2 is better than tt_tensor and dmrg_cross, but worse than greedy2_cross. Note that Algorithm 8.4.2 have not utilized higher-level linear algebra subroutines (e.g., BLAS3) or parallel processors.

8.5.3 *Comparison Algorithm 8.2.1 with Different "take_max"*

Example 8.5.8 In this example, we analyze the difference between Algorithm 8.2.1 and Algorithm 8.2.1A (our Algorithm 8.2.1 with **Procedure A**) via some testing matrices. The testing examples **baart**, **deriv2**, **foxgood**, **heat**, **phillips**, and **shaw** are from Hansen Tools [34].

For a given $\mathbf{A} \in \mathbb{R}^{I\times J}$ with $I \geq J$, when we apply Algorithm 8.2.1 (or Algorithm 8.2.1A) to find a columnwise orthogonal matrix $\mathbf{Q}$ such that

$$\left\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^\top\mathbf{A}\right\|_F \leq \varepsilon, \tag{8.5.3}$$

we use "Num_c", "CPU_t" and "Err" (or "ANum_c", "ACPU_t" and "AErr") to denote the number of the columns of $\mathbf{Q}$, CPU times and error $\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^\top\mathbf{A}\|_F$, respectively. We assume that $\varepsilon = 10^{-10}$ and $I = J = 512$. The results are shown in Table 8.5.

8.6 Conclusions and Further Research

In this chapter, we develop several adaptive randomized algorithms for the low multilinear rank approximations and the tensor train approximations. The idea originates from randomized algorithms for computing partial matrix decompositions. For a given tensor in $\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, the main computational cost for randomized algorithms is the tensor-vector products.

As we know, the tensor train format is a special case of the Hierarchical Tucker format. We can use the idea of Algorithm 8.4.2 for the tensor train

approximation to design an adaptive randomized algorithm for the Hierarchical Tucker approximation, and its probabilistic error bound can be also derived from Theorem 8.3.3.

As illustrated by Example 8.5.8, Algorithm 8.2.1 with "take_max=True" is superior to Algorithm 8.2.1 with "take_max=False": CPU times and storage size. However, we can not prove that rigorously.

For the low multilinear rank approximations of real tensors, as shown in [55], in terms of CPU time, Tucker-SVD is faster than Algorithm 8.2.2; as shown in [56], in terms of relative error, Tucker-pSVD is better than Algorithm 8.2.2. More randomized algorithms for the low multilinear rank approximations of real tensors should be investigated in the near future.

References

1. N. Halko, P. Martinsson, J. Tropp, Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. **53**(2), 217–288 (2011)
2. L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(r_1, r_2, ..., r_n) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000)
3. L. Eldén, B. Savas, A Newton-Grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor. SIAM J. Matrix Anal. Appl. **31**(2), 248–271 (2009)
4. M. Ishteva, P. Absil, S. Van Huffel, L. De Lathauwer, Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. SIAM J. Matrix Anal. Appl. **32**(1), 115–135 (2011)
5. B. Savas, L. Lim, Quasi-Newton methods on Grassmannians and multilinear approximations of tensors. SIAM J. Sci. Comput. **32**(6), 3352–3393 (2010)
6. C. Navasca, L. De Lathauwer, Low multilinear rank tensor approximation via semidefinite programming, in *IEEE 17th European Signal Processing Conference*, pp. 520–524 (2009)
7. S. Goreinov, I. Oseledets, D. Savostyanov, Wedderburn rank reduction and Krylov subspace method for tensor approximation. Part 1: Tucker case. SIAM J. Sci. Comput. **34**(1), A1–A27 (2012)
8. B. Savas, L. Eldén, Krylov-type methods for tensor computations I. Linear Algebra Appl. **438**(2), 891–918 (2013)
9. L. Grasedyck, D. Kressner, C. Tobler, A literature survey of low-rank tensor approximation techniques. GAMM-Mitt. **36**(1), 53–78 (2013)
10. T. Kolda, B. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
11. C. Caiafa, A. Cichocki, Generalizing the column-row matrix decomposition to multi-way arrays. Linear Algebra Appl. **433**(3), 557–573 (2010)
12. P. Drineas, M. Mahoney, A randomized algorithm for a tensor-based generalization of the singular value decomposition. Linear Algebra Appl. **420**(2-3), 553–571 (2007)
13. S. Goreinov, E. Tyrtyshnikov, The maximal-volume concept in approximation by low-rank matrices. Struct. Matrices Math. Comput. Sci. Eng. I **280**, 47–51 (2001)
14. M. Mahoney, M. Maggioni, P. Drineas, Tensor-CUR decompositions for tensor-based data. SIAM J. Matrix Anal. Appl. **30**(3), 957–987 (2008)
15. I. Oseledets, D. Savostianov, E. Tyrtyshnikov, Tucker dimensionality reduction of three-dimensional arrays in linear time. SIAM J. Matrix Anal. Appl. **30**(3), 939–956 (2008)
16. I. Oseledets, D. Savostianov, E. Tyrtyshnikov, Cross approximation in tensor electron density computations. Numer. Linear Algebra Appl. **17**(6), 935–952 (2010)

17. P. Michael, H. Kathrin, Computing non-negative tensor factorizations. Optim. Methods Softw. **23**(4), 631–647 (2008)
18. Y. Zhang, G. X. Zhou, Q. Zhao, A. Cichocki, X. Wang, Fast nonnegative tensor factorization based on accelerated proximal gradient and low-rank approximation. Neurocomputing **198**, 148–154 (2016)
19. G. Zhou, A. Cichocki, S. Xie. Fast nonnegative matrix/tensor factorization based on low-rank approximation. IEEE Trans. Signal Proces. **60**(6), 2928–2940 (2012)
20. G. Zhou, A. Cichocki, Q. Zhao, S. Xie, Efficient nonnegative Tucker decompositions: Algorithms and uniqueness. IEEE Trans. Image Process. **24**(12), 4990–5003 (2015)
21. I. Oseledets, Tensor-train decomposition. SIAM J. Sci. Comput. **33**(5), 2295–2317 (2011)
22. S. Holtz, T. Rohwedder, R. Schneider, The alternating linear scheme for tensor optimization in the tensor train format. SIAM J. Sci. Comput. **34**(2), A683–A713 (2012)
23. S. Dolgov, B. Khoromskij, I. Oseledets, D. Savostyanov, Computation of extreme eigenvalues in higher dimensions using block tensor train format. Comput. Phys. Commun. **185**(4), 1207–1216 (2014)
24. D. Kressner, M. Steinlechner, A. Uschmajew, Low-rank tensor methods with subspace correction for symmetric eigenvalue problems. SIAM J. Sci. Comput. **36**(5), A2346–A2368 (2014)
25. N. Lee, A. Cichocki, Estimating a few extreme singular values and vectors for large-scale matrices in tensor train format. SIAM J. Matrix Anal. Appl. **36**(3), 994–1014 (2015)
26. I. Oseledets, S. Dolgov, Solution of linear systems and matrix inversion in the TT-format. SIAM J. Sci. Comput. **34**(5), A2718–A2739 (2012)
27. A. Cichocki, *Tensor Networks for Big Data Analytics and Large-Scale Optimization Problems* (2014). ArXiv preprint:1407.3124
28. G. Golub, C. Van Loan, *Matrix Computations*, 4th edn. (Johns Hopkins University Press, Baltimore, 2013)
29. T. Chan, Rank revealing QR factorizations. Linear Algebra Appl. **88/89**, 67–82 (1987)
30. N. Nguyen, P. Drineas, T. Tran, Tensor sparsification via a bound on the spectral norm of random tensors. Inf. Inference J. IMA **4**(3), 195–229 (2015)
31. N. Vervliet, O. Debals, L. Sorber, L. De Lathauwer, Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis. IEEE Signal Process. Mag. **31**(5), 71–79 (2014)
32. G. Zhou, A. Cichocki, S. Xie, Decomposition of big tensors with low multilinear rank (2014). ArXiv preprint:1412.1885v1
33. M. Reynolds, A. Doostan, G. Beylkin, Randomized alternating least squares for canonical tensor decompositions: Application to a PDE with random data. SIAM J. Sci. Comput. **38**(5), A2634–A2664 (2015)
34. P. Hansen, Regularization Tools version 4.0 for Matlab 7.3. Numer. Algorithms **46**(2), 189–194 (2007)
35. L. Tucker, Some mathematical notes on three-mode factor analysis. Psychometrika **31**, 279–311 (1966)
36. I. Oseledets, E. Tyrtyshnikov, Breaking the curse of dimensionality, or how to use SVD in many dimensions. SIAM J. Sci. Comput. **31**(5), 3744–3759 (2009)
37. L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000)
38. L. Lim, Singular values and eigenvalues of tensors: a variational approach, in *IEEE CAMSAP 2005: First International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (IEEE, Piscataway, 2005), pp. 129–132
39. I. Oseledets, E. Tyrtyshnikov, TT-cross approximation for multidimensional arrays. Linear Algebra Appl. **432**(1), 70–88 (2010)
40. J. Ballani, L. Grasedyck, M. Kluge, Black box approximation of tensors in Hierarchical Tucker format. Linear Algebra Appl. **438**(2), 639–657 (2013)
41. L. Grasedyck, Hierarchical singular value decomposition of tensors. SIAM J. Matrix Anal. Appl. **31**(4), 2029–2054 (2010)

42. W. Hackbusch, S. Kühn, A new scheme for the tensor representation. J. Fourier Anal. Appl. **15**(5), 706–722 (2009)
43. A. Saibaba, HOID: higher order interpolatory decomposition for tensors based on Tucker representation. SIAM J. Matrix Anal. Appl. **37**(3), 1223–1249 (2016)
44. N. Vannieuwenhoven, R. Vandebril, K. Meerbergen, A new truncation strategy for the higher-order singular value decomposition. SIAM J. Sci. Comput. **34**(2), A1027–A1052 (2012)
45. P. Drineas, M.W. Mahoney, RandNLA: randomized numerical linear algebra. Comm. ACM **59**(6), 80–90 (2016)
46. D.P. Woodruff, Sketching as a tool for numerical linear algebra. Found. Trends Theor. Comput. Sci. **10**, 1–157 (2014)
47. R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, in Y.C. Eldar, G. Kutyniok (eds.) by *Compressed Sensing: Theory and Practice* (Cambridge University, Cambridge, 2012), pp. 210–268
48. B. Bader, T. Kolda, *Matlab Tensor Toolbox Version 2.6* (2015). http://www.sandia.gov/~tgkolda/TensorToolbox/
49. N. Vervliet, O. Debals, L. Sorber, M. Van Barel, L. De Lathauwer, *Tensorlab 3.0* (2016). http://tensorlab.net
50. I. Oseledets, *TT-toolbox version 2.2: Fast Multidimensional Array Operations in MATLAB* (2009–2013). http://github.com/oseledets/TT-Toolbox/
51. D. Savostyanov, I. Oseledets, Fast adaptive interpolation of multi-dimensional arrays in tensor train format, in *The 2011 International Workshop on Multidimensional (nD) Systems* (2011), pp. 1–8. http://dx.doi.org/10.1109/nDS.2011.6076873
52. D. Savostyanov, Quasioptimality of maximum-volume cross interpolation of tensors. Linear Algebra Appl. **458**, 217–244 (2014)
53. J. Nie, L. Wang, Semidefinite relaxations for best rank-1 tensor approximations. SIAM J. Matrix Anal. Appl. **35**(3), 1155–1179 (2014)
54. G. Beylkin, M. Mohlenkamp, Algorithms for numerical analysis in high dimensions. SIAM J. Sci. Comput. **26**(6), 2133–2159 (2005)
55. M. Che, Y. Wei, H. Yan, *Randomized Algorithms for the Low Multilinear Rank Approximations of Tensors* (2019). ArXiv:1908.11031
56. M. Che, Y. Wei, H. Yan, The computation for low multilinear rank approximations of tensors via power scheme and random projection. SIAM J. Matrix Anal. Appl. (2020, to appear)

Index

M. Che, Y. Wei, *Theory and Computation of Complex Tensors and its Applications*,
https://doi.org/10.1007/978-981-15-2059-4