Front. Math. China 2007, 2(4): 501–526
DOI 10.1007/s11464-007-0031-4

SURVEY ARTICLE

Numerical multilinear algebra and its applications*

QI Liqun[1], **SUN Wenyu**[2], **WANG Yiju**[3,1]

1 Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China
2 School of Mathematics and Computer Science, Nanjing Normal University, Nanjing 210097, China
3 School of Operations Research and Management Science, Qufu Normal University, Rizhao 276800, China

Abstract Numerical multilinear algebra, in which instead of matrices and vectors the higher-order tensors are considered in numerical viewpoint, is a new branch of computational mathematics. Although it is an extension of numerical linear algebra, it has many essential differences from numerical linear algebra and more difficulties than it. In this paper, we present a survey on the state of the art knowledge on this topic, which is incomplete, and indicate some new trends for further research. Our survey also contains a detailed bibliography as its important part. We hope that this new area will be receiving more attention of more scholars.

Keywords Numerical multilinear algebra, higher order tensor, tensor decomposition, lower rank approximation of tensor, multi-way data analysis
MSC 15A18, 15A69, 65F30

1 Introduction

Numerical multilinear algebra is a new branch of computational mathematics. It is concerned with treating higher-order tensors in numerical way by replacing matrix. It involves various computational topics of higher-order tensors, such as tensor decomposition, computation of tensor rank, computation of tensor eigenvalues, lower-rank approximations of tensors, numerical stability and perturbation analysis of tensor computation, and so on. This

* Received August 4, 2007; accepted August 24, 2007

Corresponding author: SUN Wenyu, E-mail: wysun@njnu.edu.cn

new branch has a strong practical background and wide applications in digital image restorations, psychometrics, chemometrics, econometrics, multiway data analysis, blind source separation and blind deconvolution in signal processing, high-order statistics, and so on (see Refs. [3,4,6,8,11,16,41,77,79]).

The classical multilinear algebra is a branch of abstract algebra and stresses how the operations work between tensors with R-module (over a commutative ring). It discusses associated algebra, exterior algebra of a module, symmetric algebra of a module, coalgebra and Hopf algebras, and so on (see Refs. [24,54,58]). However, the modern multi-way data analysis and signal processing need to know more properties about a higher-order tensor. In particular, how to decompose a tensor into the sum of products of vectors, how to approximate a tensor with a lower-rank tensor, how to compute the eigenvalues and singular values of a tensor, how to apply the higher order tensor in blind source separation (BSS), higher order statistics (HOS), blind deconvolution (BD), semidefinite programming (SDP), and others. All these constitute a new branch of computational mathematics—numerical multilinear algebra.

Although it is a very young discipline, recently the numerical multilinear algebra attracts much attention and gets dramatic development since there are strong motivations from the practical background and applications. Various experts and scholars in numerical linear algebra and engineering put their energy in this topic. Several international workshops and conferences on this new branch were held in USA, France, Switzerland, etc. For example, during July 19–23, 2004, Golub, Kolda, Nagy and Van Loan organized Workshop on Tensor Decomposition at the American Institute of Mathematics in Palo Alto, California. About 35 people—computer scientists, mathematicians, and a broad range of scholars who use tensor decompositions in their research—had come from eleven countries to participate in the week-long workshop. See the webpage of the workshop [21] and the SIAM News article on this workshop [29].

De Lathauwer and Comon further organized Workshop on Tensor Decompositions and Applications at CIRM, Luminy, Marseille, France during August 29—September 2, 2005. About 43 people from thirteen countries participated in the workshop. Topics discussed at the workshop include large-scale problems, topological properties of tensor spaces, exact or approximate tensor decompositions, mathematical properties of tensor decompositions, independent component analysis, applications in telecommunications, pattern analysis and statistical modelling, diagnostics in data analysis, sensor array processing. See the webpage of that workshop [45].

To bridge the gap among numerical linear algebra, theoretical computer science, and data applications, Golub, Mahony, Drineas, Lim organized Workshop for Modern Massive Data Set at Stanford University during June 21–24, 2006. The workshop, with 45 talks and 24 poster presentations, attracted 232 participants. See webpage [22], SIAM News and NA articles [23,46]. The theme of the last day of the workshop was tensor-based data applications. At his ending speech of the workshop, Golub pointed out that a new branch

of applied mathematics has been developed.

Recently, Golub, Comon, De Lathauwer and Lim organized an ICIAM Minisymposium on 'Numerical Multilinear Algebra: A New Beginning' at Zurich, Switzerland, during July 16–20, 2007. Golub wrote: "The name 'numerical multilinear algebra' is not as yet in common usage. We broadly define this as the study and use of tensors/multilinear algebra, symmetric tensors/symmetric algebra, alternating tensors/exterior algebra, spinors/ Clifford algebra in the computational mathematics. ... It is our hope that this minisymposium would constitute an important step towards the definition and development of this new discipline in the computational mathematics".

The aim of this paper is to give the development and a survey for numerical multilinear algebra. In fact, it can be regarded as a tutorial overview to this new area. Our survey and coverage will necessarily be incomplete and biased. Therefore, we also refer the readers to a survey paper by Comon [13] which discusses the tensor decompositions, and the survey paper by Lathauwer [44] which mainly discusses the canonical decomposition algorithms of a higher order tensor. Those people who would like to keep abreast of recent development of this new branch can attend the related webs, for example, Refs. [21,22,45].

The rest of the paper is organized as follows. In the next section, we give some basic notations, definitions and preliminary knowledge needed to the numerical multilinear algebra. In Section 3 we describe and survey the higher order tensor decompositions which include exact and approximate canonical decomposition, and higher order singular value decomposition, etc. Section 4 is devoted to the topic on the best rank-1 and rank-r approximation problem to a higher order tensor. In Section 5 we consider the theory and computation of eigenvalues of higher order tensor, which is an interesting topic and a generalization of eigenvalues of a matrix. Section 6 discusses typical applications of tensors on BSS, BD, SDP, and multivariate polynomial optimization. Finally, Section 7 contains some concluding remarks.

2 Notations and basic definitions

A tensor of order N is an N-way array, i.e., its entries are accessed via N indices, say $i_1, i_2, \ldots, i_N$ with each i_k ranging from 1 to I_k. For example, a vector is of a tensor of order 1, and a matrix is a tensor of order 2.

Throughout this paper, unless specified otherwise, variables take their values in the real field although all the statements hold true in the complex field. Vectors will be denoted by bold lowercase letters (e.g. $\mathbf{u}$), while matrices will be defined by uppercase letters (e.g. A). Higher-order tensors will be denoted by calligraphic, uppercase letters (e.g. $\mathscr{A}$). The entries of arrays are scalar quantities and are denoted with plain letters, such as u_i or $\mathscr{T}_{i,j,\ldots,l}$, the $(i, j, \ldots, l)$-element of a tensor $\mathscr{T}$.

A tensor of order N enjoys the multilinearity property after a change of

the coordinate system. In order to identify the ideas, consider a third order tensor $\mathscr{T}$ with entries $\mathscr{T}_{i_1i_2i_3}$, and a change of coordinates defined by three squares invertible matrices $A,\ B$ and C with elements $A_{i_1j_1},\ B_{i_2j_2}$ and $C_{i_3j_3}$, respectively. The tensor $\mathscr{T}$ can be written as

$$\mathscr{T}_{i_1i_2i_3} = \sum_{j_1=1}^{J_1}\sum_{j_2=1}^{J_2}\sum_{j_3=1}^{J_3} \mathscr{S}_{j_1j_2j_3}A_{i_1j_1}B_{i_2j_2}C_{i_3j_3}, \tag{2.1}$$

and is denoted by

$$\mathscr{T} = A \overset{\mathscr{S}}{\star} B \overset{\mathscr{S}}{\star} C, \tag{2.2}$$

where $\mathscr{S}$ is a $(J_1 \times J_2 \times J_3)$-tensor, $A,\ B$ and C are matrices of order $I_1 \times J_1,\ I_2 \times J_2$ and $I_3 \times J_3$, respectively. Representation (2.1)-(2.2) of tensors is referred to as the Tucker model (or Tucker product) [77], which is widely used in factor analysis, multi-way data processing and psychometrics.

The outer product of two vectors is defined as

$$C = \mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathrm{T}} \tag{2.3}$$

with element $C_{ij} = u_iv_j$, which is a matrix. The outer product of N vectors, $\mathbf{u}^{(i)}\ (i = 1, \ldots, N)$, is defined as

$$\mathscr{T} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)} \tag{2.4}$$

with $(i_1, i_2, \ldots, i_N)$-element

$$\mathscr{T}_{i_1i_2\cdots i_N} = u^{(1)}_{i_1}u^{(2)}_{i_2}\cdots u^{(N)}_{i_N},$$

which is said to be a rank-1 tensor of order N. In general, for two tensors $\mathscr{A}$ and $\mathscr{B}$ of order M and N respectively, one can define outer product $\mathscr{C} = \mathscr{A} \circ \mathscr{B}$ as a tensor of order $M + N$:

$$\mathscr{C}_{i_1\cdots i_Mj_1\cdots j_N} = \mathscr{A}_{i_1\cdots i_M}\mathscr{B}_{j_1\cdots j_N}. \tag{2.5}$$

Given two tensors, $\mathscr{A} = \{\mathscr{A}_{i_1\cdots i_M}\}$ and $\mathscr{B} = \{\mathscr{B}_{j_1\cdots j_N}\}$, having the same first dimension $(I_1 = J_1)$, one can define the mode-1 contraction product (or inner product)

$$(\mathscr{A} \bullet \mathscr{B})_{i_2\cdots i_Mj_2\cdots j_N} = \sum_{i_1=1}^{I_1} \mathscr{A}_{i_1\cdots i_M}\mathscr{B}_{j_1\cdots j_N}. \tag{2.6}$$

This product is induced from standard matrix multiplication. In fact, for given two matrices A and B, if the column number of A and the row number of B are the same $(= p)$, then the standard matrix product is

$$(AB)_{ij} = \sum_{k=1}^{p} A_{ik}B_{kj}, \tag{2.7}$$

which can be written in mode-1 contraction product

$$AB = A^{\mathrm{T}} \bullet B \tag{2.8}$$

with element

$$(AB)_{ij} = (A^{\mathrm{T}} \bullet B)_{ij} = \sum_{k=1}^{p} (A^{\mathrm{T}})_{ki} B_{kj}. \tag{2.9}$$

Similarly, we can define mode-p contraction product as long as tensors $\mathscr{A}$ and $\mathscr{B}$ have the same p-th dimension. Therefore, the Tucker product (2.1) is also regarded as a contraction product and sometimes is denoted by

$$\mathscr{T} = \mathscr{S} \bullet A \bullet B \bullet C \tag{2.10}$$

or

$$\mathscr{T} = \mathscr{S} \times_1 A \times_2 B \times_3 C, \tag{2.11}$$

where $\times_k$ denotes summing on i_k. This representation also induces several higher-order rank-factorization and higher-order singular value decomposition.

Given two tensors $\mathscr{A}$ and $\mathscr{B}$ of order N with the same dimensions. One defines their Hadamard product $\mathscr{C} = \mathscr{A} \odot \mathscr{B}$ with element as

$$\mathscr{C}_{i_1 \cdots i_N} = \mathscr{A}_{i_1 \cdots i_N} \mathscr{B}_{i_1 \cdots i_N}.$$

It says that the Hadamard product of two tensors gives a tensor with the same order and same dimension with $\mathscr{A}$ and $\mathscr{B}$.

As usual, the Kronecker product of two vectors $\mathbf{u}$ and $\mathbf{v}$ of $m \times 1$ and $n \times 1$, respectively, is defined as the vector of $mn \times 1$ with all the possible cross-product as follows:

$$\mathbf{a} \otimes \mathbf{b} := \begin{pmatrix} a_1 \mathbf{b} \\ a_2 \mathbf{b} \\ \vdots \\ a_m \mathbf{b} \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ \vdots \\ a_1 b_n \\ \vdots \\ a_m b_n \end{pmatrix}.$$

The Kronecker product of two matrices A and B of $m \times n$ and $p \times q$, respectively, is defined as the matrix of $mp \times nq$:

$$A \otimes B := (a_{ij} B)_{mp \times nq}, \quad i = 1, \ldots, m;\ j = 1, \ldots, n.$$

The operator that transforms a matrix to a vector is known as the vec operator. If the $m \times n$ matrix A has $\mathbf{a}_i$ as its i-th column, then $\mathrm{vec}(A)$ is the $mn \times 1$ vector given by

$$\mathrm{vec}(A) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

We can give some definitions of supersymmetric tensor, tensor scalar product and tensor Frobenius norm, which are the generalization of matrix case in a straightforward way.

Definition 2.1 (supersymmetric tensor) A tensor is called supersymmetric if its entries are invariant under any permutation of their indices.

Definition 2.2 (tensor scalar product) The scalar product $\langle \mathscr{S}, \mathscr{T} \rangle$ of two tensors $\mathscr{S}$ and $\mathscr{T}$ with the same order N and same dimensions is defined by

$$\langle \mathscr{S}, \mathscr{T} \rangle = \sum_{i_1,\ldots,i_N} \mathscr{S}_{i_1\cdots i_N} \mathscr{T}_{i_1\cdots i_N}. \tag{2.12}$$

Definition 2.3 (Frobenius norm of a tensor) The Frobenius norm of a tensor $\mathscr{T}$ of order N is defined as

$$\|\mathscr{T}\|_F = \langle \mathscr{T}, \mathscr{T} \rangle = \left(\sum_{i_1,\cdots,i_N} \mathscr{T}^2_{i_1\cdots i_N} \right)^{1/2}. \tag{2.13}$$

For further detailed notation and definitions of matrix and tensor computation, we refer to Refs. [15,20,40,67,74].

3 Rank and factorization: From matrix to tensor

Bergman [2] and Harshman [25] first noticed that it was difficult to extend the concept of rank from matrices to higher-order arrays. However, there are still several ways to extend the concept of rank of matrix to higher-order tensor.

Let $m\times n$ matrix $A \in \mathbb{R}_r^{m\times n}$ be of rank r, and let $B \in \mathbb{R}_r^{m\times r}$ and $C \in \mathbb{R}_r^{r\times n}$ be full-rank matrices, where $\mathbb{R}_r^{m\times n}$ denotes the space of all $m\times n$ real matrices of rank r. We have

$$A = BC = [\mathbf{b}_1, \ldots, \mathbf{b}_r] \begin{bmatrix} \mathbf{c}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{c}_r^{\mathrm{T}} \end{bmatrix} = \sum_{i=1}^{r} \mathbf{b}_i \mathbf{c}_i^{\mathrm{T}} = \sum_{i=1}^{r} \mathbf{b}_i \circ \mathbf{c}_i, \tag{3.1}$$

which is called the rank-1 factorization of matrix A (or BC factorization). Here r is the rank of matrix A and the number of rank-1 matrix whose linear combination yields A (see Ref. [26]).

Now, this idea can be extended to higher-order tensor case. Let $\mathscr{T}$ be a tensor of order N of dimensions $I_1\times\cdots\times I_N$. The tensor $\mathscr{T}$ can be decomposed into a sum of the outer products of vectors. The tensor rank r is defined as the minimal number of rank-1 tensors whose linear combination yields $\mathscr{T}$, i.e.,

$$\mathscr{T} = \sum_{i=1}^{r} \mathbf{u}_i^{(1)} \circ \mathbf{u}_i^{(2)} \circ \cdots \circ \mathbf{u}_i^{(N)}, \tag{3.2}$$

where $\mathbf{u}_i^{(1)}, \mathbf{u}_i^{(2)}, \ldots, \mathbf{u}_i^{(N)}$ are $I_1, I_2, \ldots, I_N$-dimensional vectors, respectively.

The alternative definition of tensor rank is called mode-k rank of a tensor which is defined by mode-k matrix (or k-th matrix unfolding).

Definition 3.1 Assume that N-th order tensor $\mathscr{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The mode-$k$ matrix (or k-th matrix unfolding) $A_{(k)} \in C^{I_k \times (I_{k+1} I_{k+2} \cdots I_N I_1 \cdots I_{k-1})}$ is a matrix containing the element $a_{i_1 i_2 \cdots i_N}$.

For example, for tensor $\mathscr{A} \in \mathbb{R}^{3 \times 2 \times 3}$ of order 3, where

$$I_1 = 3, \quad I_2 = 2, \quad I_3 = 3,$$

we have

$$A_{(1)} \in \mathbb{R}^{I_1 \times (I_2 I_3)}, \quad A_{(2)} \in \mathbb{R}^{I_2 \times (I_3 I_1)}, \quad A_{(3)} \in \mathbb{R}^{I_3 \times (I_1 I_2)},$$

and

$$A_{(1)} = \begin{pmatrix} a_{111} & a_{112} & a_{113} & a_{121} & a_{122} & a_{123} \\ a_{211} & a_{212} & a_{213} & a_{221} & a_{222} & a_{223} \\ a_{311} & a_{312} & a_{313} & a_{321} & a_{322} & a_{323} \end{pmatrix},$$

which is a 3×6 matrix,

$$A_{(2)} = \begin{pmatrix} a_{111} & a_{211} & a_{311} & a_{112} & a_{212} & a_{312} & a_{113} & a_{213} & a_{313} \\ a_{121} & a_{221} & a_{321} & a_{122} & a_{222} & a_{322} & a_{123} & a_{223} & a_{323} \end{pmatrix},$$

which is a 2×9 matrix,

$$A_{(3)} = \begin{pmatrix} a_{111} & a_{121} & a_{211} & a_{221} & a_{311} & a_{321} \\ a_{112} & a_{122} & a_{212} & a_{222} & a_{312} & a_{322} \\ a_{113} & a_{123} & a_{213} & a_{223} & a_{313} & a_{323} \end{pmatrix},$$

which is a 3×6 matrix.

For the above defined mode-k matrix $A_{(k)}$ (or k-th matrix unfolding) ($k = 1, \ldots, N$), its column vectors are said to be mode-k vectors of tensor $\mathscr{A}$.

Definition 3.2 The rank of mode-k matrix is said to be mode-k rank of tensor $\mathscr{A}$, written as

$$R_k = \operatorname{rank}_k(\mathscr{A}) = \operatorname{rank}(A_{(k)}). \tag{3.3}$$

Obviously, an N-th order tensor has N mode-k ranks and the different mode-k ranks of tensor are not necessarily the same. In addition, the rank and the mode-k rank of a same tensor are not necessarily equal even though all the mode-k ranks are equal.

Now, much discussion and research on definitions and computation of tensor rank are in processing.

Higher-order singular value decomposition (HOSVD) due to Lathauwer etc. [41] is extended from SVD of matrices and is an important part of tensor decomposition.

Let A be an $m \times n$ real matrix. Then there are orthogonal matrices

$$U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m] \in \mathbb{R}^{m\times m}, \quad V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \in \mathbb{R}^{n\times n},$$

such that

$$U^{\mathrm{T}} A V = D = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.4}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r), \quad \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0.$$

It means that

$$A = U \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ & \mathbf{0} & & \mathbf{0} \end{bmatrix} V^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}, \tag{3.5}$$

where σ_i is called the singular value of A, vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are said to be left singular vectors and right singular vectors, respectively (see Ref. [26]). If let A and D be $I_1 \times I_2$ matrices, U and V be $I_1 \times I_1$ and $I_2 \times I_2$ orthogonal matrices respectively, and set $U^{(1)} = U$, $U^{(2)} = V^{\mathrm{T}}$, then the SVD decomposition (3.5) can be written in the following form:

$$\begin{aligned} A &= UDV^{\mathrm{T}} \\ &= D \times_1 U^{(1)} \times_2 U^{(2)} \\ &= D \bullet U^{(1)} \bullet U^{(2)} \\ &=: U^{(1)} \overset{D}{\bullet} U^{(2)}, \end{aligned} \tag{3.6}$$

where the symbol $\times_k$ denotes summing on i_k. For example, for $I_1 \times I_2 \times I_3$ tensor $\mathscr{A}$ of order 3 and $J \times I_1$ matrix U, the component of product is

$$(\mathscr{A} \times_1 U)_{ji_2i_3} = \sum_{i_1=1}^{I_1} \mathscr{A}_{i_1i_2i_3} U_{ji_1}.$$

In matrix SVD, we have the following properties.

1) $\mathrm{Range}(A) = \mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, $\mathrm{Null}(A) = \mathrm{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$.

2) $\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2$.

3) The matrix SVD indicates that a given matrix can be approximated well by a lower-rank matrix. If

$$k < r = \mathrm{rank}(A), \quad A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}},$$

then

$$\min_{\mathrm{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}. \tag{3.7}$$

The above SVD (3.6) of matrices and its properties can be extended to higher-order tensor case.

Theorem 3.3 (HOSVD) *Any $(I_1 \times I_2 \times \cdots \times I_N)$-tensor $\mathscr{A}$ of order N can be expressed as*

$$\mathscr{A} = \mathscr{S} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 \cdots \times_N U^{(N)}, \tag{3.8}$$

or

$$\mathscr{A} = U^{(1)} \overset{\mathscr{S}}{\bullet} U^{(2)} \overset{\mathscr{S}}{\bullet} \cdots \overset{\mathscr{S}}{\bullet} U^{(N)} \tag{3.9}$$

with elements

$$\mathscr{A}_{i_1 i_2 \cdots i_N} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \cdots \sum_{j_N=1}^{I_N} \mathscr{S}_{j_1 j_2 \cdots j_N} U^{(1)}_{i_1 j_1} U^{(2)}_{i_2 j_2} \cdots U^{(N)}_{i_N j_N}, \tag{3.10}$$

where

1) $U^{(k)} = (\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \ldots, \mathbf{u}_N^{(k)})$ *is an orthogonal $I_k \times I_k$ matrix.*

2) $\mathscr{S}$ *is an $(I_1 \times I_2 \times \cdots \times I_N)$-tensor of the same size as $\mathscr{A}$, and its subtensors $\mathscr{S}_{i_k=\alpha}$, obtained by fixing the k-th index to α, have the properties of*

a) *all-orthogonality: two subtensors $\mathscr{S}_{i_k=\alpha}$ and $\mathscr{S}_{i_k=\beta}$ are orthogonal for any possible values of k and $\alpha \neq \beta$, in the sense that*

$$\langle \mathscr{S}_{i_k=\alpha}, \mathscr{S}_{i_k=\beta} \rangle = \sum_{i_1, \cdots, i_{k-1}, i_{k+1}, \cdots, i_N} \mathscr{S}_{i_1 \cdots i_{k-1} \alpha i_{k+1} \cdots i_N} \mathscr{S}_{i_1 \cdots i_{k-1} \beta i_{k+1} \cdots i_N} = 0; \tag{3.11}$$

b) *ordering: for all k,*

$$\|\mathscr{S}_{i_k=1}\| \geqslant \|\mathscr{S}_{i_k=2}\| \geqslant \cdots \geqslant \|\mathscr{S}_{i_k=I_k}\| \geqslant 0. \tag{3.12}$$

The Frobenius norm $\|\mathscr{S}_{i_n=i}\|$, symbolized by $\sigma_i^{(k)}$, are mode-k singular values of $\mathscr{A}$ and the vector $\mathbf{u}_i^{(k)}$ is the i-th mode-k singular vector.

Furthermore, an equivalent matrix representation of the HOSVD in (3.8) can be obtained by unfolding matrix of $\mathscr{A}$ and $\mathscr{S}$:

$$\begin{aligned} A_{(k)} &= U^{(k)} \cdot S_{(k)} \cdot (U^{(k+1)} \otimes \cdots \otimes U^{(N)} \otimes U^{(1)} \otimes \cdots \otimes U^{(k-1)})^{\mathrm{T}} \\ &=: U^{(k)} \cdot \Sigma^{(k)} \cdot \widetilde{S}_{(k)} \cdot (U^{(k+1)} \otimes \cdots \otimes U^{(N)} \otimes U^{(1)} \otimes \cdots \otimes U^{(k-1)})^{\mathrm{T}} \\ &=: U^{(k)} \cdot \Sigma^{(k)} \cdot V^{(k)\mathrm{T}}, \quad 1 \leqslant k \leqslant N, \end{aligned} \tag{3.13}$$

where

$$V^{(k)} := \widetilde{S}_{(k)} \cdot (U^{(k+1)} \otimes U^{(k+2)} \otimes \cdots \otimes U^{(N)} \otimes U^{(1)} \otimes \cdots \otimes U^{(k-1)})$$

is orthogonal and

$$\Sigma^{(k)} := \operatorname{diag}(\sigma_1^{(k)}, \sigma_2^{(k)}, \ldots, \sigma_{I_k}^{(k)}).$$

In HOSVD, we have the following properties.

1) The mode-k Range$(A_{(k)}) = \text{span}(\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{r_k}^{(k)})$.

2) The mode-k null space Null$(A_{(k)}^{\mathrm{T}}) = \text{span}(\mathbf{u}_{r_k+1}^{(k)}, \ldots, \mathbf{u}_{I_k}^{(k)})$.

3)
$$\|\mathscr{A}\|^2 = \sum_{i=1}^{r_1} (\sigma_i^{(1)})^2 = \cdots = \sum_{i=1}^{r_N} (\sigma_i^{(N)})^2 = \|\mathscr{S}\|^2.$$

4) Let $\hat{\mathscr{A}}$ be a tensor obtained by discarding the smallest mode-k singular values $\sigma_{I_k'+1}^{(k)}, \sigma_{I_k'+2}^{(k)}, \ldots, \sigma_{r_k}^{(k)}$ for given values of I_k' $(1 \leqslant k \leqslant N)$, i.e., set the corresponding parts of $\mathscr{S}$ equal to zero. Then
$$\|\mathscr{A} - \hat{\mathscr{A}}\|^2 \leqslant \sum_{i_1=I_1'+1}^{r_1} \sigma_{i_1}^{(1)^2} + \sum_{i_1=I_2'+1}^{r_2} \sigma_{i_2}^{(2)^2} + \cdots + \sum_{i_1=I_N'+1}^{r_N} \sigma_{i_N}^{(N)^2}.$$

By Theorem 3.3 and comparison of the matrix and tensor cases, we know that the left and the right singular vectors of a matrix (can be regarded as mode-1 and mode-2 singular vectors) are generalized to mode-k singular vectors $(k = 1, 2, \ldots, N)$. Next, the role of the singular values in HOSVD is taken over by Frobenius norm of the $(N-1)$-th subtensors of $\mathscr{S}$. Note that, in matrix SVD, the matrix D in (3.6) is diagonal or pseudo-diagonal. However, unfortunately, the tensor $\mathscr{S}$ in HOSVD (3.8) is a full tensor instead of being pseudo-diagonal form. Lathauwer [41] shows that in general, it is impossible to reduce higher-order tensors to a pseudo-diagonal form by means of orthogonal transformations. This conclusion reduces the value of HOSVD because pseudo-diagonal form in HOSVD is also important, like in matrix SVD. So, we now face to a challenging topic: how to make an approximate pseudo-diagonalization hold via computational methods or via 'approximation' and/or 'relaxation'.

Caroll and Chang [7] provided the first canonical decomposition algorithm (CANDecomp) of a three-way array. Comon proposed to approximately decompose a d-way array into d terms, which is referred to as independent component analysis (ICA) [9]. Leurgans et al. [47] proposed SVD-based algorithms to compute CANDecomp of the third-order tensor in large dimension. Lathauwer et al. [44] point out that the CANDecomp can be computed, in theory, by means of a matrix eigenvalue decomposition (EVD). However, one can actually reformulate CANDecomp as an orthogonal simultaneous matrix decomposition. So, the CANDecomp can be used in symmetric and unsymmetric case. It should be pointed out that the concepts of tensor rank and the decomposition of a tensor into a sum of outer-product of vectors was the product of much earlier work by Hitchcock in 1927 [27,28].

Canonical decomposition (CANDecomp) (or parallel factors model (PARAFAC)) is a decomposition of a given tensor as a linear combination of a minimal number of possible non-orthogonal rank-1 tensors. It is a multilinear generalization of diagonalizing a matrix by an equivalence or congruence transformation. Lathauwer et al. [44] reformulate the problem as a

simultaneous genaralized Schur decomposition (SGSD), and give a technique for the actual computation of the SGSD. The uniqueness of the decomposition for third-order real-valued and complex-valued tensors is derived in Refs. [34] and [71] respectively. The result is generalized to arbitrary tensor orders in Ref. [72].

The CANDecomp is defined as follows.

Definition 3.4 (CANDecomp) A canonical decomposition of an N-th order tensor $\mathscr{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is a decomposition of $\mathscr{A}$ as a linear combination of a minimum number of rank-1 terms:

$$\mathscr{A} = \sum_{i=1}^{r} \lambda_i U_i^{(1)} \circ U_i^{(2)} \circ \cdots \circ U_i^{(N)}. \tag{3.14}$$

There are several numerical methods to get CANDecomp. To look for CANDecomp, one method is that we can minimize the quadratic cost function

$$f(\hat{\mathscr{A}}) = \|\mathscr{A} - \hat{\mathscr{A}}\|^2 \tag{3.15}$$

over all mode-k rank tensors $\mathscr{A}$, which is parameterized as

$$\hat{\mathscr{A}} = \sum_{i=1}^{r} \hat{\lambda}_i \hat{\mathbf{u}}_i^{(1)} \circ \cdots \circ \hat{\mathbf{u}}_i^{(N)}. \tag{3.16}$$

It is possible to do it by use of an alternating least-squares (ALS) algorithm, in which the vector estimates are updated mode per mode. In each step, a classical linear least-squares problem is solved for different mode numbers. So, several classical methods for least-squares problems can be used, for example, Gauss-Newton method, Levenberg-Marquardt method, etc., in which all the CANDecomp factors are updated simultaneously (see Refs. [44,55,59,75]). These techniques are used for overcoming the illposedness and indeterminancy of the decomposition.

An interesting alternative procedure is used by use of a simple matrix eigenvalue decomposition (EVD) if (3.14) is exactly valid (see Ref. [47]). Otherwise, a least-squares matching of both sides of (3.14) can be initialized with the EVD result.

Consider $\mathscr{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and assume that $I_1 = I_2 = r$. The problem CANDecomp is to find vectors $\mathbf{u}_i^{(1)}$, $\mathbf{u}_i^{(2)}$, $\mathbf{u}_i^{(3)}$, and λ_i, such that

$$\mathscr{A} \simeq \sum_{i=1}^{r} \lambda_i \mathbf{u}_i^{(1)} \circ \mathbf{u}_i^{(2)} \circ \mathbf{u}_i^{(3)}. \tag{3.17}$$

The mode-k matrices of $\mathscr{A}$ are

$$A_{(k)} = U^{(k)} \cdot \Lambda (U^{(l)} \otimes U^{(m)})^{\mathrm{T}}, \tag{3.18}$$

where (k, l, m) is an arbitrary permutation of $(1, 2, 3)$.

Consider a linear transformation of the vector space $\mathbb{R}^{I_3}$ to the matrix space $\mathbb{R}^{I_1 \times I_2}$:

$$V = f_{\mathscr{A}}(W) = \mathscr{A} \times_3 W. \tag{3.19}$$

Substitution of (3.19) into (3.14) shows that the image of W can be expressed in terms of CANDecomp-components:

$$V = U^{(1)} \cdot D \cdot U^{(2)},$$

where

$$U^{(k)} := [\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{I_k}^{(k)}], \quad k = 1, 2,$$

$$D := \operatorname{diag}\{(\lambda_1, \ldots, \lambda_r)\} \cdot \operatorname{diag}\{U^{(3)\mathrm{T}} W\}.$$

Note that any matrix in the range of the mapping $f_{\mathscr{A}}$ can be diagonalized. Let the range be spanned by the matrices $V_1, V_2, \ldots, V_K$. Then we have the simultaneous decomposition:

$$\begin{aligned} V_1 &= U^{(1)} \cdot D_1 \cdot U^{(2)\mathrm{T}}, \\ V_2 &= U^{(1)} \cdot D_2 \cdot U^{(2)\mathrm{T}}, \\ &\cdots \\ V_K &= U^{(1)} \cdot D_K \cdot U^{(2)\mathrm{T}}, \end{aligned} \tag{3.20}$$

where $D_1, \ldots, D_K$ are diagonal. In this case, corresponding to (3.15), the cost function is

$$\widetilde{f}(\hat{U}^{(1)}, \hat{U}^{(2)}, \{\hat{D}_k\}) = \sum_{k=1}^{K} \|V_k - \hat{U}^{(1)} \cdot \hat{D}_k \cdot \hat{U}^{(2)\mathrm{T}}\|^2. \tag{3.21}$$

The Schur decomposition of a matrix is a generalization of the spectral decomposition in which the goal is to obtain a diagonal matrix. In the Schur decomposition, it turns out that for general case of any real square $m \times m$ matrix A, we can find a unitary matrix X, such that $X^* AX$ is a triangular matrix, i.e.,

$$X^* AX = T,$$

where T is an upper triangular matrix with the eigenvalues of A as its diagonal elements (see Ref. [26]).

If the $m \times m$ matrix A has real eigenvalues, then there exists an $m \times m$ orthogonal matrix X, such that

$$X^{\mathrm{T}} AX = T,$$

where T is an upper triangular matrix.

Now, in CANDecomp, introducing QR-factorization

$$U^{(1)} = Q^{\mathrm{T}} R', \quad U^{(2)\mathrm{T}} = R'' Z^{\mathrm{T}}$$

yields the SGSD equations:

$$\begin{aligned} Q \cdot V_1 \cdot Z &= R_1 = R' \cdot D_1 \cdot R'', \\ Q \cdot V_2 \cdot Z &= R_2 = R' \cdot D_2 \cdot R'', \\ &\cdots \\ Q \cdot V_K \cdot Z &= R_K = R' \cdot D_K \cdot R'', \end{aligned} \tag{3.22}$$

where $Q, Z \in \mathbb{R}^{r\times r}$ are orthogonal, and $R', R'', R_1, \ldots, R_K$ are upper triangular (see Refs. [44,78]). In this case the CANDecomp is to determine Q and Z such that $R_1, \ldots, R_K$ are upper triangular. Lathauwer [44] discussed related numerical results.

In addition, some supersymmetric decomposition of cumulant tensors and applications [5], decomposition of quantics in sums of powers of linear forms [10], some canonical decomposition [15], lower-rank decomposition [51], orthogonal tensors decompositions [32], and other singular value decomposition for higher-order tensors are presented.

4 Best rank-1 and rank-r approximation of tensors

The best rank-1 tensor approximation problem is stated as follows. Given an N-th order $n \times n \times \cdots \times n$ dimensional supersymmetric tensor, if there exists a scalar λ and a unit-norm vector $\mathbf{u}$, such that the rank-1 tensor

$$\hat{\mathscr{A}} := \lambda \mathbf{u}^{\star N} := \lambda \mathbf{u}^N$$

minimizes the least-squares cost function

$$\begin{aligned} &\min f(\hat{\mathscr{A}}) = \|\mathscr{A} - \hat{\mathscr{A}}\|_F^2 \\ &\text{s.t.} \quad \|\mathbf{u}\|_2 = 1 \end{aligned} \tag{4.1}$$

over the manifold of rank-1 tensors, then $\hat{\mathscr{A}} = \lambda \mathbf{u}^{\star N}$ is called the best rank-1 approximation to tensor $\mathscr{A}$.

Similarly, for general case (not necessarily supersymmetric case), if there exist a scalar λ and N unit-norm vectors $u^{(j)} \in \mathbb{R}^n,\ j = 1, \cdots, N$, such that the rank-1 tensor

$$\hat{\mathscr{A}} := \lambda \mathbf{u}^{(1)} \star \mathbf{u}^{(2)} \star \cdots \star \mathbf{u}^{(N)}$$

minimizes

$$f(\hat{\mathscr{A}}) = \|\mathscr{A} - \hat{\mathscr{A}}\|_F^2 \tag{4.2}$$

over the manifold of rank-1 tensors, then

$$\hat{\mathscr{A}} = \lambda \mathbf{u}^{(1)} \star \mathbf{u}^{(2)} \star \cdots \star \mathbf{u}^{(N)}$$

is called the best rank-1 approximation to tensor $\mathscr{A}$.

The minimization problem (4.1) is equivalent to the dual problem of maximizing

$$\begin{aligned} \max g(\mathbf{u}) &= \sum_{i_1,\ldots,i_N} \mathscr{A}_{i_1\cdots i_N} u_{i_1} u_{i_2} \cdots u_{i_N} = \langle \mathscr{A}, \mathbf{u}^{\star N} \rangle \\ \text{s.t.} \quad & \|\mathbf{u}\|_2 = 1. \end{aligned} \tag{4.3}$$

The corresponding value of λ is $\lambda = g(\mathbf{u})$.

Similar to the matrix Rayleigh quotient, we define the generalized Rayleigh quotient (or higher-order Rayleigh quotient) as

$$h(\mathbf{u}) = \frac{\langle \mathscr{A}, \mathbf{u}^{\star N} \rangle^2}{\langle \mathbf{u}, \mathbf{u} \rangle^N} = \frac{(\sum_{i_1,\ldots,i_N} \mathscr{A}_{i_1\cdots i_N} u_{i_1} u_{i_2} \cdots u_{i_N})^2}{\langle \mathbf{u}, \mathbf{u} \rangle^N}, \tag{4.4}$$

whose maximization corresponds to (4.3).

For the maximization problem (4.4), several classical methods exist and can be extended to higher-order case, for example, higher-order power method [31,39], generalized Newton method [81] and others.

The stationary point of (4.4) is the solution of the following equations:

$$\begin{aligned} \sum_{i_2,\ldots,i_N} \mathscr{A}_{i_1 i_2 \cdots i_N} u_{i_2} u_{i_3} \cdots u_{i_N} &= \lambda u_{i_1} \quad \text{for all } i_1, \\ &\cdots, \\ \sum_{i_1,\ldots,i_{N-1}} \mathscr{A}_{i_1 i_2 \cdots i_N} u_{i_1} u_{i_2} \cdots u_{i_{N-1}} &= \lambda u_{i_N} \quad \text{for all } i_N \end{aligned} \tag{4.5}$$

with $\lambda = \langle \mathscr{A}, \mathbf{u}^{\star N} \rangle$.

We have proved the following property. For a nonzero N-th order n-dimensional supersymmetric tensor $\mathscr{A}$ with $N = 2l$ for some positive integer l, there exist l unit-norm vectors $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(l)}$, such that

$$\lambda = \mathscr{A}(\mathbf{u}^{(1)})^2 \circ (\mathbf{u}^{(2)})^2 \circ \cdots \circ (\mathbf{u}^{(l)})^2.$$

For problem (4.5), the higher-order power method (HOPM) is suggested in Refs. [39,42]. In general case, let $\mathscr{A}$ be a tensor with order N and dimension $I_1 \times I_2 \times \cdots \times I_N$. Given an initial unit-norm I_k-vector $\mathbf{u}_0^{(k)}$, the general iteration is defined as follows. For $i = 1, 2, \ldots$, let

$$\begin{aligned} \widetilde{\mathbf{u}}_i^{(k)} &= \mathbf{u}_i^{(1)} \overset{\mathscr{A}}{\star} \cdots \overset{\mathscr{A}}{\star} \mathbf{u}_i^{(k-1)} \overset{\mathscr{A}}{\star} I \overset{\mathscr{A}}{\star} \mathbf{u}_{i-1}^{(k+1)} \overset{\mathscr{A}}{\star} \cdots \overset{\mathscr{A}}{\star} \mathbf{u}_{i-1}^{(N)}, \\ \lambda_i^{(k)} &= \|\widetilde{\mathbf{u}}_i^{(k)}\|, \\ \mathbf{u}_i^{(k)} &= \frac{\widetilde{\mathbf{u}}_i^{(k)}}{\lambda_i^{(k)}}. \end{aligned} \tag{4.6}$$

Then we obtain

$$\hat{\mathscr{A}} = \lambda \mathbf{u}^{(1)} \overset{\mathscr{A}}{\star} \mathbf{u}^{(2)} \overset{\mathscr{A}}{\star} \cdots \overset{\mathscr{A}}{\star} \mathbf{u}^{(N)}, \tag{4.7}$$

which is a best rank-1 approximation to tensor $\mathscr{A}$.

In supersymmetric case, the general iteration is defined as follows. For $i = 1, 2, \ldots,$ let

$$\begin{aligned} \widetilde{\mathbf{u}}_i &= I \overset{\mathscr{A}}{\star} (\mathbf{u}_{i-1})^{\overset{\mathscr{A}}{\star}(N-1)}, \\ \lambda_i &= \|\widetilde{\mathbf{u}}_i\|, \\ \mathbf{u}_i &= \frac{\widetilde{\mathbf{u}}_i}{\lambda_i}. \end{aligned} \tag{4.8}$$

Then we obtain

$$\hat{\mathscr{A}} = \lambda \mathbf{u}^{\overset{\mathscr{A}}{\star} N}. \tag{4.9}$$

In addition, for system (4.5), Newton's method can be applied, which is called generalized Rayleigh-Newton iteration in Ref. [81]. The competitive numerical results are reported about generalized power method and generalized Rayleigh-Newton method for the best rank-1 tensor approximation problem.

Using a sequence of the best rank-1 approximation, we can consider the decomposition of a supersymmetric tensor into the sums of supersymmetric rank-1 tensors. We proposed a greedy method to decompose a higher-order supersymmetric tensor into successive supersymmetric rank-1 tensors (see Ref. [80]). Several approximations to tensors and their applications in signal processing, and their perturbation analysis are discussed in Refs. [30,43].

5 Eigenvalues of tensors and their applications

In this section we discuss the eigenvalues of N-th order n-dimensional tensor $\mathscr{A}$.

Recently, Qi defined eigenvalues and eigenvectors of a real supersymmetric tensor, and explored their applications in determining positiveness of an even order tensor, and in finding the best rank-1 approximation to a supersymmetric tensor. This work is an extension of the classical concept of eigenvalues of square matrices (see Refs. [57,61,62]). These works have been generalized to the nonsymmetric case in Ref. [63].

Let $\mathscr{A}$ be a real N-th order n-dimensional supersymmetric tensor. The tensor $\mathscr{A}$ defines an N-degree homogeneous polynomial $f(x)$,

$$f(x) := \mathscr{A}x^N := \sum_{i_1,\ldots,i_N=1}^{n} \mathscr{A}_{i_1\cdots i_N} x_{i_1} \cdots x_{i_N}, \tag{5.1}$$

where x^N can be regarded as an N-th order n-dimensional rank-1 tensor with entries $x_{i_1} \cdots x_{i_N}$. Qi [61] defines the concepts of H-eigenvalue (eigenvalue) and Z-eigenvalue (E-eigenvalue) of a real supersymmetric tensor $\mathscr{A}$.

For a vector $x \in \mathbb{R}^n$ with components x_i, we use $x^{[N]}$ to denote a vector in $\mathbb{R}^n$ such that

$$x^{[N]} := [x_1^N, \ldots, x_n^N]^{\mathrm{T}}, \quad \text{i.e.,} \quad x_i^{[N]} = x_i^N$$

for all i. Let $\mathscr{A}x^{N-1}$ be a vector in $\mathbb{R}^n$ whose i-th component is

$$(\mathscr{A}x^{N-1})_i = \sum_{i_2,\ldots,i_N=1}^{n} \mathscr{A}_{ii_2\cdots i_N} x_{i_1}\cdots x_{i_N}.$$

A real number λ is called an H-eigenvalue of $\mathscr{A}$ if it and a real nonzero vector x satisfy the following homogeneous polynomial equation:

$$\mathscr{A}x^{N-1} = \lambda x^{[N-1]}. \tag{5.2}$$

If a complex number λ and a complex vector x satisfy (5.2), they are called eigenvalue and eigenvector of $\mathscr{A}$, respectively.

A real number λ and a real nonzero vector $x \in \mathbb{R}^n$ are called a Z-eigenvalue and a Z-eigenvector of $\mathscr{A}$, respectively, if they satisfy the following system:

$$\begin{cases} \mathscr{A}x^{N-1} = \lambda x, \\ x^{\mathrm{T}}x = 1. \end{cases} \tag{5.3}$$

If a number $\lambda \in \mathbb{C}$ and a nonzero vector $x \in \mathbb{C}^n$ satisfy (5.3), then we say that they are E-eigenvalue and E-eigenvector, respectively.

Theorem 5.1 *H-eigenvalues and Z-eigenvalues always exist for an even order supersymmetric tensor. An even order supersymmetric tensor is positive definite* (*semidefinite*) *if and only if all of its H-eigenvalues or all of its Z-eigenvalues are positive* (*nonnegative*).

Theorem 5.2 a) *A number $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathscr{A}$ if and only if it is a root of the characteristic polynomial $\phi(\lambda) = \det(\mathscr{A} - \lambda\mathscr{I})$.*

b) *The number of eigenvalues of $\mathscr{A}$ is $d = n(N-1)^{n-1}$. Their product is equal to* $\det(\mathscr{A})$. *The sum of all the eigenvalues of $\mathscr{A}$ is* $(N-1)^{n-1}\mathrm{tr}(\mathscr{A})$.

c) *The $n(N-1)^{n-1}$ eigenvalues are distributed in n disks in $\mathbb{C}$. The centers and radii of these n disks are the diagonal elements, and the sums of the absolute values of the corresponding off-diagonal elements, of that supersymmetric tensor.*

d) *E-eigenvalues are invariant under orthogonal transformations.*

In addition, the Gerschgorin-type theorem also holds for eigenvalues of supersymmetric tensors. These properties do not hold for E-eigenvalues of higher order supersymmetric tensors. However, the invariant property does not hold for eigenvalues of higher order supersymmetric tensors, but it does hold for E-eigenvalues. It was shown that the E-eigenvalues are invariant under orthogonal transformation. It was also shown that, when the E-characteristic polynomial is converted to a monic polynomial, the coefficients of that monic polynomial are invariants of that tensor, i.e., they are invariant under coordinate system changes. Therefore, in Ref. [63], the E-eigenvalues of supersymmetric tensor are extended to nonsymmetric case and the invariant property in nonsymmetric case is established.

By use of recession vector of a tensor, we can define the rank of a tensor. Suppose that $\mathscr{A}$ is an N-th order tensor. We call the vector y a recession vector of $\mathscr{A}$ if

$$y \cdot \mathscr{A} = 0,$$

where '$\cdot$' means the inner product. All recession vectors of $\mathscr{A}$ form the recession space V_{R} of $\mathscr{A}$ which is a linear subspace. Let

$$V_{\mathrm{B}} = \{x \mid y \cdot x = 0,\ \forall\ y \in V_{\mathrm{R}}\}$$

and let the dimension of V_{B} be r. We say that the rank of tensor $\mathscr{A}$ is r.

Independently, with a variational approach, Lim [50] defines eigenvalues of tensors in the real field. The l_2 eigenvalues of tensors defined in Ref. [50] are just Z-eigenvalues in Ref. [61], while the l_N eigenvalues of tensors defined in Ref. [50] are H-eigenvalues in Ref. [61]. Notably, Lim [50] proposed a multilinear generalization of Perron-Frobenius theorem based upon the notion of l_N eigenvalues (H-eigenvalues) of tensors.

In fact, in the matrix case, eigenvalues and eigenvectors of a symmetric matrix A are the critical values and critical points of its Raleigh quotient $x^{\mathrm{T}}Ax/\|x\|_2^2$, equivalently, the critical values and points of the problem

$$\begin{aligned} &\min x^{\mathrm{T}}Ax \\ &\text{s.t.} \quad \|x\|_2 = 1. \end{aligned} \tag{5.4}$$

Introduce the Lagrange function $L(x, \lambda)$ with Lagrange multiplier λ,

$$L(x, \lambda) = x^{\mathrm{T}}Ax - \lambda(\|x\|_2^2 - 1). \tag{5.5}$$

The vanishing of ∇L yields the definition of eigenpairs

$$Ax = \lambda x. \tag{5.6}$$

Similarly, singular values and singular vectors are the critical values and critical points of $x^{\mathrm{T}}Ay/(\|x\|_2\,\|y\|_2)$. Introduce the Lagrange function

$$L(x, y, \sigma) = x^{\mathrm{T}}Ay - \sigma(\|x\|_2\,\|y\|_2 - 1). \tag{5.7}$$

The first order condition yields

$$\frac{Ay_c}{\|y_c\|_2} = \frac{\sigma x_c}{\|x_c\|_2}, \quad \frac{A^{\mathrm{T}}x_c}{\|x_c\|_2} = \frac{\sigma y_c}{\|y_c\|_2}$$

at a critical point (x_c, y_c, σ_c). Writing

$$u_c = \frac{x_c}{\|x_c\|_2}, \quad v_c = \frac{y_c}{\|y_c\|_2},$$

we get

$$Av_c = \sigma_c u_c, \quad A^{\mathrm{T}}u_c = \sigma_c v_c, \tag{5.8}$$

which give the singular values and singular vectors.

We extend the above variational approach to the eigenvalues and singular values of tensors. Let $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_N}$,

$$\mathscr{A}(x_1, \ldots, x_N) = \sum_{i_1=1}^{d_1} \cdots \sum_{i_N=1}^{d_N} \mathscr{A}_{i_1 \cdots i_N} x_{i_1}^{(1)} \cdots x_{i_N}^{(N)}, \tag{5.9}$$

$$\nabla_{x_i} \mathscr{A}(x_1, \ldots, x_N) = \mathscr{A}(x_1, \ldots, x_{i-1}, I_{d_i}, x_{i+1}, \ldots, x_N). \tag{5.10}$$

We define the singular values and singular vectors of $\mathscr{A}$ as the critical values and critical points of

$$\frac{\mathscr{A}(x_1, \ldots, x_N)}{\|x_1\|_{p_1} \cdots \|x_N\|_{p_N}}, \tag{5.11}$$

where $\|\cdot\|_{p_i}$ denotes the l_{p_i}-norm, $i = 1, \ldots, N$.

Introduce Lagrange function

$$L(x_1, \ldots, x_N, \sigma) := \mathscr{A}(x_1, \ldots, x_N) - \sigma(\|x_1\|_{p_1}, \ldots, \|x_N\|_{p_N} - 1). \tag{5.12}$$

Vanishing the gradient,

$$\nabla L = (\nabla_{x_1} L, \ldots, \nabla_{x_N} L, \nabla_\sigma L) = (\mathbf{0}, \ldots, \mathbf{0}, 0)$$

and setting

$$\|x_1\|_{p_1} = \cdots = \|x_N\|_{p_N} = 1,$$

give: when $p_1 = \cdots p_N = 2$,

$$\mathscr{A}(x_1, \ldots, x_{i-1}, I_{d_i}, x_{i+1}, \ldots, x_N) = \sigma x_i, \quad i = 1, \ldots, N; \tag{5.13}$$

when $p_1 = \cdots = p_N = N$,

$$\mathscr{A}(x_1, \ldots, x_{i-1}, I_{d_i}, x_{i+1}, \ldots, x_N) = \sigma x_i^{[N-1]}, \quad i = 1, \ldots, N; \tag{5.14}$$

(5.13) and (5.14) give the l_2-singular values and l_N-singular values, respectively.

Let $\|\cdot\|_{\alpha_i}$ be a norm on $\mathbb{R}^{d_i}, i = 1, \ldots, N$. The norm of $\mathscr{A}$ is defined by

$$\|\mathscr{A}\|_{\alpha_1, \ldots, \alpha_N} := \sup \frac{|\mathscr{A}(x_1, \ldots, x_N)|}{\|x_1\|_{\alpha_1} \cdots \|x_N\|_{\alpha_N}}, \tag{5.15}$$

where the supremum is taken over all nonzero $x_i \in \mathbb{R}^{d_i}$ (see Ref. [18]). By use of the above norm, we have

$$\sigma_{\max}(\mathscr{A}) = \|\mathscr{A}\|_{p_1, \ldots, p_N}. \tag{5.16}$$

In particular, in the case of supersymmetric tensor, (5.12) is reduced to

$$L(x, \lambda) = \mathscr{A}(x, \ldots, x) - \lambda(\|x\|_p^N - 1). \tag{5.17}$$

Vanishing the gradient,

$$\nabla L = (\nabla_x L, \nabla_\lambda L) = (\mathbf{0}, 0)$$

gives that when $p = 2$,

$$\mathscr{A}(In, x, \ldots, x) = \lambda x, \quad \|x\|_2 = 1 \tag{5.18}$$

and when $p = N$,

$$\mathscr{A}(In, x, \ldots, x) = \lambda x^{[N-1]}, \tag{5.19}$$

and in this case the unit-norm constraint is superfluous. Obviously, the l_2-eigenvalue defined by (5.18) is just Z-eigenvalue defined in (5.3); and the l_N-eigenvalue defined by (5.19) is just the H-eigenvalue in (5.2) (see Ref. [61]).

Z-eigenvalues/Z-singular values and H-eigenvalues/H-singular values have wide applications. In §4, we have discussed that we can solve the rank-1 approximation of tensors by using Z-eigenvalues. In Refs. [1,17], Z-eigenvalues of supersymmetric tensor are used for approximating solutions of constraint satisfaction problems. Lim [50] gives a generalization of the Perron-Frobenius theorem by use of H-eigenvalues.

Theorem 5.3 *Let $\mathscr{A} \in \mathbb{R}^{n\times\cdots\times n}$ be irreducible and $\mathscr{A} \geqslant 0$. Then $\mathscr{A}$ has a positive real H-eigenvalue with its eigenvector x^* that may be chosen to have all entries non-negative. In fact, x^* is unique and has all entries positive.*

6 Applications of numerical multilinear algebra

The blind source separation (BSS) problem in signal processing has wide applications. For example, mobile communications and the surveillance of radio-communications in civil context, interception and classification in military applications resort to BSS. The BSS problem usually uses high-order statistics which are tensor objects. The BSS problem is to separate and recover a number of statistically independent sources from a set of sensor observations without a priori knowledge of either the sources or the channels, in other words, the BSS problem is to separate and recover statistically independent stochastic process by using linear mixing form

$$\mathbf{y} = A\mathbf{x} + \mathbf{v}, \tag{6.1}$$

where $\mathbf{y}$ is an observed random variable of dimension n, A: $m \times n$ a mixing matrix, $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ a source vector, $\mathbf{v}$ a noise vector and independent of $\mathbf{x}$. The usual method to recover a source is to project the observed value $\mathbf{y}$ to an $m \times 1$ vector $\mathbf{u}$ which is selected such that the source estimate $z = \mathbf{u}^{\mathrm{T}}\mathbf{y}$ is absolutely maximized (see Refs. [9,12,53,76]). Lathauwer et al. [39] shows that this maximization is equivalent to the best approximation to a forth-order cumulant tensor by another rank-1 tensor. Since such a tensor

is supersymmetric, the symmetric higher-order power method (S-HOPM) can be used. Reference [39] proves that when

$$g(\mathbf{u}) = \sum_{i_1,\ldots,i_N} \mathscr{A}_{i_1\cdots i_N} \mathbf{u}_{i_1} \cdots \mathbf{u}_{i_N}$$

is convex (or concave), S-HOPM converges to a (local) maximum of $|g|$ on unit sphere. The requirement of convexity is satisfied in BSS. The computational expenses of S-HOPM is N times less than HOPM. In Ref. [52], Luo and Lu formulate BSS problem as an optimization problem using the mutual information criterion, and solve the resulting optimization problems with an extended Newton's method.

The BSS problem is also known as the independent component analysis (ICA) problem, while the problem of blind separation of convolutive mixtures is also known as blind deconvolution (BD) problem. The problems of BD and BSS have met with fruitful interplay in recent years. The two problems share a common setting: one is given a mixture of independent random variables or 'sources', where the mixture may be temporal or spatial, and the goal is to restore one of the sources.

In blind deconvolution in the nonlinear channel, we consider the case where non-i.i.d. sources (i.e., mutually independent) and/or nonlinear channels are presented. We consider a multi-source communication channel, in which the received signal $\{\mathbf{u}\}$ comprises P components, the impulse response is $\{\mathbf{g}\}_{k=0}^{M}$. Introduce the fourth-order cumulant tensor as a four-way array

$$\mathscr{A}_{ijkl} = \mathrm{cum}(u_i, u_j, u_k, u_l). \tag{6.2}$$

Then the Shalvi-Weinstein cost function J_{SW} appears as the ratio

$$J_{\mathrm{SW}}(\mathbf{g}) = \sum_{i,j,k,l=1}^{(M+1)P} \mathscr{A}_{ijkl} g_i g_j g_k g_l \Big/ \sum_{i=1}^{(M+1)P} g_i^2. \tag{6.3}$$

The four-way Tucker product (outer product) will denote the rank-1 tensor

$$[\mathbf{g} \star \mathbf{g} \star \mathbf{g} \star \mathbf{g}]_{i,j,k,l} = g_i g_j g_k g_l. \tag{6.4}$$

So, the rank-1 tensor approximation problem can be used to find a unit-norm vector $\mathbf{g}$ and a scalar λ, such that

$$\min \|\mathscr{A} - \lambda \mathbf{g} \star \mathbf{g} \star \mathbf{g} \star \mathbf{g}\|_F^2, \tag{6.5}$$

which is the sum of squares of all elements of the array.

Note that, the unit-norm vector $\mathbf{g}$ and the scalar λ yield a local minimum of (6.5) if and only if $\mathbf{g}$ is a local maximizer of $|J_{\mathrm{SW}}|$ of (6.3), at which $\lambda = J_{\mathrm{SW}}(\mathbf{g})$.

So, blind deconvolution in the nonlinear channel is equivalent to rank-1 tensor approximation. Hence, the iterative methods, for example, higher-order power method (HOPM) for constructing rank-1 tensor approximation can be used for the blind deconvolution problem.

Solving the multivariate polynomial optimization is also an important topic of applications of tensors. Consider the problem

$$\inf_{x\in\Omega} P(x), \tag{6.6}$$

where $P(x)$ is a real-valued polynomial in n variables over $\mathbb{R}^n$, and Ω is $\mathbb{R}^n$ or is a compact set defined by polynomial inequalities.

For this problem, the derivative-based optimization method is to look at the first order conditions which constitute a system of polynomial equations. Various algebraic techniques can be used for determining the real solutions to this system of equations. Via successive changes of variables, the problem can be transformed into a quadratically constrained optimization problem and then a good lower bound can be obtained by solving a standard convex linear matrix inequality relation (see Ref. [70]). By adding redundant quadratic constraints, one may improve the lower bound and obtain the optimal value (see Ref. [19]).

Semidefinite programming relaxation (SDP-relaxation) for unconstrained polynomial programming has attracted a lot of attention in recent years. The sequence of optimal values of the SDP-relaxations associated with the original problem was shown to be convergent to the global minimum under relatively weak assumptions [35–37], and the convergence to global minimizer is proved [68].

The study of relationships between nonnegative and sums of squares (SOS) polynomials, initiated by Hilbert, is of real importance in numerous potential applications (see Ref. [38]).

As a special case of (6.1), we consider the following global polynomial optimization problem:

$$\begin{aligned} &\min f(x) = \mathscr{A}x^N = \sum_{i_1,\ldots,i_N=1}^{n} \mathscr{A}_{i_1\cdots i_N} x_{i_1}\cdots x_{i_N} \\ &\text{s.t.} \quad x^{\mathrm{T}}x = 1, \end{aligned} \tag{6.7}$$

where $x \in \mathbb{R}^n,\ N, n \geqslant 2,\ f$ is a homogeneous polynomial of degree N with n variables, and $\mathscr{A}$ is an N-th order n-dimensional real supersymmetric tensor.

A popular method for solving the polynomial minimization problem is the SOS method, which was first introduced by Shor [69] and further developed by Nesterov [56], Lasserre [35] and Parrilo [60], etc. The SOS method is based on semidefinite programming (SDP). Roughly speaking, there are two approaches used in SOS method. One is that the global minimum of a constrained polynomial optimization is approximated by a sequence of SDP bounds (see Ref. [35]). The other is that the ideal is to compute the largest real number such that the residual obtained by subtracting this real number from the objective polynomial function is a sum of squares of polynomials. The SOS method has good theoretical support and good numerical performance in the lower dimensional case.

The critical points of (6.7) satisfy the following system for some $\lambda \in \mathbb{R}$:

$$\begin{cases} \mathscr{A}x^{N-1} = \lambda x, \\ x^{\mathrm{T}}x = 1. \end{cases} \tag{6.8}$$

So, we can use Z-eigenvalue method for solving the polynomial optimization problem (6.7). Qi et al. [63] proposed a direct Z-eigenvalue method for the case $n = 3$ and an r-th order pseudo-canonical form method for the higher dimensional case which uses lower dimensional Z-eigenvalue method and the SOS method as subroutines. Numerical experiments show that these methods are efficient and promising.

In addition, numerical tensor methods play an important role in numerical optimization. As mentioned above, several topics in semidefinite programming (SDP) can be dealt with by tensor methods. In Refs. [48,49], the optimality condition and a nonsmooth Newton-type method for nonlinear SDP problems are proposed by tensor forms. The tensor-type quasi-Newton equation was established in Ref. [73]. Some tensor models for unconstrained optimization have been proposed by Schnabel et al. [64–66]. However, these works are still in a baby stage and need to do further research.

7 Conclusion

In this paper, we survey the motivation with the background, the development, the main branches of the numerical multilinear algebra. We have investigated tensor decompositions (for example, high-order eigenvalue decomposition, high-order singular value decomposition, canonical and pseudo-canonical decomposition, Schur decomposition of tensors, etc.), rank-1 and rank-r tensor approximation, high-order eigenvalue related problems, multivariate polynomial optimization, and typical applications of tensors (for example, BSS, BD, SDP, and other optimization problems). At present, various applications in digital image restoration, signal processing, wireless communication, psychometrics, multi-way data analysis and high-order statistics are strongly stimulating the development of this area. So far, several topics in the area just start the first step. Numerical investigation and numerical results are still limited. However, the area is exciting, and significant development will be expected in the next several years.

References

1. Alon N, de la Vega W F, Kannan R, et al. Random sampling and approximation of max-csps. J Comput System Sci, 2003, 67: 212–243
2. Bergman G M. Ranks of tensors and change of base field. J Algebra, 1969, 11: 613–621
3. Bienvenu G, Kopp L. Optimality of high-resolution array processing using the eigensystem approach. IEEE Trans ASSP, 1983, 31: 1235–1248
4. Cao X R, Liu R W. General approach to blind source separation. IEEE Trans Signal Processing, 1996, 44: 562–570

5. Cardoso J F. Super-symmetric decomposition of the forth-order cumulant tensor. Blind identification of more sources than sensors. In: Proceedings of the IEEE International Conference on Acoust, Speech, and Signal Processing (ICASSP'91), Toronto, Canada, 1991
6. Cardoso J F. High-order contrasts for independent component analysis. Neural Computation, 1999, 11: 157–192
7. Caroll J D, Chang J J. Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition. Psychometrika, 1970, 35: 283–319
8. Comon P. Tensor diagonalization, a useful tool in signal processing. In: Blanke M, Soderstrom T, eds. IFAC-SYSID, 10th IFAC Symposium on System Identification (Copenhagen, Denmark, July 4-6, 1994. Invited Session). Vol 1, 77–82
9. Comon P. Independent component analysis, a new concept? Signal Processing, Special Issue on Higher-Order Statistics, 1994, 36: 287–314
10. Comon P, Mourrain B. Decomposition of quantics in sums of powers of linear forms. Signal Processing, Special Issue on Higher-Order Statistics, 1996, 53: 96–107
11. Comon P. Block methods for channel identification and source separation. In: IEEE Symposium on Adaptive Systems for Signal Process, Commun Control (Lake Louise, Alberta, Canada, Oct 1-4, 2000. Invited Plenary). 87–92
12. Comon P, Chevalier P. Blind source separation: Models, concepts, algorithms, and performance. In: Haykin S, ed. Unsupervised Adaptive Filtering, Vol 1. New York: John Wiley, 2000
13. Comon P. Tensor decompositions: State of the art and applications. In: McWhirter J G, Proundler I K, eds. Mathematics in Signal Processing V. Oxford: Oxford University Press, 2002
14. Comon P. Canonical tensor decompositions. In: ARCC Workshop on Tensor Decompositions, Americal Institute of Mathematics (AIM), Palo Alto, California, USA, July 19–23, 2004
15. Comon P, Golub G, Lim L H, et al. Symmetric tensors and symmetric tensor rank. SIAM J Matrix and Applications, 2007 (in press)
16. Coppi R, Bolasco S, eds. Multiway Data Analysis. Amsterdam: Elsevier, 1989
17. de la Vega W F, Kannan R, Karpinski M, et al. Tensor Decomposition and Approximation Algorithms for Constraint Satisfaction Problems. New York: ACM Press, 2005, 747–754
18. Defant A, Floret K. Tensor Norms and Operator Ideals. North-Holland Mathematics Studies, No 176. Amsterdam: North-Holland, 1993
19. Ferrier C. Hilbert's 17th problem and best dual bounds in quadatic minimization. Cybernet Systems Anal, 1998, 34: 696–709
20. Golub G, Van Loan C F. Matrix Computations. 3rd ed. Baltimore: Johns Hopkins University Press, 1996
21. Golub G, Kolda T G, Nagy J, et al. Workshop on Tensor Decompositions, American Institute of Mathematics, Palo Alto, California, 19–23 July, 2004. http://www.aimath.org/WWN/tensordecomp/
22. Golub G, Mahoney M, Drinears P, et al. Workshop for Modern Massive Data Sets, 21–24 June, 2006. http://www.stanford.edu/group/mmds/
23. Golub G, Mahoney M, Drinears P, et al. Bridge the gap between numerical linear algebra, theoretical computer science, and data applications. SIAM News, 2006, 39(8). http://www.siam.org/pdf/news/1019.pdf
24. Greub W H. Multilinear Algebra. Berlin: Springer-Verlag, 1967
25. Harshman R A. Determination and proof of minimum uniqueness conditions of PARAFAC. UCLA Working Papers in Phonetics, 1972, 22: 111–117
26. He X, Sun W. Introduction to Generalized Inverses of Matrices. Nanjing: Jiangsu Sci & Tech Publishing House, 1990 (in Chinese)
27. Hitchcock F L. The expression of a tensor or a polyadic as a sum of products. J Math Physics, 1927, 6: 164–189
28. Hitchcock F L. Multiple invariants and generalized rank of a p-way matrix or tensor. J Math Physics, 1927, 7: 39–79

29. Kilmer M E, Martin C D M. Decomposing a tensor. SIAM News, 2004, 37. http://www.siam.org/siamnews/11-04/tensor.pdf
30. Kofidis E, Regalia P A. Tensor approximation and signal processing applications. In: Olshevsky V, ed. Structured Matrices in Mathematics, Computer Science and Engineering, Vol I. Contemporary Mathematics, 280. Providence: AMS, 2001
31. Kofidis E, Regalia P A. On the best rank-1 approximation of higher-order supersymmetric tensors. SIAM J Matrix Anal Appl, 2002, 23: 863–884
32. Kolda T G. Orthogonal tensor decompositions. SIAM J Matrix Anal Appl, 2001, 22: 243–255
33. Kroonenberg P M. Singular value decompositions of interactions in three-way contigency tables. In: Coppi R, Bolasco S, eds. Multiway Data Analysis. North Holland: Elsevier Science Publishers, 1989: 169–184
34. Kruskal J B. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra and Applications. 1977, 18: 95–138
35. Lasserre J B. Global optimization with polynomials and the problem of moments. SIAM J Optimization, 2001, 11: 796–817
36. Lasserre J B. Semidefinite programming vs. LP relaxations for polynomial programming. Mathematics of Operations Research, 2002, 27: 347–360
37. Lasserre J B. Polynomial programming: LP-relaxations also converge. SIAM J Optimization, 2004, 15: 383–393
38. Lasserre J B. A sum of squares approximation of nonnegative polynomials. SIAM J Optimization, 2006, 16: 751–765
39. De Lathauwer L, Comon P, De Moor B, et al. Higher-order power method — application in independent component analysis. In: Proceedings of the International Symposium on Nonlinear Theory and Its Applications (NOLTA'95), Las Vegas, NV. 1995, 91–96
40. De Lathauwer L, De Moor B. From matrix to tensor: Multilinear algebra and signal processing. In: Mathematics in Signal Processing, IMA Conference Series (Warwick, Dec 17-19, 1996), Oxford: Oxford University Press, 1996
41. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM J Matrix Anal Appl, 2000, 21: 1253–1278
42. De Lathauwer L, De Moor B, Vandewalle J. On the rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. SIAM J Matrix Anal Appl, 2000, 21: 1324–1342
43. De Lathauwer L. First-order perturbation analysis of the best rank-(R_1, R_2, R_3) approximation in multilinear algebra. J Chemometrics, 2004, 18: 2–11
44. De Lathauwer L, De Moor B, Vandewalle J. Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. SIAM J Matrix Anal Appl, 2004/05, 26: 295–327
45. De Lathauwer L, Comon P. Workshop on Tensor Decompositions and Applications, Luminy, Marseille, France, August 29–September 2, 2005. http://www.etis.ensea.fr/wtda/
46. Leon S. Workshop Report: Algorithms for Modern Massive Data Sets (MMDS). NA Digest, 2006, Vol 6, No 27
47. Leurgans S E, Ross R T, Abel R B. A decomposition for three-way arrays. SIAM J Matrix Anal Appl, 1993, 14: 1064–1083
48. Li C, Sun W. A nonsmooth Newton-type method for nonlinear semidefinite programs. Technical Report, School of Mathematics and Computer Science, Nanjing Normal University, March, 2007
49. Li C, Sun W. Filter-Successive Linearization Methods for Nonlinear Semidefinite Programs. Technical Report, School of Mathematics and Computer Science, Nanjing Normal University, December, 2006
50. Lim L H. Singular values and eigenvalues of tensors: A variational approach. In: Proceedings of the 1st IEEE International Workshop on Computational Advances of multi-sensor Adaptive Processing (CAMSAP), December 13–15, 2005. 2005, 129–132

51. Liu X, Sidiropoulos N D. Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays. IEEE Trans on Signal Processing, 2001, 49: 2074–2086
52. Luo Z Q, Lu J. On blind source separation using mutual information criterion. Mathematical Programming, Ser B, 2003, 97: 587–603
53. McCullagh P. Tensor Methods in Statistics. Monographs in Statistics and Applied Probability. London: Chapman and Hall, 1987
54. Merris R. Multilinear Algebra. The Netherland: Gordon and Breach Science Publishers, 1997
55. Moré J J. The Levenberg-Marquadt algorithm: implementation and theory. In: Watson G A, ed. Numerical Analysis. Lecture Notes in Math, Vol 630, Berlin: Springer-Verlag, 1977, 105–116
56. Nesterov Y. Squared functional systems and optimization problems. In: Frenk H, Roos K, Terlaky T, et al, eds. High Performance Optimization, Dordrecht: Kluwer, 2000, 405–440
57. Ni G, Qi L, Wang F, et al. The degree of the E-characteristic polynomial of an even order tensor. Journal of Mathematical Analysis and Applications, 2007, 329: 1218–1229
58. Northcoot D G. Multilinear Algebra. Cambridge: Cambridge University Press, 1984
59. Paatero P. A weighted non-negative least squares algorithm for three-way "PARAFAC" factor analysis. Chemometrics Intell Lab Syst, 1997, 38: 223–242
60. Parrilo P A. Semidefinite programming relaxation for semialgebraic problems. Mathematical Programming, 2003, 96: 293–320
61. Qi L. Eigenvalues of a real supersymmetric tensor. Journal of Symbolic Computation, 2005, 40: 1302–1324
62. Qi L. Rank and eigenvalues of a supersymmetric tensor, a multivariate homogeneous polynomial and an algebraic surface defined by them. Journal of Symbolic Computation, 2006, 41: 1309–1327
63. Qi L. Eigenvalues and invariants of tensors. Journal of Mathematical Analysis and Applications, 2007, 325: 1363–1377
64. Schnabel R B. Conic methods for unconstrained optimization and tensor methods for nonlinear equations. In: Bachem A, Grotschel M, Korte B, eds. Mathematical Programming, The State of the Art. Berlin: Springer-Verlag, 1983, 417–438
65. Schnabel R B, Frank P D. Tensor methods for nonlinear equations. SIAM J Numerical Analysis, 1984, 21: 815–843
66. Schnabel R B, Chow T T. Tensor methods for unconstrained optimization using second derivatives. SIAM J Optimization, 1991, 1: 293–315
67. Schott J R. Matrix Analysis for Statistics, 2nd ed. New York: John Wiley & Sons, Inc, 2005
68. Schweighofer M. Optimization of polynomials on compact semialgebraic sets. SIAM J Optimization, 2005, 15: 805–825
69. Shor N Z. Class of global minimum bounds of polynomial functions. Cybernetics, 1987, 23: 731–734
70. Shor N Z. Nondifferentiable Optimization and Polynomial Problems. Dordrecht: Kluwer, 1998
71. Sidiropoulos N, Giannakis G, Bro R. Blind PARAFAC receivers for DS-CDMA systems. IEEE Trans Signal Process, 2000, 48: 810–823
72. Sidiropoulos N, Bro R. On the uniqueness of multilinear decomposition of N-way arrays. J Chemometrics, 2000, 14: 229–239
73. Sun W. Quasi-Newton methods for nonlinear matrix equations. Technical Report, School of Mathematics and Computer Science, Nanjing Normal University, 2006
74. Sun W, Du Q, Chen J. Computational Methods. Beijing: Science Press, 2007 (in Chinese)
75. Sun W, Yuan Y. Optimization Theory and Methods: Nonlinear Programming. New York: Springer, 2006
76. Talwar S, Vilberg M, Paulraj A. Blind estimation of multiple cochannel digital signals arriving at an antenna array: Part I, algorithms. IEEE Trans Signal Process, 1996, 44: 1184–1197

77. Tucker L R. Some mathematical notes on the three-mode factor analysis. Psychometrika, 1966, 31: 279–311
78. Van Der Veen A -J, Paulraj A. An analytical constant modulus algorithm. IEEE Trans Signal Processing, 1996, 44: 1136–1155
79. Vorobyov S A, Rong Y, Sidiropoulos N D, et al. Robust iterative fitting of multilinear models. IEEE Trans on Signal Processing, 2005, 53: 2678–2689
80. Wang Y, Qi L. On the successive supersymmetric rank-1 decomposition of higher order supersymmetric tensors. Numerical Linear Algebra with Applications, 2007 (in press)
81. Zhang T, Golub G H. Rank-one approximation to high order tensors. SIAM J Matrix Anal Appl, 2001, 23: 534–550