T-positive semidefiniteness of third-order symmetric tensors and T-semidefinite programming

Meng-Meng Zheng[1] · Zheng-Hai Huang[1] · Yong Wang[1]

Abstract

The T-product for third-order tensors has been used extensively in the literature. In this paper, we first introduce first-order and second-order T-derivatives for the multi-variable real-valued function with the tensor T-product. Inspired by an equivalent characterization of a twice continuously T-differentiable multi-variable real-valued function being convex, we present a definition of the T-positive semidefiniteness of third-order symmetric tensors. After that, we extend many properties of positive semidefinite matrices to the case of third-order symmetric tensors. In particular, analogue to the widely used semidefinite programming (SDP for short), we introduce the semidefinite programming over the space of third-order symmetric tensors (T-semidefinite programming or TSDP for short), and provide a way to solve the TSDP problem by converting it into an SDP problem in the complex domain. Furthermore, we give several TSDP examples and especially some preliminary numerical results for two unconstrained polynomial optimization problems. Experiments show that finding the global minimums of polynomials via the TSDP relaxation outperforms the traditional SDP relaxation for the test examples.

Keywords T-product · T-positive semidefiniteness · T-semidefinite cone · T-semidefinite programming · Polynomial optimization

Mathematics Subject Classification 15A69 · 90C22

✉ Zheng-Hai Huang
    huangzhenghai@tju.edu.cn

    Meng-Meng Zheng
    zmm941112@tju.edu.cn

    Yong Wang
    wang_yong@tju.edu.cn

[1] School of Mathematics, Tianjin University, Tianjin 300072, People's Republic of China

1 Introduction

With the availability of inexpensive storage and advances in instrumentation, the data collected and stored now is more complex than ever before. Especially in practical problems such as psychometrics, signal processing, computer vision, data mining, graphical analysis, neuroscience and so on, it is usually necessary to store information in a multidimensional array, and then use the multidimensional structure to compress, sort, and/or manipulate the data. Among the many problems described by high-dimensional arrays (or tensors), third-order tensors have become increasingly prevalent in recent years with the emergence of the tensor T-product, which is a new type of multiplication between third-order tensors introduced by Kilmer, Martin, and Perrone [1]. The tensor T-product has shown to be a useful tool arising in a wide variety of application areas, including, but not limited to, image processing [2–7], computer vision [8–12], signal processing, low rank tensor recovery and robust tensor PCA [13–18], and data completion and denoising [19–31], because the tensor T-product provides an effective approach to transform the tensor multiplication into block diagonal matrix multiplication in the discrete Fourier domain.

Since Kilmer, Martin, and Perrone [1] introduced the new type of multiplications between two third-order tensors so as to devise new types of factorizations for tensors to be easily used in applications, the exploration of the algebraic properties of T-products has been in progress. Specifically, in [32] some factorization strategies were established for third-order tensors via the tensor T-product. In [33] and [3], the authors provided useful frameworks to better view the action of the third-order tensors upon a set of matrices. In [34], a lot of familiar tools of linear algebra were extended to the third-order tensors, including the T-Jordan canonical form, tensor decomposition theory, T-group inverse and T-Drazin inverse, and so on. In addition, Lund [35] proposed a definition of tensor functions based on the T-product of third-order F-square tensors, which was found to be of great use in stable tensor neural networks for rapid deep learning [36], and then Miao, Qi and Wei generalized the tensor T-function from F-square third order tensors to rectangular tensors in [37].

It is well known that the positive (semi)definite (P(S)D for short) matrix is an important class of matrices, which has a solid theory and fruitful applications. Actually, P(S)D matrices can be used for inequality proof, eigenvalue solving, extremum solving, system stability discrimination, and so on. As a result, P(S)D matrices have been applied in various fields, such as numerical analysis, optimization theory, probability and statistics, operations research, control theory, mechanics, electricity, information science and technology, management science and engineering, and so on. For more information about P(S)D matrices, the readers are referred to the monographs [38, 39]. In addition, the semidefinite programming (SDP for short) as one of the important applications of PSD matrices has received extensive attention. Especially with the appearance of some effective algorithms for SDP problems [40–42], SDP problems have been increasingly arisen in practical applications. There are rich and mature results for SDPs in both theory and algorithms. See [43–45] and references therein.

Motivated by that mentioned above, we extend the positive (semi)definiteness of matrices and the SDP problem to the case of third-order tensors. Our contribution is threefold:

(1) We explore the Fréchet derivatives of the multi-variable real-valued function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ based on the inner product and the tensor T-product. We establish the necessary and sufficient conditions for a multi-variable real-valued function being first-order and second-order T-differentiable, respectively, and present the exact forms of the T-derivatives. In particular, we propose a second-order condition to certify the convexity of the multi-variable real-valued function under the hypothesis that the function is twice continuously T-differentiable.

(2) We give a definition of a third-order symmetric tensor being T-positive semidefinite (T-PSD for short) inspired by the second-order T-derivative, and show that the new definition is equivalent to the one given by [3, Definition 2.7] and the one by [34, Definition 15] in real case. In particular, we show that the set of symmetric T-PSD tensors is a nonempty, closed, convex, pointed and self-dual cone, and extend many properties of PSD matrices to the case of third-order T-PSD tensors, including some results related to the T-eigenvalue decomposition, the T-roots, the T-Schur complement, and so on.

(3) As an important application of the T-positive semidefiniteness of third-order symmetric tensors, we introduce the semidefinite programming over the third-order symmetric tensor space (T-semidefinite programming or TSDP for short) and show that a TSDP problem of size $m \times m \times p$ can be transformed into an SDP problem with $\frac{p+1}{2}$ or $\frac{p+2}{2}$ blocks of matrices of $m \times m$ in the complex domain. Then we present several examples which can be modeled (or relaxed) as TSDPs, such as minimizing the maximum T-eigenvalue of a third-order tensor, minimizing the l_2 operator norm of a third-order tensor, minimizing the dual norm of the l_2 operator norm of a third-order tensor, integer quartic programming and calculating the global lower bound of a polynomial. Besides, we report preliminary numerical results for solving the unconstrained polynomial optimization problem via the TSDP relaxation, which can achieve higher accuracy and consume less time compared with the traditional SDP relaxation.

The rest of our paper is organized as follows. In Sect. 2, some notation and basic results are reviewed. In Sect. 3, we explore the T-derivatives for the multi-variable real-valued function and the relationship between a new type of tensors: the T-Hessian Tensor and convexity of the twice continuously T-differentiable multi-variable real-valued function. In Sect. 4, we give the definition of the symmetric T-PSD tensor, then discuss some characterizations and properties of symmetric T-PSD tensors, and investigate the set of T-PSD tensors. In Sect. 5, we introduce and study the TSDP; and convert the TSDP into the corresponding SDP in the complex domain. We also present several examples for applications and report preliminary numerical results. Finally, we sum up the conclusions and do some further discussions in Sect. 6.

2 Preliminary

In this section, we give some notation and basic results.

2.1 Notation

Throughout this paper, we use small letters $a, b, \ldots$ for scalars, small bold letters $\mathbf{a}, \mathbf{b}, \ldots$ for vectors, capital letters $A, B, \ldots$ for sets, capital bold letters $\mathbf{A}, \mathbf{B}, \cdots$ for matrices, and calligraphic letters $\mathscr{A}, \mathscr{B}, \ldots$ for tensors. For any positive integer n, denote $[n] := \{1, 2, \ldots, n\}$. Let $\mathbb{R}^n := \{\mathbf{x} := (x_1, x_2, \ldots, x_n)^\top : x_i \in \mathbb{R} \text{ for all } i \in [n]\}$ and $\mathbb{C}^n := \{\mathbf{x} := (x_1, x_2, \ldots, x_n)^\top : x_i \in \mathbb{C} \text{ for all } i \in [n]\}$ where $\mathbb{R}$ ($\mathbb{C}$) is the set of real (complex) numbers. Let m, n and p be positive integers. $\mathbb{R}^{m \times n}$ and $\mathbb{R}^{m \times n \times p}$ denote the sets consisting of all real matrices of size $m \times n$ and all real tensors of size $m \times n \times p$, respectively. Let $\mathbb{N}$ denote the set of nonnegative integers. For $\alpha \in \mathbb{N}^n$, denote $|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n$. For any $\mathbf{x} \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$, $\mathbf{x}^\alpha$ means $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, and $\mathbf{x}^\top$ represents the transpose of $\mathbf{x}$. For any $\mathscr{A}, \mathscr{B} \in \mathbb{R}^{m \times n \times p}$, the inner product between $\mathscr{A}, \mathscr{B}$ is denoted as $\mathscr{A} \cdot \mathscr{B} = \langle \mathscr{A}, \mathscr{B} \rangle := \sum_{i,j,k} a_{ijk} b_{ijk}$, and the Frobenius norm associated with the above inner product is $\|\mathscr{A}\| = \sqrt{\mathscr{A} \cdot \mathscr{A}}$. Specially, any matrix $\mathbf{A} \in \mathbb{R}^{n \times p}$ can be regarded as a tensor $\mathscr{A} \in \mathbb{R}^{n \times 1 \times p}$ with the i-th frontal slice of $\mathscr{A}$ being the i-th column of $\mathbf{A}$ for all $i \in [p]$.

Recall that a complex matrix $\mathbf{A}$ is said to be symmetric (Hermitian) if and only if $\mathbf{A}^\top = \mathbf{A}$ ($\mathbf{A}^H = \mathbf{A}$), where $\mathbf{A}^\top$ ($\mathbf{A}^H$) represents the transpose (conjugate transpose) of $\mathbf{A}$. We denote the set consisting of all real symmetric (complex Hermitian) matrices of size $n \times n$ as $S\mathbb{R}^{n \times n}$ ($H\mathbb{C}^{n \times n}$). For any $x \in \mathbb{C}$ and $\mathbf{X} := (x_{ij}) \in \mathbb{C}^{m \times n}$, $\bar{x}$ denotes the conjugate of x and $\overline{\mathbf{X}} := (\bar{x}_{ij})$ denotes the conjugate of the matrix $\mathbf{X}$. Let $\mathbf{U} \succeq (\succ)0$ represent that $\mathbf{U}$ is (Hermitian) positive semidefinite (positive definite) for any $\mathbf{U} \in H\mathbb{C}^{n \times n}$, and $S\mathbb{R}_{++}^{n \times n}(S\mathbb{R}_+^{n \times n})$ denote the set of all real symmetric positive (semi) definite matrices of size $n \times n$, while $H\mathbb{C}_{++}^{n \times n}(H\mathbb{C}_+^{n \times n})$ denotes the set of all complex Hermitian positive (semi)definite matrices of size $n \times n$. "$\otimes$" denotes the Kronecker product between two matrices and "$\cdot$" means standard matrix product.

2.2 Tensor T-product, transpose and inverse

For a third-order tensor $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$, a new perspective was proposed in [1, 32] based on treating $\mathscr{A}$ as a stack of frontal slices, which were denoted as $\mathbf{A}^{(k)} \in \mathbb{R}^{m \times n}$ for all $k \in [p]$. Furthermore, several operators on $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$ were introduced as follows:

$$bcirc(\mathscr{A}) := \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(p)} & \mathbf{A}^{(p-1)} & \cdots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \mathbf{A}^{(p)} & \cdots & \mathbf{A}^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{A}^{(p)} & \mathbf{A}^{(p-1)} & \cdots & \mathbf{A}^{(2)} & \mathbf{A}^{(1)} \end{bmatrix}, \quad unfold(\mathscr{A}) := \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(p)} \end{bmatrix},$$

$fold(unfold(\mathscr{A})) := \mathscr{A}$, and $bcirc^{-1}(bcirc(\mathscr{A})) := \mathscr{A}$.

With the help of the above operators, the following definitions and properties were given in [3] (see also [1, 10, 32, 34]).

Definition 1 [3, Definition 2.5] (T-product) Let $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$ and $\mathscr{B} \in \mathbb{R}^{n \times s \times p}$ be two real tensors. Then the T-product $\mathscr{A} * \mathscr{B}$ is an $m \times s \times p$ real tensor defined by $\mathscr{A} * \mathscr{B} := fold(bcirc(\mathscr{A})unfold(\mathscr{B}))$.

Definition 2 [3, Definition 2.7] (Transpose and conjugate transpose) If $\mathscr{A}$ is a third-order tensor of size $m \times n \times p$, then the transpose $\mathscr{A}^{\top}$ is obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through p. The conjugate transpose $\mathscr{A}^{H}$ is obtained by conjugate transposing each of the frontal slices then reversing the order of transposed frontal slices 2 through p.

For any $\mathscr{A} \in \mathbb{R}^{n \times n \times p}$, we say $\mathscr{A}$ is a symmetric tensor if and only if $\mathscr{A}^{\top} = \mathscr{A}$. The set consisting of all the real symmetric tensor of size $n \times n \times p$ is denoted by $S\mathbb{R}^{n \times n \times p}$.

Definition 3 [3, Definition 2.8,2.10] (Identity tensor and inverse) The $n \times n \times p$ identity tensor $\mathscr{I}_{nnp}$ is the tensor whose first frontal slice is the $n \times n$ identity matrix $\mathbf{I}_{n \times n}$, and whose other frontal slices are all zeroes. For a frontal square tensor $\mathscr{A} \in \mathbb{R}^{n \times n \times p}$, we say $\mathscr{A}$ is nonsingular if it has inverse tensor $\mathscr{B}(= \mathscr{A}^{-1})$, provided that $\mathscr{A} * \mathscr{B} = \mathscr{B} * \mathscr{A} = \mathscr{I}_{nnp}$.

It is easy to check that $\mathscr{A} * \mathscr{I}_{nnp} = \mathscr{I}_{mmp} * \mathscr{A} = \mathscr{A}$ for any $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$. In addition, it should be noticed that the invertibility of the third-order frontal square tensor $\mathscr{A}$ is equivalent to the invertibility of the matrix $bcirc(\mathscr{A})$, which can be seen from the following lemma [34, Lemma 3].

Lemma 1 *Suppose that $\mathscr{A} \in \mathbb{R}^{n \times n \times p}$ and $\mathscr{B} \in \mathbb{R}^{n \times s \times p}$. Then*

(a) $bcirc(\mathscr{A} * \mathscr{B}) = bcirc(\mathscr{A})bcirc(\mathscr{B})$,
(b) $bcirc(\mathscr{A})^{k} = bcirc(\mathscr{A}^{k})$ *for all* $k \in \{0, 1, 2, \dots\}$,
(c) $bcirc(\mathscr{A}^{\top}) = bcirc(\mathscr{A})^{\top}, bcirc(\mathscr{A}^{H}) = bcirc(\mathscr{A})^{H}$.

In the rest of this paper, for simplicity, we will use the following notation: for any $\mathscr{A}_i \in \mathbb{R}^{m_i \times n_i \times p}$ and $\mathscr{V}_i \in \mathbb{R}^{m_i \times n \times p}$ with $i \in [l]$, we denote

$$Diag(\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_l) := \begin{bmatrix} \mathscr{A}_1 & & & \\ & \mathscr{A}_2 & & \\ & & \ddots & \\ & & & \mathscr{A}_l \end{bmatrix}, \quad vec(\mathscr{V}_1, \mathscr{V}_2, \cdots, \mathscr{V}_l) := \begin{bmatrix} \mathscr{V}_1 \\ \mathscr{V}_2 \\ \vdots \\ \mathscr{V}_l \end{bmatrix},$$

and sometimes, they are abbreviated as $Diag(\mathscr{A}_i : i \in [l])$ and $vec(\mathscr{V}_i : i \in [l])$, respectively. When all $\mathscr{A}_i$ ($\mathscr{V}_i$) become matrices (or vectors or scalars), similar symbols are also used.

Recall that each circular matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be diagonalized with the normalized discrete Fourier transform (DFT) matrix [48], i.e., there exists a diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \mathbf{F}_n^H \mathbf{D} \mathbf{F}_n$, where $\mathbf{F}_n$ is the Fourier matrix of size $n \times n$ defined as

$$\mathbf{F}_n = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix} \tag{1}$$

where $\omega = e^{\frac{2\pi \mathbf{i}}{n}}$ with $\mathbf{i}^2 = -1$. Similarly, block circular matrices can be block diagonalized by using the Fourier transform. In [3], the authors showed that for any third-order tensor $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$, there exists a block diagonal matrix $Diag(\mathbf{A}_i : i \in [p])$ such that $bcirc(\mathscr{A}) = (\mathbf{F}_n^H \otimes \mathbf{I}_{m \times m}) Diag(\mathbf{A}_i : i \in [p])(\mathbf{F}_p \otimes \mathbf{I}_{n \times n})$, and pointed out the conjugate symmetry between these block matrices $\mathbf{A}_i$.

Lemma 2 [3] *Let $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$ be block diagonalized as*

$$bcirc(\mathscr{A}) = (\mathbf{F}_p^H \otimes \mathbf{I}_{m \times m}) Diag(\mathbf{A}_i : i \in [p])(\mathbf{F}_p \otimes \mathbf{I}_{n \times n}), \tag{2}$$

where $\underline{\mathbf{F}_p}$ is the Fourier matrix defined by (1). *Then, $\mathbf{A}_1 \in \mathbb{R}^{m \times n}$, $\mathbf{A}_i \in \mathbb{C}^{m \times n}$ and $\mathbf{A}_i = \overline{\mathbf{A}_{p+2-i}}$ for any $i \in [p] \backslash \{1\}$. In particular, if $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$ and $\mathbf{A}^{(i)} \in S\mathbb{R}^{n \times n}$, then $\mathbf{A}_i \in \mathbb{R}^{n \times n}$ for any $i \in [p]$ and $\mathbf{A}_i = \mathbf{A}_{p+2-i}$ for any $i \in [p] \backslash \{1\}$.*

Remark 1 (i) It should be noticed that for any $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$ which can be block diagonalized as (2), most of the matrices $\mathbf{A}_i$ ($i \in [p]$) may be complex because of the participating of Fourier matrix $\mathbf{F}_p$, and they satisfy the relationships: $\mathbf{A}_1 \in \mathbb{R}^{m \times n}$ and $\mathbf{A}_i = \overline{\mathbf{A}_{p+2-i}}$ for any $i \in [p] \backslash \{1\}$. On the other hand, it should be noticed that any $\mathbf{A}_i \in \mathbb{C}^{m \times n}$ with $i \in [p]$, which satisfy the above relationships, can lead to a real tensor by the construction as (2). This can be seen from the following discussion.

For any tensor $\mathscr{A}$ which can be block diagonalized as (2), we have that

$$\begin{cases} \mathscr{A}^{(1)} = \frac{1}{p}(\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_p), \\ \mathscr{A}^{(2)} = \frac{1}{p}(\mathbf{A}_1 + \omega^1 \mathbf{A}_2 + \cdots + \omega^{p-1}\mathbf{A}_p), \\ \quad \cdots \\ \mathscr{A}^{(p)} = \frac{1}{p}(\mathbf{A}_1 + \omega^{p-1}\mathbf{A}_2 + \cdots + \omega^{(p-1)(p-1)}\mathbf{A}_p). \end{cases} \tag{3}$$

Thus, from (3) and the fact that $\omega^{np+i} = \omega^i = \overline{\omega^{p-i}}$ for i and $n \in [p-1] \cup \{0\}$, we can see that for any given $\mathbf{A}_i \in \mathbb{C}^{m \times n}$ with $i \in [n]$, the tensor obtained by the construction as (2) is real if and only if $\mathbf{A}_1 \in \mathbb{R}^{m \times n}$ and $\mathbf{A}_i = \overline{\mathbf{A}_{p+2-i}}$ for any $i \in [p] \backslash \{1\}$.

(ii) In addition, when $m = n = 1$, Lemma 2 reduces to a well-known result for circular matrices, and the condition to guarantee that $\mathbf{A}_i \in \mathbb{R}^{m \times n}$ for any $i \in [p]$ is that $\mathscr{A}$ is symmetric. However, for the case $m = n \neq 1$, it should be noted that most of the matrices $\mathbf{A}_i$ ($i \in [p]$) may be complex even when $\mathscr{A}$ is symmetric, since the condition $\mathscr{A}$ is symmetric can only imply that $(\mathscr{A}^{(i)})^\top = \mathscr{A}^{(p+2-i)}$ for any $i \in [p] \backslash \{1\}$, but not $\mathscr{A}^{(i)} = \mathscr{A}^{(p+2-i)}$ for any $i \in [p] \backslash \{1\}$.

3 T-Hessian tensor and convexity of the twice continuously T-differentiable multi-variable real-valued function

As is well-known to us, the local curvature of a multi-variable real-valued function can be characterized by the positive semidefiniteness of its Hessian matrix, which is widely used in Newton-type methods for solving various optimization problems. In this section, we generalize the Hessian matrix to the third-order tensor: T-Hessian tensor, and study the discriminant condition of the convexity of the twice continuously T-differentiable multi-variable real-valued function.

3.1 T-Derivatives of multi-variable real-valued functions

In this subsection, we explore the derivatives of the multi-variable real-valued function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$. The tensor space is a normed linear space with inner product. In the following, we regard a matrix as a third-order tensor and derive the multi-variable real-valued function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ with the help of the inner product and tensor T-product. We adopt the Fréchet derivative [46]: Let V and W be normed vector spaces, and $U \subseteq V$ be an open subset of V. A function $f : U \to W$ is called to be Fréchet differentiable at $\mathbf{x} \in U$ if there exists a bounded linear operator $A : V \to W$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A(\mathbf{h})\|_W}{\|\mathbf{h}\|_V} = 0.$$

Recently, in [33] and [3], the authors showed that third-order tensors can be seen as linear operators on a space of matrices with the help of the newly proposed tensor T-product and obtained many good theoretical and computational results. Inspired by that, we wonder whether or not can we adopt the third-order tensor as the linear operator in the above definition of Fréchet derivative when the variety is a matrix? To this end, we introduce the following definition first.

Definition 4 Let $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ be a continuous map. Then, we say f is T-differentiable at $\mathscr{X} \in U$ if and only if there exists a third-order tensor $\nabla_{\mathscr{A}} f(\mathscr{X})$ such that

$$\lim_{\mathscr{H} \to \mathscr{O}} \frac{\|f(\mathscr{X} + \mathscr{H}) - f(\mathscr{X}) - \langle \nabla_{\mathscr{A}} f(\mathscr{X}), \mathscr{H} \rangle \|}{\|\mathscr{H}\|} = 0,$$

and we call $\nabla_{\mathscr{A}} f(\mathscr{X})$ the first-order T-derivative of f at $\mathscr{X}$; and we say f is twice T-differentiable at $\mathscr{X} \in U$ if and only if f is continuously T-differentiable and there exists a third-order tensor $\nabla^2_{\mathscr{A}} f(\mathscr{X})$ such that

$$\lim_{\mathscr{H} \to \mathscr{O}} \frac{\|\nabla_{\mathscr{A}} f(\mathscr{X} + \mathscr{H}) - \nabla_{\mathscr{A}} f(\mathscr{X}) - \nabla^2_{\mathscr{A}} f(\mathscr{X}) * \mathscr{H}\|}{\|\mathscr{H}\|} = 0,$$

and we call $\nabla^2_{\mathscr{A}} f(\mathscr{X})$ the second-order T-derivative of f at $\mathscr{X}$.

Furthermore, we say f is T-differentiable (twice T-differentiable) on U if and only if f is T-differentiable (twice T-differentiable) at every $\mathscr{X} \in U$.

Remark 2 From the fact that the tensor T-product of two tensors of size $m \times n \times p$ reduces to the matrix multiplication when $p = 1$, it follows that $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ being T-differentiable (twice T-differentiable) on U is equivalent to f being differentiable (twice differentiable) on U when $p = 1$.

Theorem 1 *Let f be a continuous map from $U \subseteq \mathbb{R}^{n \times 1 \times p}$ to $\mathbb{R}$. Then*

(i) *f is T-differentiable on U if and only if $\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]}$ exists for every $\mathscr{X} \in U$, where $\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]}$ is a vector in $\mathbb{R}^{np}$ with $(\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]})_i = \frac{\partial f(\mathscr{X})}{\partial ([unfold(\mathscr{X})]_i)}$ for any $i \in [np]$. Especially, for any $\mathscr{X} \in U$, $\nabla_{\mathscr{A}} f(\mathscr{X}) = fold[\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]}]$;*

(ii) *f is twice T-differentiable on U if and only if f is continuously T-differentiable on U and $\frac{\partial [unfold(\nabla_{\mathscr{A}} f(\mathscr{X}))]}{\partial [unfold(\mathscr{X})]}$ is a block circular matrix with each block being of size $n \times n$ for every $\mathscr{X} \in U$. In particular, $\nabla_{\mathscr{A}}^2 f((\mathscr{X})) = bcirc^{-1}[\frac{\partial [unfold(\nabla_{\mathscr{A}} f(\mathscr{X}))]}{\partial [unfold(\mathscr{X})]}]$ for any $\mathscr{X} \in U$.*

Proof (i) ($\Rightarrow$): If f is T-differentiable on U, then for any $\mathscr{X} \in U$, there exists a bounded operator $\mathscr{L}$ such that

$$\lim_{\mathscr{H} \to \mathscr{O}} \frac{\|f(\mathscr{X} + \mathscr{H}) - f(\mathscr{X}) - \langle \mathscr{L}, \mathscr{H} \rangle\|}{\|\mathscr{H}\|} = 0,$$

which, together with $\langle \mathscr{L}, \mathscr{H} \rangle = \langle unfold(\mathscr{L}), unfold(\mathscr{H}) \rangle$, implies that

$$\lim_{unfold(\mathscr{H}) \to \mathbf{O}} \frac{\|f(\mathscr{X} + \mathscr{H}) - f(\mathscr{X}) - \langle unfold(\mathscr{L}), unfold(\mathscr{H}) \rangle\|}{\|unfold(\mathscr{H})\|} = 0.$$

Furthermore, by introducing $g(unfold(\mathscr{X})) := f(\mathscr{X})$, we have that

$$\lim_{\mathscr{H} \to \mathscr{O}} \frac{\|g[unfold(\mathscr{X} + \mathscr{H})] - g(unfold(\mathscr{X})) - \langle unfold(\mathscr{L}), unfold(\mathscr{H}) \rangle\|}{\|unfold(\mathscr{H})\|} = 0,$$

which means that $\frac{\partial [g(unfold(\mathscr{X}))]}{\partial [unfold(\mathscr{X})]} = unfold(\mathscr{L})$, i.e., $\frac{\partial [f(\mathscr{X})]}{\partial [unfold(\mathscr{X})]} = unfold(\mathscr{L})$. Thus, $\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]}$ exists, and $\nabla_{\mathscr{A}} f(\mathscr{X}) = \mathscr{L} = fold[\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]}]$.

($\Leftarrow$): Reversing the above process, we can obtain that if $\frac{\partial f(\mathscr{X})}{\partial [unfold(\mathscr{X})]}$ exists, then f is T-differentiable on U.

(ii) ($\Rightarrow$): If f twice T-differentiable on U, then f is continuously T-differentiable and for any $\mathscr{X} \in U$, there exists a bounded operator $\mathscr{L} : \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}^{n \times 1 \times p}$ such that

$$\lim_{\mathscr{H} \to \mathscr{O}} \frac{\|f(\mathscr{X} + \mathscr{H}) - f(\mathscr{X}) - \langle \mathscr{L}, \mathscr{H} \rangle\|}{\|\mathscr{H}\|} = 0,$$

which, together with $\|\mathscr{X}\| = \|unfold(\mathscr{X})\|$ for any $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p}$, implies that

$$\lim_{\mathcal{H} \to \mathcal{O}} \frac{\|unfold[\nabla_{\mathcal{A}} f(\mathcal{X} + \mathcal{H})] - unfold[\nabla_{\mathcal{A}} f(\mathcal{X})] - bcirc(\mathcal{L})unfold(\mathcal{H})\|}{\|unfold(\mathcal{H})\|} = 0.$$

Denote $h : \mathbb{R}^{np} \to \mathbb{R}^{np}$ with $h[unfold(\mathcal{X})] = unfold(\nabla_{\mathcal{A}} f(\mathcal{X}))$, then we have that $\frac{\partial[unfold(\nabla_{\mathcal{A}} f(\mathcal{X}))]}{\partial[unfold(\mathcal{X})]} = \frac{\partial[h[unfold(\mathcal{X})]]}{\partial[unfold(\mathcal{X})]} = bcirc(\mathcal{L})$, which is a block circular matrix with each block being of size $n \times n$. Thus $\frac{\partial[unfold(\nabla_{\mathcal{A}} f(\mathcal{X}))]}{\partial[unfold(\mathcal{X})]}$ exists, and

$$\nabla_{\mathcal{A}}^2 f(\mathcal{X}) = \mathcal{L} = bcirc^{-1}[\frac{\partial[unfold(\nabla_{\mathcal{A}} f(\mathcal{X}))]}{\partial[unfold(\mathcal{X})]}].$$

($\Leftarrow$): Reversing the above process, we can obtain the desired result.

In Theorem 1, we establish the necessary and sufficient conditions for a general multi-variable real-valued function being T-differentiable and twice T-differentiable, respectively, and we give the specific expressions when the derivatives exist. Then what are the relationships between the derivatives obtained in this way and the derivatives in the traditional sense? We construct an example to illustrate that.

Example 3.1 Given a map $f : \mathbb{R}^{2 \times 1 \times 2} \to \mathbb{R}$ with for any $\mathcal{X} = [x_{i1j}] \in \mathbb{R}^{2 \times 1 \times 2}$,

$$f(\mathcal{X}) = x_{111}^2 + 2x_{111}x_{112} + x_{112}^2 + x_{211}^2 + x_{212}^2.$$

(1) First, we discuss the relationship between ∇f and $\nabla_{\mathcal{A}} f$. By the traditional way, we can obtain that

$$\nabla f(\mathcal{X}) = \begin{bmatrix} \frac{\partial f}{\partial x_{1j1}} & \frac{\partial f}{\partial x_{1j2}} \\ \frac{\partial f}{\partial x_{211}} & \frac{\partial f}{\partial x_{212}} \end{bmatrix} = \begin{bmatrix} 2x_{111} + 2x_{112} & 2x_{111} + 2x_{112} \\ 2x_{211} & 2x_{212} \end{bmatrix}.$$

Now, let us to seek the $\nabla_{\mathcal{A}} f$ by the procedure given in Theorem 1. Noting that $unfold(\mathcal{X}) = [x_{111}, x_{211}, x_{112}, x_{212}]^{\top}$, thus we can obtain that

$$\frac{\partial[f(\mathcal{X})]}{\partial[unfold(\mathcal{X})]} = \left[\frac{\partial f}{\partial x_{111}}, \frac{\partial f}{\partial x_{211}}, \frac{\partial f}{\partial x_{112}}, \frac{\partial f}{\partial x_{212}} \right]^{\top} = 2[x_{111} + x_{112}, x_{211}, x_{111} + x_{112}, x_{212}]^{\top},$$

furthermore, we can get

$$\nabla_{\mathcal{A}} f(\mathcal{X}) = fold[\frac{\partial f(\mathcal{X})}{\partial[unfold(\mathcal{X})]}] = \begin{bmatrix} 2x_{111} + 2x_{112} & 2x_{111} + 2x_{112} \\ 2x_{211} & 2x_{212} \end{bmatrix} = \nabla f(\mathcal{X}).$$

(2) Next, we investigate the relationship between $\nabla^2 f$ and $\nabla_{\mathcal{A}}^2 f$.

Since $\nabla f(\mathcal{X})$ is actually a matrix of size 2×2 and $\mathcal{X}$ is also a matrix of size 2×2, thus by the traditional derivative, $\nabla^2 f(\mathcal{X})$ should be a fourth-order tensor defined by:

$$\nabla^2 f(\mathcal{X}) = \frac{\partial \nabla f(\mathcal{X})}{\partial \mathcal{X}} = \frac{\partial \nabla f(\mathcal{X})_{ij}}{\partial \mathcal{X}_{pq}} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_p \otimes \mathbf{e}_q$$

where $\mathbf{e}_i$ represents the vector whose i-th elements equals one and others equal zero. That is to say, $\nabla^2 f(\mathscr{X}) \in \mathbb{R}^{2 \times 2 \times 2 \times 2}$ with $(\nabla^2 f(\mathscr{X}))_{1111} = 2$, $(\nabla^2 f(\mathscr{X}))_{1112} = 2$, $(\nabla^2 f(\mathscr{X}))_{1211} = 2$, $(\nabla^2 f(\mathscr{X}))_{1212} = 2$, $(\nabla^2 f(\mathscr{X}))_{2121} = 2$, $(\nabla^2 f(\mathscr{X}))_{2222} = 2$, and other entries being zero.

While, by Theorem 1, we can get that

$$bcirc(\nabla^2_{\mathscr{f}} f(\mathscr{X})) = \frac{\partial[unfold(\nabla_{\mathscr{f}} f(\mathscr{X}))]}{\partial[unfold(\mathscr{X})]} = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

which is a block circular matrix and $\nabla^2_{\mathscr{f}} f(\mathscr{X})$ is a tensor of size $2 \times 2 \times 2$, with the frontal slices being:

$$(\nabla^2_{\mathscr{f}} f(\mathscr{X}))^{(1)} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \text{ and } (\nabla^2_{\mathscr{f}} f(\mathscr{X}))^{(2)} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence, $\nabla^2 f$ and $\nabla^2_{\mathscr{f}} f$ are not coincide in the sizes, but it should be noticed that the entries of $\nabla^2 f$ and these of $bcirc(\nabla^2_{\mathscr{f}})f$ are the same regardless of the orders.

Remark 3 (i) $\nabla^2_{\mathscr{f}} f(\mathscr{X})$ is different from the traditional one due to the participation of the T-product. Traditionally, the second-order derivative of a multi-variable real-valued function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ is a fourth-order tensor. However, $\nabla^2_{\mathscr{f}} f(\mathscr{X})$ turns out to be a third-order tensor, which is also reasonable. Its rationality lies in that the existence of $\nabla^2_{\mathscr{f}} f(\mathscr{X})$ via the T-product needs such a condition that $\frac{\partial[unfold(\nabla f(\mathscr{X}))]}{\partial[unfold(\mathscr{X})]}$ is a block circular matrix, which implies that just getting the information of its first block column vector is enough. In other words, it is enough to express the information of $\nabla^2_{\mathscr{f}} f(\mathscr{X})$ by a third-order tensor in such case.

(ii) The $\nabla_{\mathscr{f}} f(\mathscr{X})$ is consistent with the traditional one. This is natural because $\nabla_{\mathscr{f}} f(\mathscr{X})$ and $\nabla f(\mathscr{X})$ are based on the coincide definition of inner product. So, we use $\nabla f(\mathscr{X})$ to represent $\nabla_{\mathscr{f}} f(\mathscr{X})$ in the rest of paper.

3.2 The semidefiniteness of $\nabla^2_{\mathscr{f}} f(\mathscr{X})$ and the convexity of $f(\mathscr{X})$

In this subsection, we investigate the second-order condition for any twice continuously T-differentiable function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ being (strictly) convex. The definition of the convex function is as follows in [49]:

Definition 5 [49] A function $f : U \subseteq \mathbb{R}^{n \times p} \to \mathbb{R}$ is convex (strictly convex) if U is a convex set, and for all $\mathbf{X}, \mathbf{Y} \in U$ ($\mathbf{X}, \mathbf{Y} \in U$ and $\mathbf{X} \neq \mathbf{Y}$) and any θ with $0 \leq \theta \leq 1$, $f(\theta \mathbf{X} + (1 - \theta)\mathbf{Y}) \leq (<)\theta f(\mathbf{X}) + (1 - \theta)f(\mathbf{Y})$.

Since the first-order T-derivative is consistent with the traditional one for a multi-variable real-valued function, it is not difficult to find that for any continuously

T-differentiable function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$, it is convex (strictly convex) if and only if for any $\mathcal{X}, \mathcal{Y} \in U$ ($\mathcal{X}, \mathcal{Y} \in U$ and $\mathcal{X} \neq \mathcal{Y}$),

$$f(\mathcal{Y}) \geq (>) f(\mathcal{X}) + \langle \nabla f(\mathcal{X}), \mathcal{Y} - \mathcal{X} \rangle.$$

Below, in order to establish the second-order condition for any twice continuously T-differentiable function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ being (strictly) convex, we first extend the second-order Taylor expansion of the function of one variable to the twice continuously T-differentiable multi-variable real-valued function via the tensor T-product.

Theorem 2 *Suppose that $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ is twice continuously T-differentiable on U. Then*

(i) *there exists $\beta \in (0, 1)$ such that*

$$f(\mathcal{X}) = f(\widetilde{\mathcal{X}}) + \langle \nabla f(\widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle + \frac{1}{2} \langle \nabla^2 f(\mathcal{Z}) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle,$$

  where $\mathcal{Z} = \beta \mathcal{X} + (1 - \beta) \widetilde{\mathcal{X}}$;

(ii) *the second-order T-Taylor expansion of f at $\widetilde{\mathcal{X}}$ as follows*:

$$\begin{aligned} f(\mathcal{X}) = f(\widetilde{\mathcal{X}}) &+ \langle \nabla f(\widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle \\ &+ \frac{1}{2} \langle \nabla^2 f(\widetilde{\mathcal{X}}) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle + o(\|\mathcal{X} - \widetilde{\mathcal{X}}\|^2) \end{aligned}$$

  and $o(\|\mathcal{X} - \widetilde{\mathcal{X}}\|^2)$ means a high-order infinitesimal of $\|\mathcal{X} - \widetilde{\mathcal{X}}\|^2$ as $\mathcal{X} \to \widetilde{\mathcal{X}}$.

Proof (i) Construct a function of one variable as: $\varphi(t) := f(\widetilde{\mathcal{X}} + t(\mathcal{X} - \widetilde{\mathcal{X}}))$ for any $t \in \mathbb{R}$. Then, $\varphi(0) = f(\widetilde{\mathcal{X}})$ and $\varphi(1) = f(\mathcal{X})$. It follows from the condition that f is twice continuously T-differentiable that φ is twice continuously differentiable. For any $t \in \mathbb{R}$, let $\mathcal{U} = \widetilde{\mathcal{X}} + t(\mathcal{X} - \widetilde{\mathcal{X}})$, then we have

$$\varphi'(t) = \frac{d\varphi(t)}{dt} = \frac{df(\mathcal{U})}{dt} = \left\langle \nabla f(\mathcal{U}), \frac{d\mathcal{U}}{dt} \right\rangle = \langle \nabla f(\mathcal{U}), (\mathcal{X} - \widetilde{\mathcal{X}}) \rangle$$

$$= \langle \nabla f(\widetilde{\mathcal{X}} + t(\mathcal{X} - \widetilde{\mathcal{X}})), (\mathcal{X} - \widetilde{\mathcal{X}}) \rangle;$$

$$\varphi''(t) = \frac{d\varphi'(t)}{dt} = \frac{d\langle \nabla f(\mathcal{U}), (\mathcal{X} - \widetilde{\mathcal{X}}) \rangle}{dt} = \left\langle \frac{d(\nabla f(\mathcal{U}))}{dt}, (\mathcal{X} - \widetilde{\mathcal{X}}) \right\rangle$$

$$= \left\langle \frac{\nabla_{\mathscr{J}}^2 f(\mathcal{U}) * d\mathcal{U}}{dt}, \mathcal{X} - \widetilde{\mathcal{X}} \right\rangle$$

$$= \left\langle \frac{bcirc(\nabla_{\mathscr{J}}^2 f(\mathcal{U})) \cdot unfold(d\mathcal{U})}{dt}, unfold(\mathcal{X} - \widetilde{\mathcal{X}}) \right\rangle$$

$$= \langle bcirc(\nabla_{\mathscr{J}}^2 f(\mathcal{U})) \cdot unfold(\mathcal{X} - \widetilde{\mathcal{X}}), unfold(\mathcal{X} - \widetilde{\mathcal{X}}) \rangle$$

$$= \langle unfold(\nabla_{\mathscr{J}}^2 f(\mathcal{U}) * (\mathcal{X} - \widetilde{\mathcal{X}})), unfold(\mathcal{X} - \widetilde{\mathcal{X}}) \rangle$$

$$= \langle \nabla_{\mathscr{J}}^2 f(\widetilde{\mathcal{X}} + t(\mathcal{X} - \widetilde{\mathcal{X}})) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle.$$

Thus, $\varphi'(0) = \langle \nabla f(\widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle$ and $\varphi''(0) = \langle \nabla_{\mathscr{J}}^2 f(\widetilde{\mathcal{X}}) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle$. It follows from the mean value theorem, that there exists some $\beta \in (0, 1)$ such that $\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\beta)$, which implies that

$$f(\mathcal{X}) = f(\widetilde{\mathcal{X}}) + \langle \nabla f(\widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle + \frac{1}{2}\langle \nabla_{\mathscr{J}}^2 f(\mathcal{Z}) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle,$$

where $\mathcal{Z} = \beta\mathcal{X} + (1 - \beta)\widetilde{\mathcal{X}}$, i.e., the result in (i) holds.

(ii) Denote $\mathcal{Y} := \frac{\mathcal{X} - \widetilde{\mathcal{X}}}{\|\mathcal{X} - \widetilde{\mathcal{X}}\|}$ and $\gamma' := \|\mathcal{X} - \widetilde{\mathcal{X}}\|$. Let $\psi(\gamma) := f(\widetilde{\mathcal{X}} + \gamma\mathcal{Y})$ for any $\gamma \in \mathbb{R}$, then by the same way as (i), we have $\psi(0) = f(\widetilde{\mathcal{X}})$, $\psi(\gamma') = f(\mathcal{X})$,

$$\psi'(0)\gamma' = \langle \nabla f(\widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle \quad \text{and} \quad \psi''(0)\gamma'^2 = \langle \nabla_{\mathscr{J}}^2 f(\widetilde{\mathcal{X}}) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle,$$

which, together with $\psi(\gamma') = \psi(0) + \psi'(0)\gamma' + \frac{1}{2}\psi''(0)\gamma'^2 + o(\|\gamma'\|^2)$, imply that

$$\begin{aligned} f(\mathcal{X}) &= f(\widetilde{\mathcal{X}}) + \langle \nabla f(\widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle \\ &+ \frac{1}{2}\langle \nabla_{\mathscr{J}}^2 f(\widetilde{\mathcal{X}}) * (\mathcal{X} - \widetilde{\mathcal{X}}), \mathcal{X} - \widetilde{\mathcal{X}} \rangle + o(\|\mathcal{X} - \widetilde{\mathcal{X}}\|^2), \end{aligned}$$

i.e., the result in (ii) holds. $\qquad\square$

Theorem 3 *Let $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ be a twice continuously T-differentiable function on an open convex set U. Then*

(i) *f is convex if and only if for any $\mathcal{X} \in U$, $\nabla_{\mathscr{J}}^2 f(\mathcal{X})$ satisfies*

$$\langle \mathcal{Y}, \nabla_{\mathscr{J}}^2 f(\mathcal{X}) * \mathcal{Y} \rangle \geq 0 \quad \text{for any } \mathcal{Y} \in \mathbb{R}^{n \times 1 \times p}; \tag{4}$$

(ii) *f is strictly convex if for any $\mathcal{X} \in U$, $\nabla_{\mathscr{J}}^2 f(\mathcal{X})$ satisfies*

$$\langle \mathscr{Y}, \nabla_{\mathscr{J}}^2 f(\mathscr{X}) * \mathscr{Y} \rangle > 0 \quad \text{for any } \mathscr{Y} \in \mathbb{R}^{n \times 1 \times p} \setminus \{\mathscr{O}\}.$$

Proof (i) ($\Rightarrow$): For any $\mathscr{X} \in U$ and $\mathscr{Y} \in \mathbb{R}^{n \times 1 \times p} \setminus \{\mathscr{O}\}$, it follows from U being an open set that there exists $\varepsilon > 0$ such that $\mathscr{X} + \theta \mathscr{Y} \in U$ when $\theta \in (-\varepsilon, \varepsilon)$. Since f is convex, we have that $f(\mathscr{X} + \theta \mathscr{Y}) \geq f(\mathscr{X}) + \theta \langle \nabla f(\mathscr{X}), \mathscr{Y} \rangle$. In addition, it follows from Theorem 2 that

$$f(\mathscr{X} + \theta \mathscr{Y}) = f(\mathscr{X}) + \theta \langle \nabla f(\mathscr{X}), \mathscr{Y} \rangle + \frac{1}{2} \theta^2 \langle \mathscr{Y}, \nabla_{\mathscr{J}}^2 f(\mathscr{X}) * \mathscr{Y} \rangle + o(\|\theta \mathscr{Y}\|^2).$$

Therefore, we can obtain that $\langle \mathscr{Y}, \nabla_{\mathscr{J}}^2 f(\mathscr{X}) * \mathscr{Y} \rangle + \frac{o(\|\theta \mathscr{Y}\|^2)}{\theta^2 / 2} \geq 0$. Let $\theta \to 0$, we further obtain that $\langle \mathscr{Y}, \nabla_{\mathscr{J}}^2 f(\mathscr{X}) * \mathscr{Y} \rangle \geq 0$.

($\Leftarrow$): For any $\mathscr{X}, \mathscr{Y} \in U$, it follows from Theorem 2 that

$$f(\mathscr{Y}) = f(\mathscr{X}) + \langle \nabla f(\mathscr{X}), \mathscr{Y} - \mathscr{X} \rangle + \frac{1}{2} \langle \nabla_{\mathscr{J}}^2 f(\mathscr{Z}) * (\mathscr{Y} - \mathscr{X}), \mathscr{Y} - \mathscr{X} \rangle,$$

where $\mathscr{Z} = \mathscr{X} + t(\mathscr{Y} - \mathscr{X})$ with $t \in (0, 1)$. Since U is convex, it follows that $\mathscr{Z} \in U$; and hence, by (4) we have that $\frac{1}{2} \langle \mathscr{Y}, \nabla_{\mathscr{J}}^2 f(\mathscr{Z}) * \mathscr{Y} \rangle \geq 0$. Furthermore, we have that $f(\mathscr{Y}) \geq f(\mathscr{X}) + \langle \nabla f(\mathscr{X}), \mathscr{Y} - \mathscr{X} \rangle$, which, together with U being convex, implies that f is convex. The proof of (i) is complete.

By the same way as in the proof of the sufficiency of (i), we can obtain (ii). □

Remark 4 Let $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \to \mathbb{R}$ be a twice continuously T-differentiable function on an open convex set U. Since $\nabla_{\mathscr{J}}^2 f \in \mathbb{R}^{n \times n \times p}$ has similar properties as Hessian matrix, we call $\nabla_{\mathscr{J}}^2 f$ the T-Hessian tensor.

4 Symmetric T-positive (semi)definite tensors

In this section, we first introduce a definition of the symmetric T-positive (semi) definite tensor; and then, we investigate properties of symmetric T-positive (semi) definite tensors.

4.1 Definition

In Sect. 3, we obtained that the convexity of a twice continuously T-differentiable multi-variable real-valued function on an open convex set can be characterized by some property of the T-Hessian tensor. Now we name such a property as the symmetric T-positive semidefiniteness.

Definition 6 Let $\mathscr{A} \in \mathbb{R}^{n \times n \times p}$. We say $\mathscr{A}$ is a symmetric T-positive (semi)definite tensor (T-P(S)D tensor for short), if and only if $\mathscr{A}$ is a symmetric tensor and

$$\langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle > (\geq) 0$$

holds for any $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p} \backslash \{\mathscr{O}\}$ (for any $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p}$). We denote the set consisting of all symmetric T-P(S)D tensors of size $n \times n \times p$ as $S\mathbb{R}_{++}^{n \times n \times p}$ ($S\mathbb{R}_{+}^{n \times n \times p}$).

Remark 5 (i) When $p = 1$, the T-product defined by Definition 1 reduces to the product of two matrices. In addition, when $p = 1$, $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p}$ reduces to a column vector and $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$ reduces to a square symmetric matrix. Thus, when $p = 1$, Definition 6 is exactly the definition of the symmetric P(S)D matrix. That is to say, the T-P(S)D tensor defined by Definition 6 is a higher-order extension of the P(S)D matrix.

(ii) From Definition 6 and Theorem 3, we can say that a twice continuously T-differentiable function $f : U \subseteq \mathbb{R}^{n \times 1 \times p} \rightarrow \mathbb{R}$ is convex (strictly convex) if and only if (if) the T-Hessian tensor $\nabla_{\mathscr{f}}^2 f(\mathscr{X})$ is symmetric T-PSD (T-PD) for any $\mathscr{X} \in U$.

(iii) It should be noticed that the positive semidefinite tensor defined by means of the nonnegativity of the corresponding multi-variate homogeneous polynomial in [47] is different with the one defined in Definition 6. Since the positive semidefinite tensor defined by Qi [47] vanishes when the order is odd, while the symmetric T-positive (semi)definite tensor in Definition 6 is introduced for third-order tensor.

4.2 Equivalent characterizations of symmetric T-P(S)D tensors

First, we give an equivalent description of Definition 6.

Theorem 4 *Let $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$. $\mathscr{A}$ is symmetric T-P(S)D if and only if $bcirc(\mathscr{A})$ is symmetric P(S)D.*

Proof Since $bcirc(\mathscr{A}^{\top}) = bcirc(\mathscr{A}^H) = bcirc(\mathscr{A})^H = bcirc(\mathscr{A})^{\top}$ by Lemma 1(c), then $\mathscr{A}$ is symmetric if and only if $bcirc(\mathscr{A})$ is symmetric. For any $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p}$, it follows from Definition 1 and the definition of operator *unfold* that

$$unfold(\mathscr{A} * \mathscr{X}) = bcirc(\mathscr{A})unfold(\mathscr{X}),$$

and hence,

$$\begin{aligned} \langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle &= \langle unfold(\mathscr{X}), unfold(\mathscr{A} * \mathscr{X}) \rangle \\ &= \langle unfold(\mathscr{X}), bcirc(\mathscr{A})unfold(\mathscr{X}) \rangle. \end{aligned}$$

Thus, by combining Definition 6 and the criterion of P(S)D matrix, we can easily obtain that $\mathscr{A}$ is symmetric T-P(S)D if and only if $bcirc(\mathscr{A})$ is symmetric P(S)D. □

Remark 6 From Theorem 4, we can see that lots of results that P(S)D matrix with block circular structure hold is true for T-P(S)D tensors by combining the properties of tensor T-product such as those shown in Lemma 1. Thus for convenience, in Sects. 4.3 and 4.4 we just list some ones which play important roles in Sects. 4.5 and 5 without proofs.

Next, we give another equivalent description of Definition 6.

Theorem 5 *Suppose that $\mathscr{A} \in \mathbb{R}^{n \times n \times p}$ can be block diagonalized as*

$$bcirc(\mathscr{A}) = (\mathbf{F}_p^H \otimes \mathbf{I}_{n \times n})Diag(\mathbf{A}_i : i \in [p])(\mathbf{F}_p \otimes \mathbf{I}_{n \times n}), \tag{5}$$

where $\mathbf{F}_p$ is the Fourier matrix of size $p \times p$, which is defined as (1). *Then $\mathscr{A}$ is symmetric T-P(S)D if and only if all the matrices $\mathbf{A}_i$ are Hermitian P(S)D.*

Proof By (5), $bcirc(\mathscr{A})$ is symmetric if and only if each $\mathbf{A}_i$ is Hermitian, and $\mathscr{A}$ is symmetric if and only if $bcirc(\mathscr{A})$ is symmetric as shown in Theorem 4.

($\Leftarrow$): Suppose that all the matrices $\mathbf{A}_i$ in (5) are Hermitian P(S)D, then for any $\mathbf{x}$ in $\mathbb{C}^n \backslash \{\mathbf{0}\}$ ($\mathbf{x} \in \mathbb{C}^n$) and $i \in [p]$, $\mathbf{x}^H \mathbf{A}_i \mathbf{x} > (\geq)0$. For any $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p} \backslash \{\mathscr{O}\}$, there exists $\mathbf{x}_i \in \mathbb{C}^n$ for each $i \in [p]$, which cannot be $\mathbf{0}$ at the same time, such that $bcirc(\mathscr{X}) = (\mathbf{F}_p^H \otimes \mathbf{I}_{n \times n})Diag(\mathbf{x}_i : i \in [p])\mathbf{F}_p$. Since

$$\begin{aligned}
\langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle &= \tfrac{1}{p}\langle bcirc(\mathscr{X}), bcirc(\mathscr{A})bcirc(\mathscr{X})\rangle \\
&= \tfrac{1}{p}Tr(bcirc(\mathscr{X})^H bcirc(\mathscr{A})bcirc(\mathscr{X})) \\
&= \tfrac{1}{p}Tr(\mathbf{F}_p^H Diag(\mathbf{x}_i^H \mathbf{A}_i \mathbf{x}_i : i \in [p])\mathbf{F}_p) \\
&= \tfrac{1}{p}Tr(Diag(\mathbf{x}_i^H \mathbf{A}_i \mathbf{x}_i : i \in [p])) = \tfrac{1}{p}\sum_{i=1}^{p}(\mathbf{x}_i^H \mathbf{A}_i \mathbf{x}_i) \geq 0,
\end{aligned}$$

where the fourth equality follows from the fact that similar matrices have the same traces, then $\mathscr{A}$ is symmetric T-PSD if all the matrices $\mathbf{A}_i$ are Hermitian PSD.

Since $\mathbf{x}_i \in \mathbb{C}^n$ for each $i \in [p]$ cannot be $\mathbf{0}$ at the same time, there exists at least an index $i \in [p]$ such that $\mathbf{x}_i^H \mathbf{A}_i \mathbf{x}_i > 0$ if all the matrices $\mathbf{A}_i$ are Hermitian PD. Hence $\mathscr{A}$ is symmetric T-PD if all the matrices $\mathbf{A}_i$ are Hermitian PD.

($\Rightarrow$): Suppose that $\mathscr{A} \in S\mathbb{R}_{++}^{n \times n \times p}$ ($S\mathbb{R}_+^{n \times n \times p}$), then $\langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle > (\geq)0$ for any $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p} \backslash \{\mathscr{O}\}$. Below, we divide the discussion into two cases.

Case 1: n is even. By Lemma , we have $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$, $\mathbf{A}_{\frac{p+2}{2}} \in \mathbb{R}^{n \times n}$, $\mathbf{A}_i \in \mathbb{C}^{n \times n}$ and $\mathbf{A}_i = \overline{\mathbf{A}_{p+2-i}}$ for any $i \in [p] \backslash \{1, \frac{p+2}{2}\}$. Then, for any $\mathbf{x} \in \mathbb{C}^n \backslash \{\mathbf{0}\}$, choose special $\mathscr{X}$ in $\mathbb{R}^{n \times 1 \times p}$ and satisfies that $bcirc(\mathscr{X}') = (\mathbf{F}_p^H \otimes \mathbf{I}_{n \times n})Diag(\mathbf{X}'_i : i \in [p])\mathbf{F}_p$, where $\mathbf{X}'_k = \overline{\mathbf{X}'_{p+2-k}} = \mathbf{x}$ and others $\mathbf{X}'_i = \mathbf{0}$ with k being any fixed number in $[p] \backslash \{1, \frac{p+2}{2}\}$. Then, from Remark 1 we have that $\mathscr{X} \in \mathbb{R}^{n \times 1 \times p}$. Thus, $\langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle > (\geq 0)$ by $\mathscr{A} \in S\mathbb{R}_{++}^{n \times n \times p}(\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p})$. Since $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$, $\mathbf{A}_k \in H\mathbb{C}^{n \times n}$ and $\mathbf{A}_k = \overline{\mathbf{A}_{p+2-k}}$. Thus, $\mathbf{x}^H \mathbf{A}_k \mathbf{x}$ is real, and

$$0 < (\leq)\langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle = \frac{1}{p}\sum_{i=1}^{p}((\mathbf{X}'_i)^H \mathbf{A}_i \mathbf{X}'_i) = \frac{1}{p}(\mathbf{x}^H \mathbf{A}_k \mathbf{x} + \overline{\mathbf{x}}^H \mathbf{A}_{p+2-k}\overline{\mathbf{x}}) = \frac{2}{p}\mathbf{x}^H \mathbf{A}_k \mathbf{x},$$

which implies $\mathbf{A}_k$ is Hermitian P(S)D for any $k \in [p] \backslash \{1, \frac{p+2}{2}\}$. In addition, for any $\mathbf{x} \in \mathbb{R}^n$, we can obtain that $\mathbf{A}_1$ ($\mathbf{A}_{\frac{p+2}{2}}$) is symmetric P(S)D by choosing $\mathbf{X}'_1 = \mathbf{x}$ ($\mathbf{X}'_{\frac{p+2}{2}} = \mathbf{x}$) and others $\mathbf{X}'_i = \mathbf{0}$.

Case 2: n is odd. By Lemma , we have $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$, $\mathbf{A}_i \in \mathbb{C}^{n \times n}$ and $\mathbf{A}_i = \overline{\mathbf{A}_{p+2-i}}$ for any $i \in [p] \backslash \{1\}$. Then by the same method as Case 1, we can obtain that $\mathbf{A}_k$ is Hermitian P(S)D for any $k \in [p]$.

Thereby, combining **Case 1** and **Case 2**, we can obtain that all the matrices $\mathbf{A}_i$ are Hermitian P(S)D if $\mathscr{A}$ is symmetric T-P(S)D. ◻

Remark 7 Theorem 5 shows that the judgement of the T-positive semidefiniteness of a symmetric tensor of size $n \times n \times p$ can be transformed into the judgement of positive semidefiniteness of p Hermitian matrices of size $n \times n$. Furthermore, Theorem 5 shows that the symmetric T-P(S)D tensor in Definition 6 is equivalent to the one by [3, Definition 2.7] and the one by [34, Definition 15] in real case.

4.3 T-eigenvalue decomposition of the symmetric T-P(S)D tensor

In this subsection, we aim to establish the T-eigenvalue decomposition for the symmetric T-P(S)D tensor. To do this, we give the following definition first. It should be noticed that the definition of T-eigenvalue for third-order F-square tensor was given in [34] and here we redefine it in an equivalent way for convenience.

Definition 7 (T-eigenvalue and Trace) Let $\mathscr{A} \in \mathbb{R}^{n \times n \times p}$, which can be block diagonalized as (5). Then a real number λ is said to be a T-eigenvalue of $\mathscr{A}$ if and only if it is an eigenvalue of some $\mathbf{A}_i$ for $i \in [p]$, denoted by $\lambda(\mathscr{A})$. The largest and smallest T-eigenvalues of $\mathscr{A}$ are denoted by $\lambda_{\max}(\mathscr{A})$ and $\lambda_{\min}(\mathscr{A})$, respectively. Moreover, the trace of $\mathscr{A}$, denoted by $Tr(\mathscr{A})$, is defined as $Tr(\mathscr{A}) := \sum_{i=1}^{p} Tr(\mathbf{A}_i)$.

Remark 8 By Definition 7, Theorem 5 and [44, Fact 6], it is not difficult to see that a symmetric third-order tensor $\mathscr{A}$ is T-P(S)D if and only if each T-eigenvalue of $\mathscr{A}$ is positive (nonnegative).

It is easy to establish the following properties for the T-eigenvalues and traces of tensors from the above definition and some known results in [39].

Proposition 1 *Let $\mathscr{A}$ and $\mathscr{B}$ be two tensors in $\mathbb{R}^{n \times n \times p}$, $\mathscr{C} \in \mathbb{R}^{n \times n \times p}$ be nonsingular, and $spec(\mathscr{A})$ be the set consisting of all the T-eigenvalues of $\mathscr{A}$. Then*

(a) $spec(\mathscr{A}) = spec(bcirc(\mathscr{A}))$;
(b) $Tr(\mathscr{A}) = Tr(bcirc(\mathscr{A})) = \sum_i \lambda_i(\mathscr{A}) = p \sum_{i=1}^{n} (\mathbf{A}^{(1)})_{ii}$;
(c) $Tr(\mathscr{A} * \mathscr{B}) = Tr(\mathscr{B} * \mathscr{A})$;
(d) $spec(\mathscr{C}^{-1} * \mathscr{A} * \mathscr{C}) = spec(\mathscr{A})$ and $Tr(\mathscr{C}^{-1} * \mathscr{A} * \mathscr{C}) = Tr(\mathscr{A})$.

Remark 9 (i) From Lemma 1(*a*) and Proposition 1, it is easy to see that for any $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$ and $\mathscr{B} \in S\mathbb{R}^{n \times n \times p}$,

$$p\langle \mathscr{A}, \mathscr{B} \rangle = \langle bcirc(\mathscr{A}), bcirc(\mathscr{B}) \rangle = Tr(bcirc(\mathscr{A})bcirc(\mathscr{B})) = Tr(\mathscr{A} * \mathscr{B}),$$

and $\mathscr{A}$ is T-P(S)D if and only if $Tr(\mathscr{V}^T * \mathscr{A} * \mathscr{V}) > 0$ for all nonzero $\mathscr{V} \in \mathbb{R}^{n \times 1 \times p}$ ($Tr(\mathscr{V}^T * \mathscr{A} * \mathscr{V}) \geq 0$ for all $\mathscr{V} \in \mathbb{R}^{n \times 1 \times p}$).

(ii) Let $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}$ and $\mathscr{B} \in S\mathbb{R}_+^{n \times n \times p}$. Then, it follows from [43, Lemmas 1.2.3, 1.2.4] and Proposition 1 that $\langle \mathscr{A}, \mathscr{B} \rangle \geq 0$; $\langle \mathscr{A}, \mathscr{B} \rangle = 0$ iff $\mathscr{A} * \mathscr{B} = \mathscr{O}$; and

$$p\langle \mathscr{A}, \mathscr{B} \rangle \geq \lambda_{\min}(\mathscr{A})\lambda_{\max}(\mathscr{B}) \leq \lambda_{\min}(\mathscr{A})Tr(\mathscr{B});$$
$$p\langle \mathscr{A}, \mathscr{B} \rangle \leq \lambda_{\max}(\mathscr{A})Tr(\mathscr{B}) \leq n\lambda_{\max}(\mathscr{A})\lambda_{\max}(\mathscr{B}).$$

It is known that the eigenvalue decomposition plays an important role in the study of symmetric matrices. In the following, we establish a similar decomposition for symmetric third-order tensors, especially for the T-P(S)D tensor.

Theorem 6 (T-eigenvalue decomposition) *Every $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$ can be factored as*

$$\mathscr{A} = \mathscr{U}^\top * \mathscr{S} * \mathscr{U},$$

where $\mathscr{U} \in \mathbb{R}^{n \times n \times p}$ is an orthogonal tensor and $\mathscr{S} \in \mathbb{R}^{n \times n \times p}$ is an F-diagonal tensor(That is, each frontal slice of $\mathscr{S}$ is a diagonal matrix) with all of the diagonal entries of $(\mathbf{F}_p \otimes I_{n \times n})bcirc(\mathscr{S})(\mathbf{F}_p^H \otimes I_{n \times n})$ being the T-eigenvalues of $\mathscr{A}$. In particular, if $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}$ ($\mathscr{A} \in S\mathbb{R}_{++}^{n \times n \times p}$), then all of the diagonal entries of $(\mathbf{F}_p \otimes I_{n \times n})bcirc(\mathscr{S})(\mathbf{F}_p^H \otimes I_{n \times n})$ are nonnegative (positive).

Proof The conclusion can be easily proved from [39, Corollary4.4.4,Theorem 7.6.4] and Remark 1(i). □

4.4 The T-roots of a symmetric T-PSD tensor

The following result about the roots of a symmetric T-PSD tensor also holds by Theorem 4, Lemma 1 and [39, Theorem 7.2.6].

Theorem 7 (The T-roots of a symmetric T-PSD tensor) *Let $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}$ and $k \geq 1$. Then there exists a unique $\mathscr{B} \in S\mathbb{R}_+^{n \times n \times p}$ with $\mathscr{B}^k = \mathscr{A}$.*

Corollary 1 *Let $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$ be T-PSD. Then there exists a unique positive semidefinite tensor $\mathscr{B} \in S\mathbb{R}^{n \times n \times p}$ with $\mathscr{B}^2 = \mathscr{A}$. We write such $\mathscr{B}$ as $\mathscr{A}^{\frac{1}{2}}$.*

Furthermore, the following conclusion is true.

Theorem 8 *For any $\mathscr{A} \in S\mathbb{R}^{n \times n \times p}$ with $bcirc(\mathscr{A})$ being block diagonalized as (5), (a) $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}(\mathscr{A} \in S\mathbb{R}_{++}^{n \times n \times p})$ if and only if (b) $\mathscr{A} = \mathscr{P}^\top * \mathscr{P}$ for some tensor $\mathscr{P} \in \mathbb{R}^{m \times n \times p}(\mathscr{A} = \mathscr{P}^\top * \mathscr{P}$ for some nonsingular tensor $\mathscr{P} \in \mathbb{R}^{m \times n \times p})$.*

4.5 The cone of T-PSD tensors

In this subsection, we investigate the set of T-PSD tensors. Recall that a subset C of a vector space V is called a cone (or sometimes called a linear cone) if for each $\mathbf{x}$ in C and any nonnegative scalar α, the product $\alpha\mathbf{x}$ is in C; C is called a convex cone if

for any nonnegative scalars α, β and any $\mathbf{x}$, $\mathbf{y}$ in C, it follows that $\alpha\mathbf{x} + \beta\mathbf{y}$ belongs to C; and if additionally C is a closed set, then we call C a closed, convex cone.

Proposition 2 $S\mathbb{R}^{n \times n \times p}$ *is isomorphic to* $\mathbb{R}^{\frac{pn^2+n}{2}}$ *if p is odd; and* $S\mathbb{R}^{n \times n \times p}$ *is isomorphic to* $\mathbb{R}^{\frac{pn^2}{2}+n}$ *if p is even.*

Proof This proposition can be easily proved; and we omit the proof here. $\square$

Proposition 3 $S\mathbb{R}_+^{n \times n \times p}$ *is a nonempty, closed, convex, pointed cone.*

Proof From Theorem 4 and the fact that $S\mathbb{R}_+^{n \times n}$ is nonempty and closed, it follows that $S\mathbb{R}_+^{n \times n \times p}$ is nonempty and closed. For any $\mathscr{A}, \mathscr{B} \in S\mathbb{R}_+^{n \times n \times p}$ and any two non-negative scalars α and β, we have that $bcirc(\alpha\mathscr{A} + \beta\mathscr{B}) = \alpha bcirc(\mathscr{A}) + \beta bcirc(\mathscr{B})$. Suppose that $\mathscr{A}, \mathscr{B} \in S\mathbb{R}_+^{n \times n \times p}$, then $bcirc(\mathscr{A})$ and $bcirc(\mathscr{B})$ belong to $S\mathbb{R}_+^{np \times np}$ from Theorem 4. Therefore, from the fact that $S\mathbb{R}_+^{np \times np}$ is a convex cone, it follows that $bcirc(\alpha\mathscr{A} + \beta\mathscr{B}) = \alpha bcirc(\mathscr{A}) + \beta bcirc(\mathscr{B}) \in S\mathbb{R}_+^{np \times np}$, which together with Theorem 4 implies that $\alpha\mathscr{A} + \beta\mathscr{B} \in S\mathbb{R}_+^{n \times n \times p}$. That is to say, $S\mathbb{R}_+^{n \times n \times p}$ is a convex cone. Suppose that $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}$ and $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}$, then $bcirc(\mathscr{A}) \in S\mathbb{R}_+^{np \times np}$ and $bcirc(\mathscr{A}) = -bcirc(\mathscr{A}) \in S\mathbb{R}_+^{np \times np}$. From the fact that $S\mathbb{R}_+^{np \times np}$ is pointed, it follows that $bcirc(\mathscr{A}) = \mathbf{O}$. Thus $\mathscr{A} = \mathscr{O}$, which implies that $S\mathbb{R}_+^{n \times n \times p}$ is pointed. $\square$

Remark 10 By the proof of Proposition 3, it is easy to obtain that $S\mathbb{R}_{++}^{n \times n \times p}$ is a nonempty, open, convex cone. It is also easy to show that $S\mathbb{R}_{++}^{n \times n \times p}$ is the interior of $S\mathbb{R}_+^{n \times n \times p}$. By the theory of conic optimization, it follows that $S\mathbb{R}_+^{n \times n \times p}$ ($S\mathbb{R}_{++}^{n \times n \times p}$) can induce a partial order on $S\mathbb{R}^{n \times n \times p}$, denoted by $\succeq_{\mathscr{T}}(\succ_{\mathscr{T}})$. That is, for any $\mathscr{A}, \mathscr{B} \in S\mathbb{R}^{n \times n \times p}$, $\mathscr{A} \succeq_{\mathscr{T}} (\succ_{\mathscr{T}}) \mathscr{B}$ if and only if $\mathscr{A} - \mathscr{B} \in S\mathbb{R}_+^{n \times n \times p}(S\mathbb{R}_{++}^{n \times n \times p})$.

In the following, we will use $\mathscr{A} \succeq_{\mathscr{T}} (\succ_{\mathscr{T}}) \mathscr{O}$ if $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}(S\mathbb{R}_{++}^{n \times n \times p})$. Especially, we replace $\mathscr{A} \succeq_{\mathscr{T}} (\succ_{\mathscr{T}}) \mathscr{O}$ with $\mathbf{A} \succeq (\succ)\mathbf{O}$ if $\mathbf{A} \in S\mathbb{R}_+^{n \times n}(S\mathbb{R}_{++}^{n \times n})$, as any $\mathscr{A} \in S\mathbb{R}_+^{n \times n \times p}(S\mathbb{R}_{++}^{n \times n \times p})$ reduces to the $\mathbf{A} \in S\mathbb{R}_+^{n \times n}(S\mathbb{R}_{++}^{n \times n})$ when $p = 1$.

Recall that $S\mathbb{R}_+^{n \times n}$ is a self-dual cone, which plays an important role in the widely studied semidefinite programming. In the following, we generalize this fact to T-semidefinite cone $S\mathbb{R}_+^{n \times n \times p}$. For a cone C, the polar cone (or dual cone) [43] is the set $C^* := \{\mathbf{y} : \langle \mathbf{x}, \mathbf{y} \rangle \geq 0, \text{ for any } \mathbf{x} \in C\}$.

Theorem 9 (Self-duality) $S\mathbb{R}_+^{n \times n \times p} = (S\mathbb{R}_+^{n \times n \times p})^*$.

Proof (i) $S\mathbb{R}_+^{n \times n \times p} \subseteq (S\mathbb{R}_+^{n \times n \times p})^*$: To this end, we only need to show that $\mathscr{A} \cdot \mathscr{B} \geq 0$ for any $\mathscr{A}, \mathscr{B} \succeq_{\mathscr{T}} \mathscr{O}$. Since $\mathscr{A}, \mathscr{B} \succeq_{\mathscr{T}} \mathscr{O}$, it follows from Corollary 1 that there exists $\mathscr{A}^{\frac{1}{2}}$ and $\mathscr{B}^{\frac{1}{2}}$ such that $\mathscr{A} = \mathscr{A}^{\frac{1}{2}} * \mathscr{A}^{\frac{1}{2}}$ and $\mathscr{B} = \mathscr{B}^{\frac{1}{2}} * \mathscr{B}^{\frac{1}{2}}$. Thus, we can obtain that

$$
\begin{aligned}
\mathscr{A} \cdot \mathscr{B} &= \frac{1}{p} Tr(\mathscr{A} * \mathscr{B}) = \frac{1}{p} Tr(\mathscr{A}^{\frac{1}{2}} * \mathscr{A}^{\frac{1}{2}} * \mathscr{B}^{\frac{1}{2}} * \mathscr{B}^{\frac{1}{2}}) \\
&= \frac{1}{p} Tr(\mathscr{B}^{\frac{1}{2}} * \mathscr{A}^{\frac{1}{2}} * \mathscr{A}^{\frac{1}{2}} * \mathscr{B}^{\frac{1}{2}}) = (\mathscr{A}^{\frac{1}{2}} * \mathscr{B}^{\frac{1}{2}}) \cdot (\mathscr{A}^{\frac{1}{2}} * \mathscr{B}^{\frac{1}{2}}) \geq 0.
\end{aligned}
$$

(ii) $(S\mathbb{R}_+^{n\times n\times p})^* \subseteq S\mathbb{R}_+^{n\times n\times p}$: We only need to show $\mathscr{A} \succeq_{\mathscr{T}} \mathscr{O}$ if $\mathscr{A} \in (S\mathbb{R}_+^{n\times n\times p})^*$. Suppose $\mathscr{A} \in (S\mathbb{R}_+^{n\times n\times p})^*$, then $\mathscr{A} \cdot \mathscr{B} \geq 0$ for any $\mathscr{B} \succeq_{\mathscr{T}} \mathscr{O}$. Taken $\mathscr{B} = \mathscr{D} * \mathscr{D}^{\top}$ where $\mathscr{D} \in \mathbb{R}^{n\times 1\times p}$ is an arbitrary given tensor, then we have that $\mathscr{B} \succeq_{\mathscr{T}} \mathscr{O}$ and $\mathscr{A} \cdot (\mathscr{D} * \mathscr{D}^{\top}) \geq 0$, i.e., $Tr(\mathscr{A} * (\mathscr{D} * \mathscr{D}^{\top})) \geq 0$. By Proposition 1(b), we get

$$
\begin{aligned}
Tr(\mathscr{D}^{\top} * \mathscr{A} * \mathscr{D}) &= Tr(bcirc(\mathscr{D}^{\top})bcirc(\mathscr{A})bcirc(\mathscr{D})) \\
&= Tr(bcirc(\mathscr{A})(bcirc(\mathscr{D})bcirc(\mathscr{D}^{\top}))) = Tr(\mathscr{A} * (\mathscr{D} * \mathscr{D}^{\top})) \geq 0.
\end{aligned}
$$

Hence, by Remark 9(i), we can obtain that $\mathscr{A} \succeq_{\mathscr{T}} \mathscr{O}$. □

4.6 The T-schur complement of a symmetric T-PSD tensor

In this subsection, we give a characterization of the T-positive semidefiniteness of a third-order tensor by the T-positive semidefiniteness of the T-schur complement [34].

Lemma 3 (Tensor block multiplication via T-product) [37] *Suppose* $\mathscr{A}_1 \in \mathbb{C}^{n_1\times m_1\times p}$, $\mathscr{B}_1 \in \mathbb{C}^{n_1\times m_2\times p}$, $\mathscr{C}_1 \in \mathbb{C}^{n_2\times m_1\times p}$, $\mathscr{D}_1 \in \mathbb{C}^{n_2\times m_2\times p}$, $\mathscr{A}_2 \in \mathbb{C}^{m_1\times r_1\times p}$, $\mathscr{B}_2 \in \mathbb{C}^{m_1\times r_2\times p}$, $\mathscr{C}_2 \in \mathbb{C}^{m_2\times r_1\times p}$ and $\mathscr{D}_2 \in \mathbb{C}^{m_2\times r_2\times p}$ are complex tensors, then*

$$
\begin{bmatrix} \mathscr{A}_1 & \mathscr{B}_1 \\ \mathscr{C}_1 & \mathscr{D}_1 \end{bmatrix} * \begin{bmatrix} \mathscr{A}_2 & \mathscr{B}_2 \\ \mathscr{C}_2 & \mathscr{D}_2 \end{bmatrix} = \begin{bmatrix} \mathscr{A}_1 * \mathscr{A}_2 + \mathscr{B}_1 * \mathscr{C}_2 & \mathscr{A}_1 * \mathscr{B}_2 + \mathscr{B}_1 * \mathscr{D}_2 \\ \mathscr{C}_1 * \mathscr{A}_2 + \mathscr{D}_1 * \mathscr{C}_2 & \mathscr{C}_1 * \mathscr{B}_2 + \mathscr{D}_1 * \mathscr{D}_2 \end{bmatrix}.
$$

Lemma 4 *Suppose that* $\mathscr{A}_i \in S\mathbb{R}^{n_i\times n_i\times p}$ *for any* $i \in [m]$. *Then the block diagonal tensor* $\mathscr{A} = Diag(\mathscr{A}_i : i \in [p])$ *is symmetric T-P(S)D if and only if all* $\mathscr{A}_i$ *are so.*

Proof For any nonzero $\mathscr{V} \in \mathbb{R}^{n\times 1\times p}$ with $n = \sum_{i=1}^{m} n_i$, we divided it into a block tensor, i.e., $\mathscr{V} = vec(\mathscr{V}_i : i \in [m])$ where $\mathscr{V}_i \in \mathbb{R}^{n_i\times 1\times p}$ for any $i \in [m]$. Then, by Lemma 3, we can obtain that

$$
\begin{aligned}
\mathscr{V}^{\top} * \mathscr{A} * \mathscr{V} &= vec(\mathscr{V}_i : i \in [m])^{\top} * Diag(\mathscr{A}_i : i \in [p]) * vec(\mathscr{V}_i : i \in [m]) \\
&= \sum_{i=1}^{m} \mathscr{V}_i^{\top} * \mathscr{A}_i * \mathscr{V}_i.
\end{aligned}
$$

Hence, it is not difficult to get that $\mathscr{A}$ is symmetric T-P(S)D iff all $\mathscr{A}_i$ are so. □

Besides, from [43, Proposition 1.1.7], Lemma 1 and Theorem 4, the following result holds.

Lemma 5 *Suppose that* $\mathscr{B} \in \mathbb{R}^{n\times n\times p}$ *be nonsingular. Then* $\mathscr{A} \in S\mathbb{R}_+^{n\times n\times p}$ ($S\mathbb{R}_{++}^{n\times n\times p}$) *if and only if* $(\mathscr{B}^{\top} * \mathscr{A} * \mathscr{B}) \in S\mathbb{R}_+^{n\times n\times p}$ ($S\mathbb{R}_{++}^{n\times n\times p}$).

Now, we can establish a theorem about the T-Schur complement.

Theorem 10 (T-Schur complement) *Suppose that* $\mathscr{A} \in S\mathbb{R}_{++}^{m\times m\times p}$, $\mathscr{C} \in S\mathbb{R}^{n\times n\times p}$, *and* $\mathscr{B} \in \mathbb{R}^{m\times n\times p}$. *Then*

$$\begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{B}^\top & \mathscr{C} \end{bmatrix} >_{\mathscr{T}} (\succeq_{\mathscr{T}}) \, \mathscr{O} \iff \mathscr{C} - \mathscr{B}^\top * \mathscr{A}^{-1} * \mathscr{B} >_{\mathscr{T}} (\succeq_{\mathscr{T}}) \, \mathscr{O}.$$

Proof It follows from $\mathscr{A} \in S\mathbb{R}_{++}^{m \times m \times p}$ that $\mathscr{A}$ is nonsingular. Denote the block tensor

$$\mathscr{D} := \begin{bmatrix} \mathscr{I}_{mmp} & -\mathscr{A}^{-1} * \mathscr{B} \\ \mathscr{O} & \mathscr{I}_{nnp} \end{bmatrix},$$

then we have

$$\mathscr{D}^\top * \begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{B}^\top & \mathscr{C} \end{bmatrix} * \mathscr{D} = \begin{bmatrix} \mathscr{A} & \mathscr{O} \\ \mathscr{O} & \mathscr{C} - \mathscr{B}^\top * \mathscr{A}^{-1} * \mathscr{B} \end{bmatrix}.$$

Therefore, by Lemmas 4 and 5, the theorem is proved. $\qquad\square$

5 Semidefinite programming over the third-order symmetric tensor space

In this section, we first introduce the TSDP and give its duality theory; and then, we show the transformation of TSDPs into SDPs in the complex domain. After that, we consider several problems and reformulate (or relax) them as TSDPs. Finally, we present some preliminary numerical results for solving the unconstrained polynomial optimization problem via the TSDP relaxation.

5.1 TSDP problems in primal-dual forms

In this subsection, we replace the matrix variables in the classic SDP by the tensor variables to yield the TSDP. We consider the TSDP in primal form:

$$(\text{PTSDP}) \quad \min_{\mathscr{X}} \langle \mathscr{C}, \mathscr{X} \rangle \quad \text{s.t.} \quad \mathscr{A}\mathscr{X} = [\langle \mathscr{A}_i, \mathscr{X} \rangle]_{i \in [m]} = \mathbf{b}, \quad \mathscr{X} \succeq_{\mathscr{T}} \mathscr{O},$$

where all $\mathscr{A}_i \in S\mathbb{R}^{n \times n \times p}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathscr{C} \in S\mathbb{R}^{n \times n \times p}$ are given and $\mathscr{X} \in S\mathbb{R}^{n \times n \times p}$ is the variable. $\mathscr{A}$ is a linear operator from $S\mathbb{R}^{n \times n \times p}$ into $\mathbb{R}^m$.

Just as the derivation of the dual problem of the SDP, in order to obtain the dual problem of (PTSDP), we try to find the adjoint operator of $\mathscr{A}$ at first, which is a linear operator from $\mathbb{R}^m$ into $S\mathbb{R}^{n \times n \times p}$ satisfying $\langle \mathscr{A}\mathscr{X}, \mathbf{y} \rangle = \langle \mathscr{X}, \mathscr{A}^*\mathbf{y} \rangle$ for any $\mathscr{X}$ in $S\mathbb{R}^{n \times n \times p}$ and $\mathbf{y}$ in $\mathbb{R}^m$. Since

$$\langle \mathscr{A}\mathscr{X}, \mathbf{y} \rangle = \frac{1}{p} \sum_{i=1}^m y_i Tr(\mathscr{A}_i * \mathscr{X}) = \frac{1}{p} Tr\left(\mathscr{X} * \sum_{i=1}^m y_i \mathscr{A}_i \right) = \langle \mathscr{X}, \mathscr{A}^*\mathbf{y} \rangle,$$

we have $\mathscr{A}^*\mathbf{y} = \sum_{i=1}^m y_i \mathscr{A}_i$. Now we can construct the dual of (PTSDP) by the Lagrange approach. By adding a Lagrange multiplier $\mathbf{y} \in \mathbb{R}^m$, (PTSDP) can be turned into $\inf_{\mathscr{X} \succeq_{\mathscr{T}} \mathscr{O}} \sup_{\mathbf{y} \in \mathbb{R}^m} \langle \mathscr{C}, \mathscr{X} \rangle + \langle \mathbf{b} - \mathscr{A}\mathscr{X}, \mathbf{y} \rangle$, then the dual of (PTSDP) yields through interchanging inf and sup. Note that

$$\sup_{\mathbf{y}\in\mathbb{R}^m}\inf_{\mathscr{X}\succeq_{\mathscr{T}}\mathscr{O}}\langle\mathbf{b},\mathbf{y}\rangle+\langle\mathscr{C}-\mathscr{A}^*\mathbf{y},\mathscr{X}\rangle=\begin{cases}\langle\mathbf{b},\mathbf{y}\rangle,&\text{if }\mathscr{C}-\mathscr{A}^*\mathbf{y}\in(S\mathbb{R}_+^{n\times n\times p})^*,\\-\infty,&\text{otherwise.}\end{cases}$$

This, together with Theorem 9, implies that we can write the dual problem of (PTSDP) by introducing a slack tensor $\mathscr{S}$ as:

$$(\text{DTSDP})\quad\max_{\mathbf{y},\mathscr{S}}\langle\mathbf{b},\mathbf{y}\rangle\quad\text{s.t.}\quad\mathscr{A}^*\mathbf{y}+\mathscr{S}=\mathscr{C},\ \ \mathscr{S}\succeq_{\mathscr{T}}\mathscr{O},$$

where $\mathbf{y}\in\mathbb{R}^m$ and $\mathscr{S}\in S\mathbb{R}^{n\times n\times p}$ are the variables. When $p=1$, the TSDP problems (PTSDP) and (DTSDP) are corresponding to the classic SDP problems in primal-dual forms. Denote

$$\begin{aligned}F(P)&:=\{\mathscr{X}\in S\mathbb{R}^{n\times n\times p}:\mathscr{A}\mathscr{X}=\mathbf{b},\ \mathscr{X}\succeq_{\mathscr{T}}\mathscr{O}\},\\F(D)&:=\{(\mathbf{y},\mathscr{S})\in\mathbb{R}^m\times S\mathbb{R}^{n\times n\times p}:\mathscr{A}^*\mathbf{y}+\mathscr{S}=\mathscr{C},\ \mathscr{S}\succeq_{\mathscr{T}}\mathscr{O}\},\\p^*&:=\inf\{\langle\mathscr{C},\mathscr{X}\rangle:\mathscr{X}\in F(P)\}\ \text{ and }\ d^*:=\sup\{\langle\mathbf{b},\mathbf{y}\rangle:(\mathbf{y},\mathscr{S})\in F(D)\}.\end{aligned}$$
(6)

From properties of the T-semidefinite cone obtained in Sect. 4.5 and the theory of conic optimization problems [50], it is not difficult to obtain the following results, and the proofs are omitted here.

Theorem 11 *Let $F(P)$, $F(D)$, p^* and d^* be defined as (6). Suppose that $\mathscr{X}\in F(P)$ and $(\mathbf{y},\mathscr{S})\in F(D)$. Then*

- (weak duality) $\langle b,y\rangle\leq\langle\mathscr{C},\mathscr{X}\rangle$.
- (strong duality) Suppose that (PTSDP) is bounded below and strictly feasible (respectively, (DTSDP) is bounded above and strictly feasible), then $p^*=d^*$ and (DTSDP) (respectively, (PTSDP)) is solvable.
- (complementarity slackness condition) If $p^*=d^*$, then $\mathscr{X}$ is optimal for (PTSDP) and $(\mathbf{y},\mathscr{S})$ is optimal for (DTSDP) if and only if the complementarity slackness condition holds, that is, $\langle\mathscr{X},\mathscr{S}\rangle=0$.
- (optimality condition) If $\langle\mathscr{C},\mathscr{X}\rangle=\langle b,y\rangle$, then $\mathscr{X}$ is optimal for (PTSDP), and $(\mathbf{y},\mathscr{S})$ is optimal for (DTSDP).

5.2 The transformation of TSDPs into SDPs in the complex domain

In this subsection, we present a method to solve the TSDP problem by transforming it into an SDP in the complex domain (CSDP for short).

For any $\mathscr{A}_i\in S\mathbb{R}^{n\times n\times p}$ ($i\in[m]$) in (PTSDP), $bcirc(\mathscr{A}_i)$ can be block diagonalized as $bcirc(\mathscr{A}_i)=(\mathbf{F}_p^H\otimes\mathbf{I}_{n\times n})\mathbf{A}^i(\mathbf{F}_p\otimes\mathbf{I}_{n\times n})$ with $\mathbf{A}^i=Diag(\mathbf{A}_j^i:j\in[p])$ where all $\mathbf{A}_j^i\in H\mathbb{C}^{n\times n}$ ($\mathbf{A}_j^i\in H\mathbb{C}_+^{n\times n}$ if particularly $\mathscr{A}\in S\mathbb{R}_+^{n\times n\times p}$). Note that

$$\begin{aligned}&\langle\mathscr{C},\mathscr{X}^*\rangle=\min_{\mathscr{X}}\langle\mathscr{C},\mathscr{X}\rangle\\\Leftrightarrow\ &\langle bcirc(\mathscr{C}),bcirc(\mathscr{X}^*)\rangle=\min_{bcirc(\mathscr{X})}\langle bcirc(\mathscr{C}),bcirc(\mathscr{X})\rangle\\\Leftrightarrow\ &\langle\mathbf{C},\mathbf{X}^*\rangle=\min_{\mathbf{X}}\langle\mathbf{C},\mathbf{X}\rangle,\end{aligned}$$

where $bcirc(\mathscr{C}) = (\mathbf{F}_p^H \otimes \mathbf{I}_{n \times n}) \mathbf{C} (\mathbf{F}_p \otimes \mathbf{I}_{n \times n})$, $bcirc(\mathscr{X}) = (\mathbf{F}_p^H \otimes \mathbf{I}_{n \times n}) \mathbf{X} (\mathbf{F}_p \otimes \mathbf{I}_{n \times n})$ and $bcirc(\mathscr{X}^*) = (\mathbf{F}_p^H \otimes \mathbf{I}_{n \times n}) \mathbf{X}^* (\mathbf{F}_p \otimes \mathbf{I}_{n \times n})$ with

$$\mathbf{C} = Diag(\mathbf{C}_i, i \in [p]), \mathbf{X} = Diag(\mathbf{X}_i, i \in [p]), \mathbf{X}^* = Diag(\mathbf{X}_i^*, i \in [p]) \tag{7}$$

with all $\mathbf{C}_i$, $\mathbf{X}_i$ and $\mathbf{X}^*_i$ in $H\mathbb{C}^{n \times n}$. In addition, $\mathscr{X} \succeq_{\mathscr{T}} \mathcal{O} \Leftrightarrow \mathbf{X} \succeq \mathbf{O}$ and

$$\mathscr{A}\mathscr{X} = [\langle \mathscr{A}_i, \mathscr{X} \rangle]_{i \in [m]} = \left[\frac{1}{p} Tr(bcirc(\mathscr{A}_i), bcirc(\mathscr{X})) \right]_{i \in [m]} = \left[\frac{1}{p} \langle \mathbf{A}^i, \mathbf{X} \rangle \right]_{i \in [m]}.$$

Therefore, let S denote the space of block diagonal Hermitian matrices as the form in (7), then (PTSDP) and (DTSDP) are equivalent to the following SDP problems (PCSDP) and (DCSDP) respectively:

$$\text{(PCSDP)} \qquad \min_{\mathbf{X} \in S} \frac{1}{p} \langle \mathbf{C}, \mathbf{X} \rangle \quad \text{s.t.} \quad \mathbf{A}\mathbf{X} = p\mathbf{b}, \ \mathbf{X} \succeq \mathbf{O},$$

$$\text{(DCSDP)} \qquad \max_{(\mathbf{y}, \mathbf{S}) \in \mathbb{R}^m \times S} \langle \mathbf{b}, \mathbf{y} \rangle \quad \text{s.t.} \quad \mathbf{A}^* \mathbf{y} + \mathbf{S} = \mathbf{C}, \ \mathbf{S} \succeq \mathbf{O},$$

where $\mathbf{X}$, $\mathbf{C}$ and $\mathbf{S}$ are given as (7), and $\mathbf{A}$ is a linear operator from $H\mathbb{C}^{np \times np}$ into $\mathbb{R}^m$ denoted as $\mathbf{A}\mathbf{X} = [\langle \mathbf{A}^i, \mathbf{X} \rangle]_{i \in [m]}$ with $\mathbf{A}^*$ being its adjoint operator.

It should be noted that both (DCSDP) and (PCSDP) are SDPs in the complex domain and (DCSDP) is exactly the dual problem of (PCSDP). Noting that these diagonal blocks of the complex matrices $\mathbf{X}$, $\mathbf{C}$ and $\mathbf{A}^i$ for $i \in [m]$, obtained by block diagonalizing the real tensors $\mathscr{X}$, $\mathscr{C}$ and $\mathscr{A}$, satisfy the relationships described in Lemma . So, (PCSDP) and (DCSDP) can be converted to SDPs of smaller size. For the cleanness of the paper, we only take the transformation of (PCSDP) for example, which can be divided into the following two cases.

Case 1: p is even. Let $\mathbf{X}$, $\mathbf{C}$ and $\mathbf{A}^i$ for $i \in [m]$ be the complex block diagonal matrices in (PCSDP), which are obtained by block diagonalizing the real tensors $\mathscr{X}$, $\mathscr{C}$ and $\mathscr{A}$ for $i \in [m]$ in (PTSDP), respectively. From Lemma , it follows that for any $j \in [p] \backslash \{1, \frac{p+2}{2}\}$ and $i \in [m]$,

$$\begin{cases} \mathbf{X}_1 \in \mathbb{R}^{n \times n}, \ \mathbf{X}_{\frac{p+2}{2}} \in \mathbb{R}^{n \times n}, \ \mathbf{X}_j \in \mathbb{C}^{n \times n}, \ \mathbf{X}_j = \overline{\mathbf{X}_{p+2-j}}; \\ \mathbf{C}_1 \in \mathbb{R}^{n \times n}, \ \mathbf{C}_{\frac{p+2}{2}} \in \mathbb{R}^{n \times n}, \ \mathbf{C}_j \in \mathbb{C}^{n \times n}, \ \mathbf{C}_j = \overline{\mathbf{C}_{p+2-j}}; \\ \mathbf{A}_1^i \in \mathbb{R}^{n \times n}, \ \mathbf{A}_{\frac{p+2}{2}}^i \in \mathbb{R}^{n \times n}, \ \mathbf{A}_j^i \in \mathbb{C}^{n \times n}, \ \mathbf{A}_j^i = \overline{\mathbf{A}_{p+2-j}^i}. \end{cases}$$

Thus, for any $i \in [m]$,

$$\begin{aligned} \mathbf{A}^i \cdot \mathbf{X} &= \mathbf{A}_1^i \cdot \mathbf{X}_1 + \mathbf{A}_2^i \cdot \mathbf{X}_2 + \cdots + \mathbf{A}_p^i \cdot \mathbf{X}_p \\ &= \mathbf{A}_1^i \cdot \mathbf{X}_1 + \mathbf{A}_{\frac{p+2}{2}}^i \cdot \mathbf{X}_{\frac{p+2}{2}} + \sum_{j=2}^{\frac{p}{2}} (\mathbf{A}_j^i \cdot \mathbf{X}_j + \overline{\mathbf{A}_j^i} \cdot \overline{\mathbf{X}_j}) \\ &= \mathbf{A}_1^i \cdot \mathbf{X}_1 + \mathbf{A}_{\frac{p+2}{2}}^i \cdot \mathbf{X}_{\frac{p+2}{2}} + 2 \sum_{j=2}^{\frac{p}{2}} \mathbf{A}_j^i \cdot \mathbf{X}_j, \end{aligned}$$

where the last equality follows from the fact that the inner product between two Hermitian matrices is real. Similarly, we can also obtain that

$$\mathbf{C} \cdot \mathbf{X} = \mathbf{C}_1 \cdot \mathbf{X}_1 + \mathbf{C}_{\frac{p+2}{2}} \cdot \mathbf{X}_{\frac{p+2}{2}} + 2 \sum_{j=1}^{\frac{p}{2}} \mathbf{C}_j \cdot \mathbf{X}_j.$$

Thus, by letting $\widetilde{\mathbf{X}} = Diag(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{\frac{p}{2}}, \mathbf{X}_{\frac{p+2}{2}})$, $\widetilde{\mathbf{A}}^i = Diag(\mathbf{A}_1^i, 2\mathbf{A}_2^i, \dots, 2\mathbf{A}_{\frac{p}{2}}^i, \mathbf{A}_{\frac{p+2}{2}}^i)$ for any $i \in [m]$, and $\widetilde{\mathbf{C}} = Diag(\mathbf{C}_1, 2\mathbf{C}_2, \dots, 2\mathbf{C}_{\frac{p}{2}}, \mathbf{C}_{\frac{p+2}{2}})$, it follows that (PCSDP) is equivalent to

$$(\text{P}'\text{CSDP}) \quad \min_{\widetilde{\mathbf{X}} \in S} \frac{1}{p} \langle \widetilde{\mathbf{C}}, \widetilde{\mathbf{X}} \rangle \quad \text{s.t.} \quad \widetilde{\mathbf{A}}\widetilde{\mathbf{X}} = p\mathbf{b}, \ \ \widetilde{\mathbf{X}} \succeq \mathbf{O},$$

where $\widetilde{\mathbf{A}}$ is a linear operator with $\widetilde{\mathbf{A}}\widetilde{\mathbf{X}} = [\langle \widetilde{\mathbf{A}}^i, \widetilde{\mathbf{X}} \rangle]_{i \in [m]}$.

Case 2: p is odd. By the same process as **Case 1**, it is not difficult to obtain that (PCSDP) is equivalent to

$$(\text{P}''\text{CSDP}) \quad \min_{\widetilde{\mathbf{X}} \in S} \frac{1}{p} \langle \widetilde{\mathbf{C}}, \widetilde{\mathbf{X}} \rangle \quad \text{s.t.} \quad \widetilde{\mathbf{A}}\widetilde{\mathbf{X}} = p\mathbf{b}, \ \ \widetilde{\mathbf{X}} \succeq \mathbf{O},$$

where $\widetilde{\mathbf{A}}^i = Diag(\mathbf{A}_1^i, 2\mathbf{A}_2^i, \cdots, 2\mathbf{A}_{\frac{p+1}{2}}^i)$ for any $i \in [m]$; $\widetilde{\mathbf{C}} = Diag(\mathbf{C}_1, 2\mathbf{C}_2, \cdots, 2\mathbf{C}_{\frac{p+1}{2}})$; $\widetilde{\mathbf{X}} = Diag(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{\frac{p+1}{2}})$; and $\widetilde{\mathbf{A}}$ is a linear operator with $\widetilde{\mathbf{A}}\widetilde{\mathbf{X}} = [\langle \widetilde{\mathbf{A}}^i, \widetilde{\mathbf{X}} \rangle]_{i \in [m]}$.

As can be seen from the above discussion, we provide a way to deal with (PTSDP) of size $n \times n \times p$ by transforming it into a CSDP with block diagonal structure of size $n(\frac{p+2}{2}) \times n(\frac{p+2}{2})$ as (P'CSDP) or $n(\frac{p+1}{2}) \times n(\frac{p+1}{2})$ as (P''CSDP), which are almost half the size of (PCSDP) with block diagonal structure of size $np \times np$ when $p > 2$.

5.3 Some applications of TSDPs

In this subsection, we show several applications which can be formulated as TSDP problems.

Application 1. Minimizing the maximum T-eigenvalue of a third-order symmetric tensor. The T-eigenvalue was first proposed for third-order symmetric tensors in [34] by Miao, Qi and Wei. Suppose that $\mathcal{M}(\mathbf{z}) \in S\mathbb{R}^{n \times n \times p}$ is a third-order symmetric tensor, which depends linearly on a vector $\mathbf{z}$. Since $\lambda_{max}(\mathcal{M}(\mathbf{z})) \le \eta$ if and only if $\lambda_{max}(\mathcal{M}(\mathbf{z}) - \eta \mathcal{I}_{nnp}) \le 0$, i.e., $\lambda_{min}(\eta \mathcal{I}_{nnp} - \mathcal{M}(\mathbf{z})) \ge 0$, which and Remark 8 imply that $\eta \mathcal{I}_{nnp} - \mathcal{M}(\mathbf{z}) \succeq_{\mathcal{T}} \mathcal{O}$. Therefore, the problem of minimizing the maximum T-eigenvalue of $\mathcal{M}(\mathbf{z})$ can be transformed as the following TSDP problem:

$$\max_{\eta, \mathbf{z}} \ -\eta \quad \text{s.t.} \quad \eta \mathcal{I}_{nnp} - \mathcal{M}(\mathbf{z}) \succeq_{\mathcal{T}} \mathcal{O}.$$

Application 2. Minimizing the l_2 operator norm of a third-order tensor induced by the matrix l_2 operator norm and T-product. Analogue to the induction of the matrix l_2 operator norm, for any $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$, the l_2 operator norm of $\mathscr{A}$ with T-product, denoted by $\|\mathscr{A}\|_2$, is defined as

$$\|\mathscr{A}\|_2 = \max_{\|\mathbf{X}\|_2 \le 1} \|\mathscr{A} * \mathbf{X}\|_2,$$

which is shown to equal the largest singular value of $\mathscr{A}$ (see [18, 30]). It should be pointed out that the above l_2 operator norm and its dual norm were named as the tensor spectral norm and tensor nuclear norm, respectively, in some of the literature. These are not appropriate. For correct definitions of tensor spectral norm and tensor nuclear norm can referee to [51, 52].

It is known that the l_2 operator norm of a third-order tensor plays an important role in the proof of the optimal conditions for the relative problems in [18, 30]. Thus, in this part, we consider the problem of minimizing the l_2 operator norm of a third-order. Suppose that $\mathscr{P}(\mathbf{z}) \in \mathbb{R}^{m \times n \times p}$ is a third-order real tensor, which depends linearly on a vector $\mathbf{z}$. Noting that $\eta \geq \|\mathscr{P}(\mathbf{z})\|_2$ if and only if $\eta^2 \geq \lambda_{max}(\mathscr{P}(\mathbf{z})^\top * \mathscr{P}(\mathbf{z}))$; and by Theorem 10, the latter is equivalent to

$$\begin{bmatrix} \eta \mathscr{I}_{mmp} & \mathscr{P}(\mathbf{z}) \\ \mathscr{P}(\mathbf{z})^\top & \eta \mathscr{I}_{nnp} \end{bmatrix} \succeq_{\mathscr{T}} \mathscr{O}.$$

Therefore, the problem of minimizing $\|\mathscr{P}(\mathbf{z})\|_2$ can be transformed as the following TSDP problem:

$$\max_{\eta, \mathbf{z}} \ -\eta \quad \text{s.t.} \quad \begin{bmatrix} \eta \mathscr{I}_{mmp} & \mathscr{P}(\mathbf{z}) \\ \mathscr{P}(\mathbf{z})^\top & \eta \mathscr{I}_{nnp} \end{bmatrix} \succeq_{\mathscr{T}} \mathscr{O}.$$

Application 3. Minimizing the dual norm of the l_2 operator norm of a third-order tensor. Recall that the the dual norm of the l_2 operator norm of any $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$, denoted by $\|\mathscr{A}\|_*$, is defined as the sum of singular values of the first frontal slice in [18]. It has been showed in [18] that the tensor robust principal component analysis problem can be transformed into the minimization problem with such third-order tensor norm $\| \cdot \|_*$, which can be solved with the help of the theory of tensor decomposition. In this part, we consider the following two cases.

Case 1. Minimizing the third-order tensor norm $\| \cdot \|_*$ without constraint. Recall that for a given norm $\| \cdot \|$ in the inner product space consisting of three-order tensors, its dual norm $\| \cdot \|_d$ is defined as $\|\mathscr{X}\|_d := \sup\{\langle \mathscr{X}, \mathscr{Y} \rangle : \|\mathscr{Y}\| \leq 1\}$. Since the norm $\|\mathscr{A}\|_*$ of tensor $\mathscr{A}$ is the dual norm of the l_2 operator norm $\|\mathscr{A}\|_2$, according to the relationship that for any $\mathscr{X} \in \mathbb{R}^{m \times n \times p}$,

$$\|\mathscr{X}\|_2 \leq 1 \iff \begin{bmatrix} \mathscr{I}_{mmp} & \mathscr{X} \\ \mathscr{X}^\top & \mathscr{I}_{nnp} \end{bmatrix} \succeq_{\mathscr{T}} \mathscr{O}$$

as shown in **Application 2** by taking $\eta = 1$, we can obtain that the problem of computing the norm $\| \cdot \|_*$ of $\mathscr{A} \in \mathbb{R}^{m \times n \times p}$ is equivalent to the following TSDP model:

$$\max_{\mathscr{X}} \ \langle \mathscr{A}, \mathscr{X} \rangle = \frac{1}{2} \left\langle \begin{bmatrix} \mathscr{O} & \mathscr{A} \\ \mathscr{A}^\top & \mathscr{O} \end{bmatrix}, \begin{bmatrix} \mathscr{I}_{mmp} & \mathscr{X} \\ \mathscr{X}^\top & \mathscr{I}_{nnp} \end{bmatrix} \right\rangle \quad \text{s.t.} \quad \begin{bmatrix} \mathscr{I}_{mmp} & \mathscr{X} \\ \mathscr{X}^\top & \mathscr{I}_{nnp} \end{bmatrix} \succeq_{\mathscr{T}} \mathscr{O}. \tag{8}$$

Noting that for any $\mathscr{W}_1 \in \mathbb{R}^{m \times m \times p}$ and $\mathscr{W}_2 \in \mathbb{R}^{n \times n \times p}$,

$$\left\langle \begin{bmatrix} \mathscr{W}_1 & \mathscr{O} \\ \mathscr{O}^\top & \mathscr{W}_2 \end{bmatrix}, \begin{bmatrix} \mathscr{I}_{mmp} & \mathscr{X} \\ \mathscr{X}^\top & \mathscr{I}_{nnp} \end{bmatrix} \right\rangle = \frac{1}{p}[Tr(\mathscr{W}_1) + Tr(\mathscr{W}_2)],$$

thus, the dual problem of (8) can be formulated as

$$\min_{\mathscr{W}_1, \mathscr{W}_2} \frac{1}{2p}[Tr(\mathscr{W}_1) + Tr(\mathscr{W}_2)] \quad \text{s.t.} \quad \begin{bmatrix} \mathscr{W}_1 & \mathscr{A} \\ \mathscr{A}^\top & \mathscr{W}_2 \end{bmatrix} \succeq_{\mathscr{T}} \mathscr{O}. \tag{9}$$

It is not difficult to show that there is no duality gap between (8) and (9). We omit the proof here.

Case 2. Minimizing the third-order tensor norm $\|\cdot\|_*$ with an equality constraint. In this part, we investigate the problem of minimizing the norm $\|\cdot\|_*$ of $\mathscr{X} \in \mathbb{R}^{m \times n \times p}$ over a given affine subspace. Usually, the subspace is described by a linear equations of the form $\mathscr{A}\mathscr{X} = \mathbf{b}$ as discussed in Sect. 5.1. This problem can be formulated as a convex optimization in the following form:

$$\min_{\mathscr{X}} \|\mathscr{X}\|_* \quad \text{s.t.} \quad \mathscr{A}\mathscr{X} = \mathbf{b}. \tag{10}$$

Then, by using the TSDP characterization of the dual norm of the l_2 operator norm given in (9), we can rewrite (10) as

$$\min_{\mathscr{X}, \mathscr{W}_1, \mathscr{W}_2} \frac{1}{2p}[Tr(\mathscr{W}_1) + Tr(\mathscr{W}_2)] \quad \text{s.t.} \quad \begin{bmatrix} \mathscr{W}_1 & \mathscr{X} \\ \mathscr{X}^\top & \mathscr{W}_2 \end{bmatrix} \succeq_{\mathscr{T}} \mathscr{O}, \quad \mathscr{A}\mathscr{X} = \mathbf{b}.$$

The above discussion demonstrates that the minimization problem of the third-order tensor norm $\|\cdot\|_*$ can also be solved by dealing with the corresponding TSDP problem. As is known to us, a lot of practical problems are usually turned out to be low-rank models with third-order tensors, and most of them can be solved by the minimization problem with another related tensor norm defined in [10], which has been shown to be widely applied in some practical problems, such as image processing, tensor principal component analysis, tensor completion, and so on. Then it is worthwhile to investigate these forms of convex relaxation by replacing the norm defined in [10] by the one given in [18], and the TSDP provides another path to achieve solutions of low-rank recovery problems with third-order tensors appearing in the real-life applications.

Application 4. Integer quartic programming. Analogue to the classic SDP relaxation of integer quadratic programming [53, 54], we investigate the TSDP relaxation for the following integer quartic programming:

$$\max_{\mathscr{X}} \langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle \quad \text{s.t.} \quad \mathbf{X} = \mathbf{x}\mathbf{x}^\top, \ x_i \in \{+1, -1\}, \ \forall i \in [n], \tag{11}$$

where $\mathscr{A} \in S\mathbb{R}^{n \times n \times n}$ is given, $\mathbf{x} := (x_i : i \in [n])^\top \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times n}$ and $\mathscr{X} \in \mathbb{R}^{n \times 1 \times n}$ is the corresponding tensor of the matrix $\mathbf{X}$.

Noting that for any $i \in [n]$, $x_i \in \{+1, -1\}$ if and only if $x_i^2 = 1$, if and only if $x_i^2(x_1^2 + x_2^2 + \cdots + x_n^2) = n$. Thus, problem (11) is equivalent to

$$\max_{\mathscr{X}} \ \langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle \quad \text{s.t.} \quad \mathbf{X} = \mathbf{x}\mathbf{x}^\top, \ x_i^2(x_1^2 + x_2^2 + \cdots + x_n^2) = n, \ \forall \, i \in [n]. \tag{12}$$

Now, we try to arrive at the TSDP relaxation of (12) by using the Lagrangian multiplier method. By adding Lagrangian multiplier $y_i \in \mathbb{R}$ to each equality constraint in (12), we can obtain the Lagrangian function for (12):

$$\begin{aligned} L(\mathscr{X}, \mathbf{y}) \ &:= \ \langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle - \sum_{i=1}^{n} y_i(x_i^2(x_1^2 + x_2^2 + \cdots + x_n^2) - n) \\ &= \ \langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle - \langle \mathscr{X}, \mathscr{D}iag(\mathbf{y}) * \mathscr{X} \rangle + n\mathbf{e}^\top \mathbf{y}, \end{aligned}$$

where $\mathscr{D}iag(\mathbf{y})$ represents the F-diagonal tensor induced by $\mathbf{y} := (y_i : i \in [n])^\top$ with $\mathscr{D}iag(\mathbf{y})^{(1)}$ being the diagonal matrix $Diag(y_i : i \in [n])$ and other frontal slices being zeroes, and $\mathbf{e} := (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$. Since

$$\begin{aligned} \max_{\mathscr{X}} L(\mathscr{X}, \mathbf{y}) \ &= \ \max_{\mathscr{X}}[\langle \mathscr{X}, \mathscr{A} * \mathscr{X} \rangle - \langle \mathscr{X}, \mathscr{D}iag(\mathbf{y}) * \mathscr{X} \rangle + n\mathbf{e}^\top \mathbf{y}] \\ &= \ -\min_{\mathscr{X}}[\langle \mathscr{X}, (\mathscr{D}iag(\mathbf{y}) - \mathscr{A}) * \mathscr{X} \rangle - n\mathbf{e}^\top \mathbf{y}] \\ &= \ \begin{cases} n\mathbf{e}^\top \mathbf{y}, & \text{if } \mathscr{D}iag(\mathbf{y}) - \mathscr{A} \succeq_{\mathscr{T}} \mathscr{O}, \\ +\infty, & \text{otherwise,} \end{cases} \end{aligned}$$

the Lagrangian dual problem of (12) turns out to be

$$\min_{\mathbf{y}} \ n\mathbf{e}^\top \mathbf{y} \quad \text{s.t.} \quad \mathscr{D}iag(\mathbf{y}) - \mathscr{A} \succeq_{\mathscr{T}} \mathscr{O},$$

which is a TSDP model and its dual problem is

$$\max_{\mathscr{X}} \ \mathscr{A} \cdot \mathscr{X} \quad \text{s.t.} \quad \mathscr{X}_{ii1} = n, \ i \in [n], \ \mathscr{X} \succeq_{\mathscr{T}} \mathscr{O}.$$

Previously, we have converted some optimization problems over tensor space and matrix space into TSDP problems, respectively. Next, we show that some optimization problems over vector space can also be solved by TSDP problem models, such as polynomial optimization problems.

Application 5. Calculating the global lower bound of a polynomial of even degree. Consider the polynomial optimization problem:

$$f^{uc} := \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = f_0 + \sum_{\alpha \in U_{2d}^n} f_\alpha \mathbf{x}^\alpha, \tag{13}$$

where $U_{2d}^n = \{\alpha \in \mathbb{N}^n : 0 < |\alpha| \le 2d\}$. As is well-known to us, problem (13) can be solved by a relaxation into the following model through the sums of squares (SOS for short) method [55]:

$$f_{sos}^{uc} := \max_{(\mathbf{x}, \gamma) \in \mathbb{R}^n \times \mathbb{R}} \ \gamma \quad \text{s.t.} \quad f(\mathbf{x}) - \gamma \text{ is SOS.} \tag{14}$$

Now, we consider the SDP relaxation of (14) by fixing the SOS polynomial in the right hand side of the constraint in (14) to be of degree $2d$. Actually, define

$$[\mathbf{x}]_d = (1, x_1, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_1 x_n, x_2^2, x_2 x_3, \ldots, x_n^2, \ldots x_1^d, \ldots, x_n^d)^\top,$$

which is a column vector of size C_{n+d}^d consisting of all monomials whose degrees are no more than $d = \frac{deg(f)}{2}$, then (14) can be transformed into a standard SDP as [56]:

$$f_{sdp}^{uc} := \max_{\mathbf{X}} \ \mathbf{C} \cdot \mathbf{X} \quad \text{s.t.} \quad \mathbf{AX} = \mathbf{b}, \ \mathbf{X} \succeq \mathbf{O}, \tag{15}$$

where $\mathbf{b} = (f_\alpha)_{\alpha \in U_{2d}^n}$ whose dimension is $C_{n+2d}^{2d} - 1$, and $\mathbf{A}$ is a linear operator denoted as $\mathbf{AX} = (\mathbf{A_\alpha} \cdot \mathbf{X})_{\alpha \in U_{2d}^n}$ with $\mathbf{A_\alpha}$ and $\mathbf{C}$ being constant symmetric matrices such that $[\mathbf{x}]_d [\mathbf{x}]_d^\top = \mathbf{C} + \sum_{\alpha \in U_{2d}^n} \mathbf{A_\alpha} \mathbf{x}^\alpha$. In addition, $f_{sos}^{uc} = f_0 - f_{sdp}^{uc}$.

As is known to us, in the process of solving the polynomial optimization problem with the SDP relaxation, one of the challenges is that the size of the SDP model increases significantly with the increasing of the number of variables or the degree of polynomial. Next, we show that the minimization problem of f can be relaxed into a standard TSDP problem by rearranging the above monomial vector $[\mathbf{x}]_d$ into a third-order tensor form and then exploiting the properties of T-PSD tensors, which can be dealt with by solving an SDP problem of smaller size than the one in (15).

Suppose that $C_{n+d}^d = mp$, then $[\mathbf{x}]_d \in \mathbb{R}^{mp}$. Let $[\mathcal{X}]_d$ be a tensor in $\mathbb{R}^{m \times 1 \times p}$ with $[\mathcal{X}]_d = fold([\mathbf{x}]_d)$. Then by Theorem 8, a tensor $\mathcal{A} \in S\mathbb{R}^{m \times m \times p}$ is T-PSD iff there exists some tensor $\mathcal{P} \in S\mathbb{R}^{l \times m \times p}$ such that $\mathcal{A} = \mathcal{P}^\top * \mathcal{P}$. Thus, $f(\mathbf{x}) - \gamma$ must be SOS (and then be nonnegative) if there exists $\mathcal{Z} \in S\mathbb{R}^{m \times m \times p}$ such that

$$f(\mathbf{x}) - \gamma = \frac{1}{p} Tr([\mathcal{X}]_d^\top * \mathcal{Z} * [\mathcal{X}]_d) = \mathcal{Z} \cdot ([\mathcal{X}]_d * [\mathcal{X}]_d^\top), \quad \mathcal{Z} \succeq_{\mathcal{T}} \mathcal{O}. \tag{16}$$

Therefore, finding the minimum value of f can be relaxed into the following problem:

$$f_{sos-tsdp}^{uc} := \max \ \gamma \quad \text{s.t.} \quad f(\mathbf{x}) - \gamma = \mathcal{Z} \cdot ([\mathcal{X}]_d * [\mathcal{X}]_d^\top), \ \mathcal{Z} \succeq_{\mathcal{T}} \mathcal{O}. \tag{17}$$

Define constant symmetric tensors $\mathcal{C}$ and $\mathcal{A}_\alpha$ such that

$$[\mathcal{X}]_d * [\mathcal{X}]_d^\top = \mathcal{C} + \sum_{\alpha \in U_{2d}^n} \mathcal{A}_\alpha \mathbf{x}^\alpha,$$

then (16) can be expressed as follows:

$$f(\mathbf{x}) - \gamma = \mathcal{C} \cdot \mathcal{Z} + \sum_{\alpha \in U_{2d}^n} (\mathcal{A}_\alpha \cdot \mathcal{Z}) \mathbf{x}^\alpha, \quad \mathcal{Z} \succeq_{\mathcal{T}} \mathcal{O}.$$

Noting that $f(\mathbf{x}) = f_0 + \sum_{\alpha \in U_{2d}^n} f_\alpha \mathbf{x}^\alpha$, hence γ is feasible for (17) if and only if there exists $\mathcal{Z} \succeq_{\mathcal{T}} \mathcal{O}$ such that $\mathcal{C} \cdot \mathcal{Z} + \gamma = f_0$, and $\mathcal{A}_\alpha \cdot \mathcal{Z} = f_\alpha$ for any $\alpha \in U_{2d}^n$. Define a linear operator from $S\mathbb{R}^{m \times m \times p}$ into $\mathbb{R}^{(C_{n+2d}^{2d}-1)}$ as $\mathcal{AZ} = [\mathcal{A}_\alpha \cdot \mathcal{Z}]_{\alpha \in U_{2d}^n}$. Then up to a constant, the problem (17) is equivalent to the TSDP problem:

$$f_{tsdp}^{uc} := \max_{\mathcal{Z}} \ \mathcal{C} \cdot \mathcal{Z} \quad \text{s.t.} \quad \mathcal{AZ} = \mathbf{b}, \quad \mathcal{Z} \succeq_{\mathcal{T}} \mathcal{O}, \tag{18}$$

and $f_{sos-tsdp}^{uc} = f_0 - f_{tsdp}^{uc}$.

Besides, from the above discuss, it is easy to see that (17) is both relaxation of (13) and (14) when $p \neq 1$, i.e., $f^{uc} \geq f_{sos}^{uc} \geq f_{sos-tsdp}^{uc}$, and (17) reduces to (14) when $p = 1$. Thereby, it should be noticed that (18) with $p \neq 1$ is a further relaxation of (15). Then *whether or not is there a case where the relaxation of* (18) *with $p \neq 1$ can achieve the same effect as* (15)? Below, we answer this question by giving a theorem to show the necessary and sufficient condition of $f_{sdp}^{uc} = f_{tsdp}^{uc}$ under the assumption that $p \neq 1$.

Theorem 12 *Suppose that $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is a 2d-degree polynomial, whose global lower bound can be solved by both an SDP relaxation as* (15) *and a TSDP relaxation as* (18) *with $p \neq 1$. Then $f_{sdp}^{uc} = f_{tsdp}^{uc}$ if and only if there exists an optimal solution $\mathbf{X}^*$ of* (15) *which is p-block circular.*

Proof Recall that (18) is equivalent to (17), that is

$$f_0 - f_{tsdp}^{uc} := \max \ \gamma \quad \text{s.t.} \quad f(\mathbf{x}) - \gamma = \mathscr{Z} \cdot ([\mathscr{X}]_d * [\mathscr{X}]_d^\top), \quad \mathscr{Z} \succeq_{\mathscr{T}} \mathscr{O}; \quad (19)$$

and (15) is equivalent to (14), which is further equivalent to the following problem:

$$f_0 - f_{sdp}^{uc} := \max \ \gamma \quad \text{s.t.} \quad f(\mathbf{x}) - \gamma = \mathbf{X} \cdot ([\mathbf{x}]_d \cdot [\mathbf{x}]_d^\top), \quad \mathbf{X} \succeq \mathbf{O}. \quad (20)$$

From

$$\begin{aligned}
\mathscr{Z} \cdot ([\mathscr{X}]_d * [\mathscr{X}]_d^\top) &= \frac{1}{p} Tr([bcirc([\mathscr{X}]_d)]^\top \cdot bcirc(\mathscr{Z}) \cdot bcirc([\mathscr{X}]_d)) \\
&= \frac{1}{p} bcirc([\mathscr{X}]_d) \cdot [bcirc(\mathscr{Z}) \cdot bcirc([\mathscr{X}]_d)] \\
&= \left(\frac{1}{p} \cdot p\right) unfold([\mathscr{X}]_d) \cdot [bcirc(\mathscr{Z}) \cdot unfold([\mathscr{X}]_d)] \\
&= [\mathbf{x}]_d \cdot [bcirc(\mathscr{Z}) \cdot [\mathbf{x}]_d] \\
&= bcirc(\mathscr{Z}) \cdot ([\mathbf{x}]_d \cdot [\mathbf{x}]_d^\top),
\end{aligned}$$

it is easy to see that if $\mathscr{Z}^*$ is an optimal solution of (19), then $bcirc(\mathscr{Z}^*)$ must be a feasible solution of (20) and $f_{sdp}^{uc} \leq f_{tsdp}^{uc}$.

On one hand, suppose that $\mathscr{Z}^*$ is an optimal solution of (19), if $f_{sdp}^{uc} = f_{tsdp}^{uc}$, then $bcirc(\mathscr{X}^*)$ is exact an optimal solution of (20) and (15), which is a p-block circular matrix. On the other hand, if there exists an optimal solution $\mathbf{X}^*$ of (15) which is p-block circular, then by choosing $\mathscr{Z}^* = bcirc^{-1}(\mathbf{X}^*)$ we can find that $\mathscr{Z}^*$ is a feasible solution of (19) and the corresponding value of γ is $f_0 - f_{sdp}^{uc}$ when $\mathscr{Z} = \mathscr{Z}^*$ in (19). Noting that the optimal value of (19) is $f_0 - f_{tsdp}^{uc}$, then we have that $f_0 - f_{sdp}^{uc} \leq f_0 - f_{tsdp}^{uc}$, which together with $f_{sdp}^{uc} \leq f_{tsdp}^{uc}$ implies that $f_{sdp}^{uc} = f_{tsdp}^{uc}$. □

Theorem 12 states that if there exists an optimal solution $\mathbf{X}^*$ of (15) is p-block circular, then the relaxation problem (18) can achieve not worse effect than (15). Below, we show that there are some benefits in computation costs and storage costs of (18) compared with (15) for the cases where the unconstrained

polynomial optimization can be solved by both TSDP relaxation as (18) and SDP relaxation as (15):

– *The number of entries of the variable $\mathscr{X}$ in* (18) *is less than the one of* **X** *in* (15). Since $[\mathscr{X}]_d \in \mathbb{R}^{m \times 1 \times p}$ and $[\mathbf{x}]_d \in \mathbb{R}^{mp}$, it is easy to find that **X** in (15) is a matrix with $mp \times mp$ entries, while $\mathscr{X}$ in (15) turns out to be a tensor in $S\mathbb{R}^{m \times m \times p}$ which contains only $m \times m \times p$ entries.
– (18) *can be solved by transformed into a CSDP problem with block diagonal sparse structure, which is not available for* (15) *even if the variable of* (15) *possess special structure since the constant matrice* $\mathbf{A}_\alpha$ *is fixed and do not possess the block diagonal structures.* Specifically speaking, if we adopt the method described in Sect. 5.2 to deal with (18), we only need to solve a CSDP with $\frac{p+1}{2}$ or $\frac{p+2}{2}$ blocks of size $m \times m$ under some equality constraints, while (15) can be seen as dealing with an ordinary SDP with 1 block of size $mp \times mp$ under the same number of equality constraints. Hence, for the minimization problem of a real polynomial of even degree, solving via the TSDP relaxation built the above would lead to lower computation costs than the corresponding SDP relaxation, which can also be seen from the numerical experiments in the next subsection.
– *The constant symmetric tensors* $\mathscr{A}_\alpha$ *for* $\alpha \in U_{2d}^n$ *in* (18) *spend no more storage cost than those matrices* $\mathbf{A}_\alpha$ *in* (15). Since the number of entries in the tensor $[\mathscr{X}]_d * [\mathscr{X}]_d^\top$ is $\frac{1}{p}$ of those in the moment matrix $[\mathbf{x}]_d[\mathbf{x}]_d^\top$ and each entry of $[\mathscr{X}]_d * [\mathscr{X}]_d^\top$ is consisted of a linear combination of p monomials whose degrees are no more than $2d$. Thus it seems that we need to take as the same storage as those for matrices $\mathbf{A}_\alpha$ to store these tensors $\mathscr{A}_\alpha$. However, noting that some entries of $[\mathscr{X}]_d * [\mathscr{X}]_d^\top$ may be consisted of a linear combination of p same monomials whose degrees are no more than $2d$, therefore the constant symmetric tensors $\mathscr{A}_\alpha$ for $\alpha \in U_{2d}^n$ in (18) could sometimes save some storage than those matrices $\mathbf{A}_\alpha$ in (15). In addition, the storage capacity does not change in the transformation of TSDPs into CSDPs as shown in Sect. 5.2.

5.4 Numerical computation

In this subsection, we report preliminary numerical results for solving TSDPs by the method shown in Sect. 5.2. Taking Application 5 discussed in Sect. 5.3 for example, we consider two polynomial optimization problems and implement these problems in Matlab R2016a on our PC via SDPNAL+ [42]. The computation is performed on a Dell Laptop with CPU of 3.2 GHz and RAM of 4.0 GB. As pointed in Application 5 of Sect. 5.2, the TSDP relaxation of unconstrained polynomial optimization can achieve the same optimal value as the SDP relaxation sometimes, thus here we confirm this conclusion by two simple examples and illustrate the benefits of the TSDP relaxation furthermore.

Example 1 Minimize the following polynomial:

$$f(\mathbf{x}) = (x_1 + x_2^3 + x_1^2 x_2)^2 + (x_1 + x_1^2 + x_2^3)^2 + (x_1 + x_1^2 + x_2^2)^2 \\ + (x_1^2 + x_2^2 + x_1^2 x_2)^2 + (x_2^2 + x_1^2 x_2 + x_2^3)^2.$$

It is obvious that the global minimum value $f^* = 0$, the number of variables $n = 2$, $d = \frac{deg(f)}{2} = 3$ and $f(\mathbf{x}) - f^*$ can be expressed as

$$f(\mathbf{x}) - f^* = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \\ x_1^3 \\ x_1^2 x_2 \\ x_1 x_2^2 \\ x_2^3 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 2 & 0 & 1 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 3 & 0 & 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 3 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 2 & 0 & 3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \\ x_1^3 \\ x_1^2 x_2 \\ x_1 x_2^2 \\ x_2^3 \end{bmatrix},$$

where the square matrix is 5-block circular. Therefore, by Theorem 12, we know that the TSDP relaxation with $p = 5$ can achieve not worse effect than the SDP relaxation. Specifically, to solve this problem via the TSDP relaxation as described from (13) to (18), we first find the corresponding $\mathscr{A}_\alpha$, $\mathscr{C}$ and $\mathbf{b}$ in (18), then transform $\mathscr{A}_\alpha$ and $\mathscr{C}$ into the corresponding $\mathbf{A}_\alpha$ and $\mathbf{C}$ as the procedure from (PTSDP) to (P″CSDP), and finally, call the package SDPNAL+ to solve the derived SDP. In addition, we also solve this problem via the traditional SDP relaxation as described from (13) to (15) for comparison.

Numerical results for Example 1 by SDP and TSDP via SDPNAL+ are shown in Table 1, where the size of the TSDP is replaced by the size of the corresponding CSDP with block structure and *in each pair "(blk, N, m)", blk, N and m mean the* number of blocks, the length of each block and the number of equality constraints, respectively; **"opt" means** the computed optimal value of the corresponding relaxation; **"cpu1" means** the time in seconds spent for finding the corresponding matrices $\mathbf{A}_\alpha$ of the SDP or the CSDP; **"cpu2" means** the time in seconds spent for solving corresponding SDP problem or CSDP problem via SDPNAL+; and **"cpu3" means** the total time in seconds spent for solving Example 1 by the SDP relaxation or the TSDP relaxation.

Table 1 Numerical results for Example 1 by the SDP and the TSDP via SDPNAL+

	n	d	p	(*blk*, *N*, *m*)	opt	cpu1(s)	cpu2(s)	cpu3(s)
SDP	2	3	1	(1,10,28)	$2.1545e^{-09}$	0.021	0.969	1.008
TSDP	2	3	5	(5,2,28)	$-1.1183e^{-15}$	0.037	0.426	0.483

From 1, we can see that the TSDP relaxation performs well for **Example** 1. Especially, for the TSDP relaxation, we obtain that the global minimum value $f_{tsdp}^* = -1.1183e^{-15}$ by solving a CSDP problem with block diagonal structure, which only takes less than one-twice of the time of the corresponding SDP relaxation. In addition, the TSDP relaxation has higher accuracy than the SDP relaxation in this example.

Example 2 Minimize the following polynomial:

$$f(\mathbf{x}) = 1 + x_1^{10}x_2^4 + x_1^8x_2^{12} + x_1^{24}x_2^2 + x_1^{24}x_2^6 + x_1^{32}x_2^2 + x_1^8x_2^{28} + x_1^{28}x_2^{12} + x_1^{10}x_2^{32}$$
$$+ x_1^{42}x_2^4 + x_1^{30}x_2^{18} + x_1^{20}x_2^{30} + x_1^{12}x_2^{40} + x_1^6x_2^{48} + x_1^2x_2^{54} + x_2^{58}.$$

It is obvious that the global minimum value $f^* = 1$, the number of variables $n = 2$, and $d = \frac{deg(f)}{2} = 29$. We test $p = 3$ and $p = 15$ for the TSDP relaxation of this problem, respectively. Especially, by the TSDP relaxation with $p = 15$, we obtain that the global minimum value $f_{tsdp}^* = 1 + 6.1507e^{-08}$ by solving a CSDP problem with block diagonal structure, which takes only 2.5 seconds; while the corresponding SDP relaxation takes 90.3 seconds. Meanwhile, the TSDP relaxation has higher accuracy than the SDP relaxation in this example.

Numerical results for **Example** 2 by the SDP relaxation and the TSDP relaxation are shown in Table 2, where "(blk, N, m)", "opt" , "cpu1","cpu2" and "cpu3" are same as those in Table 1.

From Tables 1 and 2, we can see that the TSDP relaxation has better performances both in time cost and precision than the traditional SDP relaxation for these test examples. Especially in the time cost, solving the CSDP problem transformed from the corresponding TSDP problem saves a lot of time compared with solving the corresponding SDP problem; while the time for finding the corresponding complex matrices $\mathbf{A}_\alpha$ in the CSDP problem is almost the same as the time for finding the corresponding matrices $\mathbf{A}_\alpha$ in the SDP problem. In fact, it should be noticed that the time for finding the corresponding matrices $\mathbf{A}_\alpha$ can be removed from the total time of solving above polynomial optimization problems via the TSDP or the SDP relaxation because these $\mathbf{A}_\alpha$ are fixed as long as n, d and p are given, and we show them in Tables 1 and 2 just to make the time spent for the TSDP relaxation and the SDP relaxation more clear.

Table 2 Numerical results for Example 2 by the SDP and the TSDP via SDPNAL+

	n	d	p	(blk, N, m)	opt	cpu1(s)	cpu2(s)	cpu3(s)
SDP	2	29	1	(1,465,1769)	$1 - 1.1897e^{-07}$	109.453	90.303	199.896
TSDP	2	29	3	(2,155,1769)	$1 - 4.3220e^{-08}$	107.834	61.426	170.272
TSDP	2	29	15	(8,31,1769)	$1 + 6.1507e^{-08}$	103.302	2.515	110.483

6 Concluding remarks

In this paper, we aimed to generalize the SDP problem to the third-order tensor case. For this purpose, we first introduced the T-positive semidefiniteness of third-order symmetric tensors from the second-order discrimination condition of the convexity of the twice continuously T-differentiable multi-variable real-valued function, and then, extended some of useful characterizations and properties of the symmetric PSD matrix to the third-order symmetric T-PSD tensor. After that, we replaced the variable in the classic SDP by the third-order symmetric tensor to introduce the TSDP, which was dealt with by converting it to a CSDP with the block structure. Finally, several examples with respect to the TSDP problem were shown and some numerical results of minimizing two polynomials via the TSDP relaxation were reported, which demonstrated that our method by the TSDP relaxation performs better than the traditional SDP relaxation for the test examples.

Some issues need to be studied in the future.

(i)   In Sect. 5.4, we have just done some preliminary numerical experiments to test the feasibility and effectiveness of solving TSDPs by dealing with the corresponding CSDPs. Surprisingly, we find that for some unconstrained polynomial optimization, the method of the TSDP relaxation has good performance sometimes. Then what kind of polynomial optimization problems does the TSDP relaxation work well for? It deserves further study.

(ii) It is known to us that the SDP has shown great power in a very wide range of areas. In Sect. 5.3, we have just presented a few simple transformations from some other models into TSDPs. We believe that more problems can be modeled (or relaxed) as TSDPs. It is also known that the T-product between third-order tensors promotes the emergence of many algorithms with good performance in many practical problems. It is possible that more efficient algorithms for TSDP problems can be designed by making use of the characteristics of the T-product and special-structures of practical problems. Furthermore, it deserves to study how to design efficient algorithms for solving large-scale realistic problems.

References

1. Kilmer, M.E., Martin, C.D., Perrone, L.: A third-order generalization of the matrix SVD as a product of third-order tensors. Technical Report TR-2008-4, Tufts University, Department of Computer Science, October (2008)
2. Khaleel, H.S., Sagheer, S.V.M., Baburaj, M., et al.: Denoising of rician corrupted 3D magnetic resonance images using tensor-SVD. Biomed. Signal Process. **44**, 82–95 (2018)

3. Kilmer, M.E., Braman, K., Hao, N., Hoover, R.C.: Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. SIAM J. Matrix Anal. Appl. **34**, 148–172 (2013)
4. Martin, C.D., Shafer, R., Larue, B.: An order-p tensor factorization with applications in imaging. SIAM J. Sci. Comput. **35**, A474–A490 (2013)
5. Soltani, S., Kilmer, M.E., Hansen, P.C.: A tensor-based dictionary learning approach to tomographic image reconstruction. BIT **56**, 1425–1454 (2016)
6. Tarzanagh, D.A., Michailidis, G.: Fast randomized algorithms for t-product based tensor operations and decompositions with applications to imaging data. SIAM J. Imaging Sci. **11**, 2629–2664 (2018)
7. Zhou, P., Lu, C., Lin, Z., Zhang, C.: Tensor factorization for low-rank tensor completion. IEEE Trans. Image Process. **27**, 1152–1163 (2018)
8. Baburaj, M., George, S.N.: Tensor based approach for inpainting of video containing sparse text. Multimed. Tools Appl. **78**(2), 1805–1829 (2019)
9. Semerci, O., Hao, N., Kilmer, M.E., Miller, E.L.: Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. IEEE Trans. Image Process. **23**, 1678–1693 (2014)
10. Hao, N., Kilmer, M.E., Braman, K., Hoover, R.C.: Facial recognition using tensor decompositions. SIAM J. Imaging Sci. **6**, 437–463 (2013)
11. Xie, Y., Tao, D., Zhang, W., Liu, Y., Zhang, L., Qu, Y.: On unifying multi-view self-representations for clustering by tensor multi-rank minimization. Int. J. Comput. Vis. **126**, 1157–1179 (2018)
12. Yin, M., Gao, J., Xie, S., Guo, Y.: Multiview subspace clustering via tensorial t-product representation. IEEE Trans. Neural Netw. Learn. Syst. **30**(3), 851–864 (2019)
13. Chan, T., Yang, Y., Hsuan, Y.: Polar n-complex and n-bicomplex singular value decomposition and principal component pursuit. IEEE Trans. Signal Process. **64**, 6533–6544 (2016)
14. Liu, Y., Chen, L., Zhu, C.: Improved robust tensor principal component analysis via low-rank core matrix. IEEE J-STSP **12**, 1378–1389 (2018)
15. Long, Z., Liu, Y., Chen, L., et al.: Low rank tensor completion for multiway visual data. Signal Process. **155**, 301–316 (2019)
16. Sun, W., Huang, L., So, H.C., et al.: Orthogonal tubal rank-1 tensor pursuit for tensor completion. Signal Process. **157**, 213–224 (2019)
17. Kong, H., Xie, X., Lin, Z.: t-Schatten-p norm for low-rank tensor recovery. IEEE J-STSP **12**, 1405–1419 (2018)
18. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: Tensor robust principal component analysis with a new tensor nuclear norm. IEEE T. Pattern Anal. **42**, 925–938 (2020)
19. Ely, G., Aeron, S., Hao, N., et al.: 5D seismic data completion and denoising using a novel class of tensor decompositions. Geophysics **80**, V83–V95 (2015)
20. Hu, W., Tao, D., Zhang, W., Xie, Y., Yang, Y.: The twist tensor nuclear norm for video completion. IEEE Trans. Neural Netw. Learn. Syst. **28**, 2961–2973 (2017)
21. Hu, W., Yang, Y., Zhang, W., Xie, Y.: Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition. IEEE Trans. Image Process. **26**, 724–737 (2017)
22. Liu, X., Aeron, S., Aggarwal, V., Wang, X., Wu, M.: Adaptive sampling of RF fingerprints for fine-grained indoor localization. IEEE Trans. Mobile Comput. **15**, 2411–2423 (2016)
23. Madathil, B., George, S.N.: Twist tensor total variation regularized-reweighted nuclear norm based tensor completion for video missing area recovery. Inform. Sci. **423**, 376–397 (2018)
24. Madathil, B., George, S.N.: DCT based weighted adaptive multi-linear data completion and denoising. Neurocomputing **318**, 120–136 (2018)
25. Qin, B., Jin, M., Hao, D., et al.: Accurate vessel extraction via tensor completion of background layer in X-ray coronary angiograms. Pattern Recogn. **87**, 38–54 (2019)
26. Wang, A., Lai, Z., Jin, Z.: Noisy low-tubal-rank tensor completion. Neurocomputing **330**, 267–279 (2019)
27. Yang, L., Huang, Z.H., Hu, S., Han, J.: An iterative algorithm for third-order tensor multi-rank minimization. Comput. Optim. Appl. **63**, 169–202 (2016)
28. Zhang, C., Hu, W., Jin, T., et al.: Nonlocal image denoising via adaptive tensor nuclear norm minimization. Neural Comput. Appl. **29**, 3–19 (2018)
29. Zhang, Z., Ely, G., Aeron, S., Hao, N., Kilmer, M.: Novel methods for multilinear data completion and de-noising based on tensor-SVD, pp. 3842–3849. IEEE, CVPR (2014)
30. Zhang, Z., Aeron, S.: Exact tensor completion using t-SVD. IEEE Trans. Signal Process. **65**, 1511–1526 (2017)

31. Zhang, Z.: A novel algebraic framework for processing multidimensional data: theory and application. Tufts University, Ph.D thesis, (2017)
32. Kilmer, M.E., Martin, C.D.: Factorization strategies for third-order tensors. Linear Algebra Appl. **435**, 641–658 (2011)
33. Braman, K.: Third-order tensors as linear operators on a space of matrices. Linear Algebra Appl. **433**, 1241–1253 (2010)
34. Miao, Y., Qi, L., Wei, Y.M.: T-Jordan canonical form and T-Drazin inverse based on the T-Product. Commun. Appl. Math. Comput. (2020). https://doi.org/10.1007/s42967-019-00055-4
35. Lund, K.: The tensor t-function: a definition for functions of third-order tensors. Numer Linear Algebra Appl. **27**, e2288 (2020)
36. Newman, E., Horesh, L., Avron, H., Kilmer, M.: Stable tensor neural networks for rapid deep learning. arXiv preprint arXiv:1811.06569 (2018)
37. Miao, Y., Qi, L., Wei, Y.: Generalized tensor function via the tensor singular value decomposition based on the T-product. Linear Algebra Appl. **590**, 258–303 (2020)
38. Bhatia, R.: Positive Definite Matrices. Princeton University Press, Princeton (2009)
39. Horn, A.R., Johnson, C.R.: Matrix Analysis, 2nd edn. Cambridge University Press, Cambridge (2013)
40. Sturm, J.F.: Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. Optim. Method. Softw. **11**, 625–653 (1999)
41. Toh, K.C., Todd, M.J., Tutuncu, R.H.: SDPT3-a Matlab software package for semidefinite programming. Optim. Method. Softw. **11**, 545–581 (1999)
42. Sun, D.F., Toh, K.C., Yuan, Y.C., Zhao, X.Y.: SDPNAL+: A Matlab software for semidefinite programming with bound constraints. Optim. Method. Softw. (version 1.0) **35**, 87–115 (2020)
43. Helmberg, C.: Semidefinite programming for combinatorial optimization. Technical Report 00-34, Konrad-Zuse-Zentrum für Informationstechink Berlin, Takustraße 7, D-14195 Berlin, Germany, (2000)
44. Todd, M.J.: Semidefinite optimization. Acta Numer. **10**, 515–560 (2001)
45. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**, 49–95 (1996)
46. Coleman, R.: Calculus on Normed Vector Spaces, Universitext. Springer, New York (2012)
47. Qi, L., Luo, Z.Y.: Tensor analysis: spectral theory and special tensors, Society for Industrial and Applied Mathematics, Philadelphia, (2017)
48. Golub, G.H., Van Loan, C.F.: Matrix Computations, 4th edn. Johns Hopkins University Press, Baltimore (2013)
49. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2011)
50. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization, Analysis, Algorithms, and Engineering Applications. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, (2001)
51. Nie, J.: Symmetric tensor nuclear norms. SIAM J. Appl. Algebra Geometry **1**, 599–625 (2017)
52. Friedland, S., Lim, L.H.: Nuclear norm of higher-order tensors. Math. Comput. **87**, 1255–1281 (2018)
53. Shor, N.Z.: Dual quadratic estimates in polynomial and Boolean programming. Ann. Oper. Res. **25**, 163–168 (1990)
54. Poljak, S., Rendl, F., Wolkowicz, H.: A recipe for semidefinite relaxation for $(0, 1)$-quadratic programming. J. Global Optim. **7**, 51–73 (1995)
55. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Math. Program. **96**, 293–320 (2003)
56. Nie, J., Wang, L.: Regularization methods for SDP relaxations in large-scale polynomial optimization. SIAM J. Matrix Anal. Appl. **22**, 408–428 (2012)