CONTINUITY, DIFFERENTIABILITY AND SEMISMoothNESS OF GENERALIZED TENSOR FUNCTIONS

XIA LI
School of Mathematics, Tianjin University
Tianjin 300350, China

YONG WANG* and ZHENG-HAI HUANG
School of Mathematics, Tianjin University
Tianjin 300350, China

(Communicated by Liqun Qi)

Abstract. A large number of real-world problems can be transformed into mathematical problems by means of third-order real tensors. Recently, as an extension of the generalized matrix function, the generalized tensor function over the third-order real tensor space was introduced with the aid of a scalar function based on the T-product for third-order tensors and the tensor singular value decomposition; and some useful algebraic properties of the function were investigated. In this paper, we show that the generalized tensor function can inherit a lot of good properties from the associated scalar function, including continuity, directional differentiability, Fréchet differentiability, Lipschitz continuity and semismoothness. These properties provide an important theoretical basis for the studies of various mathematical problems with generalized tensor functions, and particularly, for the studies of tensor optimization problems with generalized tensor functions.

1. Introduction. Matrix optimization has been studied extensively during the past two decades because a lot of real-world problems can be modeled as matrix optimization problems, such as positive semidefinite programming, positive semidefinite complementarity problem, matrix completion, matrix rank minimization and matrix norm minimization. Several specific matrix functions were used in these matrix optimization problems. With the development of matrix optimization, the theory and algorithms of matrix optimization problems with general matrix functions have been investigated, in which different matrix functions were defined, such as that defined by Jordan canonical form [17], by spectral decomposition [6], by singular value decomposition [16] or others [31]. Lots of useful properties of matrix functions have been achieved. For example, Chen et al. [9] showed that a class of symmetric matrix functions induced by a scalar function inherits some properties from the associated scalar function, including continuity, differentiability and semismoothness. Yang [42] expanded the results obtained in [9] to the case of the

2020 Mathematics Subject Classification. Primary: 15A69, 49J50, 49J52.

Key words and phrases. T-product, tensor singular value decomposition, generalized tensor function, semismoothness, differentiability.

* Corresponding author: Yong Wang.
generalized matrix function defined in [16]. More discussions about matrix functions can be found in [1, 3, 4, 5, 18, 30, 34]. All these properties provide important theoretical and algorithmic basis for the studies of matrix optimization problems with matrix functions.

In the era of big data, problems in reality are becoming more and more sophisticated. It is not enough to use matrices to describe data from real-world problems. The tensor, as an extension of the matrix, has become an efficient tool to describe complicated data from real-world problems. In the last decade, a lot of mathematical problems using tensors have drawn widespread attention, including tensor optimization problems, such as tensor completion, tensor decomposition, tensor norm minimization and low rank tensor recovery (see [12, 13, 22, 33, 40, 41, 44, 45] and the references therein). Several specific tensor functions were used in these tensor optimization problems. Recently, a few classes of general tensor functions over the space of third-order real tensors were introduced and some basic properties of these tensor functions were investigated. With the development of the research on mathematical problems modeled by tensors, it is necessary to further investigate other properties of general tensor functions.

The class of third-order real tensors is an important class of tensors and a lot of problems can be described by using third-order real tensors, such as 3-person noncooperative game [19], color images [22] and grey-scale videos. Recently, a useful tool to study third-order tensors, called the ‘tensor T-product’, has been developed [7, 21]. According to [7, 21], each third-order tensor can be uniquely reconstructed as a block-circulant matrix. What is more, block-circulant matrices can be diagonalizable by the fast Fourier Transformation (FFT) [8], which leads to rapid computing performance. Then, by the tensor T-product, the multiplication of two third-order tensors can be effectively dealt with to obtain a new third-order tensor. With these advantages, the tensor T-product has been used in many fields, such as computer vision [15, 36, 43], data completion [46, 47, 48], image processing [20, 26, 38], low rank minimization and robust tensor PCA [23, 24, 37, 39].

Several kinds of tensor functions defined by using tensor T-product have been investigated. For example, Lund [25] introduced the tensor T-function for frontal-square third-order tensors and showed that the tensor T-function has similar properties to matrix functions in many cases. Newman et al. [29] used a class of tensor functions, defined by applying the tube-wise of a tensor, in tensor neural networks to study rapid deep learning. Miao et al. [28] recently studied the tensor similarity, proposed the T-Jordan canonical form and discussed its properties; besides, the same group of authors [27] also extended the generalized matrix function to the generalized tensor function via the tensor singular value decomposition and discussed some algebraic properties of the function, such as the properties about the generalized tensor powers and some structure properties preserved by generalized tensor functions. However, it should be mentioned that the information we have gotten about the properties of these tensor functions is very limited. More useful properties of these functions need to be detected in order to study various mathematical problems using these functions.

It is known that for a function, (Lipschitz) continuity, (directional) differentiability and semismoothness, are basic properties, which are useful for the studies for various mathematical problems with such functions, especially for optimization problems with such functions. In this paper, inspired by the studies of the related
properties for matrix functions in [9, 42], we investigate continuity, differentiability, directional differentiability, continuous differentiability, Lipschitz continuity and semismoothness of the generalized tensor functions proposed in [27]. All these properties provide an important foundation of theoretical research and algorithm design for tensor optimization problems and other tensor applications with generalized tensor functions.

The structure of this paper is as follows. In Section 2, we recall some conclusions of the matrix singular value decomposition, the tensor T-product and the generalized tensor function for our subsequent studies. In Section 3, we show that the generalized tensor function induced by a scalar function inherits continuity, directional differentiability, Fréchet differentiability, continuous differentiability, Lipschitz continuity and semismoothness from the associated scalar function. Conclusions are given in Section 4.

In this paper, we will use the following notations. Without special explanations, we use small letters $a, b, c, \ldots$ to represent scalars, capital letters $A, B, C, \ldots$ to represent matrices and calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$ to represent third-order tensors. We always assume that $i, j, k, l, m, n, p, q$ are positive integers. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. $\mathbb{R}, \mathbb{R}_+, \mathbb{C}, \mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ stand for the sets of real numbers, nonnegative real numbers, complex numbers, $m \times n$ real matrices and $m \times n$ complex matrices, respectively. For a block-diagonal matrix $A \in \mathbb{R}^{mp \times np}$ with $p$ diagonal blocks, we use $A_i \in \mathbb{R}^{m \times n}$ to represent the $i$-th diagonal block, then we simply denote it by $A = \text{diag}(A_1^T, A_2^T, \ldots, A_p^T)$. In particular, when $A_i = a_i \in \mathbb{R}$, $A = \text{diag}(a_1, a_2, \ldots, a_p)$ is a $p \times p$ diagonal matrix. “$\otimes$” represents the Kronecker product, i.e., for matrices $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ and $B = (b_{kl}) \in \mathbb{R}^{p \times q}$,

$$A \otimes B := \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp \times nq}.$$  

Denote

$$F_{p,n} := F_p \otimes I_n,$$  

(1)

where $I_n$ is the $n \times n$ identity matrix and $F_p$ is the discrete Fourier transform matrix of size $p \times p$, which is defined as

$$F_p := \frac{1}{\sqrt{p}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{p-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(p-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{p-1} & \omega^{2(p-1)} & \cdots & \omega^{(p-1)(p-1)} \end{bmatrix},$$

where $\omega = e^{-\frac{2\pi i}{p}}$ with $i^2 = -1$. A third-order $(m \times n \times p)$-dimensional real tensor $X$ can be written as $X = (x_{ijk})$, where $x_{ijk} \in \mathbb{R}$ for any $i \in [m], j \in [n], k \in [p]$. The set of all third-order $(m \times n \times p)$-dimensional real tensors is denoted by $\mathbb{R}^{m \times n \times p}$. For any $A \in \mathbb{R}^{m \times n \times p}$, we always assume that $m \geq n$ throughout this paper. Furthermore, we use $A^T$ and $A^H$ to denote the transpose and conjugate transpose of $A$, respectively. Additionally, we denote a zero matrix by $O$ and a zero tensor by $\mathcal{O}$. “$\otimes$” represents the Hadamard product, i.e., for two $(m \times n \times p)$-dimensional real tensors $X = (x_{ijk})$ and $Y = (y_{ijk})$,

$$X \circ Y := (x_{ijk}y_{ijk}) \in \mathbb{R}^{m \times n \times p}.$$
Without special explanations, \( \| \cdot \| \) denotes the Frobenius norm or the operation norm. Let \( \mathbb{K} \) denote one of \( \mathbb{R}^{m \times n} \), \( \mathbb{C}^{m \times n} \) and \( \mathbb{R}^{m \times n \times p} \), then for any \( \xi \in \mathbb{K} \) and \( \epsilon > 0 \), \( \mathbb{B}(\xi, \epsilon) \) stands for the closed ball \( \mathbb{B}(\xi, \epsilon) := \{ \zeta \in \mathbb{K} : \| \zeta - \xi \| \leq \epsilon \} \). Besides, \( \xi = o(t) \) means that \( \frac{\| \xi \|}{t} \to 0 \) as \( t \to 0 \) and \( \xi = O(t) \) means that \( \frac{\| \xi \|}{t} \) is uniformly bounded as \( t \to 0 \).

2. Preliminary. Now let us review some definitions and results about some concepts for our subsequent studies.

2.1. The matrix singular value decomposition. From matrix analysis, one knows that each rectangular matrix has a singular value decomposition (SVD). In the following, we list some essential results that would be used in this paper.

**Proposition 1.** [14, Theorem 2.4.1(or Section 2.4.4)] If \( A \in \mathbb{R}^{m \times n} \) (or \( \mathbb{C}^{m \times n} \)), then there exist orthogonal (or unitary) matrices \( U \in \mathbb{R}^{m \times m} \) (or \( \mathbb{C}^{m \times m} \)) and \( V \in \mathbb{R}^{n \times n} \) (or \( \mathbb{C}^{n \times n} \)) such that
\[
U^T AV (\text{or } U^H AV) = \left( \begin{array}{c} \text{diag}(\sigma_1, \ldots, \sigma_n) \\ O \end{array} \right) \in \mathbb{R}^{m \times n},
\]
where \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0 \).

**Lemma 2.1.** [14, Corollary 8.6.2] If \( A \) and \( A + E \) are in \( \mathbb{R}^{m \times n} \) with \( m \geq n \), then for any \( k \in [n] \),
\[
|\sigma_k(A + E) - \sigma_k(A)| \leq \sigma_1(E),
\]
where \( \sigma_k(M) \) denotes the \( k \)-th largest singular value of the matrix \( M \).

**Corollary 1.** If \( A \) and \( A + E \) are in \( \mathbb{R}^{m \times n} \) with \( m \geq n \), then for any \( k \in [n] \),
\[
|\sigma_k(A + E) - \sigma_k(A)| \leq \| E \|.
\]

**Proof.** Since the Frobenius norm of the matrix \( E \) is \( \| E \| = \sqrt{\sum_{k=1}^n \sigma_k^2(E)} \), it is obvious that \( \sigma_1(E) \leq \| E \| \). So, the result holds. \( \square \)

Define the set
\[
\mathbb{D} := \{ D \in \mathbb{R}^{n \times n} : D \text{ is a diagonal matrix} \}.
\]
and the set
\[
\bar{\mathbb{D}} := \{ \bar{D} \in \mathbb{R}^{m \times n} : \bar{D} = (D \ O)^\top \text{ and } D \in \mathbb{D} \}.
\]
Let \( \mathbb{O} \) denote the space of orthogonal matrices. For each real symmetric matrix \( X \in \mathbb{R}^{n \times n} \), we define
\[
\mathbb{O}_X := \{ P \in \mathbb{O} : P^\top XP \in \mathbb{D} \text{ with nonincreasing diagonal entries} \},
\]
i.e., the set of orthogonal matrices whose columns consist of the eigenvectors of \( X \).

**Lemma 2.2.** [10, Lemma 3] For any symmetric real matrix \( X \in \mathbb{R}^{n \times n} \), there exist scalars \( \eta, \epsilon > 0 \) such that
\[
\min_{P \in \mathbb{O}_X} \| P - Q \| \leq \eta \| X - Y \|, \quad \forall Y \in \mathbb{B}(X, \epsilon), \quad \forall Q \in \mathbb{O}_Y.
\]

**Proposition 2.** Let \( A \in \mathbb{R}^{m \times n} \) and \( A = U_A \Sigma_A V_A^\top \) be an SVD of \( A \), then there exist scalars \( \eta_1, \eta_2, \epsilon > 0 \) such that
\[
\min_{U_A} \| U_A - U_B \| \leq \eta_1 \| A - B \| \quad \text{and} \quad \min_{V_A} \| V_A - V_B \| \leq \eta_2 \| A - B \|
\]
hold for any matrix \( B \in \mathbb{B}(A, \epsilon) \) with \( B = U_B \Sigma_B V_B^\top \) being an SVD of \( B \).
Proof. Since \( A = U_A \Sigma_A V_A^T \) is an SVD of \( A \), we have that \( AA^\top = U_A(\Sigma_A \Sigma_A^\top)U_A^\top \) is an SVD of \( AA^\top \). Since \( B \in \mathcal{B}(A, \epsilon) \), it is trivial that there exists a scalar \( \epsilon' \) depending on \( \epsilon \) such that \( BB^\top \in \mathcal{B}(AA^\top, \epsilon') \). Then by Lemma 2.2, there exists a scalar \( \eta > 0 \) such that

\[
\min_{U_A} \| U_A - U_B \| \leq \eta \| AA^\top - BB^\top \|. \tag{4}
\]

Let \( H := A - B \), then

\[
\| AA^\top - BB^\top \| = \| AA^\top - (A - H)(A - H)^\top \| = \| AH^\top + HA^\top - HH^\top \| \\
\leq 2 \| A \| \| H \| + \| H \|^2.
\]

This, together with (4), implies that

\[
\min_{U_A} \| U_A - U_B \| \leq \eta (2 \| A \| \| H \| + \| H \|^2).
\]

Since for any given \( A \), \( \| A \| \) is a constant and \( \| H \| \leq \epsilon \), there exists a scalar \( \eta_1 > 0 \) such that

\[
\min_{U_A} \| U_A - U_B \| \leq \eta_1 \| A - B \|.
\]

Similarly, it holds that

\[
\min_{V_A} \| V_A - V_B \| \leq \eta_2 \| A - B \|.
\]

Hence we obtain the conclusions. \( \square \)

**Proposition 3.** Let \( A \in \mathbb{C}^{m \times n} \) and \( A = U_A \Sigma_A V_A^H \) be an SVD of \( A \), then there exist scalars \( \eta_1, \eta_2, \epsilon > 0 \) such that

\[
\min_{U_A} \| U_A - U_B \| \leq \eta_1 \| A - B \| \quad \text{and} \quad \min_{V_A} \| V_A - V_B \| \leq \eta_2 \| A - B \|
\]

hold for any matrix \( B \in \mathcal{B}(A, \epsilon) \) with \( B = U_B \Sigma_B V_B^H \) being an SVD of \( B \).

Proof. Suppose \( A = A_1 + iA_2, \ U_A = U_{A1} + iU_{A2}, \text{ and } V_A = V_{A1} + iV_{A2} \), where \( i^2 = -1 \) and \( A_k, U_k, V_k \) \((k = 1, 2)\) are real matrices. Then we can transform \( A \) to a real matrix \( \hat{A} \) defined by

\[
\hat{A} := \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix}.
\]

Furthermore,

\[
\hat{A} = \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} = \begin{bmatrix} U_{A1} & U_{A2} \\ U_{A2} & U_{A1} \end{bmatrix} \begin{bmatrix} \Sigma_A & 0 \\ 0 & \Sigma_A \end{bmatrix} \begin{bmatrix} V_{A1} & -V_{A2} \\ V_{A2} & V_{A1} \end{bmatrix}^\top = \hat{U}_A \hat{\Sigma}_A \hat{V}_A^\top.
\]

For any \( B \in \mathcal{B}(A, \epsilon) \), we can also assume that \( B = B_1 + iB_2, \ U_B = U_{B1} + iU_{B2}, \text{ and } V_B = V_{B1} + iV_{B2} \), where \( B_k, \ U_{Bk}, \text{ and } V_{Bk} \) \((k = 1, 2)\) are also real matrices. Similar to \( \hat{A} \), we can obtain the decomposition \( \hat{B} = \hat{U}_B \hat{\Sigma}_B \hat{V}_B^\top \). For \( \hat{U}_A, \hat{U}_B, \hat{V}_A, \hat{V}_B \), there exist permutation matrices \( P \) and \( Q \) such that the diagonal entries of \( P^\top \hat{U}_A^\top \hat{A} V_A Q \) and \( P^\top \hat{U}_B^\top \hat{B} V_B Q \) are in nonincreasing order. By Proposition 2, there exist scalars \( \hat{\eta}_1, \hat{\eta}_2 > 0 \) such that

\[
\min_{U_A} \| \hat{U}_A - \hat{U}_B \| = \min_{U_A} \| \hat{U}_A P - \hat{U}_B P \| \leq \hat{\eta}_1 \| \hat{A} - \hat{B} \|,
\]

\[
\min_{V_A} \| \hat{V}_A - \hat{V}_B \| = \min_{V_A} \| \hat{V}_A Q - \hat{V}_B Q \| \leq \hat{\eta}_2 \| \hat{A} - \hat{B} \|
\]

hold for any \( \hat{B} \in \mathcal{B}(\hat{A}, \hat{\epsilon}) \), where \( \hat{\epsilon} \) depends on \( \epsilon \).
Besides, we have
\[ \|A - B\|^2 = \|A_1 - B_1 + i(A_2 - B_2)\|^2 = \|A_1 - B_1\|^2 + \|A_2 - B_2\|^2 \]
and
\[ \|\hat{A} - \hat{B}\|^2 = \left\| \begin{array}{ccc} A_1 & -B_1 \\ A_2 & B_2 \end{array} \right\|^2 = 2\|A_1 - B_1\|^2 + 2\|A_2 - B_2\|^2, \]
which deduce that \( \|\hat{A} - \hat{B}\|^2 = 2\|A - B\|^2 \). With similar discussions, we can obtain
\[ \|\hat{U}_A - \hat{U}_B\|^2 = 2\|U_A - U_B\|^2 \] and \( \|\hat{V}_A - \hat{V}_B\|^2 = 2\|V_A - V_B\|^2 \). Then there exist \( \eta_1 = \hat{\eta}_1, \eta_2 = \hat{\eta}_2, \epsilon > 0 \) such that
\[ \min_{\hat{U}_A} \|U_A - U_B\| = \frac{1}{\sqrt{2}} \min_{\hat{U}_A} \|\hat{U}_A - \hat{U}_B\| \leq \frac{1}{\sqrt{2}} \hat{\eta}_1 \|\hat{A} - \hat{B}\| = \eta_1 \|A - B\|, \]
\[ \min_{\hat{V}_A} \|V_A - V_B\| = \frac{1}{\sqrt{2}} \min_{\hat{V}_A} \|\hat{V}_A - \hat{V}_B\| \leq \frac{1}{\sqrt{2}} \hat{\eta}_2 \|\hat{A} - \hat{B}\| = \eta_2 \|A - B\| \]
hold for any \( B \in B(A, \epsilon) \). Therefore, we obtain the conclusions.

2.2. Tensor T-product and generalized tensor functions. In the following, we use \( X(k) \in \mathbb{R}^{m \times n \times p} \) to denote the frontal slices of \( X \in \mathbb{R}^{n \times n \times p} \) for all \( k \in [p] \). \( X \) is said to be a frontal-diagonal tensor if each \( X(k) \) \( k \in [p] \) is diagonal and a frontal-square tensor if each \( X(k) \) \( k \in [p] \) is square. The operations ‘bcirc’, ‘unfold’ and ‘fold’ defined in [7, 20, 21] are given as follows:
\[
\text{bcirc}(X) := \left[ \begin{array}{cccc} X^{(1)} & X^{(p)} & X^{(p-1)} & \cdots & X^{(2)} \\ X^{(2)} & X^{(1)} & X^{(p)} & \cdots & X^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X^{(p)} & X^{(p-1)} & X^{(p-2)} & \cdots & X^{(1)} \end{array} \right], \text{unfold}(X) := \left[ \begin{array}{c} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(p)} \end{array} \right],
\]
and fold(unfold(X)) := X. Besides, the operation ‘bcirc\(^{-1}\)’ denotes the inverse operation of bcirc [27], i.e.,
\[
\text{bcirc}^{-1}(\text{bcirc}(X)) := X.
\]

**Definition 2.3.** [21, Definition 3.1] Let \( X \in \mathbb{R}^{m \times n \times p} \) and \( Y \in \mathbb{R}^{n \times l \times p} \). Then the T-product \( X \ast Y \) is an \((m \times l \times p)\)-dimensional tensor, which is given by
\[
X \ast Y := \text{fold(bcirc}(X)\text{unfold}(Y))\).
\]

**Definition 2.4.** [21, Definition 3.14] Let \( X \in \mathbb{R}^{m \times n \times p} \), then \( X^T \) is the tensor in \( \mathbb{R}^{n \times m \times p} \) obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through \( p \), i.e.,
\[
X^T := \text{fold}([X^{(1)}, X^{(p)}, X^{(p-1)}, \ldots, X^{(2)}]^T).
\]

**Definition 2.5.** [21, Definition 3.4] The \((n \times n \times p)\)-dimensional identity tensor, denoted by \( I_{nmp} \) (\( I \) for simpleness), is the tensor whose first frontal slice is the \( n \times n \) identity matrix and other frontal slices are all zero matrices.

**Definition 2.6.** [21, Definition 3.18] Let \( X \in \mathbb{R}^{n \times n \times p} \), then \( X \) is an orthogonal tensor if \( X^T \ast X = X \ast X^T = I \).

We adopt the familiar definition of the Frobenius norm of a tensor used in the literature:
Definition 2.7. For any $\mathcal{X} = (x_{ijk}) \in \mathbb{R}^{m \times n \times p}$, define the Frobenius norm $\|\mathcal{X}\|$ of $\mathcal{X}$ as

$$\|\mathcal{X}\| := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} x_{ijk}^2}.$$ 

We say $\mathcal{X} \to \mathcal{Y}$ if and only if $\|\mathcal{X} - \mathcal{Y}\| \to 0$. Besides, it is easy to see that $\|\text{bcirc}(\mathcal{X})\| = \sqrt{\|\mathcal{X}\|}$, then $\mathcal{X} \to \mathcal{Y}$ if and only if $\|\text{bcirc}(\mathcal{X}) - \text{bcirc}(\mathcal{Y})\| \to 0$.

The following definition of the generalized matrix function comes from [16, 30].

Definition 2.8. For any scalar function $f : \mathbb{R} \to \mathbb{R}$, any matrix $A \in \mathbb{R}^{m \times n}$ with $A = U \Sigma V^\top$ being an SVD of $A$, where $\Sigma = \begin{pmatrix} D & O \\ O & 0 \end{pmatrix} \in \mathbb{D}$ and $D = \text{diag}(\sigma_1, \ldots, \sigma_n)$, the generalized matrix function $\bar{f} : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ induced by $f$ is defined as

$$\bar{f}(A) := U \bar{f}(\Sigma) V^\top,$$

where

$$\bar{f}(\Sigma) := \begin{pmatrix} \text{diag}(f(\sigma_1), f(\sigma_2), \ldots, f(\sigma_n)) \end{pmatrix} \in \mathbb{D}.$$ (5)

Lemma 2.9. [21, Theorem 4.1] Let $\mathcal{X} \in \mathbb{R}^{m \times n \times p}$, then $\mathcal{X}$ can be factored as

$$\mathcal{X} = \mathcal{U} \ast \mathcal{S} \ast \mathcal{V}^\top,$$

where $\mathcal{U} \in \mathbb{R}^{m \times n \times p}$ and $\mathcal{V} \in \mathbb{R}^{m \times n \times p}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{m \times n \times p}$ is a frontal-diagonal tensor. The factorization above is called the tensor singular value decomposition (T-SVD).

From the proof of [21, Theorem 4.1], we can obtain that $\text{bcirc}(\mathcal{X}) = F_{p,m} D F_{p,n}^H$, where $F_{p,m}(F_{p,n})$ is defined by (1), $D$ is a block-diagonal matrix with $p$ principal diagonal blocks $D^i$ and $D^i = U^i \Sigma^i (V^i)^H$ is an SVD of $D^i$. What is more, it follows that

$$\text{bcirc}(\mathcal{U}) = F_{p,m} \text{diag}(U^1, U^2, \ldots, U^p) F_{p,m}^H,$$

$$\text{bcirc}(\mathcal{S}) = F_{p,m} \text{diag}(\Sigma^1, \Sigma^2, \ldots, \Sigma^p) F_{p,n}^H,$$

$$\text{bcirc}(\mathcal{V}^\top) = F_{p,n} \text{diag}(V^1, V^2, \ldots, (V^i)^H) F_{p,n}^H.$$ (6)

Besides, the diagonal entries of $\Sigma^i$, i.e., $\sigma_j^i$, $i \in [p]$, $j \in [n]$, are called the singular values of $\mathcal{X}$. It is clear that the singular values of $\mathcal{X}$ are the singular values of $\text{bcirc}(\mathcal{X})$. The following definition is from [27, Theorem 1].

Definition 2.10. Suppose $\mathcal{X} \in \mathbb{R}^{m \times n \times p}$ and $\mathcal{X} = \mathcal{U} \ast \mathcal{S} \ast \mathcal{V}^\top$ is a T-SVD of $\mathcal{X}$, where $\mathcal{S} \in \mathbb{R}^{m \times n \times p}$ can be factored as (6). Let $f : \mathbb{R} \to \mathbb{R}$, then the generalized tensor function $\bar{f}^\circ : \mathbb{R}^{m \times n \times p} \to \mathbb{R}^{m \times n \times p}$ induced by $f$ is defined as

$$\bar{f}^\circ(\mathcal{X}) := \mathcal{U} \ast \bar{f}(\mathcal{S}) \ast \mathcal{V}^\top,$$

where $\bar{f}(\mathcal{S})$ is defined by

$$\bar{f}(\mathcal{S}) := \text{bcirc}^{-1} \begin{pmatrix} \bar{f}(\Sigma^1) \\ \bar{f}(\Sigma^2) \\ \vdots \\ \bar{f}(\Sigma^p) \end{pmatrix} F_{p,n}^H,$$

and $\bar{f}(\Sigma^i)$ are defined by (5) for all $\Sigma^i \in \mathbb{D}$ with $i \in [p]$. 

SEMIMOSSMOOTHNESS OF GENERALIZED TENSOR FUNCTIONS 7
Definition 2.11. Let \( f^\circ : \mathbb{R}^{m \times n \times p} \to \mathbb{R}^{m \times n \times p} \), then we say

(i) \( f^\circ \) is continuous at \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \) if

\[
\| f^\circ(\mathcal{Y}) - f^\circ(\mathcal{X}) \| \to 0 \quad \text{as} \quad \| \mathcal{Y} - \mathcal{X} \| \to 0;
\]

and \( f^\circ \) is continuous if \( f^\circ \) is continuous at every \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \).

(ii) \( f^\circ \) is locally Lipschitz continuous at \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \) if there exist scalars \( \eta > 0 \) and \( \delta > 0 \) such that

\[
\| f^\circ(\mathcal{Y}) - f^\circ(\mathcal{Z}) \| \leq \eta \| \mathcal{Y} - \mathcal{Z} \|
\]

for all \( \mathcal{Y}, \mathcal{Z} \in \mathbb{R}^{m \times n \times p} \) with \( \| \mathcal{Y} - \mathcal{X} \| \leq \delta \) and \( \| \mathcal{Z} - \mathcal{X} \| \leq \delta \); If \( \delta \) can be taken to \( +\infty \), we say \( f^\circ \) is Lipschitz continuous with Lipschitz constant \( \eta \).

(iii) \( f^\circ \) is directionally differentiable at \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \) if

\[
(f^\circ)'(\mathcal{X}; \mathcal{H}) := \lim_{t \in \mathbb{R}, t \to 0^+} \frac{f^\circ(\mathcal{X} + t\mathcal{H}) - f^\circ(\mathcal{X})}{t}
\]

exists for all \( \mathcal{H} \in \mathbb{R}^{m \times n \times p} \); and \( f^\circ \) is directionally differentiable if \( f^\circ \) is directionally differentiable at every \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \).

(iv) \( f^\circ \) is differentiable at \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \) (in Fréchet sense) if there exists a linear mapping \( \nabla f^\circ : \mathbb{R}^{m \times n \times p} \to \mathbb{R}^{m \times n \times p} \) such that

\[
f^\circ(\mathcal{X} + \mathcal{H}) - f^\circ(\mathcal{X}) - \nabla f^\circ(\mathcal{X})\mathcal{H} = o(\|\mathcal{H}\|),
\]

where \( \mathcal{H} \in \mathbb{R}^{m \times n \times p} \); and \( f^\circ \) is differentiable if \( f^\circ \) is differentiable at every \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \).

(v) \( f^\circ \) is continuously differentiable if \( f^\circ \) is differentiable and \( \nabla f^\circ \) is continuous, where \( \nabla f^\circ \) is defined in (iv).

If \( f^\circ \) is locally Lipschitz continuous, by Rademacher’s theorem (see [11] or [32]), then \( f^\circ \) is differentiable almost everywhere. Then the generalized Jacobian \( \partial f^\circ(\mathcal{X}) \) of \( f^\circ \) at \( \mathcal{X} \) in the Clarke sense can be defined as

\[
\partial f^\circ(\mathcal{X}) := \text{conv}(\partial_B f^\circ(\mathcal{X})),
\]

where \( \text{conv}(\mathbb{K}) \) means the convex hull of set \( \mathbb{K} \), \( \partial_B f^\circ(\mathcal{X}) \) is the generalized Jacobian of \( f^\circ \) at \( \mathcal{X} \) in the Bouligand sense defined as

\[
\partial_B f^\circ(\mathcal{X}) := \left\{ \lim_{\mathcal{X}^\nu \to \mathcal{X}} \nabla f^\circ(\mathcal{X}^\nu) : f^\circ \text{ is differentiable at } \mathcal{X}^\nu \in \mathbb{R}^{m \times n \times p} \right\}.
\]

Definition 2.12. Suppose \( f^\circ : \mathbb{R}^{m \times n \times p} \to \mathbb{R}^{m \times n \times p} \) is locally Lipschitz continuous. \( f^\circ \) is called to be semismooth at \( \mathcal{X} \) if \( f^\circ \) is directionally differentiable at \( \mathcal{X} \) and for any \( \mathcal{J} \in \partial f^\circ(\mathcal{X} + \mathcal{H}) \),

\[
f^\circ(\mathcal{X} + \mathcal{H}) - f^\circ(\mathcal{X}) - \mathcal{J}\mathcal{H} = o(\|\mathcal{H}\|).
\]

\( f^\circ \) is called to be \( \gamma \)-order semismooth (\( \gamma \in (0, \infty) \)) at \( \mathcal{X} \) if \( f^\circ \) is semismooth at \( \mathcal{X} \) and for any \( \mathcal{J} \in \partial f^\circ(\mathcal{X} + \mathcal{H}) \),

\[
f^\circ(\mathcal{X} + \mathcal{H}) - f^\circ(\mathcal{X}) - \mathcal{J}\mathcal{H} = O(\|\mathcal{H}\|^{1+\gamma}).
\]

\( f^\circ \) is said to be semismooth (or \( \gamma \)-order semismooth) if \( f^\circ \) is semismooth (or \( \gamma \)-order semismooth) at every \( \mathcal{X} \in \mathbb{R}^{m \times n \times p} \).
3. Continuity, differentiability and semismoothness of generalized tensor functions. In the studies of the generalized matrix function induced by a scalar function $f$, the condition $f(0) = 0$ is necessary to guarantee the well-definedness of the generalized matrix function (see [2, Page 1869], [30, Page 438] or [42, Page 5]). Similarly, the condition $f(0) = 0$ is also necessary for the well-definedness of $f^o$. Therefore, we always assume that $f(0) = 0$ in the rest of this paper.

**Lemma 3.1.** Let $\mathcal{X} \in \mathbb{R}^{m \times n \times p}$ and $\mathcal{X} = \mathcal{U} \mathcal{X} \ast \mathcal{S} \mathcal{X} \ast \mathcal{V}_\mathcal{X}^\top$ be a T-SVD of $\mathcal{X}$, then there exist scalars $\eta_1, \eta_2, \epsilon > 0$ such that

$$\min_{\mathcal{U}_\mathcal{X}} \| \mathcal{U}_\mathcal{X} - \mathcal{U}_\mathcal{Y} \| \leq \eta_1 \| \mathcal{X} - \mathcal{Y} \| \quad \text{and} \quad \min_{\mathcal{V}_\mathcal{X}} \| \mathcal{V}_\mathcal{X} - \mathcal{V}_\mathcal{Y} \| \leq \eta_2 \| \mathcal{X} - \mathcal{Y} \|$$

hold for any $\mathcal{Y} \in \mathcal{B}(\mathcal{X}, \epsilon)$ with $\mathcal{Y} = \mathcal{U}_\mathcal{Y} \ast \mathcal{S}_\mathcal{Y} \ast \mathcal{V}_\mathcal{Y}^\top$ being a T-SVD of $\mathcal{Y}$.

**Proof.** By Lemma 2.9, we can obtain that $F_{p,m}^H \text{bcirc}(\mathcal{X})F_{p,n} = D = U \Sigma V^H$ and $F_{p,m}^H \text{bcirc}(\mathcal{Y})F_{p,n} = \tilde{D} = \tilde{U} \tilde{\Sigma} \tilde{V}^H$, where

$$D = \text{diag}(D^1, D^2, \ldots, D^p), \quad \tilde{D} = \text{diag}(\tilde{D}^1, \tilde{D}^2, \ldots, \tilde{D}^p),$$

$$U = \text{diag}(U^1, U^2, \ldots, U^p), \quad \tilde{U} = \text{diag}(\tilde{U}^1, \tilde{U}^2, \ldots, \tilde{U}^p),$$

$$V = \text{diag}(V^1, V^2, \ldots, V^p), \quad \tilde{V} = \text{diag}(\tilde{V}^1, \tilde{V}^2, \ldots, \tilde{V}^p),$$

$$\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_p), \quad \tilde{\Sigma} = \text{diag}(\tilde{\Sigma}_1, \tilde{\Sigma}_2, \ldots, \tilde{\Sigma}_p).$$

In other words, $D$ is a block-diagonal matrix and $D^i = U^i \Sigma^i (V^i)^H$ is an SVD of the block $D^i$. Similarly, $\tilde{D} = \tilde{U}^i \tilde{\Sigma}^i (\tilde{V}^i)^H$. By Proposition 3, we have

$$\min_{U^i} \| U^i - \tilde{U}^i \| \leq \eta_1^i \| D^i - \tilde{D}^i \| \quad \text{and} \quad \min_{V^i} \| V^i - \tilde{V}^i \| \leq \eta_2^i \| D^i - \tilde{D}^i \|.$$

Let $\eta_1 := \max_{i \in [p]} \{ \eta_1^i \}$ and $\eta_2 := \max_{i \in [p]} \{ \eta_2^i \}$. Then by taking summation, it follows that

$$\min_{\mathcal{U}} \| \mathcal{U} - \tilde{\mathcal{U}} \|^2 = \sum_{i=1}^p \min_{U^i} \| U^i - \tilde{U}^i \|^2 \leq \sum_{i=1}^p (\eta_1^i)^2 \| D^i - \tilde{D}^i \|^2 \leq (\eta_1)^2 \| D - \tilde{D} \|^2,$$

$$\min_{\mathcal{V}} \| \mathcal{V} - \tilde{\mathcal{V}} \|^2 = \sum_{i=1}^p \min_{V^i} \| V^i - \tilde{V}^i \|^2 \leq \sum_{i=1}^p (\eta_2^i)^2 \| D^i - \tilde{D}^i \|^2 \leq (\eta_2)^2 \| D - \tilde{D} \|^2.$$

Besides, since

$$\| D - \tilde{D} \| = \| F_{p,m}^H \text{bcirc}(\mathcal{X})F_{p,n} - F_{p,m}^H \text{bcirc}(\mathcal{Y})F_{p,n} \| = \sqrt{p} \| \mathcal{X} - \mathcal{Y} \|,$$

$$\| U - \tilde{U} \| = \| F_{p,m}^H \text{bcirc}(\mathcal{U} \mathcal{X})F_{p,n} - F_{p,m}^H \text{bcirc}(\mathcal{U} \mathcal{Y})F_{p,n} \| = \sqrt{p} \| \mathcal{U} \mathcal{X} - \mathcal{U} \mathcal{Y} \|,$$

$$\| V - \tilde{V} \| = \| F_{p,m}^H \text{bcirc}(\mathcal{V} \mathcal{X})F_{p,n} - F_{p,m}^H \text{bcirc}(\mathcal{V} \mathcal{Y})F_{p,n} \| = \sqrt{p} \| \mathcal{V} \mathcal{X} - \mathcal{V} \mathcal{Y} \|,$$

we can deduce that

$$\min_{\mathcal{U}_\mathcal{X}} \| \mathcal{U}_\mathcal{X} - \mathcal{U}_\mathcal{Y} \| \leq \eta_1 \| \mathcal{X} - \mathcal{Y} \| \quad \text{and} \quad \min_{\mathcal{V}_\mathcal{X}} \| \mathcal{V}_\mathcal{X} - \mathcal{V}_\mathcal{Y} \| \leq \eta_2 \| \mathcal{X} - \mathcal{Y} \|.$$

We obtain the conclusions. \[ \square \]

**Lemma 3.2.** Let $A, B \in \mathbb{C}^{m \times n}$ and $A = U_A \Sigma_A V_A^H, B = U_B \Sigma_B V_B^H$ be the SVDs of $A, B$ respectively. If $B \rightarrow A$, then it follows that $U_B \rightarrow U_A$ and $V_B \rightarrow V_A$.

**Proof.** We first verify that the similar result holds for real matrices, and then extend it to the case of complex matrices.

Let $X, Y \in \mathbb{R}^{m \times n}$ and $X = U \Sigma V, Y = \tilde{U} \tilde{\Sigma} \tilde{V}$ be the SVDs of $X, Y$ respectively. In what follows, we prove that if $Y \rightarrow X$ then $\tilde{U} \rightarrow U$ and $\tilde{V} \rightarrow V$. Suppose $X$ has
singular values $\sigma_1, \ldots, \sigma_n$. Let $U = [U_1 \ U_2]$, where $U_1 \in \mathbb{R}^{m \times n}$ and $U_2 \in \mathbb{R}^{m \times (m-n)}$. We define $A_X, Q_X, S_X \in \mathbb{R}^{(m+n) \times (m+n)}$ by

$$
A_X := \text{diag}(\sigma_1, \ldots, \sigma_n, -\sigma_1, \ldots, -\sigma_n, 0, \ldots, 0),
$$

$$
Q_X := \frac{1}{\sqrt{2}} \begin{bmatrix} V & V \\ U_1 & -U_1 \end{bmatrix},
$$

$$
S_X := \begin{bmatrix} O & X^T \\ X & O \end{bmatrix}.
$$

Then, by simple derivations, we know that $S_X = Q_X A_X Q_X^T$. Similarly, for the matrix $Y$, we can define $A_Y, Q_Y, S_Y$ and it follows that $S_Y = Q_Y A_Y Q_Y^T$. It is clear that $Y \rightarrow X$ is equivalent to $S_Y \rightarrow S_X$. By Lemma 2.2, if $S_Y \rightarrow S_X$ then $Q_Y \rightarrow Q_X$. This implies that if $Y \rightarrow X$ then $\hat{U} \rightarrow U$ and $V \rightarrow V$.

Now we prove the desired result. From the proof of Proposition 3, we can transform the complex matrices $A, B$ to the real matrices $\hat{A}, \hat{B}$ respectively. Moreover, it follows that $\|\hat{A} - \hat{B}\|^2 = 2\|A - B\|^2$, which implies that $B \rightarrow A$ is equivalent to $\hat{B} \rightarrow \hat{A}$. Let $A = \hat{U}_A \Sigma_A \hat{V}_A^T, B = \hat{U}_B \Sigma_B \hat{V}_B^T$ be the SVDs of $\hat{A}, \hat{B}$ respectively. From the proof of Proposition 3, it follows that $\|\hat{U}_A - \hat{U}_B\|^2 = 2\|U_A - U_B\|^2$, which means that $U_B \rightarrow U_A$ is equivalent to $\hat{U}_B \rightarrow \hat{U}_A$. Similarly, we can get that $V_B \rightarrow V_A$ is equivalent to $\hat{V}_B \rightarrow \hat{V}_A$. For the real matrices $\hat{A}$ and $\hat{B}$, we have proved in the last paragraph that if $\hat{B} \rightarrow \hat{A}$ then $\hat{U}_B \rightarrow \hat{U}_A$ and $\hat{V}_B \rightarrow \hat{V}_A$. This, together with the discussions above, implies that if $B \rightarrow A$ then $U_B \rightarrow U_A$ and $V_B \rightarrow V_A$. 

\[\square\]

**Proposition 4.** Let $f : \mathbb{R} \rightarrow \mathbb{R}$ with $f(0) = 0$ and $f^*$ be defined by Definition 2.10. Then the following results hold:

(i) $f^*$ is continuous at $\mathcal{X}$ with singular values $\sigma_j^i$, $i \in [p], j \in [n]$ if and only if $f$ is continuous at all singular values $\sigma_j^i$;

(ii) $f^*$ is continuous if and only if $f$ is continuous on $\mathbb{R}_+$.

**Proof.** (i) Sufficiency: Suppose $f$ is continuous at all singular values $\sigma_j^i$, $i \in [p], j \in [n]$. For $\epsilon > 0$ and any $\mathcal{Y} \in \mathcal{B}(\mathcal{X}, \epsilon)$, suppose $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ is a T-SVD of $\mathcal{X}$ and $\mathcal{Y} = \hat{\mathcal{U}} * \hat{\mathcal{S}} * \hat{\mathcal{V}}^T$ is a T-SVD of $\mathcal{Y}$. Here by (6), we have

$$
S = \text{bcirc}^{-1}(F_{p,m} \text{diag}(\Sigma_1, \Sigma^2, \ldots, \Sigma^p) F_{p,n}^H)
$$

and

$$
\hat{S} = \text{bcirc}^{-1}(F_{p,m} \text{diag}(\Sigma_1, \Sigma^2, \ldots, \Sigma^p) F_{p,n}^H),
$$

where $\Sigma^i = \left( \begin{array}{c} \text{diag}(\sigma_1^i, \sigma_2^i, \ldots, \sigma_n^i) \\ O \end{array} \right), \Sigma^i = \left( \begin{array}{c} \text{diag}(\hat{\sigma}_1^i, \hat{\sigma}_2^i, \ldots, \hat{\sigma}_n^i) \\ O \end{array} \right), i \in [p]$. Besides, by (5) and (7), it follows that

$$
\|\hat{f}(\hat{S}) - \hat{f}(S)\|^2 = \sum_{i=1}^p \|\hat{f}(\Sigma^i) - \hat{f}(\Sigma^i)\|^2 = \sum_{i=1}^p \sum_{j=1}^n |f(\hat{\sigma}_j^i) - f(\sigma_j^i)|^2.
$$

Therefore, as $\mathcal{Y} \rightarrow \mathcal{X}$, by Corollary 1 and the discussions about singular values after Lemma 2.9, it is easy to see that $\hat{\sigma}_j^i \rightarrow \sigma_j^i$. Then by the continuity of $f$, we have $f(\hat{\sigma}_j^i) \rightarrow f(\sigma_j^i)$, and hence $\hat{f}(\hat{S}) \rightarrow \hat{f}(S)$. Denote

$$
D := \text{diag}(D^1, \ldots, D^p) = F_{p,m}^H \text{bcirc}(\mathcal{X}) F_{p,n},
$$

$$
\hat{D} := \text{diag}(\hat{D}^1, \ldots, \hat{D}^p) = F_{p,m}^H \text{bcirc}(\mathcal{Y}) F_{p,n}.
$$

Hence, it yields that $\hat{D}^i \rightarrow D^i$ for any $i \in [p]$ as $\hat{D} \rightarrow D$. Besides, it is easy to see that $\mathcal{Y} \rightarrow \mathcal{X}$ is equivalent to $\hat{D} \rightarrow D$. Suppose for any $i \in [p], \hat{D}^i = \hat{U}_i \Sigma^i \hat{V}_i$ and $D^i = U_i \Sigma^i V_i$ are the SVDs of $\hat{D}^i$ and $D^i$, respectively. Then by Lemma 3.2, for
any $i \in [p]$ it follows that $\tilde{U}^i \rightarrow U^i$ as $\tilde{D}^i \rightarrow D^i$. From the proof of Lemma 3.1, we can obtain that $p||U - \tilde{U}||^2 = \sum_{i=1}^{p} ||U^i - \tilde{U}^i||^2$. Then we can deduce $U \rightarrow U$ as $Y \rightarrow X$. Similarly, we can also get that $\tilde{V} \rightarrow V$ as $Y \rightarrow X$.

From the definition of $f^\circ$, it yields that

$$
\|f^\circ(Y) - f^\circ(X)\| = \|\tilde{U} * \hat{f}(\tilde{S}) * \tilde{V}^\top - U * \hat{f}(S) * V^\top\|
$$

$$
= \|\tilde{U} * \hat{f}(\tilde{S}) * \tilde{V}^\top - \tilde{U} * \hat{f}(\tilde{S}) * \tilde{V}^\top + \tilde{U} * \hat{f}(\tilde{S}) * \tilde{V}^\top - U * \hat{f}(S) * V^\top
$$

$$
+ \tilde{U} * \hat{f}(\tilde{S}) * V^\top - U * \hat{f}(S) * V^\top\|
$$

$$
= \|\tilde{U} * \hat{f}(\tilde{S}) * (\tilde{V} - V)^\top + \tilde{U} * (\hat{f}(\tilde{S}) - \hat{f}(S)) * V^\top + (\tilde{U} - U) * \hat{f}(S) * V^\top\|
$$

$$
\leq \|\tilde{U} * \hat{f}(\tilde{S}) * (\tilde{V} - V)^\top\| + \|\tilde{U} * (\hat{f}(\tilde{S}) - \hat{f}(S)) * V^\top\| + \|\tilde{U} - U\| \|\hat{f}(S)\| + \|\tilde{U} - U\| \|\hat{f}(S)\|
$$

$$
\leq \sqrt{\rho}||\hat{f}(\tilde{S})||\|\tilde{V} - V\| + \|\hat{f}(\tilde{S}) - \hat{f}(S)\| + \sqrt{\rho}||\tilde{U} - U\| \|\hat{f}(S)\|
$$

where the first inequality holds from the triangular inequality of the norm, the last equality holds from the orthogonality of $\tilde{U}, \tilde{V}$ (see [21, Lemma 3.19]) and the last inequality holds from $||A*B|| \leq \sqrt{\rho}||A||B||$ for any $A, B \in \mathbb{R}^{m \times n \times p}$. This, together with the discussions above, implies that $f^\circ(Y) \rightarrow f^\circ(X)$ as $Y \rightarrow X$. Therefore, $f^\circ$ is continuous at $X$.

Necessity: Suppose $f^\circ$ is continuous at $X$ and $X = U*S*V^\top$ is a T-SVD of $X$, where $S$ is defined by (6). Fixing any $i \in [p], j \in [n]$ and recalling $\overline{D}$ defined by (3), let

$$
\overline{\Sigma}^i := \begin{pmatrix}
\text{diag}(\overline{\sigma}_1^i, \overline{\sigma}_2^i, \ldots, \overline{\sigma}_n^i) \\
O
\end{pmatrix} \in \overline{D}
$$

with

$$
\overline{\sigma}_k^i = \begin{cases}
\mu_j^i, & \text{if } k = j; \\
\overline{\sigma}_k, & \text{if } k \in [n] \text{ and } k \neq j.
\end{cases}
$$

Let $\hat{S}$ be defined by

$$
beirc(\hat{S}) := F_{p,m}(\text{diag}(\Sigma^1, \ldots, \Sigma^{i-1}, \Sigma^i, \Sigma^{i+1}, \ldots, \Sigma^p))F_{p,m}^H,
$$

and $Y = U*S*V^\top$. As $Y \rightarrow X$, it follows that $\hat{S} \rightarrow S$, and further, $\mu_j^i \rightarrow \sigma_j^i$. By the continuity of $f^\circ$, we have that $f^\circ(Y) \rightarrow f^\circ(X)$ as $Y \rightarrow X$, and further, $f(\mu_j^i) \rightarrow f(\sigma_j^i)$ as $\mu_j^i \rightarrow \sigma_j^i$. Therefore, $f$ is continuous at $\sigma_j^i$. By the arbitrariness of $i$ and $j$, we can deduce that $f$ is continuous at all singular values of $X$.

(ii) The result can be easily obtained from (i). \qed

For any $\Sigma = \text{diag}(\mu_1, \ldots, \mu_n) \in \overline{D}$, we assume $f : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable at $\mu_j, j \in [n]$, then we define the function $g : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ as

$$
(g(\Sigma))_{jk} := \begin{cases}
\frac{f(\mu_j) - f(\mu_k)}{\mu_j - \mu_k}, & \text{if } \mu_j \neq \mu_k; \\
f'(\mu_j), & \text{if } \mu_j = \mu_k
\end{cases}
$$

for any $j, k \in [n]$, where $f'(\mu_j)$ is the derivative of $f$ at $\mu_j$. When the matrices involved are symmetric, the following result is given by[42, Theorem 2.2.6].

**Lemma 3.3.** Let $f : \mathbb{R} \rightarrow \mathbb{R}$ with $f(0) = 0$ and $\tilde{f}$ be defined by Definition 2.8. Suppose symmetric matrix $X \in \mathbb{R}^{n \times n}$ has singular values $\sigma_1, \ldots, \sigma_n$, and $X = U*\Sigma*U^\top$ is an SVD of $X$. Then $f$ is differentiable at $X$ with singular values $\sigma_1, \ldots, \sigma_n$ if and only if $f$ is differentiable at $\sigma_1, \ldots, \sigma_n$. Moreover, $\nabla \tilde{f}$ is given by

$$
\nabla \tilde{f}(X)H = U(g(\Sigma) \odot (U^\top H U))U^\top
$$
for any symmetric matrix $H \in \mathbb{R}^{n \times n}$.

**Proposition 5.** Suppose $\mathcal{X} \in \mathbb{R}^{n \times n \times p}$ with $\mathcal{X} = \mathcal{X}^{\top}$ has singular values $\sigma_j^i, i \in [p], j \in [n]$ and $\mathcal{X} = \mathcal{U} \ast S \ast \mathcal{U}^{\top}$ is a T-SVD of $\mathcal{X}$. Let $f : \mathbb{R} \to \mathbb{R}$ with $f(0) = 0$ and $f^\ast$ be defined by Definition 2.10, then the following results hold:

(i) $f^\ast$ is differentiable at $\mathcal{X}$ if and only if $f$ is differentiable at $\sigma_j^i, i \in [p], j \in [n]$. What is more, let $X := \text{bcirc}(\mathcal{X})$ and $X = U_X \Sigma_X U_X^\top$ be an SVD of $X$, then $\nabla f^\ast(X)$ is given by

$$\nabla f^\ast(X) = \text{bcirc}^{-1} \left( U_X \left( g(\Sigma_X) \odot (U_X^\top H U_X) \right) U_X^\top \right) \tag{9}$$

for any nonzero $H \in \mathbb{R}^{n \times n \times p}$ with $H = H^{\top}$, where $H := \text{bcirc}(H)$;

(ii) $f^\ast$ is differentiable at any $\mathcal{X} \in \mathbb{R}^{n \times n \times p}$ with $\mathcal{X} = \mathcal{X}^{\top}$ if and only if $f$ is differentiable on $\mathbb{R}_+$.

**Proof.** (i) Sufficiency: Suppose $f$ is differentiable at $\sigma_j^i, i \in [p], j \in [n]$. By [27, Theorem 14], we know that $\text{bcirc}(f^\ast(X)) = \bar{f}(\text{bcirc}(X))$, where $\bar{f}$ is the generalized matrix function defined by Definition 2.8. Since $f$ is differentiable at each singular value $\sigma_j^i$ by Lemma 3.3, we can obtain the differentiability of $\bar{f}$ at $\text{bcirc}(X)$. Next, we show the differentiability of $f^\ast$ at $X$. To this end, it is enough to show that there exists $\nabla f^\ast$ such that

$$f^\ast(X + H) - f^\ast(X) - \nabla f^\ast(X)H = o(\|H\|).$$

Let $D_X := \text{bcirc}^{-1} \left( U_X \left( g(\Sigma_X) \odot (U_X^\top H U_X) \right) U_X^\top \right)$ and $R$ be defined as

$$R := f^\ast(X + H) - f^\ast(X) - D_X.$$

Now we aim to prove that $R = o(\|H\|)$. Taking the $\text{bcirc}$ operation on both sides, we have

$$\text{bcirc}(R) = \text{bcirc}(f^\ast(X + H) - f^\ast(X) - D_X) = \text{bcirc}(f^\ast(X + H)) - \text{bcirc}(f^\ast(X)) - \text{bcirc}(D_X) = \bar{f}(\text{bcirc}(X + H)) - \bar{f}(\text{bcirc}(X)) - \text{bcirc}(D_X).$$

It is clear that $X + H = \text{bcirc}(X) + \text{bcirc}(H) = \text{bcirc}(X + H)$, then we have

$$\text{bcirc}(R) = \bar{f}(X + H) - \bar{f}(X) = U_X \left( g(\Sigma_X) \odot (U_X^\top H U_X) \right) U_X^\top.$$

Then by Lemma 3.3 we can deduce that $\text{bcirc}(R) = o(\|H\|)$, where $o(\cdot)$ only depends on $f$ and $\mathcal{X}$. Since $\|H\| = \frac{1}{\sqrt{p}} \|\text{bcirc}(H)\| = \frac{1}{\sqrt{p}} \|H\|$ and $\|R\| = \frac{1}{\sqrt{p}} \|\text{bcirc}(R)\|$, we can deduce that $R = o(\|H\|)$. Hence $f^\ast$ is differentiable at $\mathcal{X}$ and by the uniqueness of the derivative, it follows that $\nabla f^\ast(X)H = D_X$, i.e., (9) is satisfied.

Necessity: Suppose $f^\ast$ is differentiable at $\mathcal{X}$. Now we intend to prove $f$ is differentiable at $\sigma_j^i, i \in [p], j \in [n]$. Let $H$ be the tensor defined by

$$H := \mathcal{U} * \text{bcirc}^{-1} \left( F_{p,n} \begin{bmatrix} O & & \cdots & E^i & \cdots & O \\ & \ddots & & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \vdots \end{bmatrix} F_{p,n}^H \right) * \mathcal{U}^{\top},$$

where $E^i := \text{diag}(0, \ldots, 1, \ldots, 0) \in \mathbb{R}^{n \times n}$ whose the $j$-th diagonal entry is one, and the other entries are all zeros.
Suppose on the contrary, \( f \) is not differentiable at some singular value \( \sigma_j \). Then either the directional derivative of \( f \) does not exist or the left and right derivatives of \( f \) at \( \sigma_j \) exist but are unequal. When the directional derivative does not exist, we have \( f^\circ \) is not differentiable at \( \mathcal{X} \), which leads to a contradiction. On the other hand, suppose there exist two sequences of nonzero scalars \( \{\alpha^k\} \) and \( \{\beta^k\} \) converging to zero, such that the following two limits

\[
\lim_{k \to \infty} \frac{f(\sigma_j + \alpha^k) - f(\sigma_j)}{\alpha^k} \quad \text{and} \quad \lim_{k \to \infty} \frac{f(\sigma_j + \beta^k) - f(\sigma_j)}{\beta^k}
\]

exist but are unequal. What is more, we have

\[
\lim_{k \to \infty} \frac{f^\circ(\mathcal{X} + \alpha^k \mathcal{H}) - f^\circ(\mathcal{X})}{\alpha^k} = \mathcal{U} \ast \hat{f}(\mathcal{S}_\alpha) \ast \mathcal{U}^T, \quad (10)
\]

\[
\lim_{k \to \infty} \frac{f^\circ(\mathcal{X} + \beta^k \mathcal{H}) - f^\circ(\mathcal{X})}{\beta^k} = \mathcal{U} \ast \hat{f}(\mathcal{S}_\beta) \ast \mathcal{U}^T, \quad (11)
\]

where the functions \( \hat{f}(\mathcal{S}_\alpha) \) and \( \hat{f}(\mathcal{S}_\beta) \) are given by

\[
\hat{f}(\mathcal{S}_\alpha) := \text{bcirc}^{-1} \left( F_{p,n} \text{diag}(0, \ldots, \lim_{k \to \infty} \frac{f(\sigma_j + \alpha^k) - f(\sigma_j)}{\alpha^k}, \ldots, 0) F_{p,n}^H \right),
\]

\[
\hat{f}(\mathcal{S}_\beta) := \text{bcirc}^{-1} \left( F_{p,n} \text{diag}(0, \ldots, \lim_{k \to \infty} \frac{f(\sigma_j + \beta^k) - f(\sigma_j)}{\beta^k}, \ldots, 0) F_{p,n}^H \right).
\]

Then, the above two limits, defined by (10) and (11), are not equal. It means that \( f^\circ \) is not differentiable at \( \mathcal{X} \), which leads to a contradiction. Therefore, \( f \) is differentiable at singular values \( \sigma_j \), \( i \in [p], j \in [n] \).

(ii) The result can be easily obtained from (i). \( \square \)

Now we give an example to illuminate how to use Proposition 5 to compute the derivative of a generalized tensor function.

**Example 3.1.** Let \( f(\xi) = \xi^2 \) for any \( \xi \in \mathbb{R}_+ \), then \( f'(\xi) = 2\xi \). Let \( f^\circ \) be defined by Definition 2.10. Suppose

\[
\mathcal{X}(\cdot, \cdot, 1) = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}; \quad \mathcal{X}(\cdot, \cdot, 2) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \quad \mathcal{X}(\cdot, \cdot, 3) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.
\]

It is clear that \( \mathcal{X} = \mathcal{X}^T \). To compute the singular values of \( \mathcal{X} \), we use the function ‘svd’ of MATLAB for bcirc(\( \mathcal{X} \)) and obtain that the singular values of \( \mathcal{X} \) are 6, 4, 3, 3, 1, 1. At the same time, we also obtain \( U_\mathcal{X} \).

Taking \( \mathcal{H} \) by

\[
\mathcal{H}(\cdot, \cdot, 1) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}; \quad \mathcal{H}(\cdot, \cdot, 2) = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}; \quad \mathcal{H}(\cdot, \cdot, 3) = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix},
\]

it is clear that \( \mathcal{H} = \mathcal{H}^T \). Then by (9), we obtain that

\[
(\nabla f^\circ(\mathcal{X}) \mathcal{H}) (\cdot, \cdot, 1) = \begin{bmatrix} 14 & 24 \\ 24 & 14 \end{bmatrix};
\]

\[
(\nabla f^\circ(\mathcal{X}) \mathcal{H}) (\cdot, \cdot, 2) = \begin{bmatrix} 15 & 24 \\ 28 & 15 \end{bmatrix};
\]

\[
(\nabla f^\circ(\mathcal{X}) \mathcal{H}) (\cdot, \cdot, 3) = \begin{bmatrix} 15 & 28 \\ 24 & 15 \end{bmatrix}.
\]
It can be seen from this example that $\nabla f^o(\mathcal{X}) \mathcal{H} = (\nabla f^o(\mathcal{X}) \mathcal{H})^T$. However, we do not know whether it is a general conclusion that $\nabla f^o(\mathcal{X}) \mathcal{H} = (\nabla f^o(\mathcal{X}) \mathcal{H})^T$ holds whenever $\mathcal{X} = \mathcal{X}^T$ and $\mathcal{H} = \mathcal{H}^T$. This is worth further studying.

Since the singular values of a tensor are nonnegative, in the discussions above we only use the properties of $f$ on $\mathbb{R}^+$. For the sake of discussing some results concisely and conveniently, similar to [5] and [27, Remark 1], we restrict the function to an odd function $F : \mathbb{R} \to \mathbb{R}$ as follows:

$$F(\xi) := \begin{cases} f(\xi), & \text{if } \xi \in \mathbb{R}^+, \\ -f(-\xi), & \text{if } \xi \in \mathbb{R}/\mathbb{R}^+, \end{cases} \quad (12)$$

For convenience, we still use the symbol $f$ to replace $F$. That is, unless otherwise specified, the function $f$ in the following is an odd function.

The following definition of the block tensor comes from [27, Definition 8].

**Definition 3.4.** Suppose $A \in \mathbb{R}^{m_1 \times n_1 \times p}$, $B \in \mathbb{R}^{m_2 \times n_1 \times p}$, $C \in \mathbb{R}^{m_1 \times n_2 \times p}$ and $D \in \mathbb{R}^{m_2 \times n_2 \times p}$. The block tensor $\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{R}^{(m_1+m_2) \times (n_1+n_2) \times p}$ is defined by compositing the frontal slices of four tensors.

The following result comes from [27, Remark 8] and the discussion after [27, Example 1].

**Remark 1.** For a non-frontal-square tensor $A \in \mathbb{R}^{m \times n \times p}$, if we set $B = \begin{bmatrix} O & A^T \\ A & O \end{bmatrix}$ then $B$ is a frontal-square tensor. We have that for any real-valued odd function $f$, the induced generalized tensor function satisfies

$$f^o(B) = \begin{bmatrix} O & f^o(A)^T \\ f^o(A) & O \end{bmatrix}.$$  

**Proposition 6.** Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function defined by (12) and $f^o$ be defined by Definition 2.10. Let $\hat{\mathcal{X}} := \begin{bmatrix} O & \mathcal{X}^T \\ \mathcal{X} & O \end{bmatrix}$ and $\hat{\mathcal{H}} := \begin{bmatrix} O & \mathcal{H}^T \\ \mathcal{H} & O \end{bmatrix}$ are two block tensors, where $\mathcal{X}, \mathcal{H} \in \mathbb{R}^{m \times n \times p}$. Then the differentiability of $f^o$ at $\hat{\mathcal{X}}$ is equivalent to the differentiability of $f^o$ at $\mathcal{X}$ and when $f^o$ is differentiable at $\mathcal{X}$, it follows that

$$\nabla f^o(\hat{\mathcal{X}}) \hat{\mathcal{H}} = \begin{bmatrix} O & (\nabla f^o(\mathcal{X}) \mathcal{H})^T \\ (\nabla f^o(\mathcal{X}) \mathcal{H}) & O \end{bmatrix}. \quad (13)$$

**Proof.** By Remark 1, we have that

$$f^o(\hat{\mathcal{X}}) = \begin{bmatrix} O & f^o(\mathcal{X})^T \\ f^o(\mathcal{X}) & O \end{bmatrix},$$

which means that the differentiability of $f^o$ at $\hat{\mathcal{X}}$ is equivalent to the differentiability of $f^o$ at $\mathcal{X}$.

Suppose $f^o$ is differentiable. Let $\nabla f^o(\mathcal{X})$ and $\nabla f^o(\hat{\mathcal{X}})$ denote the derivatives of $f^o$ at $\mathcal{X}$ and $\hat{\mathcal{X}}$, respectively, then it follows that

$$f^o(\mathcal{X} + \mathcal{H}) - f^o(\mathcal{X}) - \nabla f^o(\mathcal{X}) \mathcal{H} = o(\|\mathcal{H}\|),$$  

and

$$f^o(\hat{\mathcal{X}} + \hat{\mathcal{H}}) - f^o(\hat{\mathcal{X}}) - \nabla f^o(\hat{\mathcal{X}}) \hat{\mathcal{H}} = o(\|\hat{\mathcal{H}}\|).$$  

(14)

(15)
Denote $P := \nabla f(\mathcal{X})\mathcal{H}$, and let $\tilde{B} = \begin{bmatrix} O & P \end{bmatrix}^\top$, then
\[
\begin{align*}
f(\mathcal{X} + \mathcal{H}) - f(\mathcal{X}) - \tilde{B} &= f(\mathcal{X} + \mathcal{H}) - \begin{bmatrix} O & f(\mathcal{X}) \end{bmatrix} - \begin{bmatrix} O & \tilde{P} \end{bmatrix}^\top \begin{bmatrix} f(\mathcal{X}) & O \end{bmatrix} - \begin{bmatrix} f(\mathcal{X} + \mathcal{H}) - f(\mathcal{X}) & -\tilde{P} \end{bmatrix}^\top.
\end{align*}
\]
Due to (14) and $\|\tilde{\mathcal{H}}\| = \sqrt{2}\|\mathcal{H}\|$, we can obtain from above that
\[
f(\mathcal{X} + \mathcal{H}) - f(\mathcal{X}) - \tilde{B} = o(\|\tilde{\mathcal{H}}\|).
\]
Therefore, by (15) and the uniqueness of the derivative, one gets that
\[
\nabla f(\mathcal{X})\tilde{\mathcal{H}} = \tilde{B} = \begin{bmatrix} O & (\nabla f(\mathcal{X}))^\top \end{bmatrix}.
\]

The proof is complete. \qed

**Proposition 7.** Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function defined by (12) and $f^\circ$ be defined by Definition 2.10. Then the following results hold:

(i) $f^\circ$ is differentiable at $\mathcal{X} \in \mathbb{R}^{m \times n \times p}$ with singular values $\sigma_i^j$, $i \in [p]$, $j \in [n]$ if and only if $f$ is differentiable at all singular values $\sigma_i^j$;

(ii) $f^\circ$ is differentiable if and only if $f$ is differentiable.

**Proof.** (i) The sufficiency holds from Proposition 5 and the discussions in Proposition 6; and the necessity is similar to that of Proposition 5.

(ii) The result can be easily obtained from (i). \qed

**Remark 2.** Actually, when discussing the differentiability and the formula of derivative of the generalized tensor function $f^\circ$, we can refer to the results of symmetric real matrices in [9, Proposition 4.3]. Besides, we can also refer to [30, Corollary 3.10] to discuss the derivative of $f^\circ$ at real tensor $\mathcal{X}$.

The following example simply illuminates how to compute the derivative of a generalized tensor function on rectangular tensors by Proposition 7.

**Example 3.2.** Let $f$ be defined as
\[
f(\xi) = \begin{cases} \xi^2, & \text{if } \xi \in \mathbb{R}_+, \\
-\xi^2, & \text{if } \xi \in \mathbb{R}/\mathbb{R}_+.
\end{cases}
\]
then
\[
f'(\xi) = \begin{cases} 2\xi, & \text{if } \xi \in \mathbb{R}_+, \\
-2\xi, & \text{if } \xi \in \mathbb{R}/\mathbb{R}_+.
\end{cases}
\]
Let $f^\circ$ be defined by Definition 2.10. Suppose
\[
\mathcal{X}(::, 1) = \begin{bmatrix} 2 & 1 & 3 \\
0 & 1 & 0
\end{bmatrix}; \quad \mathcal{X}(::, 2) = \begin{bmatrix} 1 & 0 & 3 \\
0 & 1 & 1
\end{bmatrix}; \quad \mathcal{X}(::, 3) = \begin{bmatrix} 3 & 3 & 4 \\
1 & 2 & 2
\end{bmatrix}.
\]
Here, $\mathcal{X}$ is clearly rectangular. Therefore, we first construct $\tilde{\mathcal{X}} = \begin{bmatrix} O & \mathcal{X}^\top \\
\mathcal{X} & O
\end{bmatrix}$. Since $\tilde{\mathcal{X}} = \tilde{\mathcal{X}}^\top$, we can use MATLAB to compute the singular values of $\tilde{\mathcal{X}}$ similar to Example 3.1. By calculating, we know that the singular values of $\tilde{\mathcal{X}}$ are 13.0648,
13.0648, 3.8448, 3.8448, 3.8448, 3.8448, 2.7040, 2.7040, 1.1035, 1.1035, 1.1035, 1.1035, 0, 0, 0 (round to four decimal places). Next, taking $\mathcal{H}$ as

$$\mathcal{H}(\cdot, 1, 1) = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 0 \end{bmatrix}; \quad \mathcal{H}(\cdot, 2, 2) = \begin{bmatrix} 2 & 3 & 2 \\ 3 & 1 & 2 \end{bmatrix}; \quad \mathcal{H}(\cdot, 3, 3) = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix},$$

we construct $\tilde{\mathcal{H}} = \begin{bmatrix} 0 & \mathcal{H}^\top \\ \mathcal{H} & 0 \end{bmatrix}$. Then it follows that $\tilde{\mathcal{H}} = \mathcal{H}^\top$. For $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{H}}$, we can use (9) to compute $\nabla f^\circ(\tilde{\mathcal{X}})\tilde{\mathcal{H}}$. Finally, by (13) we obtain that (the numbers are round to two decimal places)

$$\nabla f^\circ(\mathcal{X})\mathcal{H}(\cdot, 1, 1) = \begin{bmatrix} 37.37 & 52.38 & 42.21 \\ 33.17 & 27.70 & 35.36 \end{bmatrix};$$

$$\nabla f^\circ(\mathcal{X})\mathcal{H}(\cdot, 2, 2) = \begin{bmatrix} 41.89 & 55.68 & 51.12 \\ 32.24 & 26.74 & 43.38 \end{bmatrix};$$

$$\nabla f^\circ(\mathcal{X})\mathcal{H}(\cdot, 3, 3) = \begin{bmatrix} 37.48 & 57.00 & 53.39 \\ 32.62 & 32.77 & 47.24 \end{bmatrix}.$$

**Proposition 8.** Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function defined by (12) and $f^\circ$ be defined by Definition 2.10. Then $f^\circ$ is continuously differentiable if and only if $f$ is continuously differentiable.

**Proof.** First, we consider the case when $\mathcal{X} = \mathcal{X}^\top$.

**Sufficiency:** Suppose $f$ is continuously differentiable, $\mathcal{X} \in \mathbb{R}^{n \times n \times p}$ have the singular values $\sigma_j^i$, $i \in [p]$, $j \in [n]$ and $\text{bcirc}(\mathcal{X}) = U\Sigma U^\top$ is an SVD of $\text{bcirc}(\mathcal{X})$. For any $\mathcal{H} = \mathcal{H}^\top \in \mathbb{R}^{n \times n \times p}$, let $\tilde{\mathcal{X}} := \mathcal{X} + \mathcal{H} \in \mathbb{R}^{n \times n \times p}$ with the singular values $\mu_j^i$, $i \in [p]$, $j \in [n]$ and suppose $\text{bcirc}(\tilde{\mathcal{X}}) = \tilde{U}\tilde{\Sigma}\tilde{U}^\top$ is an SVD of $\text{bcirc}(\tilde{\mathcal{X}})$. For any given $\mathcal{X}$ and $\tilde{\mathcal{X}}$, $\sigma_j^i$ and $\mu_j^i$ are finite numbers, then $f(\sigma_j^i)$ and $f(\mu_j^i)$ are bounded. Let

$$c := \max_{k, l \in [n]} \{(g(\Sigma))_{kl}\} \quad \text{and} \quad \tilde{c} := \max_{k, l \in [n]} \{(g(\tilde{\Sigma}))_{kl}\},$$

where $g(\cdot)$ is defined by (8). Thus $\|g(\Sigma)\| \leq np\tilde{c}$ and $\|g(\tilde{\Sigma})\| \leq np\tilde{c}$. Since $f$ is differentiable, then by Proposition 5, $f^\circ$ is differentiable. Let $V := \tilde{U} - U$, then by (9) we have

$$\text{bcirc}(\nabla f^\circ(\tilde{\mathcal{X}})\mathcal{H} - \nabla f^\circ(\mathcal{X})\mathcal{H})$$

$$= \tilde{U}(g(\tilde{\Sigma}) \circ (\tilde{U}^\top H\tilde{U}))\tilde{U}^\top - U(g(\Sigma) \circ (U^\top HU))U^\top$$

$$= V(g(\tilde{\Sigma}) \circ (U^\top H\tilde{U}))\tilde{U}^\top + U((g(\tilde{\Sigma}) - g(\Sigma)) \circ (\tilde{U}^\top H\tilde{U}))\tilde{U}^\top$$

$$+ U(g(\Sigma) \circ (V^\top H\tilde{U}))\tilde{U}^\top + U(g(\Sigma) \circ (U^\top HV))\tilde{U}^\top + U(g(\Sigma) \circ (U^\top HU))V^\top.$$

Then it follows that

$$\|\text{bcirc}(\nabla f^\circ(\tilde{\mathcal{X}})\mathcal{H} - \nabla f^\circ(\mathcal{X})\mathcal{H})\|$$

$$\leq \|V(g(\tilde{\Sigma}) \circ (\tilde{U}^\top H\tilde{U}))\tilde{U}^\top\| + \|U((g(\tilde{\Sigma}) - g(\Sigma)) \circ (\tilde{U}^\top H\tilde{U}))\tilde{U}^\top\|$$

$$+ \|U(g(\Sigma) \circ (V^\top H\tilde{U}))\tilde{U}^\top\| + \|U(g(\Sigma) \circ (U^\top HV))\tilde{U}^\top\|$$

$$+ \|U(g(\Sigma) \circ (U^\top HU))V^\top\|$$

$$\leq np\tilde{c}\|V\|\|H\| + \|g(\tilde{\Sigma}) - g(\Sigma)\|\|H\| + 3np\tilde{c}\|V\|\|H\|$$

$$= np(\tilde{c} + 3c)\|V\|\|H\| + \|g(\tilde{\Sigma}) - g(\Sigma)\|\|H\|,$$
where the second inequality uses that \(|g(\Sigma)| \leq npc, \ |g(\bar{\Sigma})| \leq npc; U, \bar{U} are orthogonal and \(|A \odot B| \leq \|A\|B\). Since \(|H| = \|\text{bcirc}(H)\| = \sqrt{p}\|H\|, '
\[
\|\nabla f^o(\bar{X}) - \nabla f^o(X)\| = \max_{\|H\|=1} \|\nabla f^o(\bar{X})H - \nabla f^o(X)H\|
\]
\[
= \frac{1}{\sqrt{p}} \max_{\|H\|=1} \|\text{bcirc}(\nabla f^o(\bar{X})H - \nabla f^o(X)H)\|
\]
\[
\leq \frac{1}{\sqrt{p}} \max_{\|H\|=1} \{np(c+3c)\|V\|\|H\| + \|g(\bar{\Sigma}) - g(\Sigma)\|\|H\|\}
\]
\[
\leq np(c+3c)\|V\| + \|g(\bar{\Sigma}) - g(\Sigma)\|. 
\]
By Lemma 3.1 there exists \(\tilde{U}\) related to \(U\) such that \(\|V\| = \|\tilde{U} - U\| \to 0\) as \(\tilde{X} \to X\). By the continuous differentiability of \(f\) and (8), we can deduce that \(\|g(\bar{\Sigma}) - g(\Sigma)\| \to 0\) as \(\tilde{X} \to X\). Therefore, it follows that
\[
\|\nabla f^o(\bar{X}) - \nabla f^o(X)\| \leq np(c+3c)\|V\| + \|g(\bar{\Sigma}) - g(\Sigma)\| \to 0, \text{ as } \tilde{X} \to X.
\]
Thus, \(f^o\) is continuously differentiable.

Necessity: Suppose \(f^o\) is continuously differentiable. Since \(f^o\) is differentiable, by Proposition 5, \(f\) is differentiable. Let \(\tilde{X} \in \mathbb{R}^{n \times n \times p}\) be the tensor as
\[
\tilde{X} := \text{bcirc}^{-1}(\tilde{F}_{p,n}\text{diag}(O,\ldots,O,E^i,O,\ldots,O)\hat{F}_p^H),
\]
where \(E^i := \text{diag}(0,\ldots,\xi,\ldots,0) \in \mathbb{R}^{n \times n}\) whose the \(j\)-th diagonal entry is \(\xi \in \mathbb{R}_+\), and the other entries are all zeros. Then we have that \(\nabla f^o\) is continuous at \(\tilde{X}\), which is equivalent to that \(f^o\) is continuous at \(\xi\). Since \(f\) is an odd function defined by (12), we can get that \(f^o\) is continuous at \(\xi \in \mathbb{R}\). Therefore, \(f\) is continuously differentiable.

The proof of rectangular tensor cases can be obtained by combining the discussions above with Proposition 6.

\begin{proposition}
Let \(f : \mathbb{R} \to \mathbb{R}\) be an odd function defined by (12) and \(f^o\) be defined by Definition 2.10. Then the following results hold:

(i) \(f^o\) is locally Lipschitz continuous at \(X\) with singular values \(\sigma_j^i, i \in [p], j \in [n]\) if and only if \(f\) is locally Lipschitz continuous at all singular values \(\sigma_j^i\);

(ii) \(f^o\) is locally Lipschitz continuous if and only if \(f\) is locally Lipschitz continuous;

(iii) \(f^o\) is Lipschitz continuous if and only if \(f\) is Lipschitz continuous.
\end{proposition}

\begin{proof}
(i) Sufficiency: Fix any \(X \in \mathbb{R}^{m \times n \times p}\) with singular values \(\sigma_j^i, i \in [p], j \in [n]\). Suppose \(X = \mathcal{U} \ast \mathcal{S} \ast \mathcal{V}^T\) is a T-SVD of \(X\), where \(\mathcal{S}\) is factored as (6). Suppose the scalar function \(f\) is locally Lipschitz continuous at all singular values \(\sigma_j^i, i \in [p], j \in [n]\). From the proof of [9, Proposition 4.6], there exist some compact set \(\mathcal{S} := \bigcup_{i=1}^p \bigcup_{j=1}^n [\sigma_j^i - \delta_j^i, \sigma_j^i + \delta_j^i]\) and continuously differentiable scalar functions \(f^\nu : \mathbb{R} \to \mathbb{R}, \nu = 1, 2, \ldots\) such that \(\{f^\nu\}_1^\infty\) converges uniformly to \(f\) and satisfies
\[
|f^\nu(\xi)| \leq \kappa, \quad \forall \xi \in \mathcal{S}, \forall \nu.
\]
Furthermore, there exists \(\delta := \min_{i \in [p]} \min_{j \in [n]} \delta_j^i\) such that \(\{\tilde{f}^\nu\}_1^\infty\) converges uniformly to \(\tilde{f}\) on \(B(\Sigma, \delta)\), where \(\tilde{f}\) is defined by (5) and \(\Sigma \in \tilde{\mathcal{D}}\).
\end{proof}
Next, for functions \((f^\nu)^\circ\), \(\nu = 1, 2, \ldots\) and \(f^\circ\), it follows that
\[
\|(f^\nu)^\circ(\mathcal{Y}) - f^\circ(\mathcal{Y})\|_2^2 = \|\mathcal{U} \ast \hat{f}^\nu(\mathcal{S}) \ast \mathcal{Y}^\top - \mathcal{U} \ast \hat{f}(\mathcal{S}) \ast \mathcal{Y}^\top\|_2^2
\]
\[
= \|\hat{f}^\nu(\mathcal{S}) - \hat{f}(\mathcal{S})\|_2^2 = \sum_{i=1}^p \|\hat{f}^\nu(\Sigma^i) - \hat{f}(\Sigma^i)\|_2^2.
\]
Therefore, \(\{(f^\nu)^\circ\}_{1}^{\infty}\) converges uniformly to \(f^\circ\) by virtue of the uniform convergence of \(\{\hat{f}^\nu\}_{1}^{\infty}\). Hence, fixing any \(\mathcal{Y}, \mathcal{Z} \in \mathcal{B}(\mathcal{X}, \delta)\) with \(\mathcal{Y} \neq \mathcal{Z}\), for any \(\epsilon > 0\) there exists an integer \(\nu_0\) such that for all \(\nu \geq \nu_0\),
\[
\|(f^\nu)^\circ(\mathcal{A}) - f^\circ(\mathcal{A})\| \leq \epsilon\|\mathcal{Y} - \mathcal{Z}\|, \quad \forall \mathcal{A} \in \mathcal{B}(\mathcal{X}, \delta).
\]
Besides, for any \(\mathcal{A} \in \mathcal{B}(\mathcal{X}, \delta)\), it follows that
\[
\|\nabla(f^\nu)^\circ(\mathcal{A})\| = \sup_{\|\mathcal{H}\| = 1} \|\nabla(f^\nu)^\circ(\mathcal{A})\mathcal{H}\| = \sup_{\|\mathcal{H}\| = 1} \frac{1}{\sqrt{2}} \|\nabla(f^\nu)^\circ(\tilde{\mathcal{A}})\tilde{\mathcal{H}}\|
\]
\[
= \frac{1}{\sqrt{2}} \sup_{\|\mathcal{H}\| = 1} \|\text{bcirc}^{-1}(\tilde{U}_\mathcal{A}(g^\nu(\Sigma^\mathcal{A}) \ominus (U^\top_\mathcal{A} H U^\top_\mathcal{A})) U^\top_\tilde{\mathcal{A}})\|
\]
\[
= \frac{1}{\sqrt{2}} \sup_{\|\mathcal{H}\| = 1} \|g^\nu(\Sigma^\mathcal{A}) \ominus (U^\top_\mathcal{A} H U^\top_\tilde{\mathcal{A}})\| \leq \frac{1}{\sqrt{2}} \sup_{\|\mathcal{H}\| = 1} \kappa \|U^\top_\mathcal{A} H U^\top_\mathcal{A}\|
\]
\[
= \frac{\kappa}{\sqrt{2}} \sup_{\|\mathcal{H}\| = 1} \|\tilde{\mathcal{H}}\| = \kappa \sup_{\|\mathcal{H}\| = 1} \|\mathcal{H}\| \leq \kappa,
\]
where the second equality holds due to (13), the third equality holds due to (9), the first inequality holds due to (16) and \(g^\nu\) is obtained by replacing \(f\) with \(f^\nu\) in (8). We can obtain the continuous differentiability of \((f^\nu)^\circ\) from that of \(f^\nu\) by Proposition 8. Then, it follows that
\[
\|f^\circ(\mathcal{Y}) - f^\circ(\mathcal{Z})\|
\]
\[
= \|f^\circ(\mathcal{Y}) - (f^\nu)^\circ(\mathcal{Y}) + (f^\nu)^\circ(\mathcal{Y}) - (f^\nu)^\circ(\mathcal{Z}) + (f^\nu)^\circ(\mathcal{Z}) - f^\circ(\mathcal{Z})\|
\]
\[
\leq \|f^\circ(\mathcal{Y}) - (f^\nu)^\circ(\mathcal{Y})\| + \|(f^\nu)^\circ(\mathcal{Y}) - (f^\nu)^\circ(\mathcal{Z})\| + \|(f^\nu)^\circ(\mathcal{Z}) - f^\circ(\mathcal{Z})\|
\]
\[
\leq 2\|\mathcal{Y} - \mathcal{Z}\| + \int_0^1 \|\nabla(f^\nu)^\circ(\mathcal{Z} + \tau(\mathcal{Y} - \mathcal{Z}))(\mathcal{Y} - \mathcal{Z})d\tau\|
\]
\[
\leq (2\epsilon + \kappa)\|\mathcal{Y} - \mathcal{Z}\|.
\]
Due to the arbitrariness of \(\epsilon\), it yields that
\[
\|f^\circ(\mathcal{Y}) - f^\circ(\mathcal{Z})\| \leq \kappa\|\mathcal{Y} - \mathcal{Z}\|, \quad \forall \mathcal{Y}, \mathcal{Z} \in \mathcal{B}(\mathcal{X}, \delta).
\]
Hence, \(f^\circ\) is locally Lipschitz continuous at \(\mathcal{X}\).

Necessity: Suppose \(f^\circ\) is locally Lipschitz continuous at \(\mathcal{X}\) with singular values \(\sigma_j^\mathcal{X}\), \(i \in [\bar{p}]\), \(j \in [\bar{n}]\) and \(\mathcal{X} = \mathcal{U} \ast \mathcal{S} \ast \mathcal{Y}^\top\) is a T-SVD of \(\mathcal{X}\). From (6), we know that bcirc(\(\mathcal{S}\)) = \(F_{p,m,n} \text{diag}(\Sigma^1, \Sigma^2, \ldots, \Sigma^p) F_{p,n}^\top\). For any \(i \in [\bar{p}]\), \(j \in [\bar{n}]\) and scalars \(\xi, \zeta \in [\sigma_i^\mathcal{X} - \delta, \sigma_i^\mathcal{X} + \delta]\), let \(\Sigma_i^\mathcal{X}\) and \(\Sigma_i^\circ\) be the same as \(\Sigma_i^\mathcal{X}\) except the \(j\)-th diagonal entries being \(\xi\) and \(\zeta\), respectively. Let
\[
\text{bcirc}(\tilde{\mathcal{S}}) := F_{p,m,n} \text{diag}(\bar{\Sigma}^1, \ldots, \bar{\Sigma}^{i-1}, \Sigma_i^\mathcal{X}, \bar{\Sigma}^{i+1}, \ldots, \Sigma^p) F_{p,n}^\top,
\]
\[
\text{bcirc}(\tilde{\mathcal{S}}) := F_{p,m,n} \text{diag}(\bar{\Sigma}^1, \ldots, \bar{\Sigma}^{i-1}, \Sigma_i^\mathcal{X}, \bar{\Sigma}^{i+1}, \ldots, \Sigma^p) F_{p,n}^\top.
\]
\(\mathcal{Y} := \mathcal{U} \ast \tilde{\mathcal{S}} \ast \mathcal{Y}^\top\) and \(\mathcal{Z} := \mathcal{U} \ast \tilde{\mathcal{S}} \ast \mathcal{Y}^\top\). Then, it follows that
\[
\|\mathcal{Y} - \mathcal{X}\| = |\xi - \sigma_j^\mathcal{X}| \leq \delta \quad \text{and} \quad \|\mathcal{Z} - \mathcal{X}\| = |\zeta - \sigma_j^\mathcal{X}| \leq \delta.
\]
So, we have
\[ |f(\xi) - f(\zeta)| = \|f^\circ(\mathcal{Y}) - f^\circ(\mathcal{Z})\| \leq \kappa \|\mathcal{Y} - \mathcal{Z}\| = \kappa |\xi - \zeta| . \]

Hence, \( f \) is locally Lipschitz continuous at \( \sigma_j^i \) for any \( i \in [p] \) and \( j \in [n] \).

The proof of (i) is complete.

(ii) is a direct result of (i).

(iii) Sufficiency: Suppose \( f \) is Lipschitz continuous with constant \( \kappa \). Fix any \( X \in \mathbb{R}^{m \times n \times p} \) with the singular values \( \sigma_j^i, i \in [p], j \in [n] \). For any scalar \( \delta > 0 \), define the compact set \( S \subseteq \mathbb{R} \) as
\[ S := \bigcup_{i=1}^p \bigcup_{j=1}^n [\sigma_j^i - \delta, \sigma_j^i + \delta]. \]

Similar to the proof of (i), we can obtain that (17) holds. Due to the arbitrariness of \( \delta \) and the independence of \( \kappa \), it yields that
\[ \|f^\circ(\mathcal{Y}) - f^\circ(\mathcal{Z})\| \leq \kappa \|\mathcal{Y} - \mathcal{Z}\|, \quad \forall \mathcal{Y}, \mathcal{Z} \in \mathbb{R}^{m \times n \times p}, \]

i.e., \( f^\circ \) is Lipschitz continuous with constant \( \kappa \).

Necessity: Suppose \( f^\circ \) is Lipschitz continuous with constant \( \kappa \). For any \( \xi, \zeta \in \mathbb{R} \), let
\[ \Sigma^1 := \begin{pmatrix} \text{diag}(\xi, 0, \ldots, 0) \\ O \end{pmatrix} \in \mathbb{R}^{m \times n} \text{ and } \Sigma^1 := \begin{pmatrix} \text{diag}(\zeta, 0, \ldots, 0) \\ O \end{pmatrix} \in \mathbb{R}^{m \times n}. \]

Then, with \( p - 1 \) zero matrices in \( \mathbb{R}^{m \times n} \), let
\[ X := \text{bcirc}^{-1}(F_{p,m} \text{diag}(\Sigma^1, O, \ldots, O) F^H_{p,n}) \]
and
\[ Y := \text{bcirc}^{-1}(F_{p,m} \text{diag}(\Sigma^1, O, \ldots, O) F^H_{p,n}). \]

It is clear that \( X, Y \in \mathbb{R}^{m \times n \times p} \), then we have
\[ |f(\xi) - f(\zeta)| = \|f^\circ(X) - f^\circ(Y)\| \leq \kappa \|X - Y\| = \kappa |\xi - \zeta|. \]

Therefore, \( f \) is Lipschitz continuous with constant \( \kappa \).

For any \( \Sigma = \text{diag}(\mu_1, \ldots, \mu_n) \in \mathbb{D} \) where \( \mathbb{D} \) is defined by (2) and any symmetric matrix \( D \in \mathbb{R}^{n \times n} \), suppose \( f : \mathbb{R} \to \mathbb{R} \) is directionally differentiable at all \( \mu_j, j \in [n] \), then we define the function \( g_d : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n} \) as
\[ (g_d(\Sigma; D))_{jk} := \begin{cases} \frac{f(\mu_j) - f(\mu_k)}{\mu_j - \mu_k} D_{jk}, & \text{if } \mu_j \neq \mu_k, \\ f'(\mu_j; D_{jk}), & \text{if } \mu_j = \mu_k \end{cases} \tag{18} \]
for any \( j, k \in [n] \), where \( f'(\mu_j; D_{jk}) \) is the directional derivative of \( f \) at \( \mu_j \) related to \( D_{jk} \). When the matrices involved are symmetric, the following result is given by [42, Theorem 2.3.9].

**Lemma 3.5.** Let \( f : \mathbb{R} \to \mathbb{R} \) be an odd function defined by (12) and \( \tilde{f} \) be defined by Definition 2.8. Suppose symmetric matrix \( X \in \mathbb{R}^{n \times n} \) has singular values \( \sigma_1, \ldots, \sigma_n \), and \( X = U \Sigma U^T \) is an SVD of \( X \). Then \( \tilde{f} \) is directionally differentiable at \( X \) with singular values \( \sigma_1, \ldots, \sigma_n \) if and only if \( f \) is directionally differentiable at \( \sigma_1, \ldots, \sigma_n \).

Moreover, for any nonzero symmetric matrix \( H \in \mathbb{R}^{n \times n} \),
\[ (\tilde{f})'(X; H) = U g_d(\Sigma; U^T H U) U^T. \]

**Proposition 10.** Suppose \( X \in \mathbb{R}^{n \times n \times p} \) has singular values \( \sigma_j^i, i \in [p], j \in [n] \) and \( X = X^T \). Let \( f : \mathbb{R} \to \mathbb{R} \) be an odd function defined by (12) and \( \tilde{f} \) be defined by Definition 2.10, then the following results hold:
(i) $f^\circ$ is directionally differentiable at $\mathcal{X}$ if and only if $f$ is directionally differentiable at singular values $\sigma^i_j, \ i \in [p], \ j \in [n]$. What is more, for any nonzero $H \in \mathbb{R}^{n \times n \times p}$ with $H = H^T$, $(f^\circ)'(\mathcal{X}; H)$ is given by

$$(f^\circ)'(\mathcal{X}; H) = \text{bcirc}^{-1}(U_Xg_d(\Sigma_X; U_X^T H U_X)U_X^T)$$

where $X := \text{bcirc}(\mathcal{X}), \ H := \text{bcirc}(H), \ X = U_X \Sigma_X U_X^T$ is an SVD of $X$ with $\Sigma_X = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{np})$.

(ii) $f^\circ$ is directionally differentiable at any $X \in \mathbb{R}^{n \times n \times p}$ with $\mathcal{X} = \mathcal{X}^T$ if and only if $f$ is directionally differentiable.

Proof. (i) Sufficiency: Suppose $f$ is directionally differentiable at singular values $\sigma^i_j, \ i \in [p], \ j \in [n]$. Note that the set $\{\sigma^i_j : i \in [p], \ j \in [n]\}$ is equal to the set $\{\sigma_k : k \in [np]\}$. Similar to the proof of Proposition 5, we can obtain the directional differentiability of $f^\circ$ from that of $f$. Let $D := \text{bcirc}^{-1}(U_Xg_d(\Sigma_X; U_X^T H U_X)U_X^T)$ and $\mathcal{R}$ be defined as

$$\mathcal{R} := f^\circ(\mathcal{X} + tH) - f^\circ(\mathcal{X}) - tD,$$

where $t$ is a positive scalar. We aim to prove that $\mathcal{R} = o(t)$. Taking the bcirc operation on both sides, we have

$$\begin{align*}
\text{bcirc}(\mathcal{R}) &= \text{bcirc}(f^\circ(\mathcal{X} + tH) - f^\circ(\mathcal{X}) - tD) \\
&= \text{bcirc}(f^\circ(\mathcal{X} + tH)) - \text{bcirc}(f^\circ(\mathcal{X})) - t\text{bcirc}(D) \\
&= \hat{f}(\text{bcirc}(\mathcal{X} + tH)) - \hat{f}(\text{bcirc}(\mathcal{X})) - t\text{bcirc}(D) \\
&= \hat{f}(X + tH) - \hat{f}(X) - t(U_Xg_d(\Sigma_X; U_X^T H U_X)U_X^T).
\end{align*}$$

Then by Lemma 3.5 we can deduce that $\text{bcirc}(\mathcal{R}) = o(t)$. Further, we can obtain that $\mathcal{R} = o(t)$ from $\|\mathcal{R}\| = \frac{1}{\sqrt{t}}\|\text{bcirc}(\mathcal{R})\|$. Hence it deduces that $f^\circ$ is directionally differentiable at $\mathcal{X}$ and (19) is satisfied.

Necessity: Suppose $\mathcal{X} = \mathcal{U} \ast S \ast \mathcal{U}^T$ is a T-SVD of $\mathcal{X}$ and $f^\circ$ is directionally differentiable at $\mathcal{X}$. In particular, for any given $i, j \in [n]$, let

$$\mathcal{H} := \mathcal{U} \ast \text{bcirc}^{-1}
\begin{bmatrix}
O & & \\
& \ddots & \\
& & F_{p,n}^H
\end{bmatrix}
\begin{bmatrix}
D^i & & \\
& \ddots & \\
& & O
\end{bmatrix}^T \ast \mathcal{U}^T,$$

where $D^i := \text{diag}(0, \ldots, d^i_j, \ldots, 0) \in \mathbb{R}^{n \times n}$ whose the $j$-th diagonal entry is $d^i_j \in \mathbb{R}$, and the other entries are all zeros. Since $f^\circ$ is directionally differentiable at $\mathcal{X}$ respect to $\mathcal{H}$, we can obtain that the directional derivative of $f$ is well defined and $f$ is directionally differentiable at singular value $\sigma^i_j$. Then by the arbitrariness of $i$ and $j$, $f$ is directionally differentiable at all singular values of $\mathcal{X}$.

(ii) The result can be easily obtained from (i).

Proposition 11. Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function defined by (12) and $f^\circ$ be defined by Definition 2.10. Then the following results hold:

(i) Suppose $\mathcal{X} \in \mathbb{R}^{m \times n \times p}$ has singular values $\sigma^i_j, \ i \in [p], \ j \in [n]$, then $f^\circ$ is directionally differentiable at $\mathcal{X}$ if and only if $f$ is directionally differentiable at all singular values $\sigma^i_j$;

(ii) $f^\circ$ is directionally differentiable if and only if $f$ is directionally differentiable.
Proposition 12. For any $j,k$ and nonempty. What is more, for any $J \in X\mu$, where $\tilde{X}$ can obtain the boundedness of for all $H \in (i)$ The sufficiency can be obtained from Proposition 10 combined with the discussions in Proposition 6 and

$$(f^\circ)'(\tilde{X}; \tilde{H}) = \begin{bmatrix} \mathcal{O} & ((f^\circ)'(X; H))^\top \\ (f^\circ)'(X; H) & \mathcal{O} \end{bmatrix},$$

where $\tilde{X} = \begin{bmatrix} \mathcal{O} & X^\top \\ X & \mathcal{O} \end{bmatrix}$ and $\tilde{H} = \begin{bmatrix} \mathcal{O} & H^\top \\ H & \mathcal{O} \end{bmatrix}$.

The necessity is similar to that of Proposition 10.

(ii) The result can be easily obtained from (i).

For any $\Sigma = \text{diag}(\mu_1, \ldots, \mu_n) \in \mathbb{D}$, suppose $f : \mathbb{R} \to \mathbb{R}$ is locally Lipschitz continuous at all $\mu_j$, $j \in [n]$, then we define the function $g_B : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ as

$$(g_B(\Sigma))_{jk} := \begin{cases} f(\mu_j) - f(\mu_k) & \text{if } \mu_j \neq \mu_k, \\ \alpha_j, & \text{if } \mu_j = \mu_k \end{cases}$$

for any $j, k \in [n]$, where $\alpha_j \in \partial f(\mu_j)$ and $\partial f(\mu_j)$ denotes the generalized Jacobian of $f$ at $\mu_j$.

Proposition 12. Let $f : \mathbb{R} \to \mathbb{R}$ be a locally Lipschitz continuous odd function defined by (12) and $f^\circ$ be defined by Definition 2.10. Suppose $X \in \mathbb{R}^{n \times n \times p}$ with $X^\top = X^\top$ has singular values $\sigma^i_j$, $i \in [p]$, $j \in [n]$ and $X = U \circ S \circ U^\top$ is a T-SVD of $X$. Then the generalized Jacobian of $f^\circ$ at $X$ in the Bouligand sense is well defined and nonempty. What is more, for any $J \in \partial_B f^\circ(X)$, we have

$$JH = \text{bcirc}^{-1}(U_X(g_B(\Sigma_X) \odot (U_X^\top HU_X))U_X^\top)$$

for any $H \in \mathbb{R}^{n \times n \times p}$ with $H = H^\top$, where $X = \text{bcirc}(X)$, $H = \text{bcirc}(H)$ and $X = U_X \Sigma_X U_X^\top$ is an SVD of $X$.

Proof. Suppose $f$ is locally Lipschitz continuous, by Proposition 9, $f^\circ$ is also locally Lipschitz continuous, then as we discussed in Section 2, the generalized Jacobian of $f^\circ$ in the Bouligand sense is well defined. Fix any $J \in \partial_B f^\circ(X)$, by definition, there exists a sequence $X^\nu \in \mathbb{R}^{n \times n \times p}$ with $X^\nu = (X^\nu)^\top$ such that $f^\circ$ is differentiable at $X^\nu$ for all $\nu = 1, 2, \ldots$, and $\lim_{\nu \to \infty} \nabla f^\circ(X^\nu) = J$. Suppose $X^\nu = U^\nu \circ S^\nu \circ (U^\nu)^\top$ is a T-SVD of $X^\nu$ and $\text{bcirc}(X^\nu) = U_X \Sigma_X (U_X)^\top$ is an SVD of $\text{bcirc}(X^\nu)$. Let $\sigma^i_j$ and $(\sigma^i_j)^i_j$ $(i \in [p], j \in [n])$ represent the singular values of $X$ and $X^\nu$ respectively. By Lemma 3.1, for sufficiently large $\nu$, there exists a scalar $\eta > 0$ and $U^\nu$ such that $\|U - U^\nu\| \leq \eta\|X - X^\nu\|$. When necessary, we can assume that this holds for all $\nu$ and thus $\{U^\nu\}$ converges to $U$. By Corollary 1, we can deduce that $(\sigma^i_j)^i_j \to \sigma^i_j$, $i \in [p], j \in [n]$. Since $f^\circ$ is differentiable at $X^\nu$ for all $\nu$, then by Proposition 5, we have that $f$ is differentiable at all $(\sigma^i_j)^i_j$ and

$$\nabla f^\circ(X^\nu)H = \text{bcirc}^{-1}(U_X^\nu (g(\Sigma_X^\nu) \odot ((U_X^\nu)^\top HU_X^\nu)) (U_X^\nu)^\top)$$

for all $H \in \mathbb{R}^{n \times n \times p}$ with $H = H^\top$. Since $f$ is locally Lipschitz continuous, we can obtain the boundedness of $\{(g_B(\Sigma_X^\nu))_{kl}\}$ for all $k, l \in [np]$. What is more, since $X^\nu \to X$, by [42, Proposition 2.3.7], we can obtain that $\{(g_B(\Sigma_X))_{kl}\} \to (g_B(\Sigma_X))_{kl}$, for all $k, l \in [np]$. Then taking limits on both sides of (21), we obtain (20) where $U$ is the limit of $\{U^\nu\}$ and $g_B(\Sigma_X)$ is the limit of $(g_B(\Sigma_X))$. Then we can deduce the conclusions. \hfill \Box
Proposition 13. Let \( f : \mathbb{R} \to \mathbb{R} \) be a locally Lipschitz continuous odd function defined by (12) and \( f^o \) be defined by Definition 2.10. Then for any \( X \in \mathbb{R}^{m \times n \times p} \) the generalized Jacobian of \( f^o \) at \( X \) in the Bouligand sense is well defined and nonempty.

Proof. The results can be obtained from Proposition 12 combined with the discussions of Proposition 6 and

\[
\lim_{X^o \to X} \nabla f^o(X^o)\hat{H} = \left[ \begin{array}{cc} \lim_{X^o \to X} \nabla f^o(X^o)H & \lim_{X^o \to X} (\nabla f^o(X^o)H)^T \end{array} \right].
\]

\( \square \)

Similar to [9, Lemma 4.9] and [35, Theorem 3.7], we can obtain the following lemma.

Lemma 3.6. If the generalized tensor function \( f^o : \mathbb{R}^{n \times n \times p} \to \mathbb{R}^{n \times n \times p} \) defined by Definition 2.10 is locally Lipschitz continuous and directionally differentiable in a neighborhood of \( X \) with \( X = X^\top \). Then for any \( \gamma \in (0, \infty) \), the following statements are equivalent:

(i) For any \( H = H^\top \) and any \( J \in \partial f^o(X + H) \),
\[
f^o(X + H) - f^o(X) - JH = O(||H||^{1+\gamma}) \quad \text{or} \quad o(||H||)
\]

(ii) For any \( H = H^\top \) such that \( f^o \) is differentiable at \( X + H \),
\[
f^o(X + H) - f^o(X) - \nabla f^o(X + H)H = O(||H||^{1+\gamma}) \quad \text{or} \quad o(||H||).
\]

Proof. By [27, Theorem 14], we know that \( \mathrm{bcirc}(f^o(X)) = \bar{f}(\mathrm{bcirc}(X)) \), where \( \bar{f} \) is the generalized matrix function defined by Definition 2.8. Because \( f^o \) is locally Lipschitz continuous and directionally differentiable in a neighborhood of \( X \), by Proposition 9 and Proposition 11, it follows that \( f \) is locally Lipschitz continuous and directionally differentiable in some neighborhoods of the singular values of \( X \). Then by [42, Theorem 2.3.5] and Lemma 3.5, we can deduce that \( \bar{f} \) is locally Lipschitz continuous and directionally differentiable in a neighborhood of \( \mathrm{bcirc}(X) \). Since \( \mathcal{X} \in \mathbb{R}^{n \times n \times p} \) with \( X = X^\top \), we have that \( \mathrm{bcirc}(\mathcal{X}) \in \mathbb{R}^{np \times np} \) with \( \mathrm{bcirc}(\mathcal{X}) = \mathrm{bcirc}(\mathcal{X})^\top \). Hence, by [35, Theorem 3.7], we can obtain that for any \( \gamma \in (0, \infty) \), the following two statements are equivalent:

(a) For any \( \mathrm{bcirc}(H) = \mathrm{bcirc}(H)^\top \) and any \( J \in \partial \bar{f}(\mathrm{bcirc}(X + H)) \),
\[
\bar{f}(\mathrm{bcirc}(X + H)) - \bar{f}(\mathrm{bcirc}(X)) - \mathrm{Jbcirc}(H) = O(||\mathrm{bcirc}(H)||^{1+\gamma});
\]

(b) For any \( \mathrm{bcirc}(H) = \mathrm{bcirc}(H)^\top \) such that \( \bar{f} \) is differentiable at \( \mathrm{bcirc}(X + H) \),
\[
\bar{f}(\mathrm{bcirc}(X + H)) - \bar{f}(\mathrm{bcirc}(X)) - \nabla \bar{f}(\mathrm{bcirc}(X + H))\mathrm{bcirc}(H) = O(||\mathrm{bcirc}(H)||^{1+\gamma}).
\]

Similarly, we can also get that (a) and (b) are equivalent if \( O(||\mathrm{bcirc}(H)||^{1+\gamma}) \) is replaced by \( o(||\mathrm{bcirc}(H)||) \).

Noticing \( ||X|| = \frac{1}{\sqrt{p}}||\mathrm{bcirc}(X)|| \) and \( \mathrm{bcirc}(f^o(X)) = \bar{f}(\mathrm{bcirc}(X)) \), from the equivalence between (a) and (b) we can obtain that for any \( \gamma \in (0, \infty) \), the following two statements are equivalent:

(i) For any \( H = H^\top \) and any \( J \in \partial f^o(X + H) \),
\[
f^o(X + H) - f^o(X) - \partial H = O(||H||^{1+\gamma});
\]

(ii) For any \( H = H^\top \) such that \( f^o \) is differentiable at \( X + H \),
\[
f^o(X + H) - f^o(X) - \nabla f^o(X + H)H = O(||H||^{1+\gamma}).
\]
Similarly, we can also get that (i) and (ii) are equivalent if $O(||H||^{1+\gamma})$ is replaced by $o(||H||)$. \hfill \Box

The following result comes from [42, Theorem 2.3.11].

**Lemma 3.7.** Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function defined by (12) and $\tilde{f}$ be defined by Definition 2.8. Suppose $X \in \mathbb{R}^{m \times n}$ has singular values $\sigma_1, \ldots, \sigma_n$. If $f$ is (1-order) semismooth at $\sigma_1, \ldots, \sigma_n$, then $\tilde{f}$ is (1-order) semismooth at $X$.

**Proposition 14.** Let $f : \mathbb{R} \to \mathbb{R}$ be an odd function defined by (12) and $f^\circ$ be defined by Definition 2.10. Then the following results hold:

(i) $f^\circ$ is semismooth if and only if $f$ is semismooth;

(ii) If $f$ is 1-order semismooth then $f^\circ$ is 1-order semismooth.

**Proof.** (i) We only prove the semismoothness of $f^\circ$ when $X = X^\top$, then it can be extended to rectangular tensor cases similar to the discussions in Proposition 6.

Sufficiency: By the definition of semismoothness, we can obtain that if $f$ is semismooth then $f$ is locally Lipschitz continuous and directionally differentiable. Then by Proposition 9 and Proposition 10, $f^\circ$ is locally Lipschitz continuous and directionally differentiable. By Lemma 3.6, we only need to prove when $H = H^\top$, it follows that

$$f^\circ(X + H) - f^\circ(X) - \nabla f^\circ(X + H)H = o(||H||).$$

The rest proof can be obtained by combining Lemma 3.7 with the discussions in Proposition 5.

Necessity: Suppose $f^\circ$ is semismooth, then $f^\circ$ is locally Lipschitz continuous and directionally differentiable. Then by Proposition 9 and Proposition 10, we have that $f$ is locally Lipschitz continuous and directionally differentiable. Besides, for any $\xi, \zeta$ such that $f$ is differentiable at $\xi + \zeta$, by Proposition 7, $f^\circ$ is differentiable at $X + \tilde{H}$ where $X = \xi I$ and $\tilde{H} = \zeta I$. By the semismoothness of $f^\circ$ and Lemma 3.6, we have

$$f^\circ(X + \tilde{H}) - f^\circ(X) - \nabla f^\circ(X + \tilde{H})\tilde{H} = o(||\tilde{H}||),$$

which is equivalent to

$$f(\xi + \zeta) - f(\xi) - f'(\xi + \zeta)\zeta = o(||\zeta||)$$

Therefore, $f$ is semismooth.

(ii) Suppose $f$ is 1-order semismooth, similar to the proof of the sufficiency above, by Lemma 3.7, we can obtain the result by replacing $o(||H||)$ with $O(||H||^2)$ in (22). Therefore, $f^\circ$ is 1-order semismooth. \hfill \Box

4. **Conclusion.** In this paper, we concentrated on studying the generalized tensor function, which is defined by a scalar function, based on the tensor singular value decomposition and tensor T-product. By tensor T-product, any third-order tensor can correspond to a block-circulant matrix uniquely and vice versa. Then, with the aid of the generalized matrix function related to the corresponding block-circulant matrix, we studied the properties of the generalized tensor function and showed that the generalized tensor function has similar properties to the associated scalar function. Such properties include continuity, Fréchet differentiability, directional differentiability, continuous differentiability, Lipschitz continuity and semismoothness. All these properties provide an important foundation for studying the mathematical problems or the tensor optimization problems with such generalized tensor functions. Based on the research of these properties, one of the topics for
future research is to investigate the theory and algorithms for the tensor optimization problems with generalized tensor functions. In addition, there also exist other definitions of tensor functions, so it is worthwhile to study the similar properties to those in this paper for other tensor functions in the future.

Acknowledgments. The authors wish to express the gratitude to the anonymous referees for their valuable comments and advice which contributed to a major improvement of the first version of the manuscript. The third author’s work was supported by the National Natural Science Foundation of China [grant number 11871051].

REFERENCES

[1] B. P. W. Ames and H. S. Sendov, Derivatives of compound matrix valued functions, *Journal of Mathematical Analysis and Applications*, 433 (2016), 1459–1485.
[2] F. Andersson, M. Carlsson and K. M. Perfekt, Operator-Lipschitz estimates for the singular value functional calculus, *Proceedings of the American Mathematical Society*, 144 (2016), 1867–1875.
[3] F. Arrigo, M. Benzi and C. Fenu, Computation of generalized matrix functions, *SIAM Journal on Matrix Analysis and Applications*, 37 (2016), 836–860.
[4] J. L. Aurentz, A. P. Austin, M. Benzi and V. Kalantzis, Stable computation of generalized matrix functions via polynomial interpolation, *SIAM Journal on Matrix Analysis and Applications*, 40 (2019), 210–234.
[5] M. Benzi and R. Huang, Some matrix properties preserved by generalized matrix functions, *Special Matrices*, 7 (2019), 27–37.
[6] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.
[7] K. Braman, Third-order tensors as linear operators on a space of matrices, *Linear Algebra and its Applications*, 433 (2010), 1241–1253.
[8] R. H. F. Chan and X. Q. Jin, *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia, PA, 2007.
[9] X. Chen, H. Qi and P. Tseng, Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems, *SIAM Journal on Optimization*, 13 (2003), 960–985.
[10] X. Chen and P. Tseng, Non-interior continuation methods for solving semidefinite complementarity problems, *Mathematical Programming*, 95 (2003), 431–474.
[11] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Second edition, SIAM, Philadelphia, PA, 1990.
[12] S. Gandy, B. Recht and I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, *Inverse Problems*, 27 (2011), 025010, 19 pp.
[13] D. Goldfarb and Z. Qin, Robust low-rank tensor recovery: Models and algorithms, *SIAM Journal on Matrix Analysis and Applications*, 35 (2014), 225–253.
[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th edition, Johns Hopkins University Press, Baltimore, MD, 2013.
[15] N. Hao, M. E. Kilmer, K. Braman and R. C. Hoover, Facial recognition using tensor-tensor decompositions, *SIAM Journal on Imaging Sciences*, 6 (2013), 437–463.
[16] J. B. Hawkins and A. Ben-Israel, On generalized matrix functions, *Linear and Multilinear Algebra*, 1 (1973), 163–171.
[17] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, PA, 2008.
[18] A. Hjorungnes and D. Gesbert, Complex-valued matrix differentiation: Techniques and key results, *IEEE Transactions on Signal Processing*, 55 (2007), 2740–2746.
[19] Z. H. Huang and L. Qi, Formulating an n-person noncooperative game as a tensor complementarity problem, *Computational Optimization and Applications*, 66 (2017), 557–576.
[20] M. E. Kilmer, K. Braman, N. Hao and R. C. Hoover, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging, *SIAM Journal on Matrix Analysis and Applications*, 34 (2013), 148–172.
[21] M. E. Kilmer and C. D. Martin, Factorization strategies for third-order tensors, *Linear Algebra and its Applications*, 435 (2011), 641–658.
[22] J. Liu, P. Musialski, P. Wonka and J. Ye, Tensor completion for estimating missing values in visual data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35** (2013), 208–220.

[23] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin and S. Yan, Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 5249–5257.

[24] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin and S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42** (2020), 925–938.

[25] K. Lund, The tensor t-function: a definition for functions of third-order tensors, *Numerical Linear Algebra with Applications*, **27** (2020), e2288.

[26] C. D. Martin, R. Shafer and B. LaRue, An order-p tensor factorization with applications in imaging, *SIAM Journal on Scientific Computing*, **35** (2013), A474–A490.

[27] Y. Miao, L. Qi and Y. Wei, Generalized tensor function via the tensor singular value decomposition based on the T-product, *Linear Algebra and its Applications*, **590** (2020), 258–303.

[28] Y. Miao, L. Qi and Y. Wei, T-Jordan canonical form and T-Drazin inverse based on the T-product, *Communications on Applied Mathematics and Computation*, (2020).

[29] E. Newman, L. Horesh, H. Avron and M. E. Kilmer, Stable tensor neural networks for rapid deep learning, preprint, *arXiv:1811.05659*

[30] V. Noferini, A formula for the Fréchet derivative of a generalized matrix function, *SIAM Journal on Matrix Analysis and Applications*, **38** (2017), 434–457.

[31] R. F. Rinehart, The equivalence of definitions of a matric function, *American Mathematical Monthly*, **62** (1955), 395–414.

[32] T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, Heidelberg, 2009.

[33] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis and C. Faloutsos, Tensor decomposition for signal processing and machine learning, *IEEE Transactions on Signal Processing*, **65** (2017), 3551–3582.

[34] D. Sun and J. Sun, Semismooth matrix-valued functions, *Mathematics of Operations Research*, **27** (2002), 150–169.

[35] L. Yang, Z. H. Huang and X. Shi, A fixed point iterative method for low n-rank tensor pursuit, *IEEE Transactions on Signal Processing*, **61** (2013), 2952–2962.

[36] Z. Zhang and S. Aeron, Exact tensor completion using t-SVD, *IEEE Transactions on Signal Processing*, **65** (2017), 1511–1526.
[47] Z. Zhang, G. Ely, S. Aeron, N. Hao and M. E. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), 3842–3849.

[48] P. Zhou, C. Lu, Z. Lin and C. Zhang, Tensor factorization for low-rank tensor completion, IEEE Transactions on Image Processing, 27 (2018), 1152–1163.

Received February 2020; revised May 2020.

E-mail address: lixia666@tju.edu.cn
E-mail address: wang.yong@tju.edu.cn
E-mail address: huangzhenghai@tju.edu.cn