Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 19, No. 2, pp. 205–235, 2023

STOCHASTIC CONDITIONING OF TENSOR FUNCTIONS BASED ON THE TENSOR-TENSOR PRODUCT

Yun Miao, Tianru Wang* and Yimin Wei†

Abstract: The conditioning of matrix functions is one of the fundamental topics in linear algebra. In this paper, we extend matrix case to third order tensor functions based on the tensor-tensor product. We first give the deterministic perturbation bounds for the tensor Moore-Penrose inverse based on the tensor-tensor product and generalized them to T-Total Least Squares problem. Then, we discuss the bound of the stochastic perturbations of third order tensors. We investigate the stochastic conditioning problem for general tensor function if random noises are input. We define the Fréchet derivative of the generalized tensor function and give the upper bound of stochastic condition number and compare it with the deterministic condition number in the first order estimation. The stochastic conditioning will be better than the deterministic conditioning. Finally, we present a numerical example of the tensor least squares problem to show the effectiveness of our stochastic error estimate.

Key words: *tensor, optimization*

Mathematics Subject Classification: *15A18, 15A69, 65F15, 90C33*

---

1 Introduction

In numerical linear algebra [18], one of the fundamental topics is the sensitivity of matrix functions under the perturbation. A matrix function is a mapping $F : \Omega \to \mathbb{R}^{p\times q}$, where Ω is an open subset of $\mathbb{R}^{m\times n}$. The goal of perturbation analysis is to quantify the effect that uncertainties might have on the computation of matrix functions [2, 23]. When the perturbation is sufficiently small, the worst error bounds are explored by Higham and Al-Mohy [24].

As a special kind of the matrix function, the perturbation analysis upon the linear least squares problem, to find a vector $x \in \mathbb{R}^n$ subject to

$$\min_x \|b - Ax\|_2 ,$$

has been well developed. The error bounds of the linear least squares problem have been developed by Stewart [48, 49] and Wedin [58], recently improved in [12, 32]. Golub and Van Loan [19] establish the estimations to the total least squares problem $Ax \approx b$,

$$\min_{E,r} \|[E\ r]\|_F ,$$

*T. Wang is supported by the National Natural Science Foundation of China under grant 12271108.
†Corresponding author. Y. Wei is supported by Innovation Program of Shanghai Municipal Education Committee and the National Natural Science Foundation of China under grant 12271108.

$$\text{s.t. } b + r \in \mathcal{R}(A + E),$$

by the singular value decomposition (SVD). Randomized algorithms for total least squares problems are presented in [62].

The sensitivity analysis of the matrix function $F(A)$ is deterministic, which leads to the expression of first order estimation [16]

$$\|F(A + H) - F(A)\| \leq \kappa \|H\| + o(\|H\|),$$

where κ is the deterministic condition number defined as [20]

$$\kappa = \lim_{\delta \to 0} \frac{c_\delta}{\delta} = \lim_{\delta \to 0} \sup_{\|H\| \leq \delta} \frac{\|F(A + H) - F(A)\|}{\delta}.$$

Rice [46] shows that if F is Fréchet differentiable at A, then κ is the operator norm of the Fréchet derivative of F at A. However, few attempts have been made to cases when random noises exist. In fact, the random noise exists in many cases. Some automatic error analysis software perform random perturbations and the experimental data obtained by containing random errors which confirm certain kinds of distribution. Therefore, it is worthwhile to gain the error bounds by the stochastic analysis. We can trace the original idea to Turing [53], who first considers the function $F(A) = A^{-1}b$ for a nonsingular matrix $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. He ignores the $o(\|H\|)$ term of the high order error in the equation

$$(A + H)^{-1}b - A^{-1}b = -A^{-1}HA^{-1}b + o(\|H\|),$$

and gives an expression for the root-mean-square of $\left\|A^{-1}HA^{-1}b\right\|_2$. Stewart [50] considers general cases with random noises, and derives the expression of expectation of the error bound. Fletcher [16] obtains the similar results independently and provides the error expectation, when the input data satisfies the standard normal distribution. In recent years, Gratton and Titley-Peloquin [20] propose the stochastic condition number which determines the sensitivity of the matrix function with random noises. They utilize the stochastic condition number to present the first-order estimation (FOE) of a matrix function and compare it with deterministic cases to show the effectiveness or their stochastic error estimation. Breiding and Vannieuwenhoven [3] investigate the average condition number of tensor rank decompositions.

Unfortunately, the linear least squares problems and the error estimation theories have not been fully extended to tensor functions due to the multiplications of tensors have not been well-defined yet.

Recently, a lot of tensor multiplication methods and recent results come into the world [11, 43, 60]. There are two important kinds of products between tensors, which is the tensor Einstein product and the tensor-tensor product [29]. Under both kinds of products, the set of tensors forms a ring structure and inherits nice properties from matrices. The tensor Moore-Penrose inverse and tensor least squares problems have already been established for the Einstein product in [37, 51]. They also discuss about the minimum-norm least squares solution with the Einstein product. Few papers emphasize the conditioning of tensor functions based on the tensor-tensor product.

The tensor-tensor product is introduced by Kilmer et al. [29], which has been proved to be of great help in many areas, such as, the image processing [29, 38, 47, 52], the computer vision [22, 63], the signal processing [5, 34, 35], the low rank tensor recovery and robust tensor PCA [31, 34], the data completion and denoising [15, 26, 35, 54] and random tensors [6, 7, 8, 9]. An approach of linearization is provided by the tensor-tensor product to transfer tensor multiplication, to the matrix multiplication by the discrete Fourier transformation and the block circulant matrices [4, 28]. Due to the importance of the tensor-tensor product, Lund [36] gives the definition of tensor functions based on the tensor-tensor product of third order F-square tensors in her Ph.D thesis in 2018. The definition of T-function is given by

$$f^{\diamondsuit}(\mathcal{A}) = \mathrm{fold}(f(\mathrm{bcirc}(\mathcal{A}))\widehat{E_1}^{np\times n}),$$

where 'bcirc($\mathcal{A}$)' is the block circulant matrix [28] defined by the F-square tensor $\mathcal{A} \in \mathbb{C}^{n\times n\times p}$ and $\widehat{E_1}^{np\times n} = \hat{e}_k^p \otimes I_n$, where $\hat{e}_k^p \in \mathbb{C}^p$ is the vector of all zeros except for the k-th entry and I_n is the identity matrix, '$\otimes$' is the matrix Kronecker product [25].

The T-function has been proved to be useful in stable tensor neural networks for rapid deep learning [42]. Special kinds of T-function, such as tensor power used in Arnoldi methods to compute the tensor eigenvalues and diagonal tensor canonical form [17] is also proposed. Miao, Qi and Wei [40, 41] investigate the generalized tensor functions, the tensor Jordan canonical forms and the tensor generalized inverses, which gives the classification of tensors based on the tensor-tensor product. The tensor neural network models based on the tensor singular value decomposition (T-SVD) is presented in [57]. Quantum tensor singular value decomposition with applications to recommendation systems is given in [56]. Xu et al. [64] developed tensor-tensor product based nonlocal tensor sparse representation model

$$\min_{\mathcal{D},\mathcal{S}} \left\{ \frac{1}{2} \|\mathcal{X} - \mathcal{D} * \mathcal{S}\|_F^2 + \lambda \|\mathcal{S}\|_1 \right\}.$$

for hyperspectral image super-resolution. Ekanadham et al. [14] also used L_1 regularization method to solve automatic neural spike identification problem. Tensor Tikhonov methods for ill-posed problems with the tensor-tensor product structure can be found in [1, 21, 44, 45].

In this paper, we dedicate to investigate the conditioning of tensor functions based on the tensor-tensor product and show the effectiveness of stochastic condition number. In the preliminaries, we recall the basic notations and the definition of tensor-tensor product and give the definition of tensor norms, tensor rank, tensor range space and tensor null space based on the tensor-tensor product. Properties of standard tensor functions are collected. In the main part, we discuss the bound of the stochastic perturbation for third order tensors. As an application, we obtain the perturbation bounds for the solution to the T-least squares problem and the T-Total least squares problem based on the tensor-tensor product. The stochastic conditioning problem of general tensor T-functions is taken into consideration. We give the definition of Fréchet derivative of tensor functions. Then we obtain the estimation of the upper bound of stochastic condition number by the basic probability inequalities. Comparisons are made with the deterministic condition number in the first-order estimation (FOE) of tensor T-functions in stochastic cases. To illustrate the above results, a numerical test is presented to compare the FOE of the tensor least squares problem by using the deterministic condition number with the stochastic condition number.

2 Notation and Preliminaries

2.1 Notation

A new concept is proposed by Kilmer et al. [29, 30] for multiplying third order tensors, viewing a tensor as a stack of frontal slices. Suppose that we have two tensors $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ and $\mathcal{B} \in \mathbb{R}^{n\times s\times p}$ and denote their frontal slices respectively as $A^{(k)} \in \mathbb{R}^{m\times n}$ and $B^{(k)} \in$

$\mathbb{R}^{n\times s}$, $(k = 1, 2, \ldots, p)$. We can also define the operations bcirc, unfold and fold as [22, 29, 30],

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

and $\text{fold}(\text{unfold}(\mathcal{A})) := \mathcal{A}$.

We can define the inverse operation $\text{bcirc}^{-1} : \mathbb{R}^{mp\times np} \to \mathbb{R}^{m\times n\times p}$ such that $\text{bcirc}^{-1}(\text{bcirc}(\mathcal{A})) = \mathcal{A}$.

In view of the large number of symbols used in this paper, the meanings of the main symbols are listed in the following table.

Symbols	Meaning
$*$	T-product
bcirc	block circulant operator
F_p	DFT matrix
$\otimes$	matrix Kronecker product
$\mathcal{R}(\cdot)$	Range space of tensors
$\mathcal{N}(\cdot)$	Null space of tensors
$\mathbb{E}\{\cdot\}$	Mathematical expectation
$\|\|\cdot\|\|_S$	Tensor stochastic norm
Prob$(\cdot)$	Probability
Med$(\cdot)$	Median of random variable
$\mathcal{J}$	Jacobi tensor

Table 1: Notations and Symbols

2.2 The Tensor-Tensor Product

The following definitions and properties are adopted in [22, 29, 30].

Definition 2.1. (Tensor-tensor product) Let $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ and $\mathcal{B} \in \mathbb{R}^{n\times s\times p}$ be two real tensors. Then the tensor-tensor product $\mathcal{A} * \mathcal{B}$ is an $m \times s \times p$ real tensor defined by

$$\mathcal{A} * \mathcal{B} := \text{fold}(\text{bcirc}(\mathcal{A})\text{unfold}(\mathcal{B})).$$

We introduce definitions of transpose, identity and orthogonal of tensors as follows.

Definition 2.2. (Transpose and conjugate transpose) If $\mathcal{A}$ is a third order tensor of size $m \times n \times p$, then the transpose $\mathcal{A}^\top$ is obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n. The conjugate transpose $\mathcal{A}^H$ is obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n.

Definition 2.3. (Identity tensor) The $n \times n \times p$ identity tensor $\mathcal{I}_{nnp}$ is the tensor whose first frontal slice is the $n \times n$ identity matrix, and whose other frontal slices are all zeros.

It is easy to check that $\mathcal{A} * \mathcal{I}_{nnp} = \mathcal{I}_{mmp} * \mathcal{A} = \mathcal{A}$ for $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$.

Definition 2.4. (Orthogonal and unitary tensors) An $n \times n \times p$ real-valued tensor $\mathcal{P}$ is orthogonal if $\mathcal{P}^\top * \mathcal{P} = \mathcal{P} * \mathcal{P}^\top = \mathcal{I}$. An $n \times n \times p$ complex-valued tensor $\mathcal{Q}$ is unitary if $\mathcal{Q}^H * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^H = \mathcal{I}$.

For a frontal square tensor $\mathcal{A}$ of size $n \times n \times p$, it has inverse tensor $\mathcal{B}$ $(= \mathcal{A}^{-1})$, provided that

$$\mathcal{A} * \mathcal{B} = \mathcal{I}_{nnp} \quad \textit{and} \quad \mathcal{B} * \mathcal{A} = \mathcal{I}_{nnp}.$$

It should be noticed that invertible third order tensors of size $n \times n \times p$ forms a group, since the invertibility of tensor $\mathcal{A}$ is equivalent to the invertibility of the matrix $\mathrm{bcirc}(\mathcal{A})$, and the set of invertible matrices forms a group. Also, the orthogonal tensors via the tensor-tensor product also forms a group, since $\mathrm{bcirc}(\mathcal{Q})$ is an orthogonal matrix.

The concept of T-range space, T-null space, tensor norm, and T-Moore-Penrose inverse are defined as follows [40].

Definition 2.5. Let $\mathcal{A}$ be an $m \times n \times p$ real-valued tensor.
(1) The T-range space of $\mathcal{A}$, $\mathcal{R}(\mathcal{A}) := \mathrm{Ran}((F_p \otimes I_m)\mathrm{bcirc}(\mathcal{A})(F_p^H \otimes I_n))$, 'Ran' means the range space of the matrix,
(2) The T-null space of $\mathcal{A}$, $\mathcal{N}(\mathcal{A}) := \mathrm{Null}((F_p \otimes I_m)\mathrm{bcirc}(\mathcal{A})(F_p^H \otimes I_n))$, 'Null' represents the null space of the matrix,
(3) The tensor norm $\|\mathcal{A}\| := \|\mathrm{bcirc}(\mathcal{A})\|$,
(4) The tensor Moore-Penrose inverse $\mathcal{A}^\dagger = \mathrm{bcirc}^{-1}((\mathrm{bcirc}(\mathcal{A}))^\dagger)$.

In detail, let $x \in \mathbb{R}^{m \times 1 \times p}$ and $\mathcal{A} \in \mathbb{R}^{m \times n \times p}$ be two real tensors. They have the following factorization, respectively,

$$\begin{cases} \mathrm{bcirc}(x) \\ = (F_p^H \otimes I_m) \begin{bmatrix} \mathbf{x}_1 & & & \\ & \mathbf{x}_2 & & \\ & & \ddots & \\ & & & \mathbf{x}_p \end{bmatrix} (F_p \otimes 1) = (F_p^H \otimes I_m)\mathrm{diag}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}(F_p \otimes 1), \\ \mathrm{bcirc}(\mathcal{A}) \\ = (F_p^H \otimes I_m) \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p \end{bmatrix} (F_p \otimes I_n) = (F_p^H \otimes I_m)\mathrm{diag}\{A_1, A_2, \dots, A_p\}(F_p \otimes I_n), \end{cases}$$

where F_n is the discrete Fourier matrix of size $n \times n$, which is defined as [4]

$$F_n = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix},$$

where $\omega = e^{-2\pi \mathbf{i}/n}$ is the primitive n-th root of unity in which $\mathbf{i} = \sqrt{-1}$. F_p^H is the conjugate transpose of F_p. $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p \in \mathbb{C}^m$, $A_1, A_2, \dots, A_p \in \mathbb{C}^{m \times n}$.

Then

$$\|x\|_2 = \max_{1\le i\le p} \|\mathbf{x}_i\|_2\,, \quad \|\mathcal{A}\|_F = \sqrt{\sum_{i=1}^{p} \|A_i\|_F^2}, \quad \|\mathcal{A}\|_2 = \max_{1\le i\le p} \|A_i\|_2\,,$$

where $\|\mathbf{x}_i\|_2 = \left(\sum_{j=1}^{m} (\mathbf{x}_i)_j^2\right)^{1/2}$ and $\|A_i\|_2 = \max\limits_{\|y\|_2=1} \|A_i y\|_2\,.$

It is easy to check that $\|\mathcal{A}\|_2 = \|\mathrm{bcirc}(\mathcal{A})\|_2$ and $\|\mathcal{A}\|_F = \|\mathrm{bcirc}(\mathcal{A})\|_F$ due to the fact that the discrete Fourier transformation is a unitary transformation.

Remark 2.6. It should be noticed that

$$\|\mathcal{A}\|_F^2 \neq \sum_{i=1}^{p} \left\|A^{(i)}\right\|_F^2,$$

where $A^{(i)}$ are the frontal slices of $\mathcal{A}$, since $\|\mathcal{A}\|_F^2 := \|\mathrm{bcirc}(\mathcal{A})\|_F = p\sum_{i=1}^{p} \left\|A^{(i)}\right\|_F^2.$

It indicates the Frobenius norm of a third order tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ is defined to be p times the sum of the squares of all entries.

We call $x \in \mathbb{R}^{m\times 1\times p}$ is in the range space of $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$, denoted by $x \in \mathcal{R}(\mathcal{A})$, if and only if for all $\mathbf{x}_i \in \mathcal{R}(A_i)$, $i = 1, 2, \ldots, p$.

Similarly, we call $x \in \mathbb{R}^{m\times 1\times p}$ is in the null space of $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$, denoted by $x \in \mathcal{N}(\mathcal{A})$, if and only if for all $\mathbf{x}_i \in \mathcal{N}(A_i)$, $i = 1, 2, \ldots, p$.

Remark 2.7. We call a third order tensor $\mathcal{A}$ is T-full column (row) rank if and only if each A_i is of full column (row) rank.

Remark 2.8. We call a third order F-square tensor $\mathcal{A}$ is T-nonsingular if and only if each square matrix A_i is nonsingular.

The definition of tensor rank based on tensor-tensor product is given as follows.

Definition 2.9. (Tensor rank) Let $\mathcal{A}$ be an $m \times n \times p$ real-valued tensor. If we have the factorization,

$$(F_p \otimes I_m)\mathrm{bcirc}(\mathcal{A})(F_p^H \otimes I_n) = \mathrm{diag}\{A_1, A_2, \ldots, A_p\},$$

then we call the rank of $\mathcal{A}$ is

$$\mathrm{rank}(\mathcal{A}) = (r_1, r_2, \ldots, r_p),$$

where $r_i = \mathrm{rank}(A_i)$, $i = 1, 2, \ldots, p$.

Especially, if $r_1 = r_2 = \cdots = r_p = r$, we call the rank of $\mathcal{A}$ is r, denoted by $\mathrm{rank}(\mathcal{A}) = r$.

Che and Wei [10] present the randomized algorithms for the approximations of Tucker and the tensor train decompositions.

2.3 Tensor T-Function

By using the tensor-tensor product, the matrix function can be generalized to tensors of size $n \times n \times p$. Assume that we have tensors $\mathcal{A} \in \mathbb{C}^{n\times n\times p}$ and $\mathcal{B} \in \mathbb{C}^{n\times s\times p}$, then the tensor T-function of $\mathcal{A}$ is defined by [36]

$$f(\mathcal{A}) * \mathcal{B} := \mathrm{fold}(f(\mathrm{bcirc}(\mathcal{A})) \cdot \mathrm{unfold}(\mathcal{B})),$$

or equivalently

$$f(\mathcal{A}) := \text{fold}(f(\text{bcirc}(\mathcal{A}))\widehat{E_1}^{np\times n}).$$

There is another way to express $\widehat{E_1}^{np\times n}$:

$$\widehat{E_1}^{np\times n} = \begin{bmatrix} I_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \otimes I_n = \text{unfold}(\mathcal{I}_{n\times n\times p}).$$

Note that f on the right-hand side of the equation is merely the matrix function defined above, the T-function is well defined.

From this definition, we could see that for a tensor $\mathcal{A} \in \mathbb{C}^{n\times n\times p}$, bcirc($\mathcal{A}$) is a block circulant matrix of size $np \times np$. The frontal faces of $\mathcal{A}$ are the block entries of $A\widehat{E_1}^{np\times n}$, then $\mathcal{A} = \text{fold}(A\widehat{E_1}^{np\times n})$, where $A = \text{unfold}(\mathcal{A})$.

To obtain further properties of generalized tensor functions, we make reviews of the results on block circulant matrices and the tensor-tensor product.

Lemma 2.10 ([4])**.** *Suppose that $A, B \in \mathbb{C}^{np\times np}$ are block circulant matrices with $n \times n$ blocks. Let $\{\alpha_j\}_{j=1}^k$ be scalars. Then $A^\top$, A^H, $\alpha_1 A + \alpha_2 B$, AB, $q(A) = \sum_{j=1}^k \alpha_j A^j$ and A^{-1} are also block circulant.*

Lemma 2.11 ([36])**.** *Suppose that tensors $\mathcal{A} \in \mathbb{C}^{n\times n\times p}$ and $\mathcal{B} \in \mathbb{C}^{n\times s\times p}$. Then*
(1) $\text{unfold}(\mathcal{A}) = \text{bcirc}(\mathcal{A})\widehat{E_1}^{np\times n}$,
(2) $\text{bcirc}(\text{fold}(\text{bcirc}(\mathcal{A})\widehat{E_1}^{np\times n})) = \text{bcirc}(\mathcal{A})$,
(3) $\text{bcirc}(\mathcal{A} * \mathcal{B}) = \text{bcirc}(\mathcal{A})\text{bcirc}(\mathcal{B})$,
(4) $\text{bcirc}(\mathcal{A})^j = \text{bcirc}(\mathcal{A}^j)$, *for all $j = 0, 1, \ldots$, where $\mathcal{A}^j$ denotes the j-th power of $\mathcal{A}$ under T-product.*
(5) $(\mathcal{A} * \mathcal{B})^H = \mathcal{B}^H * \mathcal{A}^H$,
(6) $\text{bcirc}(\mathcal{A}^\top) = (\text{bcirc}(\mathcal{A}))^\top$, $\text{bcirc}(\mathcal{A}^H) = (\text{bcirc}(\mathcal{A}))^H$.

3 Main Results

3.1 Upper-bounds for $\left\|\overline{\mathcal{A}}^\dagger\right\|$

The pseudo-inverse [55, 61] (or Moore-Penrose inverse) of a third order tensor $\mathcal{A}$ based on the tensor-tensor product can be defined as the unique tensor $\mathcal{A}^\dagger$ satisfying the following four equations,

$$\mathcal{A}^\dagger * \mathcal{A} * \mathcal{A}^\dagger = \mathcal{A}^\dagger, \quad \mathcal{A} * \mathcal{A}^\dagger * \mathcal{A} = \mathcal{A}, \quad (\mathcal{A} * \mathcal{A}^\dagger)^H = \mathcal{A} * \mathcal{A}^\dagger, \quad (\mathcal{A}^\dagger * \mathcal{A})^H = \mathcal{A}^\dagger * \mathcal{A}. \tag{3.1}$$

The tensor Moore-Penrose inverse and the generalized tensor functions of third order tensors based on the tensor-tensor product have been investigated by Miao, Qi and Wei [40]. It is proved that if a third order tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ has T-SVD [29] as

$$\mathcal{A} = \mathcal{U} * \Sigma * \mathcal{V}^\top,$$

then the unique T-Moore-Penrose inverse could be expressed by $\mathcal{A}^\dagger = \mathcal{V} * \Sigma^\dagger * \mathcal{U}^\top$.

The orthogonal projections to the subspaces $\mathcal{R}(\mathcal{A})$ and $\mathcal{R}(\mathcal{A}^\top)$ are defined as

$$\mathcal{P}_\mathcal{A} = \mathcal{A} * \mathcal{A}^\dagger, \quad \mathcal{R}_\mathcal{A} = \mathcal{A}^\dagger * \mathcal{A}.$$

Let $\overline{\mathcal{A}} = \mathcal{A}+\Delta\mathcal{A}$ be the perturbed tensor from the original tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$. Pursuing the upper bound for $\|\overline{\mathcal{A}}^\dagger\|$ plays an important role in the perturbation behaviour of the tensor least square problems [17, 40].

Since an arbitrarily small perturbation of a tensor can change the T-rank of $\mathcal{A}$, which will cause arbitrarily large perturbation of the T-Moore-Penrose inverse, it is no wonder that we should put some constraints of the way that $\mathcal{A}$ is perturbed. Stewart [48, 49] and Wedin [58] independently constrained the acute perturbation and the similar idea could be put on the tensor cases.

Definition 3.1 (Acute perturbation of tensors)**.** A tensor $\overline{\mathcal{A}} = \mathcal{A}+\Delta\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ is called an acute perturbation of tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ if $\|\mathcal{P}_\mathcal{A} - \mathcal{P}_{\overline{\mathcal{A}}}\|_2 < 1$ and $\|\mathcal{R}_\mathcal{A} - \mathcal{R}_{\overline{\mathcal{A}}}\|_2 < 1$.

Definition 3.2 (Stable perturbation of tensors)**.** A tensor $\overline{\mathcal{A}} = \mathcal{A}+\Delta\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ is called a stable perturbation of tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ if $\mathcal{R}(\overline{\mathcal{A}}) \cap \mathcal{R}^\perp(\mathcal{A}) = \{0\}$, where $\mathcal{R}^\perp(\mathcal{A})$ denotes the orthogonal complementary subspace of $\mathcal{R}(\mathcal{A})$.

Similar to the matrix cases, it can be proved that if $\overline{\mathcal{A}}$ is an acute perturbation of $\mathcal{A}$, then $\overline{\mathcal{A}}$ is a stable perturbation of $\mathcal{A}$. But $\mathcal{A}$ is not necessarily an acute perturbation of $\overline{\mathcal{A}}$, when $\overline{\mathcal{A}}$ is a stable perturbation of $\mathcal{A}$. Fortunately, if $\|\Delta\mathcal{A}\|_2$ is small enough, then the stable perturbation implies acute perturbation, as the following lemma.

Lemma 3.3. *Let $\overline{\mathcal{A}} = \mathcal{A}+\Delta\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ be a perturbation of $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$. If*

$$\left\|\mathcal{A}^\dagger\right\|_2 \|\Delta\mathcal{A}\|_2 < 1,$$

then the following statements are equivalent:
(1) *$\overline{\mathcal{A}}$ is a stable perturbation of $\mathcal{A}$.*
(2) *$\overline{\mathcal{A}}$ is an acute perturbation of $\mathcal{A}$.*
(3) $\mathrm{rank}(\mathcal{A}) = \mathrm{rank}(\overline{\mathcal{A}})$.

As a generalization of the estimation of Stewart [48], we obtain the following theorem.

Theorem 3.4. *Let $\overline{\mathcal{A}} = \mathcal{A}+\Delta\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ be an acute perturbation of $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ and $\left\|\mathcal{A}^\dagger\right\|_2 \|\Delta\mathcal{A}\|_2 < 1$, then*

$$\left\|\overline{\mathcal{A}}^\dagger\right\|_2 \leq \frac{\left\|\mathcal{A}^\dagger\right\|_2}{1 - \|\mathcal{A}^\dagger\|_2 \|\Delta\mathcal{A}\|_2}. \tag{3.2}$$

Proof. Without loss of generality, we only prove this theorem with the tensor two-norm. By the definition of the tensor norm,

$$\begin{aligned}\|\mathcal{A}\|_2 = \|\mathrm{bcirc}(\mathcal{A})\|_2 &= \left\|(F_p^H \otimes I_m)\mathrm{bcirc}(\mathcal{A})(F_p \otimes I_n)\right\|_2 \\ &= \|\mathrm{diag}\{A_1, A_2, \ldots, A_p\}\|_2 \\ &= \max_{1\le i\le p} \|A_i\|_2 .\end{aligned}$$

Since $\mathrm{bcirc}(\mathcal{A}^\dagger) = (F_p^H \otimes I_m)\mathrm{diag}\{A_1^\dagger, A_2^\dagger, \ldots, A_p^\dagger\}(F_p \otimes I_n)$, we obtain that

$$\left\|\mathcal{A}^\dagger\right\|_2 = \left\|\mathrm{bcirc}(\mathcal{A}^\dagger)\right\|_2 = \max_{1\le i\le p} \left\|A_i^\dagger\right\|_2 .$$

On the other hand, $\overline{\mathcal{A}} = \mathcal{A} + \Delta\mathcal{A}$,

$$\mathrm{bcirc}(\overline{\mathcal{A}}) = (F_p^H \otimes I_m)\mathrm{diag}\{\overline{A_1}, \overline{A_2}, \ldots, \overline{A_p}\}(F_p \otimes I_n).$$

By Lemma 2.11, we have

$$\begin{aligned}\mathrm{bcirc}(\overline{\mathcal{A}}) &= \mathrm{bcirc}(\mathcal{A} + \Delta\mathcal{A}) = \mathrm{bcirc}(\mathcal{A}) + \mathrm{bcirc}(\Delta\mathcal{A}) \\ &= (F_p^H \otimes I_m)\left(\mathrm{diag}\{A_1, A_2, \ldots, A_p\} + \mathrm{diag}\{\Delta A_1, \Delta A_2, \ldots, \Delta A_p\}\right)(F_p \otimes I_n),\end{aligned}$$

which is equivalent to $\overline{A_i} = A_i + \Delta A_i$, $(i = 1, 2, \ldots, p)$.

By the results of Stewart [48], it turns out that

$$\left\|\overline{\mathcal{A}}^{\dagger}\right\|_2 = \left\|\mathrm{bcirc}(\overline{\mathcal{A}})\right\|_2 = \max_{1\le i\le p}\left\|\overline{A_i}^{\dagger}\right\|_2 \le \max_{1\le i\le p}\frac{\left\|A_i^{\dagger}\right\|_2}{1 - \left\|A_i^{\dagger}\right\|_2\|\Delta A_i\|_2}.$$

The condition $0 < \left\|\mathcal{A}^{\dagger}\right\|_2\|\Delta\mathcal{A}\|_2 < 1$ shows that

$$0 < \left\|A_i^{\dagger}\right\|_2\|\Delta A_i\|_2 < 1, \quad i = 1, 2, \ldots, p.$$

Then we have

$$\left\|\overline{\mathcal{A}}^{\dagger}\right\|_2 \le \frac{\max\limits_{1\le i\le p}\left\|A_i^{\dagger}\right\|_2}{1 - \max\limits_{1\le i\le p}\left\|A_i^{\dagger}\right\|_2\|\Delta A_i\|_2} = \frac{\left\|\mathcal{A}^{\dagger}\right\|_2}{1 - \|\mathcal{A}^{\dagger}\|_2\|\Delta\mathcal{A}\|_2}.$$

□

We can also obtain the sharper upper bound for $\left\|\overline{\mathcal{A}}^{\dagger}\right\|_2$ as follows,

$$\left\|\overline{\mathcal{A}}^{\dagger}\right\|_2 \le \mu\frac{\|\mathcal{A}^{\dagger}\|_2}{1 - \left\|\mathcal{A}_2^{\dagger}\right\|\|\Delta\mathcal{A}\|_2} \quad \text{with} \quad \mu \le 1.$$

Furthermore,

$$\mu < 1 \quad \text{if and only if} \quad \mathcal{R}(\overline{\mathcal{A}}) \cap \mathcal{R}(\mathcal{A}) = \{0\} \quad \text{and} \quad \mathcal{R}(\overline{\mathcal{A}}^{\top}) \cap \mathcal{R}(\mathcal{A}^{\top}) = \{0\}.$$

The proof can be referred to Li, Xu, and Wei [32, Theorem 2].

3.2 T-Total Least Squares Problem

The tensor T-Least Squares problem is to minimize $\|\mathcal{A} * x - b\|_2$, which can be rewrite as follows [18, 63]:

$$\min_{b+r\in\mathcal{R}(\mathcal{A})}\|r\|_2.$$

By using the generalized inverses of tensors, Jin et al. [27] obtained the explicit expression of the solution to the tensor least squares. If there is error on the tensor $\mathcal{A}$, it is natural to consider the following problem:

$$\min_{b+r\in\mathcal{R}(\mathcal{A}+\mathcal{E})}\|[\mathcal{E}\ \ r]\|_F. \tag{3.3}$$

This is the tensor T-Total Least Squares problem (T-TLS). If we can get the solution which can minimize the Frobenius norm of $[\mathcal{E}_0 \;\; r_0]$, then we call the tensor x_0 satisfying $(\mathcal{A}+\mathcal{E}_0)*x_0 = b+r_0$ to be the solution of the T-TLS problem. By using the Frobenius norm, Golub and Van Loan [19] generalized the matrix TLS problem. For tensors, we can introduce the weight tensors $\mathcal{D} \in \mathbb{R}^{m\times m\times p}, \mathcal{T} \in \mathbb{R}^{(n+k)\times(n+k)\times p}$, which are F-diagonal reversible tensors. Then we can raise the multidimensional T-TLS problem [39, 65] as follows:

$$\min_{\mathcal{B}+\mathcal{R}\in\mathcal{R}(\mathcal{A}+\mathcal{E})} \|\mathcal{D} * [\mathcal{E} \;\; \mathcal{R}] * \mathcal{T}\|_F ,$$

Here $\mathcal{E} \in \mathbb{R}^{m\times n\times p}$ and $\mathcal{R} \in \mathbb{R}^{m\times k\times p}$. If $[\mathcal{E}_0 \;\; \mathcal{R}_0]$ is the solution to the above problem, then any tensor $\mathcal{X} \in \mathbb{R}^{n\times k\times p}$ satisfying

$$(\mathcal{A} + \mathcal{E}_0) * \mathcal{X} = \mathcal{B} + \mathcal{R}_0 \tag{3.4}$$

is called the solution to the T-TLS problem. We can use the T-SVD [30] to solve this problem as the following theorem.

Theorem 3.5. *Let $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ and $\mathcal{B} \in \mathbb{R}^{m\times k\times p}$ be third order tensors. $\mathcal{D} \in \mathbb{R}^{m\times m\times p}$ and $\mathcal{T} \in \mathbb{R}^{(n+k)\times(n+k)\times p}$ are two reversible F-diagonal tensors. if $m \geq n+k$ and the T-SVD of*

$$\mathcal{C} = \mathcal{D} * [\mathcal{A} \;\; \mathcal{B}] * \mathcal{T} = [\mathcal{C}_1 \;\; \mathcal{C}_2], \quad \mathcal{C}_1 \in \mathbb{R}^{m\times n\times p}, \quad \mathcal{C}_2 \in \mathbb{R}^{m\times k\times p},$$

takes the form $\mathcal{U}^\top * \mathcal{C} * \mathcal{V} = \Sigma$, where $\mathcal{U}$, $\mathcal{V}$, Σ can be written as:

$$\begin{cases} \mathcal{U} = [\mathcal{U}_1 \;\; \mathcal{U}_2], \quad \mathcal{U}_1 \in \mathbb{R}^{m\times n\times p}, \quad \mathcal{U}_2 \in \mathbb{R}^{m\times k\times p}, \\ \mathcal{V} = \begin{bmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} \\ \mathcal{V}_{21} & \mathcal{V}_{22} \end{bmatrix}, \quad \mathcal{V}_{11} \in \mathbb{R}^{n\times n\times p},\ \mathcal{V}_{12} \in \mathbb{R}^{n\times k\times p},\ \mathcal{V}_{21} \in \mathbb{R}^{k\times n\times p},\ \mathcal{V}_{22} \in \mathbb{R}^{k\times k\times p}, \\ \Sigma = \begin{bmatrix} \Sigma_1 & \mathcal{O} \\ \mathcal{O} & \Sigma_2 \end{bmatrix}, \quad \Sigma_1 \in \mathbb{R}^{n\times n\times p},\ \Sigma_2 \in \mathbb{R}^{k\times k\times p}. \end{cases}$$

Suppose tensor $\mathcal{C}$ satisfies

$$\mathrm{bcirc}(\mathcal{C}) = (F_p^H \otimes I_m)\mathrm{diag}\{C_1, C_2, \dots, C_p\}(F_p \otimes I_{n+k}),$$

and the singular values of matrix C_i are $\sigma_1^{(i)}(\mathcal{C}), \sigma_2^{(i)}(\mathcal{C}) \dots, \sigma_{n+k}^{(i)}(\mathcal{C})$, $i = 1, 2, \dots, p$. Meanwhile,

$$\mathrm{bcirc}(\mathcal{C}_1) = (F_p^H \otimes I_m)\mathrm{diag}\{C_1^1, C_2^1, \dots, C_p^1\}(F_p \otimes I_n),$$

and the singular values of C_i^1 are $\sigma_1^{(i)}(\mathcal{C}_1), \sigma_2^{(i)}(\mathcal{C}_1) \dots, \sigma_n^{(i)}(\mathcal{C}_1)$, $i = 1, 2, \dots, p$.

If $i = 1, 2, \dots p$, we have $\sigma_n^{(i)}(\mathcal{C}_1) > \sigma_{n+1}^{(i)}(\mathcal{C})$, then the block tensor $[\mathcal{E}_0 \;\; \mathcal{R}_0]$ is the solution to the equation

$$\mathcal{D} * [\mathcal{E}_0 \;\; \mathcal{R}_0] * \mathcal{T} = -\mathcal{U}_2 * \Sigma_2 * [\mathcal{V}_{12}^\top \;\; \mathcal{V}_{22}^\top].$$

Moreover, if the F-diagonal tensor $\mathcal{T}$ can be written as

$$\mathcal{T} = \begin{bmatrix} \mathcal{T}_1 & \mathcal{O} \\ \mathcal{O} & \mathcal{T}_2 \end{bmatrix}, \quad \mathcal{T}_1 \in \mathbb{R}^{n\times n\times p},\ \mathcal{T}_2 \in \mathbb{R}^{k\times k\times p},$$

then the tensor

$$\mathcal{X}_{TLS} = -\mathcal{T}_1 * \mathcal{V}_{12} * \mathcal{V}_{22}^{-1} * \mathcal{T}_2^{-1} \tag{3.5}$$

exists, and it is the unique solution to the T-TLS problem $(\mathcal{A} + \mathcal{E}_0) * \mathcal{X} = \mathcal{B} + \mathcal{R}_0$.

Proof. From equation $\mathcal{C} * \mathcal{V} = \mathcal{U} * \Sigma$, we get

$$\mathcal{C}_1 * \mathcal{V}_{12} + \mathcal{C}_2 * \mathcal{V}_{22} = \mathcal{U}_2 * \Sigma_2.$$

We need to prove $\mathcal{V}_{22}$ is a reversible tensor. Suppose there exists a tensor $x \in \mathbb{R}^{k\times 1\times p}$, $\|x\|_2 = 1$, satisfying

$$\mathcal{V}_{22} * x = \mathcal{O}.$$

From

$$\mathcal{V}_{12}^{\top} * \mathcal{V}_{12} + \mathcal{V}_{22}^{\top} * \mathcal{V}_{22} = \mathcal{I}$$

we have $\|\mathcal{V}_{12} * x\|_2 = 1$. However

$$\sigma_{n+1}^{(i)}(\mathcal{C}) \geq \|\mathcal{U}_2 * \Sigma_2 * x\|_2 = \|\mathcal{C}_1 * \mathcal{V}_{12} * x\|_2 \geq \sigma_n^{(i)}(\mathcal{C}),$$

which comes to a contradiction. Thus $\mathcal{V}_{22}$ must be a reversible tensor. From the seperation properties of singular values, we have $\sigma_n^{(i)}(\mathcal{C}) \geq \sigma_n^{(i)}(\mathcal{C}_1)$. Therefore,

$$\sigma_n^{(i)}(\mathcal{C}) \geq \sigma_n^{(i)}(\mathcal{C}_1) > \sigma_{n+1}^{(i)}(\mathcal{C}).$$

Now we begin to prove our theorem. If $\mathcal{R}(\mathcal{B}+\mathcal{R}) \subseteq \mathcal{R}(\mathcal{A}+\mathcal{E})$, then there exists a tensor $\mathcal{X} \in \mathbb{R}^{n\times k\times p}$ satisfying $(\mathcal{A}+\mathcal{E}) * \mathcal{X} = \mathcal{B}+\mathcal{R}$, that is

$$\{\mathcal{D} * [\mathcal{A} \ \ \mathcal{B}] * \mathcal{T} + \mathcal{D} * [\mathcal{E} \ \ \mathcal{R}] * \mathcal{T}\} * \mathcal{T}^{-1} * \begin{bmatrix} \mathcal{X} \\ -\mathcal{I}_k \end{bmatrix} = \mathcal{O}.$$

Therefore, the rank of the tensor in the brace is at most n. On the other hand, we have

$$\|\mathcal{D} * [\mathcal{E} \ \ \mathcal{R}] * \mathcal{T}\|_F^2 \geq \sum_{j=n+1}^{n+k} \sigma_j^{(i)}(\mathcal{C})^2.$$

Moreover, equality can be achieved when $[\mathcal{E} \ \ \mathcal{R}] = [\mathcal{E}_0 \ \ \mathcal{R}_0]$. By using the seperation property of singular values, we have $\sigma_n^{(i)}(\mathcal{C}) > \sigma_{n+1}^{(i)}(\mathcal{C})$. Thus we have $[\mathcal{E}_0 \ \ \mathcal{R}_0]$ is the unique solution to let the equality to be achieved.

In order to get to the solution $\mathcal{X}_{TLS}$, first it can be observed that the kernel of the tensor

$$\{\mathcal{D} * [\mathcal{A} \ \ \mathcal{B}] * \mathcal{T} + \mathcal{D} * [\mathcal{E} \ \ \mathcal{R}] * \mathcal{T}\} = \mathcal{U}_1 * \Sigma_1 * [\mathcal{V}_{11}^{\top} \ \ \mathcal{V}_{21}^{\top}]$$

is the range space of the tensor $\begin{bmatrix} \mathcal{V}_{12} \\ \mathcal{V}_{22} \end{bmatrix}$. Therefore, there exists a tensor $\mathcal{S} \in \mathbb{R}^{k\times k\times p}$ satisfying

$$\mathcal{T}^{-1} * \begin{bmatrix} \mathcal{X} \\ -\mathcal{I}_k \end{bmatrix} = \begin{bmatrix} \mathcal{V}_{12} \\ \mathcal{V}_{22} \end{bmatrix} * \mathcal{S}.$$

From equations $\mathcal{T}_1^{-1} * \mathcal{X} = \mathcal{V}_{12} * \mathcal{S}$ and $-\mathcal{T}_2^{-1} = \mathcal{V}_{22} * \mathcal{S}$, it can be obtained that

$$\mathcal{S} = -\mathcal{V}_{22}^{-1} * \mathcal{T}_2^{-1}.$$

Therefore,

$$\mathcal{X} = \mathcal{T}_1 * \mathcal{V}_{12} * \mathcal{S} = -\mathcal{T}_1 * \mathcal{V}_{12} * \mathcal{V}_{22}^{-1} * \mathcal{T}_2^{-1} = \mathcal{X}_{TLS},$$

which comes to the proof. $\square$

Specially, if we take $p = k = 1$, then the T-TLS problem degenerates to the matrix total least squares problem. The following result was given by Golub and Van Loan [19]:

Corollary 3.6. *Suppose matrices $A, E \in \mathbb{R}^{m\times n}$ and vectors $b, r \in \mathbb{R}^{m\times 1}$ satisfies*

$$\min_{E,r} \|D[E \quad r]T\|_F$$

$$\text{s.t. } b + r \in \mathcal{R}(A + E),$$

where

$$\begin{cases} D = \operatorname{diag}(d_1, d_2, \dots, d_m),\ d_i > 0,\ i = 1, 2, \dots, m, \\ T = \operatorname{diag}(t_1, t_2, \dots, t_{n+1}) = \begin{bmatrix} T_1 & O \\ O & t_{n+1} \end{bmatrix},\ t_i > 0,\ i = 1, 2, \dots, n+1, \end{cases}$$

are reversible weight matrices.

Denote $C = D[A \quad b]T$, and it has singular value decomposition as follows

$$U^\top CV = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_{n+1}),$$

$$\hat{A} = DAT_1,\ \hat{b} = Db,\ \lambda = t_{n+1},$$

and the singular value decomposition of matrix $\hat{A}$ is

$$\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^\top = \operatorname{diag}(\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n).$$

If $\hat{\sigma}_n > \sigma_{n+1}$, then the solution x_{TLS} exists and it is the unique solution to the matrix TLS problem. Moreover,

$$\begin{cases} x_{TLS} = T_1(\hat{A}^\top \hat{A} - \sigma_{n+1}^2 I)^{-1}\hat{A}^\top \hat{b}, \\ \sigma_{n+1}^2 \left[\frac{1}{\lambda^2} + \sum_{i=1}^n \frac{c_i^2}{\hat{\sigma}_i^2 - \sigma_{n+1}^2}\right] = \rho_{LS}^2, \end{cases}$$

where

$$\begin{cases} c = (c_1, c_2, \dots, c_m)^\top = \hat{U}^\top \hat{b}, \\ \rho_{LS}^2 = \min \|D(b - Ax)\|_2^2 = \min \|D(b - Ax_{LS})\|_2^2. \end{cases}$$

By using the same kind of technique as Wei [59], we directly give the perturbation bound of the TLS problem.

Corollary 3.7. *For the TLS problem (3.4), we assume that the conditions in Theorem 3.5 hold. Partition the tensor $\mathcal{V}$ as*

$$\mathcal{V} = \begin{bmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} \\ \mathcal{V}_{21} & \mathcal{V}_{22} \end{bmatrix}$$

where $\mathcal{V}_{11} \in \mathbb{R}^{n\times q\times p}$, $\mathcal{V}_{12} \in \mathbb{R}^{n\times(n+d-q)\times p}$, $\mathcal{V}_{21} \in \mathbb{R}^{d\times q\times p}$, $\mathcal{V}_{22} \in \mathbb{R}^{d\times(n+d-q)\times p}$. Let $\mathcal{A}' \in \mathbb{C}^{m\times n\times p}$, $\mathcal{B}' \in \mathbb{C}^{m\times d\times p}$ and $[\mathcal{A}' \ \mathcal{B}'] = [\mathcal{A} \ \mathcal{B}] + \mathcal{E}$ with $\eta^{(k)} = \frac{1}{6}(\overline{\sigma}_p^{(k)} - \sigma_{n+1}^{(k)})$, and the T-SVD for $\mathcal{A}'$, $[\mathcal{A}' \ \mathcal{B}']$ be

$$\overline{\mathcal{U}}' * \mathcal{A} * \overline{\mathcal{V}}' = \overline{\Sigma}',$$

where $\overline{\Sigma}' = \operatorname{diag}(\overline{\sigma}_1^{(1)'}, \dots, \overline{\sigma}_n^{(p)'})$ and $\Sigma = \operatorname{diag}(\sigma_1^{(1)'}, \dots, \sigma_{n+d}^{(p)\ '})$. Suppose that all elements of $\overline{\sigma}_q^{(k)}$ after 'bcirc' operation and fast Fourier transformation is greater than all elements *of $\sigma_{q+1}^{(k)}$ after 'bcirc' operation and fast Fourier transformation for all k. Partition $\mathcal{V}'$ conformally with $\mathcal{V}$, and replace $\mathcal{V}_{ij}$ by $\mathcal{V}'_{ij}$ for $i, j = 1, 2$. The perturbed solution comes to be $\mathcal{X}'_{TLS} = ((\mathcal{V}'_{11})^H)^\dagger * (\mathcal{V}'_{21})^H$. Then we get the following estimates:*

$$\|\mathcal{X}_{TLS} - \mathcal{X}'_{TLS}\| \le \max_k \frac{\eta^{(k)} + \sigma_{n+1}^{(k)}}{\overline{\sigma}_q^{(k)} - \sigma_{n+1}^{(k)}} (3 + 5\,\|\mathcal{X}_{TLS}\|).$$

The condition numbers and algorithms for the total least squares problem can be found in [33, 39, 62, 65].

3.3 Stochastic Perturbation Bound for $\mathcal{A}^\dagger$

In this section, we approach the tensor perturbation from a probabilistic point of view. Without loss of generality, for a tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$ $(m \ge n)$, we assume that the perturbation tensor $\mathcal{S} \in \mathbb{R}^{m\times n\times p}$ is a random tensor whose distributions of elements satisfies the independently identically distribution (*i.i.d.*) $\mathcal{N}(0, \sigma^2)$, where the variance σ^2 is sufficiently small.

Under the above assumptions the perturbed tensor $\overline{\mathcal{A}} = \mathcal{A} + \mathcal{S}$ has rank n almost surely, it is no wonder that we suppose the tensor $\mathcal{A}$ always has T-rank n in this section.

Definition 3.8 (Stochastic norm)**.** The stochastic norm of a tensor $\|\cdot\|_S$ is defined by

$$\|\cdot\|_S = \sqrt{\mathbb{E}\left(\|\cdot\|_F^2\right)},$$

where $\|\cdot\|_F$ is the Frobenius norm of a tensor.

Lemma 3.9. *For a third order tensor $\mathcal{A} \in \mathbb{C}^{m\times n\times p}$, we have*

$$\|\mathcal{A}\|_S^2 = \|\mathrm{bcirc}(\mathcal{A})\|_S^2\,, \tag{3.6}$$

where $\|\mathrm{bcirc}(\mathcal{A})\|_S$ is the matrix stochastic norm of $\mathrm{bcirc}(\mathcal{A})$ defined by Stewart in [50].

Applying the discrete Fourier transformation to $\overline{\mathcal{A}}$, we obtain

$$\begin{aligned}\mathrm{bcirc}(\overline{\mathcal{A}}) &= \mathrm{bcirc}(\mathcal{A} + \mathcal{S})\\ &= \mathrm{bcirc}(\mathcal{A}) + \mathrm{bcirc}(\mathcal{S})\\ &= (F_p^H \otimes I_m)\left(\mathrm{diag}\{A_1, A_2, \dots, A_p\} + \mathrm{diag}\{S_1, S_2, \dots, S_p\}\right)(F_p \otimes I_n),\end{aligned}$$

and

$$\mathrm{bcirc}(\overline{\mathcal{A}}) = (F_p^H \otimes I_m)\mathrm{diag}\{\overline{A_1}, \overline{A_2}, \dots, \overline{A_p}\}(F_p \otimes I_n).$$

It is easy to find that $\overline{A_i} = A_i + S_i$, $(i = 1, 2, \dots, p)$ and S_i are random matrices. The following lemma characterizes the probability distribution of S_i.

Lemma 3.10. *Suppose that S_1 is a real matrix. The entries of S_1 satisfies the i.i.d. $\mathcal{N}(0, p\sigma^2)$. S_i, $(i = 2, 3, \dots, p)$ are complex matrices. The real part and imaginary part of S_i, $(i = 2, 3, \dots, p)$ satisfies the i.i.d. $\mathcal{N}(0, \frac{1}{2}p\sigma^2)$, respectively. The real and imaginary part are two independent Gaussian matrices.*

Proof. Denote the k-th slice of tensor $\mathcal{A}$ as $A^{(k)} = (a_{ij}^{(k)})_{m\times n}$, $(k = 1, 2, \ldots, p)$ and define $S^{(k)}$ similarly from $\mathcal{S}$. $S^{(k)}$ are independent matrices with *i.i.d.* Gaussian elements.

For $i = 1$, we obtain that $S_1 = \sum_{k=1}^{p} S^{(k)}$.

It can be deduced from the distribution of entries of $\mathcal{S}$ and the properties of normal distribution that the elements of S_1 satisfies the *i.i.d.* Gaussian distribution with variance $p\sigma^2$.

Now we turn to $i = 2, 3, \ldots, p$. Similarly we have $S_i = \sum_{k=1}^{p} S^{(p+2-k)}\omega^{(i-1)k}$, where $\omega^q = \exp\left\{\frac{2q\pi}{p}\mathbf{i}\right\}$ is the q-th unit root of order p and $\mathbf{i} \equiv \sqrt{-1}$ which represents the imaginary units.

Let R_k be the real part of S_k and denote $\theta_q = \frac{2q\pi}{p}\mathbf{i}$. We arrive at the expression $R_i = \sum_{k=1}^{p} S^{(p+2-k)}\cos(k\theta_{i-1})$, from the corresponding formula of S_k. It implies that the elements of R_k satisfies the *i.i.d.* Gaussian distribution with mean zero, and the variance of elements is

$$\begin{aligned}\sum_{k=1}^{p}\cos^2(k\theta_{i-1}) &= \begin{cases}\frac{p}{2}, & \text{if} \quad (i-1)\nmid p,\\ \left(\frac{p}{i-1}\right)\left(\frac{i-1}{2}\right), & \text{if} \quad (i-1)\mid p,\end{cases}\\ &= \frac{p}{2}.\end{aligned}$$

The distribution of the imaginary part of S_i, $(i = 2, 3, \ldots, p)$ denoted by W_i can be derived in the same way. The variance of the elements w_{jl} of W_i is

$$\begin{aligned}\sum_{k=1}^{p}\sin^2(k\theta_{i-1}) &= \begin{cases}\frac{p}{2}, & \text{if} \quad (i-1)\nmid p,\\ \left(\frac{p}{i-1}\right)\left(\frac{i-1}{2}\right), & \text{if} \quad (i-1)\mid p,\end{cases}\\ &= \frac{p}{2}.\end{aligned}$$

Finally, we need to prove the independence of R_i and W_i. Only the uncorrelation of the entries in the corresponding positions are needed to prove since the both matrices have *i.i.d.* Gaussian elements. It comes to

$$\begin{aligned}\mathbb{E}(r_{jl}w_{jl}) &= \mathbb{E}\left(\sum_{k=1}^{p} s_{jl(p+2-k)}\cos(k\theta_{i-1})\right)\left(\sum_{k=1}^{p} s_{jl(p+2-k)}\sin(k\theta_{i-1})\right)\\ &= \mathbb{E}\left(\sum_{k=1}^{p}\cos(k\theta_{i-1})\sin(k\theta_{i-1})s_{jl(p+2-k)}^2\right)\\ &= p\sigma^2\sum_{k=1}^{p}\cos(k\theta_{i-1})\sin(k\theta_{i-1})\\ &= \frac{p}{2}\sigma^2\sum_{k=1}^{p}\sin(2k\theta_{i-1})\\ &= 0,\end{aligned}$$

which completes the proof. $\square$

The explicit formula of the perturbed tensor Moore-Penrose inverse

$$\begin{aligned}\operatorname{bcirc}(\overline{\mathcal{A}^{\dagger}}) &= \operatorname{bcirc}((\mathcal{A}+\mathcal{S})^{\dagger})\\ &= (F_p^H\otimes I_n)\left(\operatorname{diag}\left\{(A_1+S_1)^{\dagger}, (A_2+S_2)^{\dagger}, \ldots, (A_p+S_p)^{\dagger}\right\}\right)(F_p\otimes I_m),\end{aligned}$$

implies that we can bound the tensor perturbation error by analyzing the deviations between the diagonal block before and after perturbed. Thus we arrive at the main theorem.

Theorem 3.11. *Let $\mathcal{A}$ be an $m \times n \times p$ tensor and $\mathcal{S}$ be the Gaussian perturbation tensor with i.i.d. $\mathcal{N}(0, \sigma^2)$ entries. If the following two conditions hold,*
(i) $\left\|\mathcal{A}^\dagger * \mathcal{S}\right\|_S^2 \ll 1$,
(ii) $\sigma^2(m-n)\left\|\mathcal{A}^\dagger\right\|_F^2 \ll 1$,
then the perturbation bound of the stochastically perturbed tensor based on the stochastic norm is

$$\left\|\overline{\mathcal{A}}^\dagger - \mathcal{A}^\dagger\right\|_S^2 \le p\sigma^2(m-n+\sqrt{n})\left\|\mathcal{A}^\dagger\right\|_F^2\left\|\mathcal{A}^\dagger\right\|_2^2. \tag{3.7}$$

Proof. The perturbation bound of the tensor Moore-Penrose inverse relies on the perturbation expansion of the matrix Moore-Penrose inverse, i.e., for an $m \times n$ matrix A of full column rank, that is $\operatorname{rank}(A) = n$ $(m \ge n)$, with a perturbation E. Dropping high-order terms, we have [13]

$$(A+E)^\dagger = A^\dagger - A^\dagger E A^\dagger + (A^H A)^{-1} E^H P,$$

where $P = I - AA^\dagger$ is the projection (see [48, 49, 50]).

Because of the fact that

$$\begin{aligned}\operatorname{bcirc}(\overline{\mathcal{A}^\dagger}) &= \operatorname{bcirc}((\mathcal{A}+\mathcal{S})^\dagger)\\ &= (F_p^H \otimes I_n)\left(\operatorname{diag}\left\{(A_1+S_1)^\dagger, (A_2+S_2)^\dagger, \ldots, (A_p+S_p)^\dagger\right\}\right)(F_p \otimes I_m),\end{aligned}$$

we have

$$\left\|\overline{\mathcal{A}}^\dagger\right\|_S^2 = \left\|(\mathcal{A}+\mathcal{S})^\dagger\right\|_S^2 = \sum_{i=1}^p \left\|(A_i+S_i)^\dagger\right\|_S^2.$$

Hence we may bound them separately.

By the perturbation expansion of the Moore-Penrose inverse and Lemma 3.10, firstly we have

$$\begin{aligned}(A_k+S_k)^\dagger - A_k^\dagger &= -A_k^\dagger S_k A_k^\dagger + (A_k^H A_k)^{-1} S_k^H P_k\\ &= -A_i^\dagger (R_k + \mathbf{i} W_k) A_k^\dagger + (A_k^\top A_k)^{-1}(R_k - \mathbf{i} W_k)^\top P_k,\end{aligned}$$

for $S_k = R_k + \mathbf{i} W_k, P_k = I - A_k A_k^\dagger$ and $k = 2, 3, \ldots, p$.

It is the two technical conditions in the theorems and the discussion in [50] (Section 2.4 and Section 3.1.2 of [50]) that ensure the rationality of truncating the higher-order terms $O(\|S_k\|_S^2)$.

According to the fact that

$$(A_k^\dagger S_k A_k^\dagger)((A_k^H A_k)^{-1} S_k^H P_k)^H = A_k^\dagger S_k [A_k^\dagger (I - A_k A_k^\dagger)] S_k (A_k^H A_k)^{-1} \equiv O,$$

we derive that

$$\begin{aligned}\left\|\overline{A}_k^\dagger - A_k^\dagger\right\|_S^2 &= \mathbb{E}\left(\left\|\overline{A}_k^\dagger - A_k^\dagger\right\|_F^2\right) = \mathbb{E}\left(\left\|(A_k+S_k)^\dagger - A_k^\dagger\right\|_F^2\right)\\ &= \mathbb{E}\left(\left\|-A_k^\dagger S_k A_k^\dagger + (A_k^H A_k)^{-1} S_k^H P_k\right\|_F^2\right)\\ &= \mathbb{E}\left(\left\|A_k^\dagger S_k A_k^\dagger\right\|_F^2\right) + \mathbb{E}\left(\left\|(A_k^H A_k)^{-1} S_k^H P_k\right\|_F^2\right).\end{aligned}$$

By Theorem 2.3 and 2.4 in [50], we obtain that

$$\begin{aligned}
&\mathbb{E}\left(\left\|A_k^\dagger S_k A_k^\dagger\right\|_F^2\right) \\
&= \mathbb{E}\left(\left\|A_k^\dagger (R_k + \mathbf{i}W_k) A_k^\dagger\right\|_F^2\right) \\
&= \mathbb{E}\left(\left\|A_k^\dagger R_k A_k^\dagger\right\|_F^2\right) + \mathbb{E}\left(\left\|A_k^\dagger W_k A_k^\dagger\right\|_F^2\right) \quad \text{(by Lemma 3.10)} \\
&= \mathbb{E}\left[\mathbf{tr}\left((A_k^\dagger)^H R_k^\top (A_k^\dagger)^H A_k^\dagger R_k A_k^\dagger\right)\right] + \mathbb{E}\left[\mathbf{tr}\left((A_k^\dagger)^H W_k^\top (A_k^\dagger)^H A_k^\dagger W_k A_k^\dagger\right)\right] \\
&= \mathbf{tr}\left((A_k^\dagger)^H \mathbb{E}[R_k^\top (A_k^\dagger)^H A_k^\dagger R_k] A_k^\dagger\right) + \mathbf{tr}\left((A_k^\dagger)^H \mathbb{E}[W_k^\top (A_k^\dagger)^H A_k^\dagger W_k] A_k^\dagger\right) \\
&= \frac{p}{2}\sigma^2 \left\|A_k^\dagger\right\|_F^2 \mathbf{tr}\left((A_k^\dagger)^H A_k^\dagger\right) + \frac{p}{2}\sigma^2 \left\|A_k^\dagger\right\|_F^2 \mathbf{tr}\left((A_k^\dagger)^H A_k^\dagger\right) \ \text{(by [50, Theorem2.3])} \\
&= p\sigma^2 \left\|A_k^\dagger\right\|_F^4 .
\end{aligned}$$

Similarly, we can derive

$$\begin{aligned}
\mathbb{E}\left(\left\|(A_k^H A_k)^{-1} S_k^H P_k\right\|_F^2\right) &= \mathbb{E}\left(\left\|(A_k^H A_k)^{-1} (R_k - \mathbf{i}W_k)^\top P_k\right\|_F^2\right) \\
&= \mathbb{E}\left(\left\|(A_k^H A_k)^{-1} R_k^\top P_k\right\|_F^2\right) + \mathbb{E}\left(\left\|(A_k^H A_k)^{-1} W_k^\top P_k\right\|_F^2\right) \\
&= \frac{p}{2}\sigma^2 \left\|P_k\right\|_F^2 \mathbf{tr}\left(((A_k^H A_k)^{-1})^H (A_k^H A_k)^{-1}\right) \\
&+ \frac{p}{2}\sigma^2 \left\|P_k\right\|_F^2 \mathbf{tr}\left(((A_k^H A_k)^{-1})^H (A_k^H A_k)^{-1}\right) \\
&= p\sigma^2 (m-n) \left\|(A_k^H A_k)^{-1}\right\|_F^2 .
\end{aligned}$$

Using the results above and in light of $\left\|(A_k^H A_k)^{-1}\right\|_F^2 = \left\|(A_k^\dagger)^H A_k^\dagger\right\|_F^2 \le \left\|A_k^\dagger\right\|_2^2 \left\|A_k^\dagger\right\|_F^2$, we obtain that

$$\begin{aligned}
\left\|\overline{A}_k^\dagger - A_k^\dagger\right\|_S^2 &= p\sigma^2 \left(\left\|A_k^\dagger\right\|_F^4 + (m-n)\left\|(A_k^H A_k)^{-1}\right\|_F^2\right) \\
&\le p\sigma^2 (m - n + \sqrt{n}) \left\|A_k^\dagger\right\|_2^2 \left\|A_k^\dagger\right\|_F^2 ,
\end{aligned}$$

which is a concise form of perturbation estimation of the blocks $A_k^\dagger$ $(k \ge 2)$.

In the same way (even easier), we can derive the perturbation bound of $A_1^\dagger$ for resembling form, that is,

$$\left\|\overline{A}_1^\dagger - A_1^\dagger\right\|_S^2 \le p\sigma^2 (m - n + \sqrt{n}) \left\|A_1^\dagger\right\|_2^2 \left\|A_1^\dagger\right\|_F^2 .$$

Combining the two aforementioned inequalities, it turns out that

$$\begin{aligned}
\left\|\overline{\mathcal{A}}^\dagger - \mathcal{A}^\dagger\right\|_S^2 &= \sum_{k=1}^{p} \left\|\overline{A}_k^\dagger - A_k^\dagger\right\|_S^2 \\
&\le p\sigma^2 (m - n + \sqrt{n}) \sum_{k=1}^{p} \left\|A_k^\dagger\right\|_2^2 \left\|A_k^\dagger\right\|_F^2 \\
&\le p\sigma^2 (m - n + \sqrt{n}) \left\|\mathcal{A}^\dagger\right\|_F^2 \left(\max_{1\le k\le p} \left\|A_k^\dagger\right\|_2^2\right) \\
&\le p\sigma^2 (m - n + \sqrt{n}) \left\|\mathcal{A}^\dagger\right\|_F^2 \left\|\mathcal{A}^\dagger\right\|_2^2 .
\end{aligned}$$

□

3.4 Stochastic Conditioning of Tensor Functions

In this subsection, we investigate how sensitive are the tensor functions to stochastic perturbations in their input. Miao, Qi and Wei [40] present the definition of generalized tensor function of non-F-square tensors based on the T-SVD decomposition. Now we explore the more general cases, that is $F : \Omega \to \mathbb{R}^{t\times s\times p}$, where Ω is an open subset of $\mathbb{R}^{m\times n\times p}$. To our best knowledge, there is few literature to rigorously quantify the sensitivity of tensor functions to random noises. As for the random perturbation, these represent uncertainties in the data or rounding errors arising from computations in the finite precision arithmetic. The goal is to quantify the effect that such uncertainties have on the computed function value. First, we generalize the concept of Fréchet derivative to tensor functions based on the T-product.

Theorem 3.12 (Fréchet derivative)**.** *Let $\mathcal{A}, \mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be third order tensors and the tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ be Fréchet differentiable at $\mathcal{A}$. Then the Fréchet derivative of F at $\mathcal{A}$ is the unique bounded operator $F'(\mathcal{A})$ given by the relation*

$$F(\mathcal{A}+\mathcal{H}) = F(\mathcal{A}) + F'(\mathcal{A})(\mathcal{H}) + \mathcal{R}(\mathcal{H}), \quad \lim_{\mathcal{H}\to 0} \frac{\|\mathcal{R}(\mathcal{H})\|}{\|\mathcal{H}\|} = 0. \tag{3.8}$$

Proof. Suppose

$$(F_p \otimes I_m)\mathrm{bcirc}(\mathcal{A})(F_p^H \otimes I_n) = \mathrm{diag}\{A_1, A_2, \dots, A_p\},$$

and

$$(F_p \otimes I_m)\mathrm{bcirc}(\mathcal{H})(F_p^H \otimes I_n) = \mathrm{diag}\{H_1, H_2, \dots, H_p\}.$$

Then the tensor function satisfies,

$$(F_p \otimes I_t)\mathrm{bcirc}(F(\mathcal{A}))(F_p^H \otimes I_s) = \mathrm{diag}\{F(A_1), F(A_2), \dots, F(A_p)\},$$

and

$$(F_p \otimes I_t)\mathrm{bcirc}(F(\mathcal{A}+\mathcal{H}))(F_p^H \otimes I_s) = \mathrm{diag}\{F(A_1+H_1), F(A_2+H_2), \dots, F(A_p+H_p)\},$$

by Fréchet derivative of matrix function [20], we have

$$F(A_i+H_i) = F(A_i) + F'(A_i)(H_i) + R(H_i), \quad \lim_{H_i\to 0} \frac{\|R(H_i)\|}{\|H_i\|} = 0.$$

Then it turns out that

$$\begin{aligned}
\mathrm{bcirc}(F(\mathcal{A}+\mathcal{H})) &= (F_p^H \otimes I_t)\mathrm{diag}\{F(A_1), F(A_2), \dots, F(A_p)\}(F_p \otimes I_s) \\
&\quad + (F_p^H \otimes I_t)\mathrm{diag}\{F'(A_1)(H_1), F'(A_2)(H_2), \dots, F'(A_p)(H_p)\}(F_p \otimes I_s) \\
&\quad + (F_p^H \otimes I_t)\mathrm{diag}\{R(H_1), R(H_2), \dots, R(H_p)\}(F_p \otimes I_s) \\
&= \mathrm{bcirc}(F(\mathcal{A})) + \mathrm{bcirc}(F'(\mathcal{A})(\mathcal{H})) + \mathrm{bcirc}(R(\mathcal{H})),
\end{aligned}$$

which is equivalent to

$$F(\mathcal{A}+\mathcal{H}) = F(\mathcal{A}) + F'(\mathcal{A})(\mathcal{H}) + R(\mathcal{H}), \quad \lim_{\mathcal{H}\to 0} \frac{\|\mathcal{R}(\mathcal{H})\|}{\|\mathcal{H}\|} = 0.$$

□

In addition, we have $\mathrm{vec}(F'(A_i)H_i) = J_{A_i}\mathrm{vec}(H_i)$ is the matrix representation of each derivative $F'(A_i)$, where J_{A_i} is the Jacobian matrix in the standard coordinate.

In order to get further results, we need the following lemmas.

Lemma 3.13. (Basic probability inequalities) *If α and β are random variables such that $\alpha \le \beta$, then for any $\tau \in \mathbb{R}$, we have*

$$\mathrm{Prob}\{\alpha \ge \tau\} \le \mathrm{Prob}\{\beta \ge \tau\}, \quad \mathrm{Prob}\{\beta \le \tau\} \le \mathrm{Prob}\{\alpha \ge \tau\}.$$

For random variables α and β and any $\tau, \epsilon \in \mathbb{R}$,

$$\mathrm{Prob}\{\alpha + \beta \ge \tau\} \le \mathrm{Prob}\{\alpha \ge \tau(1-\epsilon)\} + \mathrm{Prob}\{\beta \ge \tau\epsilon\}.$$

Definition 3.14. (Covariance tensor) Let $x \in \mathbb{R}^{m\times 1\times p}$ be a third order tensor satisfing

$$(F_p \otimes I_m)\mathrm{bcirc}(x)(F_p^H \otimes 1) = \mathrm{diag}\{x_1, x_2, \dots, x_p\}.$$

We say $\Sigma \in \mathbb{R}^{m\times m\times p}$ is the covariance tensor of x if and only if

$$(F_p \otimes I_m)\mathrm{bcirc}(\Sigma)(F_p^H \otimes I_n) = \mathrm{diag}\{\Sigma_1, \Sigma_2, \dots, \Sigma_p\},$$

and each covariant matrix of x_i is Σ_i.

In addition, if the mean value of each x_i is 0, then the distribution of x is denoted by $x \sim (0, \Sigma)$. If the mean value of the above each $\mathrm{vec}(H_i)$ is 0, and covariant matrix is Σ_i, then the distribution of $\mathcal{H}$ is denoted by $\mathrm{vec}(\mathcal{H}) \sim (0, \Sigma)$.

Lemma 3.15 (Quadratic forms)**.** (1) *Let $\mathcal{M} \in \mathbb{R}^{m\times n\times p}$ and $x \in \mathbb{R}^{n\times 1\times p}$ be third order tensors. The distribution of x is $x \sim (0, \Sigma)$, $\Sigma \in \mathbb{R}^{n\times n\times p}$. Then we have*

$$\mathbb{E}\left(\|\mathcal{M} * x\|_F\right) \le \left\|\mathcal{M} * \Sigma^{1/2}\right\|_F = \|\mathcal{M} * x\|_S.$$

(2) *Let $\mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be a third order tensor and* $\mathrm{vec}(\mathcal{H}) \sim (0, \Sigma')$*, $\Sigma' \in \mathbb{R}^{mn\times mn\times p}$. Then we have*

$$\mathbb{E}\left(\|\mathcal{H}\|_F\right) \le \left\|\Sigma'^{1/2}\right\|_F.$$

Proof. (1).

$$\begin{aligned}
\mathbb{E}\left(\|\mathcal{M} * x\|_F\right) &= \mathbb{E}\left(\|\mathrm{diag}\{M_1x_1, M_2x_2, \dots, M_px_p\}\|_F\right) \\
&= \mathbb{E}\left(\sqrt{\|\mathrm{diag}\{M_1x_1, M_2x_2, \dots, M_px_p\}\|_F^2}\right) \\
&\le \sqrt{\mathbb{E}\left(\|\mathrm{diag}\{M_1x_1, M_2x_2, \dots, M_px_p\}\|_F^2\right)} \\
&= \sqrt{\left\|\mathrm{diag}\{M_1\Sigma_1^{1/2}, M_2\Sigma_2^{1/2}, \dots, M_p\Sigma_p^{1/2}\}_F^2\right\|} \\
&= \left\|\mathcal{M} * \Sigma^{1/2}\right\|_F = \sqrt{\|\mathcal{M} * x\|_S^2} = \|\mathcal{M} * x\|_S.
\end{aligned}$$

(2).

$$\begin{aligned}
\mathbb{E}\left(\|\mathcal{H}\|_F\right) &= \mathbb{E}\left(\|\mathrm{diag}\{H_1, H_2, \dots, H_p\}\|_F\right) = \mathbb{E}\left(\sqrt{\|\mathrm{diag}\{H_1, H_2, \dots, H_p\}\|_F^2}\right) \\
&\le \sqrt{\mathbb{E}\left(\|\mathrm{diag}\{H_1, H_2, \dots, H_p\}\|_F^2\right)} = \sqrt{\mathbb{E}\left(\|\mathrm{diag}\{\mathrm{vec}(H_1), \mathrm{vec}(H_2), \dots, \mathrm{vec}(H_p)\}\|_F^2\right)} \\
&= \sqrt{\left\|\mathrm{diag}\{\Sigma_1^{1/2}, \Sigma_2^{1/2}, \dots, \Sigma_p^{1/2}\}\right\|_F^2} = \left\|\Sigma^{1/2}\right\|_F.
\end{aligned}$$

□

Lemma 3.16. *Let $\mathcal{A}, \mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be third order tensors. The tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is Fréchet differentiable at $\mathcal{A}$. $\mathcal{H}$ satisfies the distribution* $\mathrm{vec}(\mathcal{H}) \sim (0, \sigma^2\Sigma)$*, where $\Sigma \in \mathbb{R}^{mn\times mn\times p}$. Then*

$$\|F'(\mathcal{A})(\mathcal{H})\|_S = \sigma \left\| \mathcal{J}_{\mathcal{A}} * \Sigma^{1/2} \right\|_F,$$

where $\mathcal{J}_{\mathcal{A}}$ is the Jacobi tensor whose block entries are J_{A_i}, $(i = 1, 2, \ldots, p)$ after the block circulant and the discrete Fourier transformation.

Proof.

$$\begin{aligned}
\mathbb{E}\left(\|F'(\mathcal{A})(\mathcal{H})\|_F^2\right) &= \mathbb{E}\left(\|\mathrm{diag}\{F'(A_1)(H_1), F'(A_2)(H_2), \ldots, F'(A_p)(H_p)\}\|_F^2\right) \\
&= \mathbb{E}\left(\|\mathrm{diag}\{\mathrm{vec}(F'(A_1)(H_1)), \mathrm{vec}(F'(A_2)(H_2)), \ldots, \mathrm{vec}(F'(A_p)(H_p))\}\|_F^2\right) \\
&= \mathbb{E}\left(\left\|\mathrm{diag}\{J_{A_1}(\mathrm{vec}(H_1)), J_{A_2}(\mathrm{vec}(H_2)), \ldots, J_{A_p}(\mathrm{vec}(H_p))\}\right\|_F^2\right) \\
&= \sigma^2 \left\|\mathrm{diag}\{J_{A_1}\Sigma_1^{1/2}, J_{A_2}\Sigma_2^{1/2}, \ldots, J_{A_p}\Sigma_p^{1/2}\}\right\|_F^2 \\
&= \sigma^2 \left\| \mathcal{J}_{\mathcal{A}} * \Sigma^{1/2} \right\|_F^2.
\end{aligned}$$

□

The idea of stochastic condition number can be traced back to Turing [53]. Fletcher [16] assumes that the elements of H are independent random variables with mean 0 and variance $\sigma^2 A_{ij}^2$. He calls

$$\sqrt{\mathbb{E}(\|A^{-1}HA^{-1}b\|_2^2)} = \sigma \left\|[A^{-1}][A][A^{-1}b]\right\|_2^{1/2}$$

the expected condition number of $F(A) = A^{-1}b$, where '$[A]$' denotes the matrix whose elements are the squares of A. Stewart [50] derives the similar result independently.

Similarly, this idea can be generalized to tensor functions based on the T-product. We consider the following first-order expansion according to Theorem 3.12,

$$F(\mathcal{A} + \mathcal{H}) - F(\mathcal{A}) = F'(\mathcal{A})(\mathcal{H}) + o(\|\mathcal{H}\|),$$

and we call

$$\|F'(\mathcal{A})(\mathcal{H})\|_S$$

the expected condition number of the tensor function F. Suppose the stochastic perturbation satisfies the distribution

$$\mathrm{vec}(\mathcal{H}) \sim (0, \sigma^2\Sigma),$$

where $\sigma \in \mathbb{R}$ and $\Sigma \in \mathbb{R}^{mn\times mn\times p}$ is F-symmetric semi-positive definite tensor. Then from Lemma 3.16, we have

$$\|F'(\mathcal{A})(\mathcal{H})\|_S = \sigma \left\| \mathcal{J}_{\mathcal{A}} * \Sigma^{1/2} \right\|_F.$$

By ignoring the $o(\|H\|)$ terms in the Taylor expansion, we obtain that

$$\|F(\mathcal{A} + \mathcal{H}) - F(\mathcal{A})\|_S \approx \|F'(\mathcal{A})(\mathcal{H})\|_S = \sigma \left\| \mathcal{J}_{\mathcal{A}} * \Sigma^{1/2} \right\|_F.$$

In order to guarantee the above approximation is tight at least when σ is small, we need the following definition.

Definition 3.17. (Entire tensor function) The tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is called an entire function if and only if each component f_{ijk} of F has a Taylor series [40] that is absolutely convergent.

Similar to the Theorem 3.2 of Gratton [20], we give the following theorem without proof.

Theorem 3.18. *Let $\mathcal{A}, \mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be third order tensors. The tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is an entire function. $\mathcal{H}$ satisfies the distribution* $\mathrm{vec}(\mathcal{H}) \sim (0, \sigma^2\Sigma)$, *where $\Sigma \in \mathbb{R}^{mn\times mn\times p}$. Additionally, the elements of $\mathcal{H}$ being random variables whose kth moments are bounded by $c\sigma^k$. Then we have*

$$\lim_{\sigma\to 0} \frac{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_S}{\sigma} = \left\| \mathcal{J}_{\mathcal{A}} * \Sigma^{1/2} \right\|_F, \tag{3.9}$$

where $\mathcal{J}_{\mathcal{A}}$ is the Jacobi tensor at $\mathcal{A}$.

The requirement of the above estimation is rather restrictive, since many interesting tensor functions based on the tensor-tensor product are Fréchet differentiable by not entire. This motivates us to another kind of estimation and condition number which only require the tensor function is Fréchet differentiable. We present the definition of stochastic condition number for tensor functions as follows.

Definition 3.19. (Stochastic condition number) Let $\mathcal{A}, \mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be third order tensors. $\mathcal{H}$ satisfies the distribution $\mathrm{vec}(\mathcal{H}) \sim (0, \sigma^2\Sigma)$, where $\Sigma \in \mathbb{R}^{mn\times mn\times p}$. The tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is Fréchet differentiable at $\mathcal{A}$, we call

$$\tilde{\kappa}_\Sigma = \limsup_{\sigma\to 0} \frac{\mathrm{Med}\{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F\}}{\sigma}, \tag{3.10}$$

the stochastic condition number of F at $\mathcal{A}$ with respect to $(0, \sigma^2\Sigma)$ perturbations. Here $\mathrm{Med}\{\alpha\}$ is the median of random variable α defined as

$$\mathrm{Med}\{\alpha\} = \sup\left\{ \tau \middle| \mathrm{Prob}\{\alpha \le \tau\} \ge \frac{1}{2} \text{ and } \mathrm{Prob}\{\alpha \ge \tau\} \ge \frac{1}{2} \right\}. \tag{3.11}$$

In order to obtain the bound for the stochastic condition number, we derive the following theorem.

Theorem 3.20. *Let $\mathcal{A}, \mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be third order tensors. $\mathcal{H}$ satisfies the distribution* $\mathrm{vec}(\mathcal{H}) \sim (0, \sigma^2\Sigma)$, *where $\Sigma \in \mathbb{R}^{mn\times mn\times p}$. The tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is Fréchet differentiable at $\mathcal{A}$. Then for any $\tau > 0$, we have*

$$\limsup_{\sigma\to 0} \mathrm{Prob}\left\{ \frac{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F}{\sigma} \ge \tau \right\} \le \frac{\left\| \mathcal{J}_{\mathcal{A}} * \Sigma^{1/2} \right\|_F}{\tau}. \tag{3.12}$$

Proof. Denote

$$\mathcal{P}_{\tau\sigma} = \mathrm{Prob}\left\{ \frac{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F}{\sigma} \ge \tau \right\}.$$

From the basic probability inequality in Lemma 3.13, for any $\epsilon \in (0,1)$, we have

$$\begin{aligned} \mathcal{P}_{\tau\sigma} &= \mathrm{Prob}\left\{ \frac{\|F'(\mathcal{A})(\mathcal{H}) + \mathcal{R}(\mathcal{H})\|_F}{\sigma} \ge \tau \right\} \\ &\le \mathrm{Prob}\left\{ \frac{\|F'(\mathcal{A})(\mathcal{H})\|_F}{\sigma} \ge \tau(1-\epsilon) \right\} + \mathrm{Prob}\left\{ \frac{\|\mathcal{R}(\mathcal{H})\|_F}{\sigma} \ge \tau\epsilon \right\}. \end{aligned}$$

From the distribution of $\mathcal{H}$ and the Markov inequality, the first term has the estimation

$$\begin{aligned}
\operatorname{Prob}\left\{\frac{\|F'(\mathcal{A})(\mathcal{H})\|_F}{\sigma} \geq \tau(1-\epsilon)\right\} &\leq \frac{\mathbb{E}\left(\|F'(\mathcal{A})(\mathcal{H})\|_F\right)}{\sigma\tau(1-\epsilon)} \\
&\leq \frac{\sqrt{\mathbb{E}\left(\|F'(\mathcal{A})(\mathcal{H})\|_F^2\right)}}{\sigma\tau(1-\epsilon)} \\
&= \frac{\left\|\mathcal{J}_{\mathcal{A}} * \Sigma^{1/2}\right\|_F}{\tau(1-\epsilon)}.
\end{aligned}$$

We show that the second term could be controlled by ϵ, if σ is sufficiently small. For any $\beta > 0$,

$$\begin{aligned}
\mathcal{R}_{\tau\sigma\epsilon} &= \operatorname{Prob}\left\{\frac{\|\mathcal{R}(\mathcal{H})\|_F}{\sigma} \geq \tau\epsilon\right\} \\
&= \operatorname{Prob}\left\{\frac{\|\mathcal{R}(\mathcal{H})\|_F}{\sigma} \geq \tau\epsilon \bigcap \|\mathcal{H}\|_F < \beta\right\} + \operatorname{Prob}\left\{\frac{\|\mathcal{R}(\mathcal{H})\|_F}{\sigma} \geq \tau\epsilon \bigcap \|\mathcal{H}\|_F \geq \beta\right\} \\
&\leq \operatorname{Prob}\left\{\frac{\|\mathcal{R}(\mathcal{H})\|_F}{\sigma} \geq \tau\epsilon \bigcap \|\mathcal{H}\|_F < \beta\right\} + \operatorname{Prob}\left\{\|\mathcal{H}\|_F \geq \beta\right\}.
\end{aligned}$$

For any $\alpha > 0$, there exists β such that

$$\frac{\|\mathcal{R}(\mathcal{H})\|_F}{\|\mathcal{H}\|_F} \leq \alpha, \quad \text{when } \|\mathcal{H}\|_F \leq \beta.$$

Therefore, for any $\alpha > 0$, there is a corresponding β such that

$$\begin{aligned}
\mathcal{R}_{\tau\sigma\epsilon} &\leq \operatorname{Prob}\left\{\frac{\|\mathcal{R}(\mathcal{H})\|_F}{\sigma} \geq \tau\epsilon\right\} + \operatorname{Prob}\left\{\|\mathcal{H}\|_F \geq \beta\right\} \\
&\leq \frac{\alpha\mathbb{E}\left(\|\mathcal{H}\|_F\right)}{\sigma\tau\epsilon} + \frac{\mathbb{E}\left(\|\mathcal{H}\|_F\right)}{\beta} \\
&\leq \frac{\alpha\left\|\Sigma^{1/2}\right\|_F}{\tau\epsilon} + \frac{\sigma\left\|\Sigma^{1/2}\right\|_F}{\beta}.
\end{aligned}$$

Set $\alpha = \frac{\tau\epsilon^2}{2\left\|\Sigma^{1/2}\right\|_F}$. Since β only depends on α, for all $\sigma \leq \frac{\epsilon\beta}{2\left\|\Sigma^{1/2}\right\|_F}$,

$$\mathcal{R}_{\tau\sigma\epsilon} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

It comes out

$$\mathcal{P}_{\tau\sigma} = \frac{\left\|\mathcal{J}_{\mathcal{A}} * \Sigma^{1/2}\right\|_F}{\tau(1-\epsilon)} + \epsilon = \gamma(\epsilon), \quad \text{if } \sigma \leq \frac{\epsilon\beta}{2\left\|\Sigma^{1/2}\right\|_F}.$$

Finally, it comes to

$$\limsup_{\sigma\to 0} \mathcal{P}_{\tau\sigma} \leq \gamma(0) = \frac{\left\|\mathcal{J}_{\mathcal{A}} * \Sigma^{1/2}\right\|_F}{\tau}.$$

□

By setting τ to be the median of $\frac{\|F(\mathcal{A}+\mathcal{H})-F(\mathcal{A})\|_F}{\sigma}$, it can be bounded that

$$\frac{1}{2} \leq \frac{\left\|\mathcal{J}_{\mathcal{A}} * \Sigma^{1/2}\right\|_F}{\operatorname{Med}\left\{\frac{\|F(\mathcal{A}+\mathcal{H})-F(\mathcal{A})\|_F}{\sigma}\right\}(1-\epsilon)} + \epsilon.$$

Therefore, we have the following bound for the stochastic condition number.

Corollary 3.21. *Let $\mathcal{A}, \mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be third order tensors. $\mathcal{H}$ satisfies the distribution $\operatorname{vec}(\mathcal{H}) \sim (0, \sigma^2\Sigma)$, where $\Sigma \in \mathbb{R}^{mn\times mn\times p}$. The tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is Fréchet differentiable at $\mathcal{A}$. Then the stochastic condition number of the tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ has the upper bound,*

$$\tilde{\kappa}_\Sigma = \limsup_{\sigma\to 0} \frac{\operatorname{Med}\{\|F(\mathcal{A}+\mathcal{H})-F(\mathcal{A})\|_F\}}{\sigma} \leq 2\left\|\mathcal{J}_{\mathcal{A}} * \Sigma^{1/2}\right\|_F.$$

Most cases in practice provide us the distribution of the original perturbation tensor $\mathcal{H} \in \mathbb{R}^{m\times n\times p}$, i.e., the distribution of $H^{(i)}$ $(i = 1, 2, \ldots, p)$, but not the distribution of the sub-matrices $H_1, H_2, \ldots, H_p$ of the block diagonal matrix $(F_p \otimes I_m)\operatorname{bcirc}(\mathcal{H})(F_p^H \otimes I_n)$. In the next part, we need to reveal the relationship between them.

For a tensor $\mathcal{A} \in \mathbb{R}^{m\times n\times p}$, $\mathcal{H} \in \mathbb{R}^{m\times n\times p}$ is supposed to be the perturbation tensor whose frontal slices $H^{(i)}$, $(i = 1, 2, \ldots, p)$ are independent and satisfies the distribution,

$$\operatorname{vec}(H^{(i)}) \sim (0, \Sigma^{(i)}).$$

By using the Lemma 4 in Miao, Qi, and Wei [41], if

$$\begin{bmatrix} H^{(1)} & H^{(p)} & H^{(p-1)} & \cdots & H^{(2)} \\ H^{(2)} & H^{(1)} & H^{(p)} & \cdots & H^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ H^{(p)} & H^{(p-1)} & \ddots & H^{(2)} & H^{(1)} \end{bmatrix} = (F_p^H \otimes I_m) \begin{bmatrix} H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_p \end{bmatrix} (F_p \otimes I_n),$$

then we have

$$\begin{cases} H_1 = \omega^0 H^{(1)} + \omega^0 H^{(p)} + \omega^0 H^{(p-1)} + \cdots + \omega^0 H^{(2)}, \\ H_2 = \omega^0 H^{(1)} + \omega^1 H^{(p)} + \omega^2 H^{(p-1)} + \cdots + \omega^{p-1} H^{(2)}, \\ \qquad\qquad \cdots \\ H_p = \omega^0 H^{(1)} + \omega^{p-1} H^{(p)} + \omega^{2(p-1)} H^{(p-1)} + \cdots + \omega^{(p-1)(p-1)} H^{(2)}, \end{cases} \tag{3.13}$$

where $\omega = e^{-2\pi\mathbf{i}/p}$ is the primitive p-th root of unity and is usually called the phase term. From the above equations, it is easy to find H_1 is a real matrix. On the other hand, since $H^{(i)}$'s are independent, satisfying the distribution $(0, \Sigma^{(i)})$, it is easy to get,

$$\operatorname{vec}(H_1) \sim \left(0, \Sigma^{(1)} + \Sigma^{(2)} + \cdots + \Sigma^{(p)}\right).$$

Because of the fact that H_t, $(t = 2, 3, \ldots, p)$ are complex matrices, we need to take complex numbers into consideration.

For $t = 2, 3, \ldots, p$, we have

$$H_t = \omega^0 H^{(1)} + \omega^{t-1} H^{(p)} + \cdots + \omega^{(p-1)(t-1)} H^{(2)},$$

where $\omega^{t-1} = \cos\left(\frac{2\pi(t-1)}{p}\right) + \mathbf{i}\sin\left(\frac{2\pi(t-1)}{p}\right)$.

Denote $\theta_t = \frac{2\pi t}{p}$ and $H^{(p+i)} = H^{(i)}$, then we have

$$H_t = \sum_{k=1}^{p} \cos[(k-1)\theta_{t-1}]H^{(p+2-k)} + \mathbf{i}\left(\sum_{k=1}^{p} \sin[(k-1)\theta_{t-1}]H^{(p+2-k)}\right).$$

From the distribution of $H^{(i)}$, it is easy to find

$$\mathbb{E}(\mathrm{vec}(H_t)_i) = 0.$$

Now we compute the covariance between the elements $\mathrm{vec}(H_t)_i$ and $\mathrm{vec}(H_t)_j$, $(i, j = 1, 2, \ldots, mn)$.

$$\begin{aligned}\mathrm{cov}(\mathrm{vec}(H_t)_i, \mathrm{vec}(H_t)_j) &= \mathbb{E}\{\mathrm{vec}(H_t)_i \cdot \overline{\mathrm{vec}(H_t)_j}\} \\ &= \mathbb{E}\Big\{\Big(\sum_{k=1}^{p} \cos[(k-1)\theta_{t-1}][\mathrm{vec}(H^{(p+2-k)})]_i \\ &\quad + \mathbf{i}\sum_{k=1}^{p} \sin[(k-1)\theta_{t-1}][\mathrm{vec}(H^{(p+2-k)})]_i\Big) \\ &\quad \cdot \Big(\sum_{k=1}^{p} \cos[(k-1)\theta_{t-1}][\mathrm{vec}(H^{(p+2-k)})]_j \\ &\quad - \mathbf{i}\sum_{k=1}^{p} \sin[(k-1)\theta_{t-1}][\mathrm{vec}(H^{(p+2-k)})]_j\Big)\Big\}.\end{aligned}$$

From the independence of $H^{(i)}$, we have

$$\mathbb{E}\left((\mathrm{vec}(H^{(t)}))_i \cdot (\mathrm{vec}(H^{(s)}))_i\right) = 0, \quad t \neq s.$$

It comes to

$$\begin{aligned}\mathrm{cov}(\mathrm{vec}(H_t)_i, \mathrm{vec}(H_t)_j) &= \mathbb{E}\left(\sum_{k=1}^{p} \mathrm{vec}[H^{(p+2-k)}]_i \cdot \mathrm{vec}[H^{(p+2-k)}]_j\right) + \mathbf{i}\mathbb{E}\{0\} \\ &= \Sigma_{ij}^{(1)} + \Sigma_{ij}^{(p)} + \Sigma_{ij}^{(p-1)} + \cdots + \Sigma_{ij}^{(2)}.\end{aligned}$$

That is equivalent to say that the covariance matrix Σ_t of H_t, $(t = 2, 3, \ldots, p)$ is

$$\Sigma^{(1)} + \Sigma^{(2)} + \cdots + \Sigma^{(p)}.$$

In conclusion, we have the following theorem:

Theorem 3.22. *Let $\mathcal{H} \in \mathbb{R}^{m\times n\times p}$ be a third order tensor. If the frontal slices $H^{(i)}$, $(i = 1, 2, \ldots, p)$ are independent and satisfy the distribution*

$$\mathrm{vec}(H^{(i)}) \sim (0, \Sigma^{(i)}),$$

then the diagonal blocks H_i $(i = 1, 2, \ldots, p)$ of $(F_p \otimes I_m)\mathrm{bcirc}(\mathcal{H})(F_p^H \otimes I_n)$ satisfies the distribution

$$\mathrm{vec}(H_i) \sim \left(0, \Sigma^{(1)} + \Sigma^{(2)} + \cdots + \Sigma^{(p)}\right). \tag{3.14}$$

In particular, if the frontal slices $\mathrm{vec}(H^{(i)})$ satisfies the same normal distribution, that is

$$\mathrm{vec}(H^{(i)}) \sim \mathcal{N}(0, \sigma^2\Sigma),$$

by the properties of standard normal distribution and the equation $\mathrm{vec}(H_i) \sim \left(0, \sigma^2(\Sigma^{(1)} + \Sigma^{(2)} + \cdots + \Sigma^{(p)})\right)$, it comes to

$$\mathrm{vec}(H_i) \sim \mathcal{N}(0, \sigma^2 p\Sigma),$$

which means H_i will also have the normal distribution.

$\boxed{3.5}$ Comparison with a Deterministic Error Estimation

Similar to Rice [46], it can be shown that if the tensor function F is Fréchet derivative at $\mathcal{A}$, then κ is the operator norm of the Fréchet derivative of F at $\mathcal{A}$. By using the Fréchet derivative of tensor functions, it can be obtained that the deterministic condition number is

$$\kappa = \lim_{\delta\to 0} \sup_{\|\mathcal{H}\|_F \le \delta} \frac{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F}{\delta} = \sup_{\|\mathcal{H}\|_F \le 1} \|F'(\mathcal{A})(\mathcal{H})\|_F = \|\mathcal{J}_{\mathcal{A}}\|_2 .$$

Therefore, the deterministic first order estimation turns out to be

$$\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F \le \|\mathcal{J}_{\mathcal{A}}\|_2 \|\mathcal{H}\|_F + o(\|\mathcal{H}\|_F).$$

For stochastic cases, the above estimation is not suitable for measuring the sensitivity of random noises, since they are based on the Frobenius norm of $\mathcal{H}$, instead of the distribution. We must turn to the stochastic condition number defined above. By Theorem 3.20 and Corollary 3.21, the stochastic estimation comes to

$$\mathrm{Med}\{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F\} \le 2\sigma \left\|\mathcal{J}_{\mathcal{A}} * \Sigma^{1/2}\right\|_F + o(\|\mathcal{H}\|_F).$$

To gain more insight to the estimation, we suppose all the frontal slices of the pertubation tensor satisfies the standard normal distribution $\mathrm{vec}(H^{(i)}) \sim \mathcal{N}(0, \sigma^2 I_{mn})$, which will lead to $\mathrm{vec}(H_i) \sim \mathcal{N}(0, \sigma^2 p I_{mn})$. Then the deterministic first-order estimation is bounded by

$$\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\| \le \sigma p\sqrt{mn}\, \|\mathcal{J}_{\mathcal{A}}\|_2 \equiv \delta_{det} + o(\|\mathcal{H}\|_F),$$

and the stochastic first-order estimation is bounded by

$$\mathrm{Med}\{\|F(\mathcal{A}+\mathcal{H}) - F(\mathcal{A})\|_F\} \le 2\sigma\sqrt{p}\, \|\mathcal{J}_{\mathcal{A}}\|_F \equiv \delta_{med} + o(\|\mathcal{H}\|_F).$$

Since $\mathcal{J}_{\mathcal{A}} \in \mathbb{R}^{ts\times mn\times p}$, $\|\mathcal{J}_{\mathcal{A}}\|_F \le \min\{\sqrt{mnp}, \sqrt{tsp}\}\, \|\mathcal{J}_{\mathcal{A}}\|_2$, it follows that

$$\frac{\delta_{med}}{\delta_{det}} = \frac{2\sigma\sqrt{p}\, \|\mathcal{J}_{\mathcal{A}}\|_F}{\sigma p\sqrt{mn}\, \|\mathcal{J}_{\mathcal{A}}\|_2} \le \frac{2\sigma\sqrt{p}\min\{\sqrt{mnp}, \sqrt{tsp}\}}{\sigma p\sqrt{mn}} = \frac{2\min\{\sqrt{mn}, \sqrt{ts}\}}{\sqrt{mn}}. \tag{3.15}$$

The absolute value of $\frac{\delta_{med}}{\delta_{det}}$ will be very small in two special cases. The first one is $\|\mathcal{J}_{\mathcal{A}}\|_F \ll \min\{\sqrt{mnp}, \sqrt{tsp}\}\, \|\mathcal{J}_{\mathcal{A}}\|_2$, i.e., there are a lot of T-singular values of the Jacobi tensor $\mathcal{J}_{\mathcal{A}}$ are very large relative to the other singular values. The other one is $ts \ll mn$, which means the domain of the tensor function $F : \Omega \subseteq \mathbb{R}^{m\times n\times p} \to \mathbb{R}^{t\times s\times p}$ is much larger than the T-range space of it.

If $\frac{\delta_{med}}{\delta_{det}} \ll 1$, there will be much difference between the first order estimation computed by the stochastic condition number and deterministic condition number.

$\boxed{3.6}$ Simulation

We give a numerical test which involves both the Moore-Penrose inverse and the first order estimation of the deterministic and the stochastic condition number for tensor functions. Our example is to analysis the following tensor least squares problem [1, 21, 44, 45].

$$F(x) = \arg\min_{x} \|b - \mathcal{A} * x\|_2 , \tag{3.16}$$

whose solution is

$$x = \mathcal{A}^{\dagger} * b.$$

Here $F : \mathbb{R}^{m\times 1\times p} \to \mathbb{R}^{t\times 1\times p}$ is a tensor function, the tensor $\mathcal{A} \in \mathbb{R}^{m\times t\times p}$ is of T-full column rank, and $\mathcal{A}^{\dagger} = (\mathcal{A}^{\top} * \mathcal{A})^{-1} * \mathcal{A}^{\top}$ is the T-Moore-Penrose inverse which can be referred to [40].

Now, we give a toy model to compare the deterministic and stochastic condition number.

We give a perturbation $h \in \mathbb{R}^{m\times 1\times p}$ on tensor $b \in \mathbb{R}^{m\times 1\times p}$, that is

$$b \to b + h,$$

where $h \sim \mathcal{N}(0, \sigma^2\Sigma)$. In this example, we fix $\sigma = 10^{-4}$ and vary m and t.

- The first test is to choose $\Sigma = I$ and $\Sigma = \mathrm{diag}(b)$. Then fix $p = 5$, the ratio of $m/t = 20$ and t varies from $10, 20, \ldots, 400$.
- The second test is to choose choose $\Sigma = I$ and $\Sigma = \mathrm{diag}(b)$. Then fix $p = 5$, $t = 10$ and the ratio m/t varies form $10, 20, \ldots, 1000$.
- The third test is to choose $\Sigma = I$ and $\Sigma = \mathrm{diag}(b)$. Then fix $t = 10$, $m/t = 20$, the number of frontal slices p varies from $10, 20, \ldots, 1000$.

It can be noticed that both the above three cases satisfies the condition $ts \ll mn$, which indicates the large difference between the first-order estimation calculated by the stochastic condition number and the deterministic condition number.

The coefficient tensor $\mathcal{A}$ is given as follows. Suppose $\mathcal{A}$ can be Fourier diagonalized as

$$(F_p \otimes I_m)\mathrm{bcirc}(\mathcal{A})(F_p^H \otimes I_n) = \mathrm{diag}\{A_1, A_2, \ldots, A_p\},$$

where $A_p = U_p S_p V_p^{\top}$, $U_p = I - 2u_p u_p^{\top}$, $V_p = I - 2v_p v_p^{\top}$, and

$$u_p = \begin{bmatrix} \cos(p) \\ \cos(2p) \\ \vdots \\ \cos(mp) \end{bmatrix}, \quad v_p = \begin{bmatrix} \sin(p) \\ \sin(2p) \\ \vdots \\ \sin(tp) \end{bmatrix}.$$

The matrix S_p is chosen to be $eye(m, n)$ in MATLAB.

Let b satisfy

$$(F_p \otimes I_m)\mathrm{bcirc}(b)(F_p^H \otimes 1) = \mathrm{diag}\{b_1, b_2, \ldots, b_p\}, \quad b_i = A_i \cdot [1, 1, \ldots, 1]^{\top}.$$

It can be verified that the Jacobi tensor at b is $\mathcal{J}_b = \mathcal{A}^{\dagger}$.

The following figures illustrate the comparison of relative error. Specifically, in each figure, the Y-coordinate represents the relative error of the above least squares problem

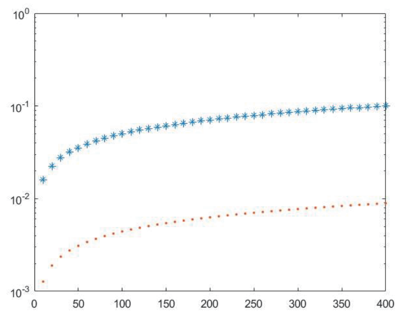

Figure 1: $\Sigma = I$, $p = 5$, $m/t = 20$, $t = 10, 20, \ldots, 400$

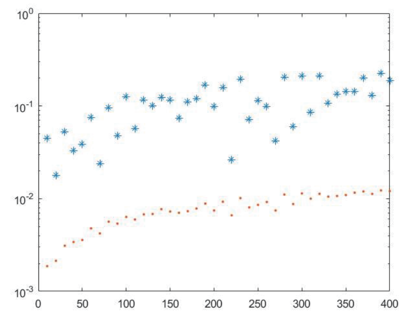

Figure 2: $\Sigma = \mathrm{diag}(b)$, $p = 5$, $m/t = 20$, $t = 10, 20, \ldots, 400$

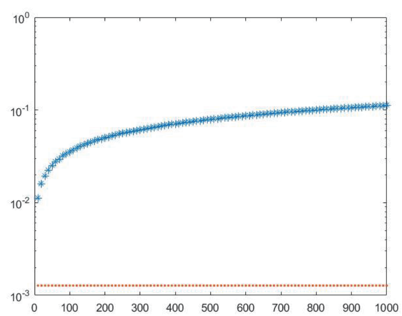

Figure 3: $\Sigma = I$, $p = 5$, $t = 10$ and $m/t = 10, 20, \ldots, 1000$

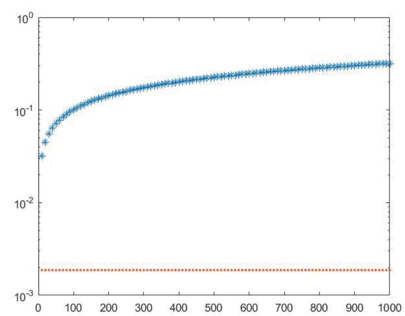

Figure 4: $\Sigma = \mathrm{diag}(b)$, $p = 5$, $t = 10$ and $m/t = 10, 20, \ldots, 1000$

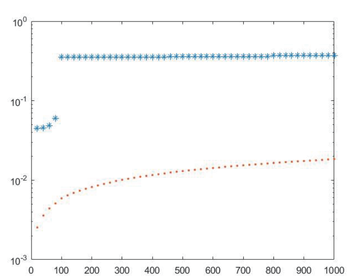

Figure 5: $\Sigma = I$, $t = 10$, $m/t = 20$ and $p = 10, 20, \ldots, 1000$

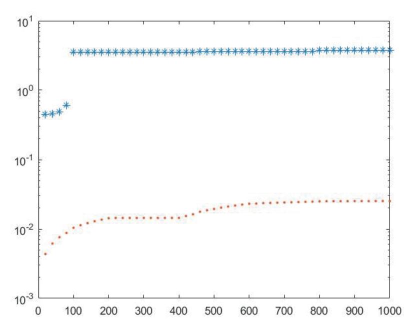

Figure 6: $\Sigma = \mathrm{diag}(b)$, $m/t = 20$ and $p = 10, 20, \ldots, 1000$

(3.16) (as a special example of tensor function), and the meaning of the X-coordinate is marked in each figure. We estimate the relative error with stochastic condition number and deterministic condition number respectively. In the figures, the former is represented by orange points and the latter by blue points.

In conclusion, we find that the stochastic first-order estimation is an excellent example of the median relative error for all values of m and the number of slice p tested. On the other hand, the deterministic first-order estimation can be several orders of magnitude larger and sometimes increasingly so with increasing m/n and p.

4 Conclusion

In this paper we extend a large number of classical results on the matrix perturbation to tensors under the T-product. We present a deterministic perturbation bound for the tensor Moore-Penrose inverse based on the T-product and then obtain a perturbation bound for the solution to the T-Total least squares problem. We focus on the behavior of tensors when they are randomly perturbed, we obtain a perturbation bound of the corresponding Moore-Penrose inverse by using the stochastic norm. Furthermore, we introduce the Fréchet derivative of the generalized tensor function and give the upper bound of stochastic condition number. Both theoretical derivation and numerical experiments show that our stochastic conditioning theory has better properties than the classical conditioning under the random perturbation.

Acknowledgements

The helpful comments of the handling editor and the referee are gratefully acknowledged. This project is supported by the Basic Theory and Algorithm Research between Fudan University and Huawei Technology Investment Co. under grant YBN 19095097.

References

[1] F. Beik, A. Ichi, K. Jbilou and R. Sadaka, Tensor extrapolation methods with applications, *Numer Algorithm.* 87 (2021) 1421-1444.

[2] M. Benzi and P. Boito, Matrix functions in network analysis, *GAMM-Mitteilungen.* 2020; 43:e202000012.

[3] P. Breiding and N. Vannieuwenhoven, On the average condition number of tensor rank decompositions, *IMA J. Numer. Anal.* 40 (2019) 1908–1936.

[4] R. Chan and X. Jin, *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia, 2007.

[5] T. Chan, Y. Yang and Y. Hsuan, Polar n-complex and n-bicomplex singular value decomposition and principal component pursuit, *IEEE Trans. Signal Process.* 64 (2016) 6533–6544.

[6] S. Chang and Y. Wei, General tail bounds for random tensors summation: majorization approach, *J. Comp. Appl. Math.* 416 (2022): 114533.

[7] S. Chang and Y. Wei, Generalized T-product tensor Bernstein bounds, *Ann. Appl. Math.* 38 (2022) 25–61.

[8] S. Chang and Y. Wei, T-product tensors part I: inequalities, *Computational and Applied Mathematics* 41 (2022): article 62.

[9] S. Chang and Y. Wei, T-product tensors part II: tail bounds for sums of random t product tensors, *Comp. Appl. Math.* 41 (2022): Article 99.

[10] M. Che and Y. Wei, Randomized algorithms for the approximations of Tucker and the tensor train decompositions, *Adv. Comput. Math.* 45 (2019) 395–428.

[11] M. Che and Y. Wei, *Theory and Computation of Complex Tensors and its Applications*, Springer, Singapore, 2020.

[12] F. Cucker, H. Diao and Y. Wei, On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems, *Math. Comput.* 76 (2007) 947–963.

[13] H. Diao and Y. Wei, On Frobenius normwise condition numbers for Moore-Penrose inverse and linear least-squares problems, *Numer. Linear Algebra Appl.* 14 (2007) 603–610.

[14] C. Ekanadham, D. Tranchina and E. P. Simoncelli, A unified framework and method for automatic neural spike identification, *Journal of Neuroscience Methods* 222 (2014) 47–55.

[15] G. Ely, S. Aeron, N. Hao and M. E. Kilmer, 5D seismic data completion and denoising using a novel class of tensor decompositions. *Geophysics* 80 (2015) V83–V95.

[16] R. Fletcher, Expected conditioning, *IMA J. Numer. Anal.* 5 (1985) 247–273.

[17] D.F. Gleich, G. Chen and J.M. Varah, The power and Arnoldi methods in an algebra of circulants, *Numer. Linear Algebra Appl.* 20 (2013) 809–831.

[18] G.H. Golub and C.F. Van Loan, Matrix Computations, 4th edition, *Johns Hopkins University Press*, Baltimore, MD, 2013.

[19] G.H. Golub and C.F. Van Loan, An analysis of the total least squares problem, *SIAM J. Numer. Anal.* 17 (1980) 883–893.

[20] S. Gratton and D.T. Peloquin, Stochastic conditioning of matrix functions, *SIAM/ASA J. Uncertainty Quantification.* 2 (2014) 763–783.

[21] M. Guide, A. EL Ichi, K. Jbilou and R. Sadaka, On tensor GMRES and Golub-Kahan methods via the T-product for color image processing, *Electronic Journal of Linear Algebra* 37 (2021) 524–543.

[22] N. Hao, M.E. Kilmer, K. Braman and R.C. Hoover, Facial recognition using tensor-tensor products, *SIAM J. Imaging Sci.* 6 (2013) 437–463.

[23] N. Higham. *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.

[24] N.J. Higham and A.H. Al-Mohy, Computing matrix functions, *Acta Numer.* 19 (2010) 159–208.

[25] A.R. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2nd edition, 2013.

[26] W. Hu, Y. Yang, W. Zhang and Y. Xie, Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition, *IEEE Trans. Image Process.* 26 (2017) 724–737.

[27] H. Jin, M. Bai, J. Bentez and X. Liu, The generalized inverses of tensors and an application to linear models, *Comput. Math. Appli.* 74 (2017) 385–397.

[28] X. Jin, *Developments and Applications of Block Toeplitz Iterative Solvers*, Science Press, Beijing and Kluwer Academic Publishers, Dordrecht, 2002.

[29] 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 J. Matrix Anal. Appl.* 34 (2013) 148–172.

[30] M.E. Kilmer and C.D. Martin, Factorization strategies for third-order tensors, *Linear Algebra Appl.* 435 (2011) 641–658.

[31] H. Kong, X. Xie and Z. Lin, t-Schatten-p norm for low-rank Tensor recovery, *IEEE Journal of Selected Topics in Signal Processing.* 12 (2018) 1405–1419.

[32] Z. Li, Q. Xu and Y. Wei, A note on stable perturbations of Moore-Penrose inverses, *Numer. Linear Algebra Appl.* 20 (2013) 18–26.

[33] Q. Liu, Z. Jia and Y. Wei, Multidimensional total least squares problem with linear equality constraints, *SIAM J. Matrix Anal. Appl.* 43 (2022) 124–150.

[34] Y. Liu, L. Chen and C. Zhu, Improved robust tensor principal component analysis via low-rank core matrix, *IEEE Journal of Selected Topics in Signal Processing* 12 (2018) 1378–1389.

[35] Z. Long, Y Liu, L. Chen and C. Zhu, Low rank tensor completion for multiway visual data, *Signal Processing* 155 (2019) 301–316.

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

[37] H. Ma, N. Li, P. Stanimirović and V. Katsikis, Perturbation theory for Moore-Penrose inverse of tensor via Einstein product, *Comp. Appl. Math.* 38 (2019): article 111.

[38] C.D. Martin, R. Shafer and B. Larue, An order-p tensor factorization with applications in imaging, *SIAM J. Sci. Comput.* 35 (2013) A474–A490.

[39] L. Meng, B. Zheng and Y. Wei, Condition numbers of the multidimensional total least squares problems having more than one solution, *Numerical Algorithms.* 84 (2020) 887–908.

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

[41] Y. Miao, L. Qi and Y. Wei, T-Jordan canonical form and T-Drazin inverse based on the T-product, *Comm. Appl. Math. Comput.* 3 (2021) 201-220.

[42] E. Newman, L. Horesh, H. Avron and M. Kilmer, Stable tensor neural networks for rapid deep learning, *arXiv preprint* arXiv:1811.06569 (2018).

[43] L. Qi and Z. Luo, *Tensor Analysis: Spectral Theory and Special Tensors*, SIAM, 2017.

[44] L. Reichel and U. Ugwu, Tensor Arnoldi-Tikhonov and GMRES-type methods for ill-posed problems with a t-product structure, *J Sci. Comput.* 90 (2022): Article 59.

[45] L. Reichel and U. Ugwu, The tensor Golub-Kahan-Tikhonov method applied to the solution of ill-posed problems with a t-product structure, *Numer. Linear Algebra Appl.* 29 (2022): e2412.

[46] J.R. Rice, A theory of condition, *SIAM J. Numer. Anal.* 3 (1966) 287–310.

[47] S. Soltani, M. E. Kilmer and P.C. Hansen, A tensor-based dictionary learning approach to tomo-graphic image reconstruction, *BIT.* 56 (2016) 1425–1454.

[48] G.W. Stewart, On the continuity of the the generalized inverse, *SIAM J. Applied Mathematics* 17 (1969) 33–45.

[49] G.W. Stewart, On the perturbation of pseudo-inverse, projections and linear least squares problems. *SIAM Review* 19 (1977) 634–662.

[50] G.W. Stewart, The stochastic perturbation theory, *SIAM Review* 32 (1990) 579–610.

[51] L. Sun, B. Zheng, C. Bu and Y. Wei, Moore-Penrose inverse of tensors via Einstein product, *Linear Multilinear Algebra* 64 (2016) 686–698.

[52] D.A. Tarzanagh and G. Michailidis, Fast randomized algorithms for t-product based tensor operations and decompositions with applications to imaging data, *SIAM J. Imag. Science* 11 (2018) 2629–2664.

[53] A. M. Turing, Rounding-off errors in matrix processes, *Quart. J. Mech. Appl. Math.* 1 (1948) 287–308.

[54] A. Wang, Z. Lai and Z. Jin, Noisy low-tubal-rank tensor completion, *Neurocomputing* 330 (2019) 267–279.

[55] G. Wang, Y. Wei and S. Qiao, *Generalized Inverses: Theory and Computations*, Singapore, Springer, Beijing, Science Press, 2018.

[56] X. Wang, L. Gu , H. Lee and G. Zhang, Quantum tensor singular value decomposition., *J. Phys. Commun.* 5 (2021): 075001

[57] X. Wang, M. Che and Y. Wei, Tensor neural network models for tensor singular value decompositions, *Comput. Optim. Appl.* 75 (2020) 753–777.

[58] P. Wedin, Perturbation theory for pseudo-inverses, *BIT.* 13 (1973) 217–232.

[59] M. Wei, The analysis for the total least squares problem with more than one solution, *SIAM J. Matrix Anal. Appl.* 13 (1992) 746–763.

[60] Y. Wei and W. Ding, *Theory and Computation of Tensors*, Elsevier, Academic Press, 2016.

[61] Y. Wei, P. Stanimirović and M. Petković, *Numerical and Symbolic Computations of Generalized Inverses*, World Scientific Publishing, Hackensack, NJ, 2018.

[62] P. Xie, H. Xiang and Y. Wei, Randomized algorithms for total least squares problems, *Numer Linear Algebra Appl.* 26 (2019): e2219.

[63] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang and Y. Qu, On unifying multi-view self-representations for clustering by tensor multi-rank minimization, *Int. J. Comput. Vis.* 126 (2018) 1157–1179.

[64] Y. Xu, Z. Wu, J. Chanussot and Z. Wei, Nonlocal patch tensor sparse representation for hyperspectral image super-resolution, *IEEE Trans. Image. Processing* 28 (2019) 3034–3047.

[65] B. Zheng, L. Meng and Y. Wei, Condition numbers of the multidimensional total least squares problem, *SIAM J. Matrix Anal. Appl.* 38 (2017) 924–948.

Manuscript received 7 November 2021
revised 6 January 2022
accepted for publication 6 February 2022

Yun Miao
Theory Lab, Shanghai R&D Center, Huawei Technologies Ltd.
Shanghai, P.R. China
E-mail address: miaoyun2@huawei.com

Tianru Wang
School of Mathematical Sciences, Fudan University, Shanghai
200433, P.R. China
E-mail address: 20110180027@fudan.edu.cn.

Yimin Wei
School of Mathematical Sciences and
Shanghai Key Laboratory of Contemporary Applied Mathematics
Fudan University, Shanghai, 200433, P.R. China
E-mail address: ymwei@fudan.edu.cn; yimin.wei@gmail.com