Ann. Appl. Math.
doi: 10.4208/aam.OA-2023-0004

Vol. **39**, No. 1, pp. 1-28
February 2023

Random Double Tensors Integrals

Shih Yu Chang[1] and Yimin Wei[2,*]

[1] *Department of Applied Data Science, San Jose State University, San Jose, CA, USA*
[2] *School of Mathematical Sciences and Shanghai Key Laboratory of Contemporary Applied Mathematics, Fudan University, Shanghai 200433, China*

Received 4 February 2023; Accepted (in revised version) 20 February 2023

Abstract. In this work, we try to build a theory for random double tensor integrals (DTI). We begin with the definition of DTI and discuss how randomness structure is built upon DTI. Then, the tail bound of the unitarily invariant norm for the random DTI is established and this bound can help us to derive tail bounds of the unitarily invariant norm for various types of two tensors means, e.g., arithmetic mean, geometric mean, harmonic mean, and general mean. By associating DTI with perturbation formula, i.e., a formula to relate the tensor-valued function difference with respect the difference of the function input tensors, the tail bounds of the unitarily invariant norm for the Lipschitz estimate of tensor-valued function with random tensors as arguments are derived for vanilla case and quasi-commutator case, respectively. We also establish the continuity property for random DTI in the sense of convergence in the random tensor mean, and we apply this continuity property to obtain the tail bound of the unitarily invariant norm for the derivative of the tensor-valued function.

AMS subject classifications: 15A09, 65F20

Key words: Einstein product, double tensor integrals (DTI), random DTI, tail bound, Lipschitz estimate, convergence in the random tensor mean, derivative of tensor-valued function.

---

*Corresponding author.
Emails: `shihyu.chang@sjsu.edu` (S. Chang), `ymwei@fudan.edu.cn` (Y. Wei)

1 Introduction

In recent years, tensors have been applied to different applications in science and engineering [1–3]. However, most of these applications assume that systems modelled by tensors are deterministic and such assumption is not always true and practical in problems involving tensor formulations. In recent years, more research results have pioneered some theories about random tensors [4–6]. One important question in random tensors is about concentration behavior of random tensors. In [7], we extend Lapalace transform method and Lieb's concavity theorem from matrices to tensors, and apply these tools to generalize the classical bounds associated with the names Chernoff, Bennett, and Bernstein from the scalar to the tensor setting. In [8], this work extends previous work by considering the tail behavior of the top k-largest singular values of a function of the tensors summation, instead of the largest/smallest singular value of the tensors summation directly (identity function) explored in [7]. Majorization and antisymmetric tensor product tools are main techniques utilized to establish inequalities for unitarily norms of multivariate tensors. Random tensors summation form discussed in [7,8] is linear form, i.e., each summand of random tensors with degree one. In works [9, 10], we extend the Hanson-Wright inequality for the maximum eigenvalue of the quadratic form of random Hermitian tensors under Einstein product. We separate the quadratic form of random tensors into diagonal summation and coupling (non-diagonal) summation parts. For the diagonal part, we can apply Bernstein inequality to bound the tail probability of the maximum eigenvalue [11] of the summation of independent random Hermitian tensors directly. For coupling summation part, we have to apply decoupling method first, i.e., decoupling inequality to bound expressions with dependent random Hermitian tensors with independent random Hermitian tensors, before applying Bernstein inequality again to bound the tail probability of the maximum eigenvalue of the coupling summation of independent random Hermitian tensors. Previous works are based on tensors with Einstein products. Since Kilmer et al. introduced the new multiplication method between two third-order tensors around 2008 and third-order tensors with such multiplication structure are also called as T-product tensors [12], T-product tensors have been applied to many fields in science and engineering, such as tensor computations [13–20], signal processing, image feature extraction, machine learning, computer vision, and the multi-view clustering problem, etc. The discussion about concentration behaviors based on T-product tensors can also be found in [21–23].

Inspired by operator mean theory (also called Kubo–Ando theory), we try to consider other operations besides + (arithmetic mean) among tensors [24]. The matrix mean for double operators can be expressed by Eq. (5:1:2) in [24], which has the same formation of double operator integral theory discussed in [25]. In this work, we begin to define double tensor integrals (DTI) and consider the tail bound for the unitarily invariant norm of random DTI, see Theorem 3.1. This bound can help us to establish tail bounds for various types of tensor means besides arithmetic mean. Since DTI can be used to express perturbation formula, i.e., an formula to relate the tensor-valued function difference with respect the difference of the function input tensors, we establish Lipschitz estimate for random tensors by the tail bound format, see Theorem 4.1. We also generalize Lipschitz estimate for random tensors with another quasi-commutator tensor, $\mathcal{D}$ by providing the tail bound for the unitarily invariant norm of $\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}$ in Theorem 4.2, where $\mathcal{A},\mathcal{B}$ are random Hermitian tensors. We also establish a continuity for random DTI in the sense of convergence in the random tensor mean. This continuity property helps us to obtain the tail bounds for the unitarily invariant norm of the derivative of the tensor-valued function under vanilla case and quasi-commutator case.

We define double tensor integrals (DTI) and randomness of DTI in Section 2. The tail bound for the unitarily invariant norm of random DTI and its applications to obtain various tail bounds for different types of double tensors means like arithmetic mean, geometric mean, harmonic mean, and general mean, are discussed in Section 3. In Section 4, we establish Lipschitz estimates for random tensors by the tail bound format for vanilla case and quasi-commutator case. We will establish continuity of random DTI based on the convergence in tensor mean of random Hermitian tensors in Section 5. The application of DTI theory to acquire the tail bound for the unitarily invariant norm of the derivative of the tensor-valued function is presented by Section 6. Finally, the conclusions are given in Section 7.

2 Random double tensor integrals

The purpose of this section is to define random double tensor integrals (DTI). We begin with the definition of DTI in Section 2.1. In Section 2.2, we will present what are randomness objects at DTI discussed at this work.

Without loss of generality, one can partition the dimensions of a tensor into two groups, say M and N dimensions, separately. Thus, for two order-$(M+N)$ tensors:

$$\mathcal{X}\stackrel{\text{def}}{=}(x_{i_1,\cdots,i_M,j_1,\cdots,j_N})\in\mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N},$$
$$\mathcal{Y}\stackrel{\text{def}}{=}(y_{i_1,\cdots,i_M,j_1,\cdots,j_N})\in\mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N},$$

according to [26, 27], the *tensor addition* $\mathcal{X}+\mathcal{Y}\in\mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N}$ is given by

$$(\mathcal{X}+\mathcal{Y})_{i_1,\cdots,i_M,j_1,\cdots,j_N}\stackrel{\text{def}}{=}x_{i_1,\cdots,i_M,j_1,\cdots,j_N}+y_{i_1,\cdots,i_M,j_1,\cdots,j_N}. \tag{2.1}$$

On the other hand, for tensors

$$\mathcal{X} \stackrel{\text{def}}{=} (x_{i_1,\cdots,i_M,j_1,\cdots,j_N}) \in \mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N},$$
$$\mathcal{Y} \stackrel{\text{def}}{=} (y_{j_1,\cdots,j_N,k_1,\cdots,k_L}) \in \mathbb{C}^{J_1\times\cdots\times J_N\times K_1\times\cdots\times K_L},$$

according to [26, 27], the *Einstein product* (or simply referred to as *tensor product* in this work) $\mathcal{X}\star_N\mathcal{Y}\in\mathbb{C}^{I_1\times\cdots\times I_M\times K_1\times\cdots\times K_L}$ is given by

$$(\mathcal{X}\star_N\mathcal{Y})_{i_1,\cdots,i_M,k_1,\cdots,k_L} \stackrel{\text{def}}{=} \sum_{j_1,\cdots,j_N} x_{i_1,\cdots,i_M,j_1,\cdots,j_N} y_{j_1,\cdots,j_N,k_1,\cdots,k_L}. \tag{2.2}$$

One can find more preliminary facts about tensors based on Einstein product in [7, 26, 27]. In the remaining of this paper, we will represent the scalar value $I_1\times\cdots\times I_N$ by $\mathbb{I}_1^N$.

In order to define *Hermitian* tensor, the *conjugate transpose operation* (or *Hermitian adjoint*) of a tensor is specified as follows.

Definition 2.1. *Given a tensor*

$$\mathcal{A} \stackrel{def}{=} (a_{i_1,\cdots,i_M,j_1,\cdots,j_N}) \in \mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N},$$

its conjugate transpose, denoted by $\mathcal{A}^H$, is defined by

$$(\mathcal{A}^H)_{j_1,\cdots,j_N,i_1,\cdots,i_M} \stackrel{def}{=} \overline{a_{i_1,\cdots,i_M,j_1,\cdots,j_N}}, \tag{2.3}$$

where the overline notion indicates the complex conjugate of the complex number $a_{i_1,\cdots,i_M,j_1,\cdots,j_N}$. If a tensor $\mathcal{A}$ satisfies $\mathcal{A}^H=\mathcal{A}$, then $\mathcal{A}$ is a Hermitian tensor. *A random* Hermitian tensor *is a* Hermitian tensor *with diagonal entries are real random variables, and non-diagonal entries are complex random variables.*

Definition 2.2. *Given a tensor*

$$\mathcal{A} \stackrel{def}{=} (a_{i_1,\cdots,i_M,j_1,\cdots,j_M}) \in \mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_M},$$

if

$$\mathcal{A}^H\star_M\mathcal{A} = \mathcal{A}\star_M\mathcal{A}^H = \mathcal{I} \in \mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_M}, \tag{2.4}$$

then $\mathcal{A}$ is a unitary tensor.

Definition 2.3. *Given a* square tensor

$$\mathcal{A} \stackrel{def}{=} (a_{i_1,\cdots,i_M,j_1,\cdots,j_M}) \in \mathbb{C}^{I_1\times\cdots\times I_M\times I_1\times\cdots\times I_M},$$

if there exists $\mathcal{X}\in\mathbb{C}^{I_1\times\cdots\times I_M\times I_1\times\cdots\times I_M}$ *such that*

$$\mathcal{A}\star_M\mathcal{X}=\mathcal{X}\star_M\mathcal{A}=\mathcal{I}, \tag{2.5}$$

then $\mathcal{X}$ *is the* inverse *of* $\mathcal{A}$. *We usually write*

$$\mathcal{X}\overset{def}{=}\mathcal{A}^{-1}$$

thereby.

We also list other crucial tensor operations here. The *trace* of a tensor is equivalent to the summation of all diagonal entries such that

$$\mathrm{Tr}(\mathcal{A})\overset{\text{def}}{=}\sum_{1\le i_j\le I_j,\, j\in[N]}\mathcal{A}_{i_1,\cdots,i_M,i_1,\cdots,i_M}. \tag{2.6}$$

The *inner product* of two tensors $\mathcal{A}$, $\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N}$ is given by

$$\langle\mathcal{A},\mathcal{B}\rangle\overset{\text{def}}{=}\mathrm{Tr}\left(\mathcal{A}^H\star_M\mathcal{B}\right). \tag{2.7}$$

2.1 Double tensor integrals

From Theorem 3.2 in [26], every Hermitian tensor $\mathcal{H}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ has the following decomposition

$$\begin{aligned}\mathcal{H}&=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{with} \quad \langle\mathcal{U}_i,\mathcal{U}_i\rangle=1 \quad \text{and} \quad \langle\mathcal{U}_i,\mathcal{U}_j\rangle=0 \quad \text{for} \quad i\neq j\\ &\overset{\text{def}}{=}\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{P}_{\mathcal{U}_i},\end{aligned} \tag{2.8}$$

where $\mathcal{U}_i\in\mathbb{C}^{I_1\times\cdots\times I_N\times 1}$, and the tensor $\mathcal{P}_{\mathcal{U}_i}$ is defined as $\mathcal{U}_i\star_1\mathcal{U}_i^H$. The values λ_i are named as *eigevalues*. A Hermitian tensor with the decomposition shown by Eq. (2.8) is named as *eigen-decomposition*. A Hermitian tensor $\mathcal{H}$ is a positive definite (or positive semi-definite) tensor if all its eigenvalues are positive (or nonnegative).

Let $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be Hermitian tensors with the following eigen-decompositions:

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H\overset{\text{def}}{=}\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{P}_{\mathcal{U}_i}, \tag{2.9}$$

and

$$\mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_i\star_1\mathcal{V}_j^H\overset{\text{def}}{=}\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{P}_{\mathcal{V}_j}. \tag{2.10}$$

We define *double tensor integrals* (DTI) with respect to tensors $\mathcal{A}$, $\mathcal{B}$ and the function $\psi:\mathbb{R}^2\to\mathbb{R}$, denoted by $T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})$, which can be expressed as

$$T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})=\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\psi(\lambda_i,\mu_j)\mathcal{P}_{\mathcal{U}_i}\star_N\mathcal{X}\star_N\mathcal{P}_{\mathcal{V}_j}. \tag{2.11}$$

Lemma 2.1. *Let $\psi,\phi:\mathbb{R}^2\to\mathbb{R}$ be two functions, we have following relationships about $T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})$: Given ψ is a constant function to one, we have*

$$(1)\quad T_{\mathcal{A},\mathcal{B},1}=\mathcal{I}. \tag{2.12}$$

We also have:

$$(2)\quad T_{\mathcal{A},\mathcal{B},\psi(\phi)}(\mathcal{X})=T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\circ T_{\mathcal{A},\mathcal{B},\phi}(\mathcal{X}), \tag{2.13}$$

where $\circ$ is the entrywise product (Hadamard product), and $\psi(\phi)$ is the composition of the function ψ and the function ϕ. Finally, we have

$$(3)\quad T_{\mathcal{A},\mathcal{B},a\psi+b\phi}(\mathcal{X})=aT_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})+bT_{\mathcal{A},\mathcal{B},\phi}(\mathcal{X}), \tag{2.14}$$

where a, b are two complex numbers.

Proof. Given a tensor $\mathcal{X}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ with orthgonal unitary tensors $\{\mathcal{U}_i\}$ and $\{\mathcal{V}_j\}$ such that

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad\text{and}\quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H,$$

we define the scaler $x_{i,j}$ associated to $\mathcal{X}$ as

$$x_{i,j}=\langle\mathcal{X}\star_N\mathcal{V}_j,\mathcal{U}_i\rangle. \tag{2.15}$$

After selecting two specific indices i' and j', we have

$$\langle T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\star_N\mathcal{V}_{j'},\mathcal{U}_{i'}\rangle=\psi(\lambda_{i'},\mu_{j'})\langle\mathcal{X}\star_N\mathcal{V}_{j'},\mathcal{U}_{i'}\rangle=\psi(\lambda_{i'},\mu_{j'})x_{i',j'}. \tag{2.16}$$

From Eq. (2.16) and the Hadamard product properties, we have properties provided by Eqs. (2.12), (2.13) and (2.14) according to the identity matrix, matrix multiplication and linearity of matrix operations properties. □

2.2 Random DTI

According to the DTI definition provided by Eq. (2.11), the random DTI considered in this work is to assume that tensors $\mathcal{A}$, $\mathcal{B}$ are random Hermitian tensors and the remaining parameters ψ and $\mathcal{X}$ are deterministics. Therefore, we have the randomness at the following terms in Eq. (2.11): $\psi(\lambda_i,\mu_j)$, $\mathcal{P}_{\mathcal{U}_i}$ and $\mathcal{P}_{\mathcal{V}_j}$. If we are provided more detailed probability density functions for entries of random Hermitian tensors $\mathcal{A}$, $\mathcal{B}$, all bounds derived in this work can be improved with more dedicated expressions associated with parameters of probability density functions.

3 Tail bound for random tensor integral norms

3.1 Unitarily invariant tensor norms

Let us represent the Hermitian eigenvalues of a Hermitian tensor $\mathcal{H} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ in decreasing order by the vector

$$\vec{\lambda}(\mathcal{H})=(\lambda_1(\mathcal{H}),\cdots,\lambda_{\mathbb{I}_1^N}(\mathcal{H})).$$

We use $\mathbb{R}_{\geq 0}(\mathbb{R}_{>0})$ to represent a set of nonnegative (positive) real numbers. Let $\|\cdot\|_\rho$ be a unitarily invariant tensor norm, i.e.,

$$\|\mathcal{H}\star_N\mathcal{U}\|_\rho=\|\mathcal{U}\star_N\mathcal{H}\|_\rho=\|\mathcal{H}\|_\rho,$$

where $\mathcal{U}$ is any unitary tensor. Let $\rho:\mathbb{R}_{\geq 0}^{\mathbb{I}_1^N}\to\mathbb{R}_{\geq 0}$ be the corresponding gauge function that satisfies Hölder's inequality so that

$$\|\mathcal{H}\|_\rho=\||\mathcal{H}|\|_\rho=\rho(\vec{\lambda}(|\mathcal{H}|)), \tag{3.1}$$

where

$$|\mathcal{H}|\overset{\text{def}}{=}\sqrt{\mathcal{H}^H\star_N\mathcal{H}}.$$

We will provide several popular tensor norm examples which can be treated as special cases of unitarily invariant tensor norm. The first one is Schatten p-norm for tensors, denoted by $\|\mathcal{X}\|_p$, is defined as:

$$\|\mathcal{X}\|_p\overset{\text{def}}{=}(\mathrm{Tr}|\mathcal{X}|^p)^{\frac{1}{p}}, \tag{3.2}$$

where $p\geq 1$. If $p=1$, it is the trace norm.

Given a Hermitian tensor $\mathcal{X}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, the second norm about such tensor $\mathcal{X}$ is k-trace norm, denoted by $\mathrm{Tr}_k[\mathcal{X}]$. It is defined in [28] as

$$\mathrm{Tr}_k[\mathcal{X}]\stackrel{\text{def}}{=}\sum_{1\le i_1<i_2<\cdots i_k\le r}\lambda_{i_1}\lambda_{i_1}\cdots\lambda_{i_k}, \tag{3.3}$$

where $1\le k\le r$ and $r=\prod_{i=1}^N I_i$. If $k=1$, $\mathrm{Tr}_k[\mathcal{X}]$ is reduced as trace norm.

The third one is Ky Fan k-norm [29] for tensors. For $k\in\{1,2,\cdots,\mathbb{I}_1^N\}$, the Ky Fan k-norm [29] for tensors $\mathcal{X}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, denoted by $\|\mathcal{X}\|_{(k)}$, is defined as:

$$\|\mathcal{X}\|_{(k)}\stackrel{\text{def}}{=}\sum_{i=1}^{k}\lambda_i(|\mathcal{X}|). \tag{3.4}$$

If k=1, the Ky Fan k-norm for tensors is the tensor operator norm, denoted by $\|\mathcal{X}\|$. In this work, we will apply the symbol $\|\mathcal{X}\|_\rho$ to represent any unitarily invariant tensor norm for the tensor $\mathcal{X}$.

In the following theorem, we will present the tail bound of unitarily invariant tensor norm for a given tensor integral $T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})$. Given two random Hermitian tensors $\mathcal{A}$ and $\mathcal{B}$, they are independent if any entry of the random tensor $\mathcal{A}$ is independent of any entry of the random tensor $\mathcal{B}$.

Theorem 3.1. *Let $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be independent random Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H,$$

then, for any $\theta>0$, we have

$$\Pr\Big(\|T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\|_\rho\ge\theta\Big)\le\frac{\left(\mathbb{I}_1^N\right)^2\|\mathcal{X}\|_\rho}{\theta}\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\mathbb{E}\left(|\psi(\lambda_i,\mu_j)|\right), \tag{3.5}$$

where $\mathbb{E}$ is the expectation.

Proof. Since we have the following norm estimation for $\|T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\|_\rho$:

$$\|T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\|_\rho=\left\|\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\psi(\lambda_i,\mu_j)\mathcal{P}_{\mathcal{U}_i}\star_N\mathcal{X}\star_N\mathcal{P}_{\mathcal{V}_j}\right\|_\rho \tag{3.6}$$

$$\begin{aligned}
&\leq_1 \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}|\psi(\lambda_i,\mu_j)|\left\|\mathcal{P}_{\mathcal{U}_i}\star_N\mathcal{X}\star_N\mathcal{P}_{\mathcal{V}_j}\right\|_\rho \\
&=_2 \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}|\psi(\lambda_i,\mu_j)|\left\|\mathcal{X}\right\|_\rho, \qquad (3.7)
\end{aligned}$$

where $\leq_1$ comes from triangle inequality of the unitarily invariant norm and $=_2$ comes from the definition of the unitarily invariant norm.

Then, we have the following bound for $\Pr(\|T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\|_\rho \geq \theta)$

$$\begin{aligned}
\Pr\left(\left\|T_{\mathcal{A},\mathcal{B},\psi}(\mathcal{X})\right\|_\rho \geq \theta\right) &\leq_1 \Pr\left(\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}|\psi(\lambda_i,\mu_j)|\left\|\mathcal{X}\right\|_\rho \geq \theta\right) \\
&\leq \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\Pr\left(|\psi(\lambda_i,\mu_j)| \geq \frac{\theta}{\left\|\mathcal{X}\right\|_\rho(\mathbb{I}_1^N)^2}\right) \\
&\leq_2 \frac{\left\|\mathcal{X}\right\|_\rho(\mathbb{I}_1^N)^2}{\theta}\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\mathbb{E}(|\psi(\lambda_i,\mu_j)|), \qquad (3.8)
\end{aligned}$$

where $\leq_1$ comes from the inequality obtained by Eq. (3.6), and the $\leq_2$ is based on Markov inequality. □

We will consider several important examples of ψ, which will represent different tensor means. Given two random Hermitian tensors

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H, \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H,$$

we use $p_{\lambda_i}(\ )$ and $p_{\mu_j}(\ )$ to represent the probability density functions for eigenvalues λ_i and μ_j, respectively. From Section 2.2. in [26], one can apply unfolding technique to convert a random Hermitian tensor into a Hermitian matrix. If we have Assumption 3.1 in [30], we are able to obtain the i-th eigenvalue distribution of a positive definite Hermitian tensor, see Corollary 3.3 in [30].

Corollary 3.1 (Arithmetic Mean)**.** *Under conditions provided by Theorem* 3.1*, if the function* ψ *has the following form:*

$$\psi(x,y)=\frac{x+y}{2}, \qquad (3.9)$$

we have

$$\Pr\left(\left\|T_{\mathcal{A},\mathcal{B},\psi=\frac{x+y}{2}}(\mathcal{X})\right\|_\rho \geq \theta\right) \leq \frac{\left(\mathbb{I}_1^N\right)^2 \|\mathcal{X}\|_\rho}{2\theta} \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}[\mathbb{E}(|\lambda_i|)+\mathbb{E}(|\mu_j|)], \tag{3.10}$$

where

$$\mathbb{E}(|\lambda_i|) = \int_0^\infty t[p_{\lambda_i}(t)+p_{\lambda_i}(-t)]\,dt, \tag{3.11}$$

and

$$\mathbb{E}(|\mu_j|) = \int_0^\infty t\left[p_{\mu_j}(t)+p_{\mu_j}(-t)\right]dt. \tag{3.12}$$

Proof. The key is to evaluate $\mathbb{E}(|\psi(\lambda_i,\mu_j)|)$, we have

$$\begin{aligned}\mathbb{E}(|\psi(\lambda_i,\mu_j)|) &= \mathbb{E}\left(\left|\frac{\lambda_i+\mu_j}{2}\right|\right) \\ &\leq \frac{1}{2}\mathbb{E}(|\lambda_i|)+\frac{1}{2}\mathbb{E}(|\mu_j|).\end{aligned} \tag{3.13}$$

This corollary is proved by the following fact for a random variable $Y=|X|$, where $|\cdot|$ is the abolute value operator:

$$F_Y(y) = \begin{cases} F_X(y)-F_X(-y), & y\geq 0, \\ 0, & y<0, \end{cases} \tag{3.14}$$

where $F_Y(y)$ and $F_X(y)$ are CDFs of random variables Y and X. The variables y in $F_Y(y)$ and $F_X(y)$ are function arguments in CDFs $F_Y(y)$ and $F_X(y)$. □

Corollary 3.2 (Geometric Mean)**.** *Under conditions provided by Theorem* 3.1 *with the assumption that $\mathcal{A}$ and $\mathcal{B}$ are random positive definite tensors, if the function ψ has the following form:*

$$\psi(x,y) = \sqrt{xy}, \tag{3.15}$$

we have

$$\Pr\left(\left\|T_{\mathcal{A},\mathcal{B},\psi=\sqrt{xy}}(\mathcal{X})\right\|_\rho \geq \theta\right) \leq \frac{\left(\mathbb{I}_1^N\right)^2 \|\mathcal{X}\|_\rho}{\theta} \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\left[\mathbb{E}\left(\sqrt{\lambda_i}\right)\mathbb{E}\left(\sqrt{\mu_j}\right)\right], \tag{3.16}$$

where

$$\mathbb{E}\left(\sqrt{\lambda_i}\right) = \int_0^\infty \sqrt{t}p_{\lambda_i}(t)dt, \quad \mathbb{E}\left(\sqrt{\mu_j}\right) = \int_0^\infty \sqrt{t}p_{\mu_j}(t)dt. \tag{3.17}$$

Proof. To evaluate $\mathbb{E}(\sqrt{\lambda_i\mu_j})$, we have

$$\begin{aligned}\mathbb{E}(|\psi(\lambda_i,\mu_j)|)=&\mathbb{E}\left(\sqrt{\lambda_i\mu_j}\right)\\=&\mathbb{E}\left(\sqrt{\lambda_i}\right)\mathbb{E}\left(\sqrt{\mu_j}\right).\end{aligned}\tag{3.18}$$

Thus, we complete the proof. □

Corollary 3.3 (Harmonic Mean). *Under conditions provided by Theorem* 3.1 *with the assumption that $\mathcal{A}$ and $\mathcal{B}$ are random positive definite tensors, if the function ψ has the following form:*

$$\psi(x,y)=\frac{2}{x^{-1}+y^{-1}},\tag{3.19}$$

we have

$$\Pr\left(\left\|T_{\mathcal{A},\mathcal{B},\psi=\frac{2}{x^{-1}+y^{-1}}}(\mathcal{X})\right\|_\rho\geq\theta\right)\leq\frac{2\left(\mathbb{I}_1^N\right)^2\|\mathcal{X}\|_\rho}{\theta}\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\int_0^\infty \mathrm{H}_{\lambda_i}(t)\mathrm{H}_{\mu_j}(t)dt,\tag{3.20}$$

where

$$\mathrm{H}_{\lambda_i}(t)=\int_0^\infty\frac{e^{-tw}}{w^2}p_{\lambda_i}(1/w)dw,\quad \mathrm{H}_{\mu_j}(t)=\int_0^\infty\frac{e^{-tz}}{z^2}p_{\mu_j}(1/z)dz.\tag{3.21}$$

Proof. Let X and Y are two positive random variables with distribution functions $f_X(x)$ and $f_Y(y)$. Then, we have

$$\begin{aligned}\mathbb{E}\left(\frac{2}{X^{-1}+Y^{-1}}\right)=_1&\mathbb{E}\left(\frac{2}{W+Z}\right)\\=&2\mathbb{E}\left(\int_0^\infty\exp(-t(W+Z))dt\right)\\=&2\int_0^\infty\mathbb{E}(\exp(-t(W+Z)))dt\\=_2&2\int_0^\infty\mathbb{E}(\exp(-tW))\mathbb{E}(\exp(-tZ))dt,\end{aligned}\tag{3.22}$$

where we set $W=X^{-1}$ and $Z=Y^{-1}$ at $=_1$ and we use independent assumptions of random variables $W=X^{-1}$ and $Z=Y^{-1}$ at $=_2$. This corollary is proved since we have following distribution functions for random variables W and Z, which will correspond to random variables λ_i and μ_j, expressed as

$$f(w)=\frac{1}{w^2}p_{\lambda_i}(1/w),\quad f(z)=\frac{1}{z^2}p_{\mu_j}(1/z).\tag{3.23}$$

Thus, we complete the proof. □

We have to prepare a lemma about the expectation of ratio between two dependent random variables before presenting the next corollary.

Lemma 3.1. *Given two random variables X and Y such that $Y \neq 0$ always, we have*

$$\mathbb{E}\left(\frac{X}{Y}\right) = \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} + \lim_{\epsilon \to 0} \sum_{i=1}^{\infty} \frac{(-1)^i \left[\mathbb{E}(X)\mathbb{E}(\acute{Y}^i) + \mathbb{E}(\acute{X}\acute{Y}^i)\right]}{\prod_{j=0}^{i} (\mathbb{E}(Y) + j\epsilon)}, \tag{3.24}$$

where we define the following random variables:

$$\acute{X} \stackrel{def}{=} X - \mathbb{E}(X), \quad \acute{Y} \stackrel{def}{=} Y - \mathbb{E}(Y). \tag{3.25}$$

Proof. We have the following expression about $\mathbb{E}(\frac{X}{Y})$, it is

$$\begin{aligned}
\mathbb{E}\left(\frac{X}{Y}\right) &= \mathbb{E}\left(\frac{\mathbb{E}(X)}{\mathbb{E}(Y)} \times \frac{1 + \frac{\acute{X}}{\mathbb{E}(X)}}{1 + \frac{\acute{Y}}{\mathbb{E}(Y)}}\right) \\
&= \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} \mathbb{E}\left(\left(1 + \frac{\acute{X}}{\mathbb{E}(X)}\right)\left(1 + \frac{\acute{Y}}{\mathbb{E}(Y)}\right)^{-1}\right) \\
&= \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} \mathbb{E}\left(\left(1 + \frac{\acute{Y}}{\mathbb{E}(Y)}\right)^{-1}\right) + \frac{1}{\mathbb{E}(Y)} \mathbb{E}\left(\acute{X}\left(1 + \frac{\acute{Y}}{\mathbb{E}(Y)}\right)^{-1}\right).
\end{aligned} \tag{3.26}$$

Given a real function g, we have the following approximation form:

$$g(a+x) = \lim_{\epsilon \to 0} \sum_{i=0}^{\infty} \frac{x^i \Delta_\epsilon^i g(a)}{\epsilon^i i!}, \tag{3.27}$$

where Δ_ϵ^i is the finite difference operator of degree i and the step size ϵ is defined as

$$\Delta_\epsilon^i g(a) = \sum_{j=0}^{i} (-1)^j \binom{i}{j} g(a + (i-j)\epsilon). \tag{3.28}$$

If we apply Eq. (3.27) to the following function at $\acute{Y} = 0$

$$g(\acute{Y}) = \left(1 + \frac{\acute{Y}}{\mathbb{E}(Y)}\right)^{-1}, \tag{3.29}$$

we will get

$$\left(1+\frac{\acute{Y}}{\mathbb{E}(Y)}\right)^{-1}=\lim_{\epsilon\to 0}\sum_{i=0}^{\infty}\frac{(-1)^i(\acute{Y})^i\mathbb{E}(Y)}{\prod\limits_{j=0}^{i}(\mathbb{E}(Y)+j\epsilon)}. \tag{3.30}$$

This lemma is proved by applying Eq (3.30) to Eq. (3.26). □

Following corollary is about the unitarily invariant norm tail bound for the two tensors mean general format. Note that we have logarithmic mean if $\alpha\to 1$ in Eq. (3.32).

Corollary 3.4 (General Mean)**.** *Under conditions provided by Theorem* 3.1 *with the assumption that $\mathcal{A}$ and $\mathcal{B}$ are random positive definite tensors, if the function ψ has the following form:*

$$\psi(x,y)=\frac{\alpha-1}{\alpha}\frac{x^{\alpha}-y^{\alpha}}{x^{\alpha-1}-y^{\alpha-1}}, \tag{3.31}$$

where $\alpha\in\mathbb{R}$ and $x\neq y$ (if $x=y$, this situation has measure zero). Then we have

$$\begin{aligned}&\Pr\left(\left\|T_{\mathcal{A},\mathcal{B},\psi=\frac{\alpha-1}{\alpha}\frac{x^{\alpha}-y^{\alpha}}{x^{\alpha-1}-y^{\alpha-1}}}(\mathcal{X})\right\|_{\rho}\geq\theta\right) \\ &\leq\frac{\left(\mathbb{I}_1^N\right)^2\|\mathcal{X}\|_{\rho}(\alpha-1)}{\theta\alpha} \\ &\quad\times\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\left\{\frac{\mathbb{E}\left(X_{\lambda_i,\mu_j}\right)}{\mathbb{E}\left(Y_{\lambda_i,\mu_j}\right)}+\lim_{\epsilon\to 0}\sum_{\imath=1}^{\infty}\frac{(-1)^{\imath}\left[\mathbb{E}(X_{\lambda_i,\mu_j})\mathbb{E}((\acute{Y}_{\lambda_i,\mu_j})^{\imath})+\mathbb{E}(\acute{X}_{\lambda_i,\mu_j}(\acute{Y}_{\lambda_i,\mu_j})^{\imath})\right]}{\prod\limits_{\jmath=0}^{\imath}(\mathbb{E}(Y_{\lambda_i,\mu_j})+\jmath\epsilon)}\right\},\end{aligned} \tag{3.32}$$

where random variables X_{λ_i,μ_j} and Y_{λ_i,μ_j} are defined by

$$X_{\lambda_i,\mu_j}=\lambda_i^{\alpha}-\mu_j^{\alpha}, \tag{3.33}$$

and

$$Y_{\lambda_i,\mu_j}=\lambda_i^{\alpha-1}-\mu_j^{\alpha-1}. \tag{3.34}$$

Proof. Since $|\psi(\lambda_i,\mu_j)|$ is

$$\psi(\lambda_i,\mu_j)=\frac{\alpha-1}{\alpha}\frac{\lambda_i^{\alpha}-\mu_j^{\alpha}}{\lambda_i^{\alpha-1}-\mu_j^{\alpha-1}}=\frac{\alpha-1}{\alpha}\frac{X_{\lambda_i,\mu_j}}{Y_{\lambda_i,\mu_j}}, \tag{3.35}$$

this corollary is proved by applying Lemma 3.1 to the expectation of Eq. (3.35). □

Note that each of the following terms $\mathbb{E}\left(X_{\lambda_i,\mu_j}\right)$, $\mathbb{E}\left(Y_{\lambda_i,\mu_j}\right)$, $\mathbb{E}((\acute{Y}_{\lambda_i,\mu_j})^{\imath})$ and $\mathbb{E}(\acute{X}_{\lambda_i,\mu_j}(\acute{Y}_{\lambda_i,\mu_j})^{\imath})$ can be evaluated exactly since we know all density distributions $p_{\lambda_i}(\ )$ and $p_{\mu_j}(\ )$.

4 Tail bounds for random Lipschitz estimates

In this section, we will try to provide tail bounds for the Lipschitz estimate for the unitarily invariant norm for a given function, which is the main result of this section. We will begin with the perturbation lemma. The vanilla case is discussed in Section 4.1. The case about considering quasi-commutator is provided by Section 4.2.

4.1 Vanilla case

We will begin by providing a perturbation formula for DTI.

Lemma 4.1. *Let* $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ *be Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \textit{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{U}_j^V.$$

Also, let $f:\mathbb{R}\to\mathbb{R}$ *such that* $f'(x)$ *exists, we define the bivariate function* $f^{[1]}$ *as*

$$f^{[1]}(x,y)=\begin{cases}\dfrac{f(x)-f(y)}{x-y}, & x\neq y,\\ f'(x), & x=y.\end{cases} \tag{4.1}$$

Then, we have

$$f(\mathcal{A})-f(\mathcal{B})=T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{A}-\mathcal{B}). \tag{4.2}$$

Proof. Since we have

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H,\quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{U}_j\star_1\mathcal{U}_j^H, \tag{4.3}$$

then, we will obtain the following

$$f(\mathcal{A})=\sum_{i=1}^{\mathbb{I}_1^N}f(\lambda_i)\mathcal{U}_i\star_1\mathcal{U}_i^H,\quad f(\mathcal{B})=\sum_{j=1}^{\mathbb{I}_1^N}f(\mu_j)\mathcal{V}_j\star_1\mathcal{V}_j^H. \tag{4.4}$$

Because, we also have

$$f(\mathcal{A})\star_N\mathcal{U}_i=f(\lambda_i)\star_N\mathcal{U}_i,\quad f(\mathcal{B})\star_N\mathcal{V}_j=f(\mu_j)\star_N\mathcal{V}_j, \tag{4.5}$$

then, we have

$$\langle (f(\mathcal{A})-f(\mathcal{B})\star_N \mathcal{V}_j,\mathcal{U}_i)\rangle = f^{[1]}(\lambda_i,\mu_j)\langle (\mathcal{A}-\mathcal{B})\star_N \mathcal{V}_j,\mathcal{U}_i\rangle. \tag{4.6}$$

By applying Eq. (2.11) definition and Eq. (2.16) to Eq. (4.6), thus this Lemma is proved. □

From perturbation formula given by Lemma 4.1, we can have the following theorem about the tail bounds of the unitarily invariant norm for the Lipschitz estimate of tensor-valued function with random tensors as inputs.

Theorem 4.1. *Let $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be independent random Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H,$$

moreover, we are given a real valued function $f(x)$ for $x\in\mathbb{R}$ such that $\left|f^{[1]}(x,y)\right|$ is bounded by a positive number denoted by $\Omega_{f^{[1]}}$. Then, for any $\theta>0$, we have

$$\Pr\Big(\|f(\mathcal{A})-f(\mathcal{B})\|_\rho\geq\theta\Big)\leq\frac{\left(\mathbb{I}_1^N\right)^2\Omega_{f^{[1]}}}{\theta}\mathbb{E}\Big(\|\mathcal{A}-\mathcal{B}\|_\rho\Big). \tag{4.7}$$

Proof. We have

$$\begin{aligned}
\|f(\mathcal{A})-f(\mathcal{B})\|_\rho &= \left\|T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{A}-\mathcal{B})\right\|_\rho \\
&= \left\|\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N} f^{[1]}(\lambda_i,\mu_j)\mathcal{P}_{\mathcal{U}_i}\star_N(\mathcal{A}-\mathcal{B})\star_N\mathcal{P}_{\mathcal{V}_j}\right\|_\rho \\
&\leq_1 \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\left|f^{[1]}(\lambda_i,\mu_j)\right|\left\|\mathcal{P}_{\mathcal{U}_i}\star_N(\mathcal{A}-\mathcal{B})\star_N\mathcal{P}_{\mathcal{V}_j}\right\|_\rho \\
&=_2 \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\left|f^{[1]}(\lambda_i,\mu_j)\right|\|\mathcal{A}-\mathcal{B}\|_\rho \\
&\leq_3 \left(\mathbb{I}_1^N\right)^2\Omega_{f^{[1]}}\|\mathcal{A}-\mathcal{B}\|_\rho,
\end{aligned} \tag{4.8}$$

where $\leq_1$ comes from triangle inequality of the unitarily invariant norm, $=_2$ comes from the definition of the unitarily invariant norm, and $\leq_3$ is due to that $\left|f^{[1]}(\lambda_i,\mu_j)\right|\leq\Omega_{f^{[1]}}$ (mean value theorem of divide difference).

Then, we have the following bound for $\Pr(\|f(\mathcal{A})-f(\mathcal{B})\|_\rho \geq \theta)$

$$\begin{aligned}\Pr\Big(\|f(\mathcal{A})-f(\mathcal{B})\|_\rho \geq \theta\Big) &\leq_1 \Pr\Big(\big(\mathbb{I}_1^N\big)^2 \Omega_{f^{[1]}} \|\mathcal{A}-\mathcal{B}\|_\rho \geq \theta\Big) \\ &\leq_2 \frac{\big(\mathbb{I}_1^N\big)^2 \Omega_{f^{[1]}}}{\theta} \mathbb{E}\Big(|\mathcal{A}-\mathcal{B}|_\rho\Big), \end{aligned} \tag{4.9}$$

where $\leq_1$ comes from the inequality obtained by Eq. (4.8), and the $\leq_2$ is based on Markov inequality. □

Since the upper bound $\Omega_{f^{[1]}}$ depends on $f(x)$, we will consider the following two corollaries about special types of the function $f(x)$.

Corollary 4.1 (Lipschitz Estimate for Polynomial Functions)**.** *Let* $\mathcal{A},\mathcal{B} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ *be independent random Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N} \lambda_i \mathcal{U}_i \star_1 \mathcal{U}_i^H \quad \textit{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N} \mu_j \mathcal{V}_j \star_1 \mathcal{V}_j^H.$$

Besides, we are given a real polynomial function $f(x)$ *with degree* m *over an interval* $[a,b]$*. For any* $\theta>0$*, we have*

$$\Pr\Big(\|f(\mathcal{A})-f(\mathcal{B})\|_\rho \geq \theta\Big) \leq \frac{1}{\theta}\big(\mathbb{I}_1^N\big)^2 \sum_{k=1}^{m} \frac{\Big|f^{(k)}(x^*_{(k)})\Big|(b-a)^k}{k!} \mathbb{E}\Big(\|\mathcal{A}-\mathcal{B}\|_\rho\Big), \tag{4.10}$$

where $x^*_{(k)}$ *is the maximizer to reach the maximum value for the function* $|f^{(k)}(x)|$*, i.e., the absolute value of the* k*-th derivative, in the interval* $[a,b]$*.*

Proof. If we perform Tayler expansion for the function f at x, we have

$$f(y)=f(x)+f'(x)(y-x)+f''(x)\frac{(y-x)^2}{2!}+f'''(x)\frac{(y-x)^3}{3!}+\cdots, \tag{4.11}$$

which is equivalent to have

$$\frac{f(y)-f(x)}{y-x}=f'(x)+f''(x)\frac{(y-x)}{2!}+f'''(x)\frac{(y-x)^2}{3!}+\cdots. \tag{4.12}$$

Because the polynomial function f has degree m, from Eq. (4.11), we have the following bound from triangle inequality:

$$\begin{aligned}\left|\frac{f(y)-f(x)}{y-x}\right| &= \left|\sum_{k=1}^{m} \frac{f^{(k)}(x)(y-x)^k}{k!}\right| \\ &\leq \sum_{k=1}^{m} \frac{\Big|f^{(k)}(x^*_{(k)})\Big|(b-a)^k}{k!}, \end{aligned} \tag{4.13}$$

where $x^*_{(k)}$ is the maximizer to reach the maximum value for the function $|f^{(k)}(x^*_{(k)})|$ in the interval $[a,b]$. This corollary is proved by Theorem 4.1. $\square$

Following corollary is about Lipschitz estimate for polygamma functions. Recall that a digamma function $\omega(x)$ is defined as

$$\omega(x)=\frac{\Gamma'(x)}{\Gamma(x)}, \tag{4.14}$$

where $\Gamma(x)=\int_0^\infty e^{-t}t^{x-1}dt$ [31]. Then polygamma functions are defined as the k-th derivative $\omega^{(k)}(x)$ for any $k\in\mathbb{N}$.

Corollary 4.2 (Lipschitz Estimate for Polygamma Functions)**.** *Let $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be independent random positive definite tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \qquad \text{and} \qquad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H.$$

Besides, we are given a polygamma function $\omega^{(k)}(x)$ for any $k\in\mathbb{N}$ and $x>0$. For any $\theta>0$, we have

$$\Pr\left(\left\|\omega^{(k)}(\mathcal{A})-\omega^{(k)}(\mathcal{B})\right\|_\rho\geq\theta\right)\leq\frac{\left(\mathbb{I}_1^N\right)^2\Omega_{\omega^{[k+1]}}}{\theta}\mathbb{E}\left(\|\mathcal{A}-\mathcal{B}\|_\rho\right), \tag{4.15}$$

where

$$\Omega_{\omega^{[k+1]}}=\max\left\{\omega^{(k+1)}\left(\frac{1}{e}\left(\frac{(b^*)^{b^*}}{(a^*)^{a^*}}\right)^{\frac{1}{b^*-a^*}}\right),\ \omega^{(k+1)}(x^*)\right\}. \tag{4.16}$$

The values a^* and b^* are the maximizers of the function

$$\omega^{(k+1)}\left(\frac{1}{e}\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}\right)$$

given $a\neq b$ and $a,b>0$. The value x^* is the maximizer of the function $\omega^{(k+1)}(x)$ given $x>0$.

Proof. This corollary is proved according to Theorem 4.1 by applying Theorem 1 from [31] to Eq. (4.14). $\square$

4.2 Quasi-commutator case

In this section, we will extend the perturbation formula provided by Lemma 4.2 to the quasi-commutator $\mathcal{D}\star\mathcal{A}-\mathcal{B}\star\mathcal{D}$. Tail bouns based on the quasi-commutator $\mathcal{D}\star\mathcal{A}-\mathcal{B}\star\mathcal{D}$ will be given in this section.

Lemma 4.2. *Let $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be random Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H.$$

Besides, we have the function $f:\mathbb{R}\to\mathbb{R}$ such that $f'(x)$ exists, then, we have

$$\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}=T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}), \tag{4.17}$$

where $f^{[1]}$ has been defined by Eq. (4.1).

Proof. Since we have

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H, \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H, \tag{4.18}$$

then, we will obtain the following

$$\mathcal{D}\star_N\mathcal{A}=T_{\mathcal{A},\mathcal{I},\lambda_i}\star_N\mathcal{D}, \quad \mathcal{B}\star_N\mathcal{D}=T_{\mathcal{I},\mathcal{B},\mu_j}\star_N\mathcal{D}. \tag{4.19}$$

By applying Lemma 2.1, we have

$$\begin{aligned}
&T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D})\\
=&T_{\mathcal{A},\mathcal{B},f^{[1]}}\left(T_{\mathcal{A},\mathcal{I},\lambda_i}\star_N\mathcal{D}-T_{\mathcal{I},\mathcal{B},\mu_j}\star_N\mathcal{D}\right)\\
=&T_{\mathcal{A},\mathcal{B},f^{[1]}(\lambda_i,\mu_j)}(\mathcal{D})\\
=&T_{\mathcal{A},\mathcal{B},f(\lambda_i)-f(\mu_j)}(\mathcal{D})\\
=&\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}.
\end{aligned} \tag{4.20}$$

This completes the proof. □

Following theorem is the tail bound for $\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}$.

Theorem 4.2. *Let $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be independent random Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H,$$

moreover, we are given a real valued function $f(x)$ for $x\in\mathbb{R}$ such that $f'(x)$ exists and $|f'(x)|$ is bounded by a positive number denoted by $\Omega_{f^{[1]}}$. Then, for any $\theta>0$, we have

$$\Pr\Big(\|\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}\|_\rho\geq\theta\Big)\leq\frac{\left(\mathbb{I}_1^N\right)^2\Omega_{f^{[1]}}}{\theta}\mathbb{E}\Big(\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho\Big). \tag{4.21}$$

Proof. From Lemma 4.2, we have

$$\begin{aligned}
&\|\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}\|_\rho\\
=&\left\|T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D})\right\|_\rho\\
=&\left\|\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}f^{[1]}(\lambda_i,\mu_j)\mathcal{P}_{\mathcal{U}_i}\star_N(\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D})\star_N\mathcal{P}_{\mathcal{V}_j}\right\|_\rho\\
\leq_1&\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\left|f^{[1]}(\lambda_i,\mu_j)\right|\left\|\mathcal{P}_{\mathcal{U}_i}\star_N(\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D})\star_N\mathcal{P}_{\mathcal{V}_j}\right\|_\rho\\
=_2&\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\left|f^{[1]}(\lambda_i,\mu_j)\right|\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho\\
\leq_3&\left(\mathbb{I}_1^N\right)^2\Omega_{f^{[1]}}\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho,
\end{aligned}\tag{4.22}$$

where $\leq_1$ comes from triangle inequality of the unitarily invariant norm, $=_2$ comes from the definition of the unitarily invariant norm, and $\leq_3$ is due to that $\left|f^{[1]}(\lambda_i,\mu_j)\right|\leq\Omega_{f^{[1]}}$ (mean value theorem of divide difference).

Then, we have the following bound for $\Pr(\|\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}\|_\rho\geq\theta)$

$$\begin{aligned}
&\Pr\Big(\|\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}\|_\rho\geq\theta\Big)\\
\leq_1&\Pr\Big(\left(\mathbb{I}_1^N\right)^2\Omega_{f^{[1]}}\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho\geq\theta\Big)\\
\leq_2&\frac{\left(\mathbb{I}_1^N\right)^2\Omega_{f^{[1]}}}{\theta}\mathbb{E}\Big(\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho\Big),
\end{aligned}\tag{4.23}$$

where $\leq_1$ comes from the inequality obtained by Eq. (4.22), and the $\leq_2$ is based on Markov inequality. □

Similar to Corollaries 4.1 and 4.2, we have following two corollaries for Lipschitz estimate for polynomial and polygamma under quasi-commutator case.

Corollary 4.3 (Lipschitz Estimate for Polynomial Functions, Quasi-Commutator Case)**.** *Let* $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ *be independent random Hermitian tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H.$$

Besides, we are given a real polynomial function $f(x)$ *with degree* m *over an interval* $[a,b]$*. For any* $\theta>0$*, we have*

$$\begin{aligned}&\mathrm{Pr}\Big(\|\mathcal{D}\star_N f(\mathcal{A})-f(\mathcal{B})\star_N\mathcal{D}\|_\rho\geq\theta\Big)\\ \leq&\frac{1}{\theta}\big(\mathbb{I}_1^N\big)^2\sum_{k=1}^{m}\frac{\Big|f^{(k)}(x^*_{(k)})\Big|(b-a)^k}{k!}\mathbb{E}\Big(\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho\Big),\end{aligned}\tag{4.24}$$

where $x^*_{(k)}$ *is the maximizer to reach the maximum value for the function* $|f^{(k)}(x^*_{(k)})|$ *in the interval* $[a,b]$*.*

Proof. If we perform Tayler expansion for the function f at x, we have

$$f(y)=f(x)+f'(x)(y-x)+f''(x)\frac{(y-x)^2}{2!}+f'''(x)\frac{(y-x)^3}{3!}+\cdots,\tag{4.25}$$

which is equivalent to have

$$\frac{f(y)-f(x)}{y-x}=f'(x)+f''(x)\frac{(y-x)}{2!}+f'''(x)\frac{(y-x)^2}{3!}+\cdots.\tag{4.26}$$

Because the polynomial function f has degree m, from Eq. (4.11), we have the following bound from triangle inequality:

$$\left|\frac{f(y)-f(x)}{y-x}\right|=\left|\sum_{k=1}^{m}\frac{f^{(k)}(x)(y-x)^k}{k!}\right|\leq\sum_{k=1}^{m}\frac{\Big|f^{(k)}(x^*_{(k)})\Big|(b-a)^k}{k!},\tag{4.27}$$

where $x^*_{(k)}$ is the maximizer to reach the maximum value for the function $|f^{(k)}(x^*_{(k)})|$ in the interval $[a,b]$. This corrollary is proved by Theorem 4.2. □

Corollary 4.4 (Lipschitz Estimate for Polygamma Functions: Quasi-Commutator Case)**.** *Let* $\mathcal{A},\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ *be independent random positive definite tensors with*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H \quad \text{and} \quad \mathcal{B}=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H.$$

Besides, we are given a polygamma function $\omega^{(k)}(x)$ *for any* $k\in\mathbb{N}$ *and* $x>0$. *For any* $\theta>0$, *we have*

$$\begin{aligned}&\Pr\left(\left\|\mathcal{D}\star_N\omega^{(k)}(\mathcal{A})-\omega^{(k)}(\mathcal{B})\star_N\mathcal{D}\right\|_\rho\geq\theta\right)\\ \leq&\frac{\left(\mathbb{I}_1^N\right)^2\Omega_{\omega^{[k+1]}}}{\theta}\mathbb{E}\left(\|\mathcal{D}\star_N\mathcal{A}-\mathcal{B}\star_N\mathcal{D}\|_\rho\right),\end{aligned}\tag{4.28}$$

where

$$\Omega_{\omega^{[k+1]}}=\max\left\{\omega^{(k+1)}\left(\frac{1}{e}\left(\frac{(b^*)^{b^*}}{(a^*)^{a^*}}\right)^{\frac{1}{b^*-a^*}}\right),\ \omega^{(k+1)}(x^*)\right\}.\tag{4.29}$$

The values a^* *and* b^* *are the maximizers of the function*

$$\omega^{(k+1)}\left(\frac{1}{e}\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}\right)$$

given $a\neq b$ *and* $a,b>0$. *The value* x^* *is the maximizer of the function* $\omega^{(k+1)}(x)$ *given* $x>0$.

Proof. This corollary is proved by applying Theorem 1 from [31] to Eq. (4.14) and Theorem 4.2. □

5 Continuity of random tensor integral

In this section, we will establish continuity of DTI. We need the following definition to define the convergence in mean for random tensors.

Definition 5.1. *We say that a sequence of random tensor* $\{\mathcal{X}_n\}$ *converges in the* r*-th mean towards the random tensor* $\mathcal{X}$ *with respect to the tensor norm* $\|\cdot\|_\rho$, *if we have*

$$\mathbb{E}\left(\|\mathcal{X}_n\|_\rho^r\right)\quad\textit{exists,}\quad\mathbb{E}\left(\|\mathcal{X}\|_\rho^r\right)\quad\textit{exists,}\tag{5.1}$$

and

$$\lim_{n\to\infty}\mathbb{E}\left(\|\mathcal{X}_n-\mathcal{X}\|_\rho^r\right)=0.\tag{5.2}$$

We adopt the notation $\mathcal{X}_n\xrightarrow{r}\mathcal{X}$ *to represent that random sequence of tensors* $\{\mathcal{X}_n\}$ *converges in the* r*-th mean to the random tensor* $\mathcal{X}$ *with respect to the tensor norm* $\|\cdot\|_\rho$.

Besides random tensor convergence definition, we also need to define triple tensor integral and second-order divide difference.

We define *triple tensor integrals* (TTI) with respect to Hermitian tensors $\mathcal{A},\mathcal{B},\mathcal{C}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ such that

$$\begin{aligned}\mathcal{A}&=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H\overset{\text{def}}{=}\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{P}_{\mathcal{U}_i},\\ \mathcal{B}&=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H\overset{\text{def}}{=}\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{P}_{\mathcal{V}_j},\\ \mathcal{C}&=\sum_{k=1}^{\mathbb{I}_1^N}\nu_k\mathcal{W}_j\star_1\mathcal{W}_j^H\overset{\text{def}}{=}\sum_{k=1}^{\mathbb{I}_1^N}\nu_k\mathcal{P}_{\mathcal{W}_k}.\end{aligned}$$

Given the function $\varphi:\mathbb{R}^3\to\mathbb{R}$, the TTI associated with tensors $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and the function φ, denoted by $T_{\mathcal{A},\mathcal{B},\mathcal{C},\varphi}(\mathcal{X},\mathcal{Y})$, can be expressed as

$$T_{\mathcal{A},\mathcal{B},\mathcal{C},\varphi}(\mathcal{X},\mathcal{Y})=\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\sum_{k=1}^{\mathbb{I}_1^N}\varphi(\lambda_i,\mu_j,\nu_k)\mathcal{P}_{\mathcal{U}_i}\star_N\mathcal{X}\star_N\mathcal{P}_{\mathcal{V}_j}\star_N\mathcal{Y}\star_N\mathcal{P}_{\mathcal{W}_j},\tag{5.3}$$

where $\mathcal{X},\mathcal{Y}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$.

The second-order divide difference for a function $f(x)$ (All functions f discussed in this work involving divide difference are assumed to be differentiable at any orders over the whole real domain), denoted by $f^{[2]}(x,y,z)$, can be defined as

$$f^{[2]}(x,y,z)\overset{\text{def}}{=}\frac{f^{[1]}(y,z)-f^{[1]}(x,y)}{z-x}.\tag{5.4}$$

Lemma 5.1. *Given three Hermitian tensors $\mathcal{A},\mathcal{B},\mathcal{C}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ such that*

$$\begin{aligned}\mathcal{A}&=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{U}_i\star_1\mathcal{U}_i^H\overset{def}{=}\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{P}_{\mathcal{U}_i},\\ \mathcal{B}&=\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{V}_j\star_1\mathcal{V}_j^H\overset{def}{=}\sum_{j=1}^{\mathbb{I}_1^N}\mu_j\mathcal{P}_{\mathcal{V}_j},\\ \mathcal{C}&=\sum_{k=1}^{\mathbb{I}_1^N}\nu_k\mathcal{W}_j\star_1\mathcal{W}_j^H\overset{def}{=}\sum_{k=1}^{\mathbb{I}_1^N}\nu_k\mathcal{P}_{\mathcal{W}_k},\end{aligned}$$

and the function $f(x)$ with $f^{[2]}(x)$ bounded by $\Omega_{f^{[2]}}$, we then have the following norm estimate for $T_{\mathcal{A},\mathcal{B},\mathcal{C},\varphi}(\mathcal{X},\mathcal{Y})$:

$$\|T_{\mathcal{A},\mathcal{B},\mathcal{C},\varphi}(\mathcal{X},\mathcal{Y})\|_\rho \leq \left(\mathbb{I}_1^N\right)^3 \Omega_{f^{[2]}} \|\mathcal{X}\|_\rho \cdot \|\mathcal{Y}\|_\rho. \tag{5.5}$$

Proof. Since we have

$$\begin{aligned}
\|T_{\mathcal{A},\mathcal{B},\mathcal{C},\varphi}(\mathcal{X},\mathcal{Y})\|_\rho &= \left\| \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\sum_{k=1}^{\mathbb{I}_1^N} \varphi(\lambda_i,\mu_j,\nu_k)\mathcal{P}_{\mathcal{U}_i}\star_N\mathcal{X}\star_N\mathcal{P}_{\mathcal{V}_j}\star_N\mathcal{Y}\star_N\mathcal{P}_{\mathcal{W}_j}\right\|_\rho \\
&\leq_1 \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\sum_{k=1}^{\mathbb{I}_1^N} |\varphi(\lambda_i,\mu_j,\nu_k)| \left\|\mathcal{P}_{\mathcal{U}_i}\star_N\mathcal{X}\star_N\mathcal{P}_{\mathcal{V}_j}\star_N\mathcal{Y}\star_N\mathcal{P}_{\mathcal{W}_j}\right\|_\rho \\
&\leq_2 \sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\sum_{k=1}^{\mathbb{I}_1^N} |\varphi(\lambda_i,\mu_j,\nu_k)| \|\mathcal{X}\|_\rho \cdot \|\mathcal{Y}\|_\rho \\
&\leq \left(\mathbb{I}_1^N\right)^3 \Omega_{f^{[2]}} \|\mathcal{X}\|_\rho \cdot \|\mathcal{Y}\|_\rho,
\end{aligned} \tag{5.6}$$

where $\leq_1$ comes from triangle inequality of the unitarily invariant norm and $\leq_2$ comes from the definition of the unitarily invariant norm and submultiplicative property of any unitarily invariant norm. □

Following theorem is about the continuity of a tensor integral.

Theorem 5.1. *Let $\mathcal{A}_n,\mathcal{A},\mathcal{B}_n,\mathcal{B} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ be random Hermitian tensors such that*

$$\mathcal{A}_n \xrightarrow{r} \mathcal{A} \quad \text{and} \quad \mathcal{B}_n \xrightarrow{r} \mathcal{B}, \tag{5.7}$$

where $1 \leq r < \infty$. Moreover, a real-valued function $f(x)$ for $x \in \mathbb{R}$ such that $f^{[2]}(x)$ exists and bounded by $\Omega_{f^{[2]}}$, respectively. Then, we have

$$T_{\mathcal{A}_n,\mathcal{B}_n,f^{[1]}}(\mathcal{X}) \xrightarrow{r} T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{X}), \tag{5.8}$$

where $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ is a fixed tensor.

Proof. From Lemma 4.1 and telescoping summation, we have the following:

$$\begin{aligned}
&\left\|T_{\mathcal{A}_n,\mathcal{B}_n,f^{[1]}}(\mathcal{X}) - T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{X})\right\|_\rho \\
=&\left\|T_{\mathcal{A}_n,\mathcal{B}_n,f^{[1]}}(\mathcal{X}) - T_{\mathcal{A},\mathcal{B}_n,f^{[1]}}(\mathcal{X}) + T_{\mathcal{A},\mathcal{B}_n,f^{[1]}}(\mathcal{X}) - T_{\mathcal{A},\mathcal{B},f^{[1]}}(\mathcal{X})\right\|_\rho \\
=_1&\left\|T_{\mathcal{A}_n,\mathcal{A},\mathcal{B}_n,f^{[2]}}(\mathcal{A}_n-\mathcal{A},\mathcal{X}) + T_{\mathcal{A},\mathcal{B}_n,\mathcal{B},f^{[2]}}(X,\mathcal{B}_n-\mathcal{B})\right\|_\rho \\
\leq&\, \Omega_{f^{[2]}} \|\mathcal{A}_n-\mathcal{A}\|_\rho \|\mathcal{X}\|_\rho + \Omega_{f^{[2]}} \|\mathcal{B}_n-\mathcal{B}\|_\rho \|\mathcal{X}\|_\rho,
\end{aligned} \tag{5.9}$$

where $=_1$ is obtained from the $f^{[2]}$ definition given by Eq. (5.4), and the $\leq$ comes from Lemma 5.1 and triangle inequality.

By raising the power r and taking the expectation at the both sides of the inequality provided by Eq. (5.9), we have proved this theorem by conditions given by Eq. (5.7) and the following inequality:

$$(a+b)^r \leq 2^r (a^r + b^r) \quad \text{given } a,b \geq 0. \tag{5.10}$$

Thus, we complete the proof. □

6 Applications of tensor integral

In this section, we will apply Theorem 5.1 and perturbation formulas provided by Lemma 4.1 and Lemma 4.2 to bound the tail probability of the derivative of tensor-valued function norm.

Given a fixed perturbation tensor $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ with respect to the random Hermitian tensor $\mathcal{A}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, and a tensor-valued function $f(x)$, we define the derivative of $f(x)$ at $\mathcal{A}$ with respect to the perturbation $\mathcal{X}$, represented by $f'_{\mathcal{X}}(\mathcal{A})$, as

$$f'_{\mathcal{X}}(\mathcal{A}) \stackrel{\text{def}}{=} \lim_{t\to 0} \frac{f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})}{t}. \tag{6.1}$$

Following theorem is about the tail bound for the norm of $f'_{\mathcal{X}}(\mathcal{A})$.

Theorem 6.1. *Given a fixed perturbation tensor $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ and a random Hermitian tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ with a tensor-valued function $f(x)$. Suppose we have*

$$\mathcal{A} = \sum_{i=1}^{\mathbb{I}_1^N} \lambda_i \mathcal{P}_{\mathcal{U}_i}.$$

Then, we have the tail bound for $\|f'_{\mathcal{X}}(\mathcal{A})\|_\rho$ as

$$\Pr\Big(\|f'_{\mathcal{X}}(\mathcal{A})\|_\rho \geq \theta\Big) \leq \frac{\big(\mathbb{I}_1^N\big)^2 \|\mathcal{X}\|_\rho}{\theta} \sum_{i=1}^{\mathbb{I}_1^N} \sum_{j=1}^{\mathbb{I}_1^N} \mathbb{E}\left(\left|f^{[1]}(\lambda_i,\lambda_j)\right|\right). \tag{6.2}$$

Proof. From Lemma 4.1, we have

$$f(\mathcal{A}+t\mathcal{X}) - f(\mathcal{A}) = T_{\mathcal{A}+t\mathcal{X},\mathcal{A},f^{[1]}}(\mathcal{A}+t\mathcal{X}-\mathcal{A}). \tag{6.3}$$

Then, we have

$$\frac{f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})}{t}=T_{\mathcal{A}+t\mathcal{X},\mathcal{A},f^{[1]}}(\mathcal{X}). \tag{6.4}$$

From Theorem 5.1, we have

$$f'_{\mathcal{X}}(\mathcal{A})=\lim_{t\to 0}\frac{f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})}{t}\xrightarrow{r} T_{\mathcal{A},\mathcal{A},f^{[1]}}(\mathcal{X}). \tag{6.5}$$

Since the convergence in the r-th mean implies the convergence in probability, we then have

$$\begin{aligned}\Pr\left(\|f'_{\mathcal{X}}(\mathcal{A})\|_\rho\geq\theta\right)=&\Pr\left(\left\|T_{\mathcal{A},\mathcal{A},f^{[1]}}(\mathcal{X})\right\|_\rho\geq\theta\right)\\ \leq&\frac{\left(\mathbb{I}_1^N\right)^2\|\mathcal{X}\|_\rho}{\theta}\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\mathbb{E}\left(\left|f^{[1]}(\lambda_i,\lambda_j)\right|\right),\end{aligned} \tag{6.6}$$

where the inequality comes from Theorem 3.1 □

The commutator of a tensor $\mathcal{A}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ with respect to the tensor $\mathcal{B}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ if we have $\mathcal{A}\star_N\mathcal{B}=\mathcal{B}\star_N\mathcal{A}$.

Given a fixed perturbation tensor $\mathcal{X}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ with respect to the random Hermitian tensor $\mathcal{A}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, and a tensor-valued function $f(x)$, we define the derivative of $f(x)$ at $\mathcal{A}$ with respect to the perturbation $\mathcal{X}$ and the commutator tensor $\mathcal{D}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ of $\mathcal{A}$, represented by $f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})$, as

$$f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})\stackrel{\text{def}}{=}\lim_{t\to 0}\frac{\mathcal{D}\star_N f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})\star_N\mathcal{D}}{t}. \tag{6.7}$$

Following theorem is about the tail bound for the norm of $f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})$.

Theorem 6.2. *Given a fixed perturbation tensor $\mathcal{X}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ and a random Hermitian tensor $\mathcal{A}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ with a tensor-valued function $f(x)$. Suppose we have*

$$\mathcal{A}=\sum_{i=1}^{\mathbb{I}_1^N}\lambda_i\mathcal{P}_{\mathcal{U}_i}$$

and the tensor $\mathcal{D}\in\mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ is the commutator of the tensor $\mathcal{A}$. Then, we have the tail bound for $\|f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})\|_\rho$ as

$$\Pr\left(\left\|f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})\right\|_\rho\geq\theta\right)\leq\frac{\left(\mathbb{I}_1^N\right)^2\|\mathcal{D}\star_N\mathcal{X}\|_\rho}{\theta}\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\mathbb{E}\left(\left|f^{[1]}(\lambda_i,\lambda_j)\right|\right). \tag{6.8}$$

Proof. From Lemma 4.2, we have

$$\mathcal{D}\star_N f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})\star_N\mathcal{D}=T_{\mathcal{A}+t\mathcal{X},\mathcal{A},f^{[1]}}(\mathcal{D}\star_N(\mathcal{A}+t\mathcal{X})-\mathcal{A}\star_N\mathcal{D}). \tag{6.9}$$

Also, from $\mathcal{D}\star_N\mathcal{A}=\mathcal{A}\star_N\mathcal{D}$, we also have

$$\frac{\mathcal{D}\star_N f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})\star_N\mathcal{D}}{t}=T_{\mathcal{A}+t\mathcal{X},\mathcal{A},f^{[1]}}(\mathcal{D}\star_N\mathcal{X}). \tag{6.10}$$

From Theorem 5.1, we have

$$f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})=\lim_{t\to 0}\frac{\mathcal{D}\star_N f(\mathcal{A}+t\mathcal{X})-f(\mathcal{A})\star_N\mathcal{D}}{t}\xrightarrow{r}T_{\mathcal{A},\mathcal{A},f^{[1]}}(\mathcal{D}\star_N\mathcal{X}). \tag{6.11}$$

Since the convergence in the r-th mean implies the convergence in probability, we then have

$$\begin{aligned}
&\mathrm{Pr}\left(\left\|f'_{\mathcal{X}|\mathcal{D}}(\mathcal{A})\right\|_\rho\geq\theta\right)\\
=&\mathrm{Pr}\left(\left\|T_{\mathcal{A},\mathcal{A},f^{[1]}}(\mathcal{D}\star_N\mathcal{X})\right\|_\rho\geq\theta\right)\\
\leq&\frac{\left(\mathbb{I}_1^N\right)^2\|\mathcal{D}\star_N\mathcal{X}\|_\rho}{\theta}\sum_{i=1}^{\mathbb{I}_1^N}\sum_{j=1}^{\mathbb{I}_1^N}\mathbb{E}\left(\left|f^{[1]}(\lambda_i,\lambda_j)\right|\right),
\end{aligned} \tag{6.12}$$

where the inequality comes from Theorem 3.1 again. □

7 Conclusions

We first define the definition of the random DTI and derive the tail bound of the unitarily invariant norm for a random DTI. This bound assists us to establish tail bounds of the unitarily invariant norm for various types of dual tensor means, e.g., arithmetic mean, geometric mean, harmonic mean, and general mean. The random DTI is also being applied to build the random Lipschitz estimate in contexts of random tensors. Finally, we derive the continuity property for random DTI in the sense of convergence in the random tensor mean, and apply this fact to obtain the tail bound of the unitarily invariant norm for the derivative of the tensor-valued function. Possible future works will be to extend DTI to multiple tensor integrals.

Acknowledgements

Dr. Wei is supported by the National Natural Science Foundation of China under grant No. 12271108, Shanghai Municipal Science and Technology Commission under grant No. 22WZ2501900 and Innovation Program of Shanghai Municipal Education Commission.

References

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

[2] Y. Wei and W. Ding, Theory and Computation of Tensors: Multi-Dimensional Arrays, Academic Press, 2016.

[3] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, Tensor decomposition for signal processing and machine learning, IEEE Trans. Signal Process., 65(13) (2017), pp. 3551–3582.

[4] R. Gurau, Random Tensors, Oxford University Press, Oxford, 2017.

[5] I. R. Klebanov and G. Tarnopolsky, Uncolored random tensors, melon diagrams, and the Sachdev-Ye-Kitaev models, Phys. Rev. D, 95(4) (2017), p. 046004.

[6] R. Vershynin, Concentration inequalities for random tensors, Bernoulli, 26(4) (2020), pp. 3139–3162.

[7] S. Y. Chang and W.-W. Lin, Convenient tail bounds for sums of random tensors, Taiwanese J. Math., 1(1) (2021), pp. 1–36.

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

[9] S. Y. Chang, Hanson-Wright inequality for random tensors under einstein product, arXiv preprint arXiv:2111.12169, (2021).

[10] S. Y. Chang, Generalized Hanson-Wright inequality for random tensors, arXiv preprint arXiv:2203.00659, (2022).

[11] Y. Wang and Y. Wei, Generalized eigenvalue for even order tensors via Einstein product and its applications in multilinear control systems, Comput. Appl. Math., 41(8) (2022), p. 30.

[12] M. E. Kilmer, C. D. Martin, and L. Perrone, A third-order generalization of the matrix svd as a product of third-order tensors, Tufts University, Department of Computer Science, Tech. Rep. TR-2008-4, (2008).

[13] M. Che and Y. Wei, An efficient algorithm for computing the approximate t-URV and its applications, J. Sci. Comput., 92(3) (2022), p. 27.

[14] M. Che, X. Wang, Y. Wei, and X. Zhao, Fast randomized tensor singular value thresholding for low-rank tensor optimization. Numer. Linear Algebra Appl., 29(6) (2022), p. e2444.

[15] J. Chen, Y. Wei, and Y. Xu, Tensor CUR decomposition under T-product and its perturbation, Numer. Functional Anal. Optim., 43(6) (2022), pp. 698–722.

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

[17] Y. Miao, L. Qi, and Y. Wei, T-Jordan canonical form and T-Drazin inverse based on the T-product, Commun. Appl. Math. Comput., 3(2) (2021), pp. 201–220.

[18] Y. Miao, T. Wang, and Y. Wei, Stochastic conditioning of tensor functions based on the tensor-tensor product, Pacific J. Optim., 19 (2023), pp. 205–235.

[19] X. Wang, M. Che, and Y. Wei, Tensor neural network models for tensor singular value decompositions, Comput. Optim. Appl., 75(3) (2020), pp. 753–777.

[20] P. Wei, X. Wang, and Y. Wei, Neural network models for time-varying tensor complementarity problems, Neurocomput., 523 (2023), pp. 18–32.

[21] S. Y. Chang and Y. Wei, T-square tensors–Part I: inequalities, Comput. Appl. Math., 41(1) (2022), p. 27.

[22] S. Y. Chang and Y. Wei, T-square tensors–Part II: tail bounds for sums of random t-product tensors, Comput. Appl. Math., 41(3) (2022), p. 32.

[23] S. Y. Chang and Y. Wei, Generalized T-product tensor Bernstein bounds, Ann. Appl. Math., 38(1) (2022), pp. 25–61.

[24] F. Hiai, Matrix analysis: matrix monotone functions, matrix means, and majorization, Interdisciplinary Information Sciences, 16(2) (2010), pp. 139–248.

[25] A. Skripka and A. Tomskova, Multilinear Operator Integrals Theory and Applications Introduction, Springer, vol. 2250, 2019.

[26] M. Liang and B. Zheng, Further results on Moore–Penrose inverses of tensors with application to tensor nearness problems, Comput. Math. Appl., 77(5) (2019), pp. 1282–1293.

[27] L. Sun, B. Zheng, C. Bu, and Y. Wei, Moore–Penrose inverse of tensors via Einstein product, Linear Multilinear Algebra, 64(4) (2016), pp. 686–698.

[28] D. Huang, Generalizing Lieb's concavity theorem via operator interpolation, Adv. Math., 369 (2020), p. 107208.

[29] K. Fan and A. J. Hoffman, Some metric inequalities in the space of matrices, Proceedings of the American Mathematical Society, 6(1) (1955), pp. 111–116.

[30] L. G. Ordóñez, D. P. Palomar, and J. R. Fonollosa, Ordered eigenvalues of a general class of Hermitian random matrices with application to the performance analysis of MIMO systems, IEEE Transactions on Signal Processing, 57(2) (2008), pp. 672–689.

[31] F. Qi, S. Guo, and S.-X. Chen, A new upper bound in the second Kershaw's double inequality and its generalizations, J. Comput. Appl. Math., 220(1-2) (2008), pp. 111–118.