6-1-2022

Convenient Tail Bounds for Sums of Random Tensors

Shih Yu Chang
San Jose State University, shihyu.chang@sjsu.edu

Wen Wei Lin
National Yang Ming Chiao Tung University

Recommended Citation

Shih Yu Chang and Wen Wei Lin. "Convenient Tail Bounds for Sums of Random Tensors" *Taiwanese Journal of Mathematics* (2022): 571-606. https://doi.org/10.11650/tjm/211201

Convenient Tail Bounds for Sums of Random Tensors

Shih Yu Chang* and Wen-Wei Lin

Abstract. This work prepares new probability bounds for sums of random, independent, Hermitian tensors. These probability bounds characterize large-deviation behavior of the extreme eigenvalue of the sums of random tensors. We extend Laplace 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. Tail bounds for the norm of a sum of random rectangular tensors are also derived from corollaries of random Hermitian tensors cases. The proof mechanism can also be applied to tensor-valued martingales and tensor-based Azuma, Hoeffding and McDiarmid inequalities are established.

1. Introduction

1.1. From random matrices to random tensors

A random matrix is a matrix-valued random variable—that is, a matrix in which some or all entries are random variables. Random matrices have played an important role in computational mathematics [13], physics [4], neuroscience [38], wireless communication [35], control [36], etc. Many important properties of scientific and engineering systems can be modelled mathematically as matrix problems. In order to consider high-dimensional system, it is often more convenient to consider tensors, or *multidimensional data*, instead of matrices (two-dimensional data).

Tensor have various applications in science and engineering [21]. In numerical applications, tensors can be applied to solve multilinear system of equations [39], high-dimensional data fitting/regression [10], tensor complementary problem [41], tensor eigenvalue problem [9], etc. In data processing applications, tensor theory applications include unsupervised separation of unknown mixtures of data signals [25, 40], data filtering [26], MIMO (multi-input multi-output) code-division [2, 8], radar, passive sensing, and communications [27, 32]. In other applications, tensors are also utilized to characterize data with *coupling effects*, e.g., network signal processing [12, 30, 31] and image processing [17, 20].

*Corresponding author.

Tensor decomposition methods have been reported recently to establish the latent-variable models, such as topic models in [3], and the connections between the orthogonal tensor decomposition and the method of moments for undertaking the *Latent Dirichlet Allocation* (LDA) in [33]. However, all these applications assumed that systems modelled by tensors are fixed and such assumption is not true and practical in solving tensors associated issues. In recent years, there are more works begin to target theory about random tensors, see [14, 18, 19, 22, 37], and references therein. In this work, we will focus on establishing a serious of tail bounds for sums of random tensors.

1.2. Technical results

Given a finite sequence of random Hermitian tensors $\{\mathcal{X}_i\} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, the main purpose of this work is to bound the following probability:

$$(1.1) \qquad \mathbb{P}\left(\lambda_{\max}\left(\sum_i \mathcal{X}_i\right) \geq \theta\right),$$

where $\lambda_{\max}$ represents the largest eigenvalue of a Hermitian tensor obtained from *eigenvalue decomposition*. We adopt the definition from [23] for the eigenvalue definition of a Hermitian tensor, see Theorem 3.2 and Lemma 4.1 in [23]. We also adopt the definition from [23] for the singular value definition of a tensor based on the Einstein product, see Theorem 3.2 in [23]. The problem posted in (1.1) are associated to the following problems: (1) the smallest and the largest singular value of a sum of random tensors with square or rectangular shapes; (2) extension random variable probability bounds, e.g., Chernoff and Bernstein bounds, to tensors settings; (3) tensor martingales and other adapted random sequences of tensors.

There are two main technical tools required by this work to build those tensor probability bounds. The first is *Laplace transform method*, which provides a systematic way to give tail bounds for the sum of scalar random variables. In [1], the authors apply Laplace transform method to bound (1.1) with the matrix setting, i.e., the tail probability for the maximum eigenvalue of the sum of Hermitian matrices is controlled by a matrix version of the moment-generating function. They proved the following:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_i \mathbf{X}_i\right) \geq \theta\right) \leq \inf_{t>0}\left\{e^{-t\theta} \mathbb{E}\operatorname{Tr}\exp\left(\sum_i t\mathbf{X}_i\right)\right\}.$$

However, the bound to $\inf_{t>0}\left\{e^{-t\theta}\mathbb{E}\operatorname{Tr}\exp\left(\sum_i t\mathbf{X}_i\right)\right\}$ is far from optimal according to [34] (see Sections 3.7 and 4.8 in [34]). In this work, we extend the Laplace transform method to tensors. The other important technique utilized in this work is to extend Lieb's concavity theorem [24] to tensors. Tensor Lieb's theorem is introduced in Section 3.2, and we

illustrate how to combine this result with the tensor Laplace transform method as our main technique to prove a serious of random tensors bounds. We utilize this mechanism as our main approach to derive a large family of probability inequalities that are essentially tight in a wide variety of scenarios. Most random matrix inequalities studied in [34] have same origin as those random tensor inequalities discussed at this work, however, we enhance main tools, i.e., Laplace transform method and Lieb's concavity theorem, from matrices to tensors, and obtain new random tensor inequalities with tensor orders and dimensions as factors.

This work will derive various inequalities based on (1.1). We will provide basic notations first before listing main inequalities investigated at this paper. The symbols $\lambda_{\min}$ and $\lambda_{\max}$ used to represent the minimum and maximum eigenvalues of a Hermitian tensor. The notation $\succeq$ is used to indicate the semidefinite ordering of tensors. $\|\mathcal{A}\|$ is defined as the largest singular value of the tensor $\mathcal{A}$.

The first category of concentration is associated to a sum of real numbers multiplied by independent random variables. If the independent random variables are drawn as normal (Gaussian) distribution or Rademacher distribution, the sum of real numbers multiplied by such random variables has sub-Gaussian tail behavior. In Section 4, we will provide the proof and its application about the following theorem.

Theorem 1.1 (Tensor Gaussian and normal series)**.** *Given a finite sequence $\mathcal{A}_i$ of fixed Hermitian tensors with dimensions as $\mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, and let $\{\alpha_i\}$ be a finite sequence of independent normal variables. We define*

$$\sigma^2 \stackrel{\text{def}}{=} \left\| \sum_{i=1}^n \mathcal{A}_i^2 \right\|.$$

If we define $\mathbb{I}_1^M \stackrel{\text{def}}{=} \prod_{m=1}^M I_m$ and let $\theta > 0$, we have

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}},$$

and

$$\Pr\left(\left\| \sum_{i=1}^n \alpha_i \mathcal{A}_i \right\| \geq \theta \right) \leq 2\mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}.$$

This theorem is also valid for a finite sequence of independent Rademacher random variables $\{\alpha_i\}$.

Chernoff bound provides an estimate on the probability of the concentration results related to the number of successes in a sequence of independent random trials. In the tensor situation, a similar theorem concerns a sum of positive-semidefinite random tensors subject to a uniform eigenvalue bound. The tensor Chernoff bound indicates that the largest (or smallest) eigenvalues of the tensor series have similar binomial random variable behavior.

Theorem 1.2 (Tensor Chernoff bound)**.** *Consider a sequence* $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ *of independent, random, Hermitian tensors that satisfy*

$$\mathcal{X}_i \succeq \mathcal{O} \quad and \quad \lambda_{\max}(\mathcal{X}_i) \leq T \quad almost\ surely.$$

Define the following two quantities:

$$\mu_{\max} \overset{\text{def}}{=} \lambda_{\max}\left(\sum_{i=1}^n \mathbb{E}\mathcal{X}_i\right) \quad and \quad \mu_{\min} \overset{\text{def}}{=} \lambda_{\min}\left(\sum_{i=1}^n \mathbb{E}\mathcal{X}_i\right).$$

Then we have the following two inequalities:

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq (1+\theta)\mu_{\max}\right) \leq \mathbb{I}_1^M \left(\frac{e^\theta}{(1+\theta)^{1+\theta}}\right)^{\mu_{\max}/T} \quad for\ \theta \geq 0;$$

and

$$\Pr\left(\lambda_{\min}\left(\sum_{i=1}^n \mathcal{X}_i\right) \leq (1-\theta)\mu_{\min}\right) \leq \mathbb{I}_1^M \left(\frac{e^{-\theta}}{(1-\theta)^{1-\theta}}\right)^{\mu_{\min}/T} \quad for\ \theta \in [0,1].$$

In Section 5, we will prove Chernoff inequality and discuss its applications.

Bernstein inequality is another inequality to bound the sum of independent, bounded random tensors by restricting the range of the maximum eigenvalue of each random tensors. Bernstein inequality can provide tighter bound compared to Hoeffding inequality. In Section 6, following tensor Bernstein bound theorem and its applications are provided.

Theorem 1.3 (Bounded $\lambda_{\max}$ tensor Bernstein bounds)**.** *Given a finite sequence of independent Hermitian tensors* $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ *that satisfy*

$$\mathbb{E}\mathcal{X}_i = 0 \quad and \quad \lambda_{\max}(\mathcal{X}_i) \leq T \quad almost\ surely.$$

Then we have the following inequalities:

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-\theta^2/2}{\sigma^2 + T\theta/3}\right);$$

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-3\theta^2}{8\sigma^2}\right) \quad for\ \theta \leq \sigma^2/T;$$

and

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-3\theta}{8T}\right) \quad for\ \theta \geq \sigma^2/T.$$

We define the total variance σ^2 as $\sigma^2 \overset{\text{def}}{=} \left\| \sum_i^n \mathcal{A}_i^2 \right\|$.

We also applied techniques developed at this work to tensor martingales in Section 7. In probability theory, the Azuma inequality gives a concentration result for the values of martingales that have bounded differences. Tensor Azuma inequality is given below.

Theorem 1.4 (Tensor Azuma inequality)**.** *Given a finite adapted sequence of Hermitian tensors $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ and a fixed sequence of Hermitian tensors $\{\mathcal{A}_i\}$ that satisfy*

$$\mathbb{E}_{i-1}\mathcal{X}_i = 0 \quad and \quad \mathcal{X}_i^2 \preceq \mathcal{A}_i^2 \quad almost\ surely,$$

where $i = 1, 2, 3, \ldots$. Then we have the following inequality:

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}.$$

In probability theory, Hoeffding's inequality builds an upper bound on the probability that the sum of bounded independent random variables drifts away from its expected value by more than a certain amount [16]. In this work, we generalize this result to the tensor setting by considering random tensors that satisfy semidefinite upper bounds in Section 7.2. In the tensor situation, the maximum eigenvalue for the sum of tensors also have subgaussian behavior.

Theorem 1.5 (Tensor Hoeffding inequality)**.** *Given a finite sequence of independent Hermitian tensors $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ and a fixed sequence of Hermitian tensors $\{\mathcal{A}_i\}$ that satisfy*

$$\mathbb{E}\mathcal{X}_i = 0 \quad and \quad \mathcal{X}_i^2 \preceq \mathcal{A}_i^2 \quad almost\ surely,$$

where $i = 1, 2, 3, \ldots$. Then we have the following inequality:

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}.$$

McDiarmid inequality is crucial in determining the stability of machine learning algorithms by applying a simple concept: small change in training set will make small change in hypothesis. In Section 7.2, we extend McDiarmid inequality from the scaler-valued function to the tensor-valued function as shown below.

Theorem 1.6 (Tensor McDiarmid inequality)**.** *Given a set of n independent random variables, i.e., $\{X_i : i = 1, 2, \ldots, n\}$, let F be a Hermitian tensor-valued function that maps these n random variables to a Hermitian tensor of dimension within $\mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$. Consider a sequence of Hermitian tensors $\{\mathcal{A}_i\}$ that satisfy*

$$\left(F(x_1, \ldots, x_i, \ldots, x_n) - F(x_1, \ldots, x_i', \ldots, x_n)\right)^2 \preceq \mathcal{A}_i^2,$$

where $x_i, x_i' \in X_i$ and $1 \leq i \leq n$. Define the total variance σ^2 as $\sigma^2 \stackrel{\text{def}}{=} \left\|\sum_i^n \mathcal{A}_i^2\right\|$. Then we have the following inequality:

$$\Pr\left(\lambda_{\max}\left(F(x_1, \ldots, x_n) - \mathbb{E}F(x_1, \ldots, x_n)\right) \geq \theta\right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}.$$

1.3. Paper organization

The remaining part of this paper is organized as follows. Basic probability theory required by our proofs and tensor notations are provided in Section 2. Main technical tools for tensor tail bounds are discussed in Section 3. Section 4 utilizes Gaussian and Rademacher series as cases study to explain tensor probability inequalities. Tensor Chernoff bound and its applications are discussed in Section 5. In Section 6, tensor Bernstein bound and its applications are provided. Several martingale results based on random tensor is presented in Section 7. Concluding remarks are given in Section 8.

2. Preliminaries of tensor and probability

In this section, we will provide a brief introduction of tensors and probability required for our future theory development. More detailed exposition about tensors can be found in [28, 29]. In [5], the author introduced those basic concepts about probability and moment-generating function.

2.1. Fundamental of tensor

2.1.1. Tensor notations

Throughout this work, scalars are represented by lower-case letters (e.g., d, e, f, ...), vectors by boldfaced lower-case letters (e.g., $\mathbf{d}$, $\mathbf{e}$, $\mathbf{f}$, ...), matrices by boldfaced capitalized letters (e.g., $\mathbf{D}$, $\mathbf{E}$, $\mathbf{F}$, ...), and tensors by calligraphic letters (e.g., $\mathcal{D}$, $\mathcal{E}$, $\mathcal{F}$, ...), respectively. Tensors are multiarrays of values which are higher-dimensional generalizations from vectors and matrices. Given a positive integer N, let $[N] \stackrel{\text{def}}{=} \{1, 2, \ldots, N\}$. An *order-N tensor* (or *N-th order tensor*) denoted by $\mathcal{X} \stackrel{\text{def}}{=} (a_{i_1,i_2,\ldots,i_N})$, where $1 \leq i_j = 1, 2, \ldots, I_j$ for $j \in [N]$, is a multidimensional array containing $I_1 \times I_2 \times \cdots \times I_N$ entries. Let $\mathbb{C}^{I_1 \times \cdots \times I_N}$ and $\mathbb{R}^{I_1 \times \cdots \times I_N}$ be the sets of the order-N $I_1 \times \cdots \times I_N$ tensors over the complex field $\mathbb{C}$ and the real field $\mathbb{R}$, respectively. For example, $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ is an order-N multiarray, where the first, second, ..., and N-th dimensions have I_1, I_2, ..., and I_N entries, respectively. Thus each entry of $\mathcal{X}$ can be represented by $a_{i_1,\ldots,i_N}$. For example, when $N = 3$, $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ is a third-order tensor containing entries a_{i_1,i_2,i_3}'s.

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: $\mathcal{X} \stackrel{\text{def}}{=} (a_{i_1,\ldots,i_M,j_1,\ldots,j_N}) \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N}$ and $\mathcal{Y} \stackrel{\text{def}}{=} (b_{i_1,\ldots,i_M,j_1,\ldots,j_N}) \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N}$, according to [23], 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

$$(\mathcal{X} + \mathcal{Y})_{i_1,\ldots,i_M,j_1 \times \cdots \times j_N} \stackrel{\text{def}}{=} a_{i_1,\ldots,i_M,j_1 \times \cdots \times j_N} + b_{i_1,\ldots,i_M,j_1 \times \cdots \times j_N}.$$

On the other hand, for tensors $\mathcal{X} \overset{\text{def}}{=} (a_{i_1,\ldots,i_M,j_1,\ldots,j_N}) \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N}$ and $\mathcal{Y} \overset{\text{def}}{=} (b_{j_1,\ldots,j_N,k_1,\ldots,k_L}) \in \mathbb{C}^{J_1 \times \cdots \times J_N \times K_1 \times \cdots \times K_L}$, according to [23], 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

$$(\mathcal{X} \star_N \mathcal{Y})_{i_1,\ldots,i_M,k_1 \times \cdots \times k_L} \overset{\text{def}}{=} \sum_{j_1,\ldots,j_N} a_{i_1,\ldots,i_M,j_1,\ldots,j_N} b_{j_1,\ldots,j_N,k_1,\ldots,k_L}.$$

Note that we will often abbreviate a tensor product $\mathcal{X} \star_N \mathcal{Y}$ to "$\mathcal{X}\mathcal{Y}$" for notational simplicity in the rest of the paper. This tensor product will be reduced to the standard matrix multiplication as $L = M = N = 1$. Other simplified situations can also be extended as tensor-vector product ($M > 1$, $N = 1$, and $L = 0$) and tensor-matrix product ($M > 1$ and $N = L = 1$). In analogy to matrix analysis, we define some basic tensors and elementary tensor operations as follows.

Definition 2.1. A tensor whose entries are all zero is called a *zero tensor*, denoted by $\mathcal{O}$.

Definition 2.2. An *identity tensor* $\mathcal{I} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N}$ is defined by

$$(\mathcal{I})_{i_1 \times \cdots \times i_N \times j_1 \times \cdots \times j_N} \overset{\text{def}}{=} \prod_{k=1}^{N} \delta_{i_k,j_k},$$

where $\delta_{i_k,j_k} \overset{\text{def}}{=} 1$ if $i_k = j_k$; otherwise $\delta_{i_k,j_k} \overset{\text{def}}{=} 0$.

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

Definition 2.3. Given a tensor $\mathcal{X} \overset{\text{def}}{=} (a_{i_1,\ldots,i_M,j_1,\ldots,j_N}) \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N}$, its conjugate transpose, denoted by $\mathcal{X}^H$, is defined by

$$(\mathcal{X}^H)_{j_1,\ldots,j_N,i_1,\ldots,i_M} \overset{\text{def}}{=} \overline{a_{i_1,\ldots,i_M,j_1,\ldots,j_N}},$$

where the overline notion indicates the complex conjugate of the number $a_{i_1,\ldots,i_M,j_1,\ldots,j_N}$. If a tensor $\mathcal{X}$ satisfies $\mathcal{X}^H = \mathcal{X}$, then $\mathcal{X}$ is a *Hermitian tensor*.

Definition 2.4. Given a tensor $\mathcal{U} \overset{\text{def}}{=} (u_{i_1,\ldots,i_M,i_1,\ldots,i_M}) \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, if

$$\mathcal{U}^H \star_M \mathcal{U} = \mathcal{U} \star_M \mathcal{U}^H = \mathcal{I} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M},$$

then $\mathcal{U}$ is a *unitary tensor*.

In this work, the symbol $\mathcal{U}$ is reserved for a unitary tensor.

Definition 2.5. Given a *square tensor* $\mathcal{X} \overset{\text{def}}{=} (a_{i_1,\ldots,i_M,j_1,\ldots,j_M}) \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, 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

$$\mathcal{X} \star_M \mathcal{X} = \mathcal{X} \star_M \mathcal{X} = \mathcal{I},$$

then $\mathcal{X}$ is the *inverse* of $\mathcal{X}$. We usually write $\mathcal{X} \overset{\text{def}}{=} \mathcal{X}^{-1}$ thereby.

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

$$\mathrm{Tr}(\mathcal{X}) \stackrel{\text{def}}{=} \sum_{1 \leq i_j \leq I_j,\ j \in [M]} \mathcal{X}_{i_1,\ldots,i_M,i_1,\ldots,i_M}.$$

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

$$(2.1) \qquad\qquad\qquad \langle \mathcal{X}, \mathcal{Y} \rangle \stackrel{\text{def}}{=} \mathrm{Tr}\left(\mathcal{X}^H \star_M \mathcal{Y}\right).$$

According to (2.1), the *Frobenius norm* of a tensor $\mathcal{X}$ is defined by

$$\|\mathcal{X}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}.$$

We use $\lambda_{\min}$ and $\lambda_{\max}$ to represent the minimum and the maximum eigenvalues of a Hermitian tensor. The notation $\succeq$ is used to indicate the semidefinite ordering of tensors. If we have $\mathcal{X} \succeq \mathcal{Y}$, this means that the difference tensor $\mathcal{X} - \mathcal{Y}$ is a positive semidefinite tensor.

2.1.2. Tensor functions

Given a function $g \colon \mathbb{R} \to \mathbb{R}$, the mapping result of a diagonal tensor by the function g is to obtain another same size diagonal tensor with diagonal entry mapped by the function g. Then the function g can be extended to allow a Hermitian tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ as an input argument as

$$g(\mathcal{X}) \stackrel{\text{def}}{=} \mathcal{U} \star_M g(\Lambda) \star_M \mathcal{U}^H, \quad \text{where } \mathcal{X} = \mathcal{U} \star_M \Lambda \star_M \mathcal{U}^H.$$

The *spectral mapping theorem* asserts that each eigenvalue of $g(\mathcal{X})$ is equal to $g(\lambda)$ for some eigenvalue λ of $\mathcal{X}$. From the semidefinite ordering of tensors, we also have

$$(2.2) \qquad f(x) \geq g(x) \ \text{ for } x \in [a,b] \quad \Longrightarrow \quad f(\mathcal{X}) \succeq g(\mathcal{X}) \ \text{ for eigenvalues of } \mathcal{X} \in [a,b];$$

where $[a,b]$ is a real interval.

Definition 2.6. Given a square tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, the *tensor exponential* of the tensor $\mathcal{X}$ is defined as

$$e^{\mathcal{X}} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{\mathcal{X}^k}{k!},$$

where $\mathcal{X}^0$ is defined as the identity tensor $\mathcal{I} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ and

$$\mathcal{X}^k = \underbrace{\mathcal{X} \star_M \mathcal{X} \star_M \cdots \star_M \mathcal{X}}_{k \text{ terms of } \mathcal{X}}.$$

Given a tensor $\mathcal{Y}$, the tensor $\mathcal{X}$ is said to be a *tensor logarithm* of $\mathcal{Y}$ if $e^{\mathcal{X}} = \mathcal{Y}$.

Several facts are about tensor exponential. First, the exponential of a Hermitian tensor is always positive-definite due to the spectral mapping theorem. Second, the trace exponential function, $\mathcal{X} \to \mathrm{Tr}\exp(\mathcal{X})$, is convex. Third, the trace exponential function follows monotone property with respect to semidefinite ordering as

$$\text{(2.3)} \qquad \mathcal{X} \succeq \mathcal{Y} \quad \implies \quad \mathrm{Tr}\exp(\mathcal{X}) \geq \mathrm{Tr}\exp(\mathcal{Y}).$$

However, different from exponential rules for scalers, the tensor exponential does not convert sums into products. The Golden–Thompson inequality for tensors [6] shows the following relationship

$$\mathrm{Tr}\, e^{\mathcal{X}+\mathcal{Y}} \leq \mathrm{Tr}\left(e^{\mathcal{X} \star_M \mathcal{Y}}\right),$$

where $\mathcal{X}, \mathcal{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ are Hermitian tensors.

For the tensor logarithm, we have the following monotone relation

$$\mathcal{X} \succeq \mathcal{Y} \quad \implies \quad \log(\mathcal{X}) \succeq \log(\mathcal{Y}).$$

Moreover, the tensor logarithm is also concave, i.e., we have

$$\text{(2.4)} \qquad t\log(\mathcal{X}_1) + (1-t)\log(\mathcal{X}_2) \preceq \log(t\mathcal{X}_1 + (1-t)\mathcal{X}_2),$$

where $\mathcal{X}_1$, $\mathcal{X}_2$ are positive-definite tensors and $t \in [0,1]$. The concavity of tensor logarithm can be derived from Hansen–Pedersen Characterizations, see [7].

If a given tensor $\mathcal{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$ is not square, we can perform *Hermitian dilation*, represented as $\mathbb{D}$, to the tensor $\mathcal{Y}$ as

$$\text{(2.5)} \qquad \mathbb{D}(\mathcal{Y}) \overset{\text{def}}{=} \begin{bmatrix} \mathcal{O} & \mathcal{Y} \\ \mathcal{Y}^H & \mathcal{O} \end{bmatrix},$$

where $\mathbb{D}(\mathcal{Y}) \in \mathbb{C}^{(I_1+J_1) \times \cdots \times (I_M+J_M) \times (I_1+J_1) \times \cdots \times (I_M+J_M)}$ is a Hermitian tensor. Since we have

$$\mathbb{D}(\mathcal{Y})^2 = \begin{bmatrix} \mathcal{Y}\mathcal{Y}^H & \mathcal{O} \\ \mathcal{O} & \mathcal{Y}^H\mathcal{Y} \end{bmatrix},$$

we have the following spectral norm relation

$$\text{(2.6)} \qquad \lambda_{\max}(\mathbb{D}(\mathcal{Y})) = \|\mathbb{D}(\mathcal{Y})\| = \|\mathcal{Y}\|,$$

where $\| \cdot \|$ is the spectral norm for a tensor and this will return the maximum singular value of its argument tensor. Hermitian dilation operation enables us to extend results from Hermitian tensors to other non-square tensors.

2.2. Tensor moments and cumulants

Since the expectation of a random tensor can be treated as convex combination, expectation will preserve the semidefinite order as

$$\mathcal{X} \succ \mathcal{Y} \text{ almost surely } \implies \mathbb{E}(\mathcal{X}) \succ \mathbb{E}(\mathcal{Y}).$$

From operator Jensen's inequality [15], we also have

$$\mathbb{E}(\mathcal{X}^2) \succeq (\mathbb{E}(\mathcal{X}))^2.$$

Suppose a random Hermitian tensor $\mathcal{X}$ having tensor moments of all orders, i.e., $\mathbb{E}(\mathcal{X}^n)$ existing for all n, we can define the tensor moment-generating function, denoted as $\mathbb{M}_{\mathcal{X}}(t)$, and the tensor cumulant-generating function, denoted as $\mathbb{K}_{\mathcal{X}}(t)$, for the tensor $\mathcal{X}$ as

$$\mathbb{M}_{\mathcal{X}}(t) \overset{\text{def}}{=} \mathbb{E}e^{t\mathcal{X}} \quad \text{and} \quad \mathbb{K}_{\mathcal{X}}(t) \overset{\text{def}}{=} \log \mathbb{E}e^{t\mathcal{X}},$$

where $t \in \mathbb{R}$. Both the tensor moment-generating function and the tensor cumulant-generating function can be expressed as power series expansions:

$$\mathbb{M}_{\mathcal{X}}(t) = \mathcal{I} + \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbb{E}(\mathcal{X}^n) \quad \text{and} \quad \mathbb{K}_{\mathcal{X}}(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} \psi_n,$$

where ψ_n is called *tensor cumulant*. The tensor cumulant ψ_n can be expressed as a polynomial in terms of tensor moments up to the order n, for example, the first cumulant is the mean and the second cumulant is the variance:

$$\psi_1 = \mathbb{E}(\mathcal{X}) \quad \text{and} \quad \psi_2 = \mathbb{E}(\mathcal{X}^2) - (\mathbb{E}(\mathcal{X}))^2.$$

Finally, we also assume that all random variables are sufficiently regular for us to compute their expectations, interchange limits, etc.

3. Trace concavity method

The main purpose of this section is to develop two important tools which will be applied intensively in the proof of probability inequalities for the maximum eigenvalue of a sum of independent random tensors. The first one is the Laplace transform method for tensors discussed in Section 3.1, and the second one is the tensor trace concavity which will be presented in Section 3.2.

3.1. Laplace transform method for tensors

We extend the Laplace transform bound from matrices to tensors based on [1]. The next lemma is given to establish the Laplace transform bound for tensors.

Lemma 3.1 (Laplace transform method for tensors). *Let $\mathcal{X}$ be a random Hermitian tensor. For $\theta \in \mathbb{R}$, we have*

$$\mathbb{P}(\lambda_{\max}(\mathcal{X}) \geq \theta) \leq \inf_{t>0} \left\{ e^{-\theta t} \mathbb{E} \operatorname{Tr} e^{t\mathcal{X}} \right\}.$$

Proof. Given a fix value t, we have

$$(3.1) \qquad \mathbb{P}(\lambda_{\max}(\mathcal{X}) \geq \theta) = \mathbb{P}(\lambda_{\max}(t\mathcal{X}) \geq t\theta) = \mathbb{P}(e^{\lambda_{\max}(t\mathcal{X})} \geq e^{t\theta}) \leq e^{-t\theta} \mathbb{E} e^{\lambda_{\max}(t\mathcal{X})}.$$

The first equality uses the homogeneity of the maximum eigenvalue map, the second equality comes from the monotonicity of the scalar exponential function, and the last relation is Markov's inequality. Because we have

$$(3.2) \qquad\qquad e^{\lambda_{\max}(t\mathcal{Y})} = \lambda_{\max}(e^{t\mathcal{Y}}) \leq \operatorname{Tr} e^{t\mathcal{Y}},$$

where the first equality used the spectral mapping theorem, and the inequality holds because the exponential of an Hermitian tensor is positive-definite and the maximum eigenvalue of a positive-definite tensor is dominated by the trace [23]. From (3.1) and (3.2), this lemma is established. $\qquad\square$

Lemma 3.1 helps us to control the tail probabilities for the maximum eigenvalue of a random Hermitian tensor by utilizing a bound for the trace of the tensor moment-generating function introduced in Section 2.2.

3.2. Tensor trace concavity

In this section, we will extended Lieb's concavity theorem to tensors and we begin with the definition about the relative entropy between two tensors.

Definition 3.2. Given two positive-definite tensors $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ and tensor $\mathcal{B} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$. The relative entropy between tensors $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$D(\mathcal{A} \parallel \mathcal{B}) \stackrel{\text{def}}{=} \operatorname{Tr} \mathcal{A} \star_M (\log \mathcal{A} - \log \mathcal{B}).$$

Given a continuous function defined over a real interval as $f \colon [a, b] \to \mathbb{R}$, and $\mathcal{T} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ as a Hermitian tensor with spectrum in $[a, b]$ and spectrum decomposition as $\mathcal{T} = \sum_{\lambda_n} \lambda_n \mathcal{U}_{\lambda_n}$, where $\mathcal{U}_{\lambda_n} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ are mutually orthogonal tensors, then the mapping for the tensor $\mathcal{T}$ by f can be defined as $f(\mathcal{T}) = \sum_{\lambda_n} f(\lambda_n) \mathcal{U}_{\lambda_n}$. The function f is called as a *tensor convex* function if $f(\lambda_n)$ is convex on the Hermitian tenor with spectrum in $[a, b]$. We apply *perspective function* [11] notion for tensor convex and introduce the following lemma about the convexity of a tensor convex function.

582 Shih Yu Chang and Wen-Wei Lin

Lemma 3.3. *Given f as a tensor convex function, two commuting tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, i.e., $\mathcal{X} \star_M \mathcal{Y} = \mathcal{Y} \star_M \mathcal{X}$, and the existence of the $\mathcal{Y}^{-1}$, then the map*

$$(3.3) \qquad h(\mathcal{X}, \mathcal{Y}) = f(\mathcal{X} \star_M \mathcal{Y}^{-1}) \star_M \mathcal{Y}$$

is jointly convex in the sense that, given $t \in [0,1]$, if $\mathcal{X} = t\mathcal{X}_1 + (1-t)\mathcal{X}_2$ and $\mathcal{Y} = t\mathcal{Y}_1 + (1-t)\mathcal{Y}_2$ with $\mathcal{X}_1 \star_M \mathcal{Y}_1 = \mathcal{Y}_1 \star_M \mathcal{X}_1$ and $\mathcal{X}_2 \star_M \mathcal{Y}_2 = \mathcal{Y}_2 \star_M \mathcal{X}_2$, we should have

$$h(\mathcal{X}, \mathcal{Y}) \le th(\mathcal{X}_1, \mathcal{Y}_1) + (1-t)h(\mathcal{X}_2, \mathcal{Y}_2).$$

Proof. Constructing tensors $\mathcal{A} = (t\mathcal{Y}_1)^{1/2} \star_M \mathcal{Y}^{-1/2}$ and $\mathcal{B} = ((1-t)\mathcal{Y}_2)^{1/2} \star_M \mathcal{Y}^{-1/2}$, then we have

$$(3.4) \qquad \mathcal{A}^H \star_M \mathcal{A} + \mathcal{B}^H \star_M \mathcal{B} = \mathcal{I}.$$

Since we have

$$\begin{aligned}
h(\mathcal{X}, \mathcal{Y}) &= f(\mathcal{X} \star_M \mathcal{Y}^{-1}) \star_M \mathcal{Y} \\
&= \mathcal{Y}^{1/2} \star_M f(\mathcal{Y}^{-1/2} \star_M \mathcal{X} \star_M \mathcal{Y}^{-1/2}) \star_M \mathcal{Y}^{1/2} \\
&= \mathcal{Y}^{1/2} \star_M f(\mathcal{A}^H \star_M \mathcal{X}_1 \star_M \mathcal{Y}_1^{-1} \star_M \mathcal{A} + \mathcal{B}^H \star_M \mathcal{X}_2 \star_M \mathcal{Y}_2^{-1} \star_M \mathcal{B}) \star_M \mathcal{Y}^{1/2} \\
&\le_1 \mathcal{Y}^{1/2} \star_M \left(\mathcal{A}^H \star_M f(\mathcal{X}_1 \star_M \mathcal{Y}_1^{-1}) \star_M \mathcal{A} + \mathcal{B}^H \star_M f(\mathcal{X}_2 \star_M \mathcal{Y}_2^{-1}) \star_M \mathcal{B} \right) \star_M \mathcal{Y}^{1/2} \\
&= (t\mathcal{Y}_1)^{1/2} f(\mathcal{X}_1 \star_M \mathcal{Y}_1^{-1})(t\mathcal{Y}_1)^{1/2} + ((1-t)\mathcal{Y}_2)^{1/2} f(\mathcal{X}_2 \star_M \mathcal{Y}_2^{-1})((1-t)\mathcal{Y}_2)^{1/2} \\
&= th(\mathcal{X}_1, \mathcal{Y}_1) + (1-t)h(\mathcal{X}_2, \mathcal{Y}_2),
\end{aligned}$$

where $\le_1$ is based on Hansen–Pedersen–Jensen inequality and the condition provided by (3.4), see Theorem 2.1 in [11]. □

The next lemma is given to establish the joint convexity property of relative entropy for tensors.

Lemma 3.4 (Joint convexity of relative entropy for tensors)**.** *The relative entropy function of two positive-definite tensors is a jointly convex function. That is,*

$$\mathbb{D}(t\mathcal{A}_1 + (1-t)\mathcal{A}_2 \parallel t\mathcal{B}_1 + (1-t)\mathcal{B}_2) \le t\mathbb{D}(\mathcal{A}_1 \parallel \mathcal{B}_1) + (1-t)\mathbb{D}(\mathcal{A}_2 \parallel \mathcal{B}_2),$$

where $t \in [0,1]$ and all the following four tensors $\mathcal{A}_1, \mathcal{B}_1, \mathcal{A}_2$ and $\mathcal{B}_2$, are positive-definite.

Proof. From Definition 3.2, we wish to show the joint convexity of the function $\mathbb{D}(\mathcal{A} \parallel \mathcal{B})$ with respect to the tensors $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$. Let us define tensor operators $\mathcal{F}(\mathcal{X}) \stackrel{\text{def}}{=} \mathcal{A} \star_M \mathcal{X}$ and $\mathcal{G}(\mathcal{X}) \stackrel{\text{def}}{=} \mathcal{X} \star_M \mathcal{B}$ for the variable tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$. Then we have $\mathcal{F}(\mathcal{X})$ and $\mathcal{G}(\mathcal{X})$ commuting on the inner product operation $\langle \mathcal{F}(\mathcal{X}), \mathcal{G}(\mathcal{X}) \rangle$

defined by (2.1), i.e., $\mathrm{Tr}(\mathcal{F}^H(\mathcal{X}) \star_M \mathcal{G}(\mathcal{X})) = \mathrm{Tr}(\mathcal{G}^H(\mathcal{X}) \star_M \mathcal{F}(\mathcal{X}))$. Since the function $f(x) = x \log x$ is tensor convex, we apply Lemma 3.3 to operators $\mathcal{F}(\cdot)$, $\mathcal{G}(\cdot)$ and the function h definition provided by (3.3) to obtain the following relation:

$$
\begin{aligned}
\langle \mathcal{I}, h(\mathcal{F}(\mathcal{I}), \mathcal{G}(\mathcal{I})) \rangle &= \langle \mathcal{I}, \mathcal{G}(\mathcal{I}) \star_M (\mathcal{F}(\mathcal{I}) \star_M \mathcal{G}^{-1}(\mathcal{I})) \log(\mathcal{F}(\mathcal{I}) \star_M \mathcal{G}^{-1}(\mathcal{I})) \rangle \\
&= \langle \mathcal{I}, \mathcal{F}(\mathcal{I})(\log \mathcal{F}(\mathcal{I}) - \log \mathcal{G}(\mathcal{I})) \rangle \\
&= \mathrm{Tr}(\mathcal{A} \log \mathcal{A} - \mathcal{A} \log \mathcal{B}) = \mathbb{D}(\mathcal{A} \parallel \mathcal{B})
\end{aligned}
$$

is jointly convex with respect to tensors $\mathcal{A}$ and $\mathcal{B}$. $\hfill\square$

Theorem 3.5 (Lieb's concavity theorem for tensors)**.** *Let $\mathcal{H}$ be a Hermitian tensor. The map*

$$
\mathcal{A} \to \mathrm{Tr}\, e^{\mathcal{H} + \log \mathcal{A}}
$$

is concave on the positive-definite cone.

Proof. From Klein's inequality for the map $t \to t \log t$ (which is strictly concave for $t > 0$) and Hermitian tensors $\mathcal{X}$, $\mathcal{Y}$, we have

$$
\mathrm{Tr}\,\mathcal{Y} \geq \mathrm{Tr}\,\mathcal{X} - \mathrm{Tr}\,\mathcal{X} \log \mathcal{X} + \mathrm{Tr}\,\mathcal{X} \log \mathcal{Y}.
$$

If we replace $\mathcal{Y}$ by $e^{\mathcal{H} + \log \mathcal{A}}$, we then have

$$
(3.5) \qquad \mathrm{Tr}\, e^{\mathcal{H} + \log \mathcal{A}} = \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\,\mathcal{X} \star \mathcal{H} - \mathbb{D}(\mathcal{X} \parallel \mathcal{A}) + \mathrm{Tr}\,\mathcal{X} \right\},
$$

where $\mathbb{D}(\mathcal{X} \parallel \mathcal{A})$ is the quantum relative entropy between two tensor operators. For real number $t \in [0, 1]$ and two positive-definite tensors $\mathcal{A}_1$, $\mathcal{A}_2$, we have

$$
\begin{aligned}
\mathrm{Tr}\, e^{\mathcal{H} + \log(t\mathcal{A}_1 + (1-t)\mathcal{A}_2)} &= \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\,\mathcal{X}\mathcal{H} - \mathbb{D}(\mathcal{X} \parallel t\mathcal{A}_1 + (1-t)\mathcal{A}_2) + \mathrm{Tr}\,\mathcal{X} \right\} \\
&\geq t \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\,\mathcal{X}\mathcal{H} - \mathbb{D}(\mathcal{X} \parallel t\mathcal{A}_1) + \mathrm{Tr}\,\mathcal{X} \right\} \\
&\qquad + (1-t) \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\,\mathcal{X}\mathcal{H} - \mathbb{D}(\mathcal{X} \parallel (1-t)\mathcal{A}_2) + \mathrm{Tr}\,\mathcal{X} \right\} \\
&= t \,\mathrm{Tr}\, e^{\mathcal{H} + \log \mathcal{A}_1} + (1-t)\,\mathrm{Tr}\, e^{\mathcal{H} + \log \mathcal{A}_2},
\end{aligned}
$$

where the first and last equalities are obtained based on the variational formula provided by (3.5), and the inequality is due to the joint convexity property of the relative entropy from Lemma 3.4. $\hfill\square$

Based on Lieb's concavity theorem for tensors, we have the following corollary.

Corollary 3.6. *Let $\mathcal{A}$ be a fixed Hermitian tensor, and let $\mathcal{X}$ be a random Hermitian tensor, then we have*

$$
\mathbb{E}\,\mathrm{Tr}\, e^{\mathcal{A} + \mathcal{X}} \leq \mathrm{Tr}\, e^{\mathcal{A} + \log\left(\mathbb{E}e^{\mathcal{X}}\right)}.
$$

Proof. Define the random tensor $\mathcal{Y} = e^{\mathcal{X}}$, we have

$$\mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\mathcal{X}} = \mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\log \mathcal{Y}} \leq \operatorname{Tr} e^{\mathcal{A}+\log(\mathbb{E}\mathcal{Y})} = \operatorname{Tr} e^{\mathcal{A}+\log\left(\mathbb{E}e^{\mathcal{X}}\right)},$$

where the inequality is based on Lieb's concavity theorem for tensors (see Theorem 3.5) and Jensen's inequality. $\qquad\square$

3.3. Tail bounds for independent sums

This section will present the tail bound for the sum of independent random tensors and several corollaries according to this tail bound for independent sums. We begin with subadditivity lemma of tensor cumulant-generating functions.

Lemma 3.7. *Given a finite sequence of independent Hermitian random tensors* $\{\mathcal{X}_i\}$, *we have*

$$\mathbb{E}\operatorname{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) \leq \operatorname{Tr}\exp\left(\sum_{i}^{n} \log \mathbb{E}e^{t\mathcal{X}_i}\right) \quad \text{for } t \in \mathbb{R}.$$

Proof. We first define the following term for the tensor cumulant-generating function for $\mathcal{X}_i$ as

$$\mathbb{K}_i(t) \stackrel{\text{def}}{=} \log(\mathbb{E}e^{t\mathcal{X}_i}).$$

Then we define the Hermitian tensor $\mathcal{H}_k$ as

$$(3.6) \qquad\qquad \mathcal{H}_k(t) = \sum_{i=1}^{k-1} t\mathcal{X}_k + \sum_{i=k+1}^{n} \mathbb{K}_i(t).$$

By applying (3.6) to Theorem 3.5 repeatedly for $k = 1, 2, \ldots, n$, we have

$$
\begin{aligned}
\mathbb{E}\operatorname{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) &=_1 \mathbb{E}_0 \cdots \mathbb{E}_{n-1} \operatorname{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + t\mathcal{X}_n\right) \\
&\leq \mathbb{E}_0 \cdots \mathbb{E}_{n-2} \operatorname{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + \log\left(\mathbb{E}_{n-1}e^{t\mathcal{X}_n}\right)\right) \\
&= \mathbb{E}_0 \cdots \mathbb{E}_{n-2} \operatorname{Tr}\exp\left(\sum_{i=1}^{n-2} t\mathcal{X}_i + t\mathcal{X}_{n-1} + \mathbb{K}_n(t)\right) \\
&\leq \mathbb{E}_0 \cdots \mathbb{E}_{n-3} \operatorname{Tr}\exp\left(\sum_{i=1}^{n-2} t\mathcal{X}_i + \mathbb{K}_{n-1}(t) + \mathbb{K}_n(t)\right) \\
&\leq \cdots \leq \operatorname{Tr}\exp\left(\sum_{i=1}^{n} \mathbb{K}_i(t)\right),
\end{aligned}
$$

where the equality $=_1$ is based on the law of total expectation by defining $\mathbb{E}_i$ as the conditional expectation given $\mathcal{X}_1, \ldots, \mathcal{X}_i$. $\qquad\square$

We are ready to present the theorem for the tail bound of independent sums.

Theorem 3.8 (Master tail bound for independent sum of random tensors)**.** *Given a finite sequence of independent Hermitian random tensors* $\{\mathcal{X}_i\}$*, we have*

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) \leq \inf_{t>0}\left\{e^{-t\theta}\operatorname{Tr}\exp\left(\sum_{i=1}^{n}\log\mathbb{E}e^{t\mathcal{X}_i}\right)\right\}.$$

Proof. Substituting Lemma 3.7 into the Laplace transform bound provided by Lemma 3.1, this theorem is established. $\qquad\square$

Several useful corollaries will be provided based on Theorem 3.8.

Corollary 3.9. *Given a finite sequence of independent Hermitian random tensors* $\{\mathcal{X}_i\}$ *with dimensions in* $\mathbb{C}^{I_1\times\cdots\times I_M\times I_1\times\cdots\times I_M}$*. If there is a function* $f\colon(0,\infty)\to[0,\infty]$ *and a sequence of non-random Hermitian tensors* $\{\mathcal{A}_i\}$ *with the following condition:*

$$(3.7)\qquad\qquad f(t)\mathcal{A}_i \succeq \log\mathbb{E}e^{t\mathcal{X}_i}\quad\text{for } t>0.$$

Then, for all $\theta\in\mathbb{R}$*, we have*

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right)\geq\theta\right)\leq\mathbb{I}_1^M\inf_{t>0}\left\{\exp\left[-t\theta+f(\theta)\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{A}_i\right)\right]\right\}.$$

Proof. From the condition provided by (3.7) and Theorem 3.8, we have

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right)\geq\theta\right)\leq e^{-t\theta}\operatorname{Tr}\exp\left(f(\theta)\sum_{i=1}^{n}\mathcal{A}_i\right)$$

$$\leq(I_1\cdots I_M)e^{-t\theta}\lambda_{\max}\left(\exp\left(f(\theta)\sum_{i=1}^{n}\mathcal{A}_i\right)\right)$$

$$=\mathbb{I}_1^M e^{-t\theta}\exp\left(f(\theta)\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{A}_i\right)\right),$$

where the second inequality holds since we bound the trace of a positive-definite tensor by the dimension size $I_1\cdots I_M$ (the multiplication of M positive integers) times the maximum eigenvalue; the last equality is based on the spectral mapping theorem since the function f is nonnegative. This theorem is proved by taking the infimum over positive t. $\qquad\square$

Corollary 3.10. *Given a finite sequence of independent Hermitian random tensors* $\{\mathcal{X}_i\}$ *with dimensions in* $\mathbb{C}^{I_1\times\cdots\times I_M\times I_1\times\cdots\times I_M}$*. For all* $\theta\in\mathbb{R}$*, we have*

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right)\geq\theta\right)\leq\mathbb{I}_1^M\inf_{t>0}\left\{\exp\left[-t\theta+n\log\lambda_{\max}\left(\frac{\sum_{i=1}^{n}\mathbb{E}e^{t\mathcal{X}_i}}{n}\right)\right]\right\}.$$

Proof. From tensor logarithm given by (2.4), we have

$$\sum_{i=1}^{n} \log \mathbb{E} e^{t\mathcal{X}_i} = n \cdot \frac{1}{n} \sum_{i=1}^{n} \log \mathbb{E} e^{t\mathcal{X}_i} \preceq n \log \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} e^{t\mathcal{X}_i} \right),$$

and from the trace exponential monotone property provided by (2.3), we have

$$\Pr \left(\lambda_{\max} \left(\sum_{i=1}^{n} \mathcal{X}_i \right) \geq \theta \right) \leq e^{-t\theta} \operatorname{Tr} \exp \left(n \log \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} e^{t\mathcal{X}_i} \right) \right)$$

$$\leq (I_1 \cdots I_M) \inf_{t>0} \left\{ \exp \left[-t\theta + n \log \lambda_{\max} \left(\frac{\sum_{i=1}^{n} \mathbb{E} e^{t\mathcal{X}_i}}{n} \right) \right] \right\},$$

where the last inequality holds since we bound the trace of a positive-definite tensor by the dimension size $I_1 \cdots I_M$ (the multiplication of M positive integers) times the maximum eigenvalue and apply spectral mapping theorem. $\square$

4. Tensor with random series

A tensor Gaussian series is one of the simplest cases of a sum of independent random tensors. For scalers, a Gaussian series with real coefficients satisfies a normal-type tail bound where the variance is controlled by the sum of squares coefficients. The main purpose of Section 4.1 is to extend this scenario to tensors. In Section 4.2, we will apply results from Section 4.1 to consider Gaussian tensor with nonuniform variances. Finally, we will provide the lower and upper bounds of tensor expectation in Section 4.3.

4.1. Tensor with Gaussian and Rademacher random series

We begin with a lemma about moment-generating functions of Rademacher and Gaussian normal random variables.

Lemma 4.1. *Suppose that the tensor $\mathcal{A}$ is Hermitian. Given a Gaussian normal random variable α and a Rademacher random variable β, then we have*

$$\mathbb{E} e^{\alpha t \mathcal{A}} = e^{t^2 \mathcal{A}^2 / 2} \quad \text{and} \quad e^{t^2 \mathcal{A}^2 / 2} \succeq \mathbb{E} e^{\beta t \mathcal{A}},$$

where $t \in \mathbb{R}$.

Proof. For the Gaussian normal random variable, because we have

$$\mathbb{E}(\alpha^{2n}) = \frac{(2i)!}{i! 2^i} \quad \text{and} \quad \mathbb{E}(\alpha^{2i+1}) = 0,$$

where $i = 0, 1, 2, \ldots$; then

$$\mathbb{E} e^{\alpha t \mathcal{A}} = \mathcal{I} + \sum_{i=1}^{\infty} \frac{\mathbb{E}(\alpha^{2i})(t\mathcal{A})^{2i}}{(2i)!} = \mathcal{I} + \sum_{i=1}^{\infty} \frac{(t^2 \mathcal{A}^2 / 2)^i}{i!} = e^{t^2 \mathcal{A}^2 / 2}.$$

For the Rademacher random variable, we have

$$\mathbb{E}e^{\beta t \mathcal{A}} = \cosh(t\mathcal{A}) \preceq e^{t^2 \mathcal{A}^2/2}.$$

Therefore, this lemma is proved. $\qquad\square$

We are ready to present the main Theorem of this section about Hermitian tensors with Gaussian and Rademacher Series.

Theorem 4.2 (Hermitian tensor with Gaussian and Rademacher series). *Given a finite sequence $\mathcal{A}_i$ of fixed Hermitian tensors with dimensions as $\mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, and let $\{\alpha_i\}$ be a finite sequence of independent normal variables. We define*

$$(4.1) \qquad \sigma^2 \stackrel{\text{def}}{=} \left\| \sum_i^n \mathcal{A}_i^2 \right\|,$$

then, for all $\theta \geq 0$, we have

$$(4.2) \qquad \mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^n \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}},$$

and

$$(4.3) \qquad \mathrm{Pr}\left(\left\| \sum_{i=1}^n \alpha_i \mathcal{A}_i \right\| \geq \theta \right) \leq 2\mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}.$$

This theorem is also valid for a finite sequence of independent Rademacher random variables $\{\alpha_i\}$.

Proof. Given a finite sequence of independent Gaussian or Rademacher random variables $\{\alpha_i\}$, from Lemma 4.1, we have

$$e^{\frac{t^2 \mathcal{A}_i^2}{2}} \succeq \mathbb{E}e^{\alpha_i t \mathcal{A}_i}.$$

From the definition in (4.1) and Corollary 3.9, we have

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^n \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M \inf_{t>0}\left\{ e^{-t\theta + \frac{t^2 \sigma^2}{2}} \right\} = \mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}.$$

This establishes (4.2). For (4.3), we have to apply the following facts: $\|\mathcal{X}\|$ for any given Hermitian tensor $\mathcal{X}$. Because Gaussian and Rademacher random variables are symmetric, we have

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^n (-\alpha_i) \mathcal{A}_i \right) \geq \theta \right) = \mathrm{Pr}\left(-\lambda_{\min}\left(\sum_{i=1}^n \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}.$$

Then we obtain (4.3) as follows:

$$\mathrm{Pr}\left(\left\| \sum_{i=1}^n \alpha_i \mathcal{A}_i \right\| \geq \theta \right) = 2\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^n \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq 2\mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}. \qquad\square$$

From the Hermitian dilation definition provided by (2.5), we can extend Theorem 4.2 from square Hermitian tensor to rectangular tensor by the following corollary.

Corollary 4.3 (Rectangular tensor with Gaussian and Rademacher series)**.** *Given a finite sequence $\mathcal{A}_i$ of fixed Hermitian tensors with dimensions as $\mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$, and let $\{\alpha_i\}$ be a finite sequence of independent normal variables. We define*

$$\sigma^2 \overset{\text{def}}{=} \max \left\{ \left\| \sum_{i=1}^{n} \mathcal{A}_i \star_M \mathcal{A}_i^H \right\|, \left\| \sum_{i=1}^{n} \mathcal{A}_i^H \star_M \mathcal{A}_i \right\| \right\}.$$

Then, for all $\theta \geq 0$, we have

$$\Pr \left(\left\| \sum_{i=1}^{n} \alpha_i \mathcal{A}_i \right\| \geq \theta \right) \leq \prod_{m=1}^{M} (I_m + J_m) e^{-\frac{\theta^2}{2\sigma^2}}.$$

This corollary is also valid for a finite sequence of independent Rademacher random variables $\{\alpha_i\}$.

Proof. Let $\{\alpha_i\}$ be a finite sequence of independent Gaussian or Rademacher random variables. Consider a finite sequence of random Hermitian tensors $\{\alpha_i \mathbb{D}(\mathcal{A}_i)\}$ with dimensions $\mathbb{C}^{(I_1+J_1) \times \cdots \times (I_M+J_M) \times (I_1+J_1) \times \cdots \times (I_M+J_M)}$, from the spectral relation of a dilation tensor provided by (2.6), we have

$$(4.4) \qquad \left\| \sum_{i}^{n} \alpha_i \mathcal{A}_i \right\| = \lambda_{\max} \left(\mathbb{D} \left(\sum_{i=1}^{n} \alpha_i \mathcal{A}_i \right) \right) = \lambda_{\max} \left(\sum_{i=1}^{n} \alpha_i \mathbb{D}(\mathcal{A}_i) \right).$$

Due to the following singular value relation:

$$(4.5) \quad \begin{aligned} \sigma^2 &= \left\| \sum_{i} \mathbb{D}(\mathcal{A}_i)^2 \right\| = \left\| \begin{bmatrix} \sum_{i=1}^{n} \mathcal{A}_i \star_M \mathcal{A}_i^H & \mathcal{O} \\ \mathcal{O} & \sum_{i=1}^{n} \mathcal{A}_i^H \star_M \mathcal{A}_i \end{bmatrix} \right\| \\ &= \max \left\{ \left\| \sum_{i=1}^{n} \mathcal{A}_i \star_M \mathcal{A}_i^H \right\|, \left\| \sum_{i}^{n} \mathcal{A}_i^H \star_M \mathcal{A}_i \right\| \right\}. \end{aligned}$$

From (4.4), (4.5) and Theorem 4.2, this corollary is proved. $\square$

4.2. A Gaussian tensor with nonuniform variances

In this section, we will apply results obtained from previous section to consider Gaussian tensor with nonuniform variances.

Corollary 4.4. *Given a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$ and a random tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$ whose entries are independent standard Gaussian normal random*

variables. Let $\circ$ used to represent a Hadamard product (entrywise) between two tensors with same dimensions. Then we have

$$\Pr\big(\|\mathcal{X} \circ \mathcal{A}\| \geq \theta\big) \leq \prod_{m=1}^{M} (I_m + J_m) e^{-\frac{\theta^2}{2\sigma^2}},$$

where

$$\sigma^2 = \max\left\{ \max_{i_1,\ldots,i_M} \|a_{i_1,\ldots,i_M,:}\|^2, \max_{j_1,\ldots,j_M} \|a_{:,j_1,\ldots,j_M}\|^2 \right\},$$

where $a_{i_1,\ldots,i_M,:}$ and $a_{:,j_1,\ldots,j_M}$ represent the row-part of the tensor $\mathcal{A}$ and column-part of the tensor $\mathcal{A}$, respectively.

Proof. Since we can decompose the tensor $\mathcal{X} \circ \mathcal{A}$ as

$$\mathcal{X} \circ \mathcal{A} = \sum_{i_1,\ldots,i_M,j_1,\ldots,j_M} x_{i_1,\ldots,i_M,j_1,\ldots,j_M} a_{i_1,\ldots,i_M,j_1,\ldots,j_M} \mathcal{E}_{i_1,\ldots,i_M,j_1,\ldots,j_M},$$

where $\mathcal{E}_{i_1,\ldots,i_M,j_1,\ldots,j_M} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$ is the tensor with all zero entries except unity at the position $i_1,\ldots,i_M,j_1,\ldots,j_M$; then we have

$$\sum_{i_1,\ldots,i_M,j_1,\ldots,j_M} \big(a_{i_1,\ldots,i_M,j_1,\ldots,j_M} \mathcal{E}_{i_1,\ldots,i_M,j_1,\ldots,j_M}\big) \overline{\big(a_{i_1,\ldots,i_M,j_1,\ldots,j_M} \mathcal{E}_{i_1,\ldots,i_M,j_1,\ldots,j_M}\big)}$$

$$= \sum_{i_1,\ldots,i_M} \left(\sum_{j_1,\ldots,j_M} \|a_{i_1,\ldots,i_M,j_1,\ldots,j_M}\|^2\right) \mathcal{E}_{i_1,\ldots,i_M,i_1,\ldots,i_M}$$

$$= \mathrm{diag}\big(\|a_{1,\ldots,1:}\|^2, \|a_{1,\ldots,2:}\|^2, \ldots, \|a_{I_1,\ldots,I_M:}\|^2\big),$$

and, similarly,

$$\sum_{i_1,\ldots,i_M,j_1,\ldots,j_M} \overline{\big(a_{i_1,\ldots,i_M,j_1,\ldots,j_M} \mathcal{E}_{i_1,\ldots,i_M,j_1,\ldots,j_M}\big)} \big(a_{i_1,\ldots,i_M,j_1,\ldots,j_M} \mathcal{E}_{i_1,\ldots,i_M,j_1,\ldots,j_M}\big)$$

$$= \sum_{j_1,\ldots,j_M} \left(\sum_{i_1,\ldots,i_M} \|a_{i_1,\ldots,i_M,j_1,\ldots,j_M}\|^2\right) \mathcal{E}_{j_1,\ldots,j_M,j_1,\ldots,j_M}$$

$$= \mathrm{diag}\big(\|a_{:,1,\ldots,1}\|^2, \|a_{:,1,\ldots,2}\|^2, \ldots, \|a_{:,J_1,\ldots,J_M}\|^2\big).$$

Therefore, we have

$$\sigma^2 = \max\big\{ \mathrm{diag}\big(\|a_{1,\ldots,1:}\|^2, \|a_{1,\ldots,2:}\|^2, \ldots, \|a_{I_1,\ldots,I_M:}\|^2\big),$$

$$\mathrm{diag}\big(\|a_{:,1,\ldots,1}\|^2, \|a_{:,1,\ldots,2}\|^2, \ldots, \|a_{:,J_1,\ldots,J_M}\|^2\big)\big\}$$

$$= \max\left\{ \max_{i_1,\ldots,i_M} \|a_{i_1,\ldots,i_M,:}\|^2, \max_{j_1,\ldots,j_M} \|a_{:,j_1,\ldots,j_M}\|^2 \right\}.$$

Finally, from Corollary 4.3, this corollary is proved. $\square$

4.3. Lower and upper bounds of tensor expectation

Given a finite sequence $\mathcal{A}_i$ of fixed Hermitian tensors with dimensions as $\mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, and let $\{\alpha_i\}$ be a finite sequence of independent normal variables. We define the following random tensor

$$\mathcal{X} = \sum_{i=1}^{n} \alpha_i \mathcal{A}_i.$$

From Theorem 4.2, we have

$$(4.6) \qquad \mathbb{E}\big(\|\mathcal{X}\|^2\big) = \int_0^{\infty} \Pr\big(\|\mathcal{X}\| > \sqrt{t}\big)\, dt \leq \int_0^{\infty} 2\mathbb{I}_1^M e^{-\frac{t}{2\sigma^2}}\, dt = 4\sigma^2 \mathbb{I}_1^M.$$

On the other hand, from Jesen's inequality, we have

$$(4.7) \qquad \mathbb{E}\big(\|\mathcal{X}\|^2\big) = \mathbb{E}\|\mathcal{X}^2\| \geq \|\mathbb{E}(\mathcal{X}^2)\| = \left\| \sum_{i=1}^{n} \mathcal{A}_i^2 \right\| = \sigma^2.$$

From both (4.6) and (4.7), we have the following relation:

$$c\sigma \leq \mathbb{E}\|\mathcal{X}\| \leq 2\sigma \sqrt{\mathbb{I}_1^M}.$$

This shows that the tensor variance parameter σ^2 controls the expected norm $\mathbb{E}\|\mathcal{X}\|$ with square root of logarithmic function for the tensor dimensions.

5. Tensor Chernoff bounds

The traditional Chernoff bounds concern the sum of independent, nonnegative, and uniformly bounded random variables. In this work, we will try to extend such Chernoff bounds under the scenario of random tensors.

5.1. Tensor Chernoff bounds derivations

We begin to present a lemma about the semidefinite relation for the tensor moment-generating function of a random positive semidefinite contraction.

Lemma 5.1. *Given a random positive semidefinite tensor with* $\lambda_{\max}(\mathcal{X}) \leq 1$*, then, for any* $t \in \mathbb{R}$*, we have*

$$\mathcal{I} + (e^t - 1)\mathbb{E}\mathcal{X} \succeq \mathbb{E}e^{t\mathcal{X}}.$$

Proof. Consider a convex function $g(x) = e^{tx}$, we have

$$1 + (e^t - 1)x \geq g(x),$$

where $x \in [0, 1]$. Since the eigenvalues of the random tensor $\mathcal{X}$ lie in the interval $[0, 1]$, from (2.2), we obtain

$$\mathcal{I} + (e^t - 1)\mathcal{X} \succeq e^{t\mathcal{X}}.$$

Then this lemma is proved by taking the expectation with respect to the random tensor $\mathcal{X}$. $\qquad\qquad\square$

Given two real values $a, b \in [0, 1]$, we define *binary information divergence* of a and b, expressed by $\mathfrak{D}(a \parallel b)$, as

$$\mathfrak{D}(a \parallel b) \overset{\text{def}}{=} a \log \frac{a}{b} + (1 - a)\frac{1 - a}{1 - b}.$$

We are ready to present tensor Chernoff inequality.

Theorem 5.2 (Tensor Chernoff bound I). *Consider a sequence* $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ *of independent, random, Hermitian tensors that satisfy*

$$\mathcal{X}_i \succeq \mathcal{O} \quad and \quad \lambda_{\max}(\mathcal{X}_i) \leq 1 \quad almost\ surely.$$

Define the following two quantities:

$$\overline{\mu}_{\max} \overset{\text{def}}{=} \lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\mathcal{X}_i\right) \quad and \quad \overline{\mu}_{\min} \overset{\text{def}}{=} \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\mathcal{X}_i\right),$$

then we have the following two inequalities:

$$(5.1) \qquad \Pr\left(\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M e^{-n\mathfrak{D}(\theta\|\overline{\mu}_{\max})} \quad for\ \overline{\mu}_{\max} \leq \theta \leq 1;$$

and

$$(5.2) \qquad \Pr\left(\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \mathcal{X}_i\right) \leq \theta\right) \leq \mathbb{I}_1^M e^{-n\mathfrak{D}(\theta\|\overline{\mu}_{\min})} \quad for\ 0 \leq \theta \leq \overline{\mu}_{\min}.$$

Proof. From Lemma 5.1, we have

$$\mathcal{I} + f(t)\mathbb{E}\mathcal{X}_i \succeq \mathbb{E}e^{t\mathcal{X}_i},$$

where $f(t) \overset{\text{def}}{=} e^t - 1$ for $t > 0$. By applying Corollary 3.10, we obtain

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \alpha\right) \leq \mathbb{I}_1^M \exp\left(-t\alpha + n\log\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n \left(\mathcal{I} + f(t)\mathbb{E}\mathcal{X}_i\right)\right)\right)$$

$$(5.3)$$

$$= \mathbb{I}_1^M \exp\left(-t\alpha + n\log\lambda_{\max}\left(\mathcal{I} + f(t)\frac{1}{n}\sum_{i=1}^n \mathbb{E}\mathcal{X}_i\right)\right)$$

$$= \mathbb{I}_1^M \exp\left(-t\alpha + n\log\left(1 + f(t)\overline{\mu}_{\max}\right)\right).$$

The last equality follows from the definition of $\overline{\mu}_{\max}$ and the eigenvalue map properties. When the value t on the right-hand side of (5.3) is

$$(5.4) \qquad t = \log \frac{\alpha}{1 - \alpha} - \log \frac{\overline{\mu}_{\max}}{1 - \overline{\mu}_{\max}},$$

we can achieve the tightest upper bound in (5.3). By substituting the value t in (5.4) into (5.3) and changing the variable $\alpha \to n\theta$, (5.1) is proved. The next goal is to prove (5.2).

If we apply Lemma 5.1 to the sequence $\{-\mathcal{X}_i\}$, we have

$$\mathcal{I} - g(t)\mathbb{E}\mathcal{X}_i \succeq \mathbb{E}e^{t(-\mathcal{X}_i)},$$

where $g(t) \stackrel{\text{def}}{=} 1 - e^t$ for $t > 0$. By applying Corollary 3.10 again, we obtain

$$\Pr\left(\lambda_{\min}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \le \alpha \right) = \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}(-\mathcal{X}_i) \right) \ge \alpha \right)$$

$$(5.5) \qquad \le \mathbb{I}_1^M \exp\left(t\alpha + n \log \lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{I} - g(t)\mathbb{E}\mathcal{X}_i \right) \right) \right)$$

$$=_1 \mathbb{I}_1^M \exp\left(t\alpha + n \log\left(1 - f(t)\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i \right) \right) \right)$$

$$= \mathbb{I}_1^M \exp\left(t\alpha + n \log\left(1 - g(t)\overline{\mu}_{\min} \right) \right),$$

where we apply the relation $\lambda_{\min}\left(-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i \right) = -\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i \right)$ at the equality $=_1$. When the value t on the right-hand side of (5.5) is

$$(5.6) \qquad t = log\frac{\overline{\mu}_{\max}}{1 - \overline{\mu}_{\max}} - \log \frac{\alpha}{1 - \alpha},$$

we can achieve the tightest upper bound in (5.5). By substituting the value t in (5.6) into (5.5) and changing the variable $\alpha \to n\theta$, (5.2) is proved also. $\qquad\square$

The tensor Chernoff bounds discussed in Theorem 5.2 is not related to $\mu_{\max}$ and $\mu_{\min}$ directly. The next theorem is another version of tensor Chernoff bounds to associate the probability range in terms of $\mu_{\max}$ and $\mu_{\min}$ directly and this format of tensor Chernoff bounds are easier to be applied.

Theorem 5.3 (Tensor Chernoff bound II). *Consider a sequence $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ of independent, random, Hermitian tensors that satisfy*

$$\mathcal{X}_i \succeq \mathcal{O} \quad and \quad \lambda_{\max}(\mathcal{X}_i) \le T \quad almost\ surely.$$

Define the following two quantities:

$$\mu_{\max} \stackrel{\text{def}}{=} \lambda_{\max}\left(\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i \right) \quad and \quad \mu_{\min} \stackrel{\text{def}}{=} \lambda_{\min}\left(\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i \right),$$

then we have the following two inequalities:

$$(5.7) \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq (1+\theta)\mu_{\max}\right) \leq \mathbb{I}_1^M\left(\frac{e^{\theta}}{(1+\theta)^{1+\theta}}\right)^{\mu_{\max}/T} \qquad \text{for } \theta \geq 0;$$

and

$$(5.8) \qquad \Pr\left(\lambda_{\min}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \leq (1-\theta)\mu_{\min}\right) \leq \mathbb{I}_1^M\left(\frac{e^{-\theta}}{(1-\theta)^{1-\theta}}\right)^{\mu_{\min}/T} \qquad \text{for } \theta \in [0,1].$$

Proof. Without loss of generality, we can assume $T = 1$ in our proof. From (5.3) and the inequality $\log(1+x) \leq x$ for $x > -1$, we have

$$(5.9) \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq t\right) \leq \mathbb{I}_1^M \exp(-\delta t + f(\delta)\mu_{\max}).$$

By selecting $\delta = \log(1+\theta)$ and $t \to (1+\theta)\mu_{\max}$, we can establish (5.7).

From (5.5) and the inequality $\log(1+x) \leq x$ for $x > -1$, we have

$$\Pr\left(\lambda_{\min}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \leq t\right) \leq \mathbb{I}_1^M \exp(-\delta t - f(\delta)\mu_{\min}).$$

By selecting $\delta = -\log(1-\theta)$ and $t \to (1-\theta)\mu_{\min}$, we can establish (5.8). Therefore, this theorem is proved. $\qquad \square$

5.2. Application of tensor Chernoff bounds

Consider a general random tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N}$, we can express the $\mathcal{X}$ as

$$\mathcal{X} = [x_1, x_2, \ldots, x_{J_1 \cdots J_N}],$$

where x_i is a family of independent random tensors in $\mathbb{C}^{I_1 \times \cdots \times I_M}$ (vector part in $\mathcal{X}$). The squared norm of $\mathcal{X}$ can be expressed as

$$(5.10) \qquad \|\mathcal{X}\|^2 = \lambda_{\max}(\mathcal{X} \star_N \mathcal{X}^H) = \lambda_{\max}\left(\sum_{i=1}^{J_1 \cdots J_N} x_i \star_0 \overline{x_i}\right).$$

Similarly, for the minimum singular value of the tensor $\mathcal{X}$, we have

$$(5.11) \qquad \text{Minimum singular value of } \mathcal{X} = \lambda_{\min}(\mathcal{X} \star_N \mathcal{X}^H) = \lambda_{\min}\left(\sum_{i=1}^{J_1 \cdots J_N} x_i \star_0 x_i^H\right).$$

From tensor Chernoff bounds provided by Theorem 5.3, we can bound both (5.10) and (5.11).

Another application of tensor Chernoff bounds is to estimate the expectation of the maximum eigenvalue of independent sum of random tensors.

Corollary 5.4 (Upper and lower bounds for the maximum eigenvalue). *Consider a sequence $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ of independent, random, Hermitian tensors that satisfy*

$$\mathcal{X}_i \succeq \mathcal{O} \quad and \quad \lambda_{\max}(\mathcal{X}_i) \leq T \quad almost\ surely.$$

Then we have

$$(5.12) \qquad \mu_{\max} \leq \mathbb{E}\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \leq C\mathbb{I}_1^M e^{-\mu_{\max}/T},$$

where the constant value of C is about 10.28.

Proof. The lower bound in (5.12) is true from the convexity of the function $\mathcal{A} \to \lambda_{\max}(\mathcal{A})$ and the Jensen's inequality.

For the upper bound, we have

$$\mathbb{E}\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) = \int_0^\infty \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \geq t\right) dt$$

$$(5.13) \qquad \leq_1 \int_0^\infty \mathbb{I}_1^M \exp(-\delta t + (e^\delta - 1)\mu_{\max}/T)\, dt$$

$$= \frac{e^{e^\delta}}{\delta}\mathbb{I}_1^M e^{-\mu_{\max}/T} \leq \frac{e^{e^{\delta_{\mathrm{opt}}}}}{\delta_{\mathrm{opt}}}\mathbb{I}_1^M e^{-\mu_{\max}/T} = C\mathbb{I}_1^M e^{-\mu_{\max}/T},$$

where the inequality $\leq_1$ comes from (5.9) with the scaling factor T. If we select θ as the solution of the following relation $e^{\delta_{\mathrm{opt}}} = \frac{1}{\delta_{\mathrm{opt}}}$ to minimize the right-hand side of (5.13), we have the desired upper bound when $\delta_{\mathrm{opt}} \approx 0.56699$. This corollary is proved. $\square$

6. Tensor Bernstein bounds

For random variables, Bernstein inequalities give the upper tail of a sum of independent, zero-mean random variables that are either bounded or subexponential. In this section, we wish to extend Bernstein bounds for a sum of zero-mean random tensors.

6.1. Tensor Bernstein bounds derivation

We will consider bounded Tensor Bernstein bounds first by considering the bounded Bernstein moment-generating function with the following lemma.

Lemma 6.1. *Given a random Hermitian tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ that satisfies*

$$\mathbb{E}\mathcal{X} = 0 \quad and \quad \lambda_{\max}(\mathcal{X}) \leq 1 \quad almost\ surely.$$

Then we have

$$e^{(e^t - t - 1)\mathbb{E}(\mathcal{X}^2)} \succeq \mathbb{E}e^{t\mathcal{X}}$$

where $t > 0$.

Proof. If we define a real function $g(x) \stackrel{\text{def}}{=} \frac{e^{tx} - tx - 1}{x^2}$, it is easy to see that this function $g(x)$ is an increasing function for $0 < x \leq 1$. From (2.2), we have

$$(6.1) \qquad\qquad\qquad\qquad g(\mathcal{X}) \preceq g(1)\mathcal{I}.$$

Moreover, we also have

$$(6.2) \qquad\qquad e^{t\mathcal{X}} = \mathcal{I} + t\mathcal{X} + g(\mathcal{X}) \star_M \mathcal{X}^2 \preceq \mathcal{I} + t\mathcal{X} + g(1)\mathcal{X}^2,$$

where $\preceq$ comes from (6.1). By taking the expectation to both sides of (6.2), we then obtain

$$\mathbb{E}e^{t\mathcal{X}} \preceq \mathcal{I} + g(1)\mathbb{E}(\mathcal{X}^2) \preceq e^{g(1)\mathbb{E}(\mathcal{X}^2)} = e^{(e^t - t - 1)\mathbb{E}(\mathcal{X}^2)}.$$

This lemma is established. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We are ready to present the tensor Bernstein bounds for random tensors with bounded $\lambda_{\max}$.

Theorem 6.2 (Bounded $\lambda_{\max}$ tensor Bernstein bounds). *Given a finite sequence of independent Hermitian tensors $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ that satisfy*

$$\mathbb{E}\mathcal{X}_i = 0 \quad \text{and} \quad \lambda_{\max}(\mathcal{X}_i) \leq T \quad \text{almost surely.}$$

Define the total variance σ^2 as $\sigma^2 \stackrel{\text{def}}{=} \big\| \sum_i^n \mathbb{E}(\mathcal{X}_i^2) \big\|$. Then we have the following inequalities:

$$(6.3) \qquad\qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M \exp\left(\frac{-\theta^2/2}{\sigma^2 + T\theta/3} \right);$$

$$(6.4) \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M \exp\left(\frac{-3\theta^2}{8\sigma^2} \right) \quad \text{for } \theta \leq \sigma^2/T;$$

and

$$(6.5) \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M \exp\left(\frac{-3\theta}{8T} \right) \quad \text{for } \theta \geq \sigma^2/T.$$

Proof. Without loss of generality, we can assume that $T = 1$ since the summands are 1-homogeneous and the variance is 2-homogeneous. From Lemma 6.1, we have

$$\mathbb{E}e^{t\mathcal{X}_i} \preceq e^{(e^t - t - 1)\mathbb{E}(\mathcal{X}_i^2)} \quad \text{for } t > 0.$$

By applying Corollary 3.9, we then have

$$(6.6) \qquad \begin{aligned} \Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i \right) \geq \theta \right) &\leq \mathbb{I}_1^M \exp\left(-t\theta + (e^t - t - 1)\lambda_{\max}\left(\sum_{i=1}^n \mathbb{E}(\mathcal{X}_i^2) \right) \right) \\ &= \mathbb{I}_1^M \exp\left(-t\theta + \sigma^2(e^t - t - 1) \right). \end{aligned}$$

The right-hand side of (6.6) can be minimized by setting $t = \log(1 + \theta/\sigma^2)$. Substituting such t and simplifying the right-hand side of (6.6), we obtain (6.3).

For $\theta \leq \sigma^2/T$, we have

$$\frac{1}{\sigma^2 + T\theta/3} \geq \frac{1}{\sigma^2 + T(\sigma^2/T)/3} = \frac{3}{4\sigma^2},$$

then we obtain (6.4). Correspondingly, for $\theta \geq \sigma^2/T$, we have

$$\frac{\theta}{\sigma^2 + T\theta/3} \geq \frac{\sigma^2/T}{\sigma^2 + T(\sigma^2/T)/3} = \frac{3}{4T},$$

then we obtain (6.5) also. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The following Theorem 6.4 is an extension of Theorem 6.2 by allowing the moments of the random tensors to grow at a controlled rate. We have to prepare subexponential Bernstein moment-generating function lemma first for later proof of Theorem 6.4.

Lemma 6.3. *Suppose that $\mathcal{X}$ is a random Hermitian tensor that satisfies*

$$\mathbb{E}\mathcal{X} = 0 \quad and \quad \mathbb{E}(\mathcal{X}^p) \preceq \frac{p!\mathcal{A}^2}{2} \quad for\ p = 2, 3, 4, \ldots.$$

Then we have

$$\exp\left(\frac{t^2\mathcal{A}^2}{2(1-t)}\right) \succeq \mathbb{E}e^{t\mathcal{X}},$$

where $0 < t < 1$.

Proof. From Taylor series of the tensor exponential expansion, we have

$$\mathbb{E}e^{t\mathcal{X}} = \mathcal{I} + t\mathbb{E}\mathcal{X} + \sum_{p=2}^{\infty} \frac{t^p\mathbb{E}(\mathcal{X}^p)}{p!} \preceq \mathcal{I} + \sum_{p=2}^{\infty} \frac{t^p\mathcal{A}^2}{2} = \mathcal{I} + \frac{t^2\mathcal{A}^2}{2(1-t)} \preceq \exp\left(\frac{t^2\mathcal{A}^2}{2(1-t)}\right).$$

Therefore, this lemma is proved. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 6.4 (Subexponential tensor Bernstein bounds)**.** *Given a finite sequence of independent Hermitian tensors $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ that satisfy*

$$\mathbb{E}\mathcal{X}_i = 0 \quad and \quad \mathbb{E}(\mathcal{X}_i^p) \preceq \frac{p!T^{p-2}}{2}\mathcal{A}_i^2,$$

where $p = 2, 3, 4, \ldots$. Define the total variance σ^2 as $\sigma^2 \stackrel{\text{def}}{=} \left\| \sum_i^n \mathcal{A}_i^2 \right\|$. Then we have the following inequalities:

$$(6.7) \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-\theta^2/2}{\sigma^2 + T\theta}\right);$$

$$(6.8) \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-\theta^2}{4\sigma^2}\right) \quad for\ \theta \leq \sigma^2/T;$$

and

$$\text{(6.9)} \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-\theta}{4T}\right) \quad \text{for } \theta \geq \sigma^2/T.$$

Proof. Without loss of generality, we can assume that $T = 1$. From Lemma 6.3, we have

$$\mathbb{E}e^{t\mathcal{X}_i} \preceq e^{\frac{t^2 \mathcal{A}_i^2}{2(1-t)}},$$

where $0 < t < 1$.

By applying Corollary 3.9, we then have

$$\text{(6.10)} \qquad \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(-t\theta + \frac{t^2}{2(1-t)}\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{A}_i^2\right)\right)$$
$$= \mathbb{I}_1^M \exp\left(-t\theta + \frac{\sigma^2 t^2}{2(1-t)}\right).$$

The right-hand side of (6.10) can be minimized by setting $t = \frac{\theta}{\theta + \sigma^2}$. Substituting such t and simplifying the right-hand side of (6.10), we obtain (6.7).

For $\theta \leq \sigma^2/T$, we have

$$\frac{1}{\sigma^2 + T\theta} \geq \frac{1}{\sigma^2 + T(\sigma^2/T)} = \frac{1}{2\sigma^2},$$

then we obtain (6.8). Similarly, for $\theta \geq \sigma^2/T$, we have

$$\frac{\theta}{\sigma^2 + T\theta} \geq \frac{\sigma^2/T}{\sigma^2 + T(\sigma^2/T)} = \frac{1}{2T}.$$

Therefore, we also obtain (6.9). $\qquad \square$

6.2. Application of tensor Bernstein bounds

The tensor Bernstein bounds can also be extended to rectangular tensors by dilation. Consider a sequence of tensors $\{\mathcal{Y}_i\} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$ satisfy the following:

$$\mathbb{E}\mathcal{Y}_i = \mathcal{O} \quad \text{and} \quad \|\mathcal{Y}_i\| \leq T \quad \text{almost surely}.$$

If the variance σ^2 is expressed as

$$\sigma^2 \overset{\text{def}}{=} \max\left\{\left\|\sum_{i=1}^{n}\mathcal{Y}_i \star_M \mathcal{Y}_i^H\right\|, \left\|\sum_{i=1}^{n}\mathcal{Y}_i^H \star_M \mathcal{Y}_i\right\|\right\},$$

we have

$$\Pr\left(\left\|\sum_{i=1}^{n}\mathcal{Y}_i\right\| \geq \theta\right) \leq \prod_{m=1}^{M}(I_m + J_m)\exp\left(\frac{-\theta^2/2}{\sigma^2 + T\theta/3}\right);$$

from Theorem 6.2.

Another application of tensor Bernstein bounds is to get upper and lower Bounds for the maximum eigenvalue with subexponential tensors. This application can relax Corollary 5.4 conditions by allowing the moments of the random tensors to grow at a controlled rate.

Corollary 6.5 (Upper and lower bounds for the maximum eigenvalue for subexponential tensors)**.** *Consider a sequence* $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ *of independent, random, Hermitian tensors that satisfy*

$$\mathcal{X}_i \succeq \mathcal{O} \quad and \quad \mathbb{E}(\mathcal{X}_i^p) \preceq \frac{p! T^{p-2}}{2} \mathcal{A}_i^2,$$

and $\sigma^2 \overset{\mathrm{def}}{=} \left\| \sum_{i=1}^n \mathcal{A}_i^2 \right\|$. *Then we have*

$$(6.11) \qquad \mu_{\max} \leq \mathbb{E}\lambda_{\max} \left(\sum_{i=1}^n \mathcal{X}_i \right) \leq 2 \mathbb{I}_1^M \left(\sigma \mathfrak{G} \left(\frac{\sigma}{2T} \right) + 2Te^{-\frac{\sigma^2}{4T^2}} \right),$$

where $\mathfrak{G} \left(\frac{\sigma}{2T} \right) \overset{\mathrm{def}}{=} \int_0^{\frac{\sigma}{2T}} e^{-s^2} \, ds$.

Proof. The lower bound in (6.11) is true from the convexity of the function $\mathcal{A} \to \lambda_{\max}(\mathcal{A})$ and the Jensen's inequality.

For the upper bound, we have

$$\mathbb{E}\lambda_{\max} \left(\sum_{i=1}^n \mathcal{X}_i \right) = \int_0^\infty \Pr \left(\lambda_{\max} \left(\sum_{i=1}^n \mathcal{X}_i \right) \geq t \right) dt$$

$$\leq_1 \mathbb{I}_1^M \int_0^{\frac{\sigma^2}{T}} \exp \left(-\frac{t^2}{4\sigma^2} \right) dt + \mathbb{I}_1^M \int_{\frac{\sigma^2}{T}}^\infty \exp \left(-\frac{t}{4T} \right) dt$$

$$= 2\mathbb{I}_1^M \left(\sigma \mathfrak{G} \left(\frac{\sigma}{2T} \right) + 2Te^{-\frac{\sigma^2}{4T^2}} \right),$$

where the inequality $\leq_1$ comes from (6.8) and (6.9). This corollary is proved by introducing Gaussian integral function $\mathfrak{G}(x) \overset{\mathrm{def}}{=} \int_0^x e^{-s^2} \, ds$. □

7. Martingale deviation bounds

In this section, we introduce concepts about tensor martingales in Section 7.1, and extend Hoeffding, Azuma, and McDiarmid inequalities to tensors context in Section 7.2.

7.1. Tensor martingales

Necessary definitions about tensor martingales will be provided here for later tensor martingale deviation bounds derivations. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a master probability space. Consider

a filtration $\{\mathfrak{F}_i\}$ contained in the master sigma algebra as

$$\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \cdots \subset \mathfrak{F}_\infty \subset \mathfrak{F}.$$

Given such a filtration, we define the conditional expectation $\mathbb{E}_i[\,\cdot\,] \stackrel{\text{def}}{=} \mathbb{E}_i[\,\cdot\,\mid\mathfrak{F}_i]$. A sequence $\{\mathcal{Y}_i\}$ of random tensors is called *adapted* to the filtration when each tensor $\mathcal{Y}_i$ is measurable with respect to $\mathfrak{F}_i$. We can think that an adapted sequence is one where the present depends only on the past.

An adapted sequence $\{\mathcal{X}_i\}$ of Hermitian tensors is named as a *tensor martingale* when

$$\mathbb{E}_{i-1}\mathcal{X}_i = \mathcal{X}_{i-1} \quad \text{and} \quad \mathbb{E}\|\mathcal{X}_i\| < \infty,$$

where $i = 1, 2, 3, \ldots$. We obtain a scalar martingale if we track any fixed entry of a tensor martingale $\{\mathcal{X}_i\}$. Given a tensor martingale $\{\mathcal{X}_i\}$, we can construct the following new sequence of tensors

$$\mathcal{Y}_i \stackrel{\text{def}}{=} \mathcal{X}_i - \mathcal{X}_{i-1} \quad \text{for } i = 1, 2, 3, \ldots.$$

We then have $\mathbb{E}_{i-1}\mathcal{Y}_i = \mathcal{O}$.

7.2. Tensor martingale deviation bounds

Two lemmas should be presented first before presenting tensor martingale deviation bounds and their proofs.

Lemma 7.1 (Tensor symmetrization). *Let $\mathcal{A}$ be a fixed Hermitian tensor, and let $\mathcal{X}$ be a random Hermitian tensor with $\mathbb{E}\mathcal{E} = \mathcal{O}$. Then*

$$\mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\mathcal{X}} \le \mathbb{E}\operatorname{Tr} e^{\mathcal{A}+2\beta\mathcal{X}},$$

where β is a Rademacher random variable.

Proof. Build an independent copy random tensor $\mathcal{Y}$ from $\mathcal{X}$, and let $\mathbb{E}_\mathcal{Y}$ denote the expectation with respect to the new random tensor $\mathcal{Y}$. Then we have

$$\mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\mathcal{X}} = \mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\mathcal{X}-\mathbb{E}_\mathcal{Y}\mathcal{Y}} \le \mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\mathcal{X}-\mathcal{Y}} = \mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\beta(\mathcal{X}-\mathcal{Y})},$$

where the first equality uses $\mathbb{E}_\mathcal{Y}\mathcal{Y} = \mathcal{O}$; the inequality uses the convexity of the trace exponential with Jensen's inequality; finally, the last equality comes from that the random tensor $\mathcal{X} - \mathcal{Y}$ is a symmetric random tensor and Rademacher is also a symmetric random variable.

This lemma is established by the following:

$$\mathbb{E}\operatorname{Tr} e^{\mathcal{A}+\mathcal{X}} \le \mathbb{E}\operatorname{Tr} \left(e^{\mathcal{A}/2+\beta\mathcal{X}} e^{\mathcal{A}/2-\beta\mathcal{Y}} \right)$$
$$\le \left(\mathbb{E}\operatorname{Tr} e^{\mathcal{A}+2\beta\mathcal{X}} \right)^{1/2} \left(\mathbb{E}\operatorname{Tr} e^{\mathcal{A}-2\beta\mathcal{Y}} \right)^{1/2} = \mathbb{E}\operatorname{Tr} e^{\mathcal{A}+2\beta\mathcal{X}},$$

where the first inequality comes from tensor Golden–Thompson inequality [6], the second inequality comes from the Cauchy–Schwarz inequality, and the last identity follows from that the two factors are identically distributed. $\qquad\square$

The other lemma is to provide the tensor cumulant-generating function of a symmetrized random tensor.

Lemma 7.2 (Cumulant-generating function of symmetrized random tensor). *Given that $\mathcal{X}$ is a random Hermitian tensor and $\mathcal{A}$ is a fixed Hermitian tensor that satisfies $\mathcal{X}^2 \preceq \mathcal{A}^2$. Then we have*

$$\log \mathbb{E}\big[e^{2\beta t \mathcal{X}} \mid \mathcal{X}\big] \preceq 2t^2 \mathcal{A}^2.$$

Proof. From Lemma 4.1, we have

$$\mathbb{E}\big[e^{2\beta t \mathcal{X}} \mid \mathcal{X}\big] \preceq e^{2t^2 \mathcal{X}^2}.$$

And, from the monotone property of logarithm, we also have

$$\log \mathbb{E}\big[e^{2t\theta \mathcal{X}} \mid \mathcal{X}\big] \preceq 2t^2 \mathcal{X}^2 \preceq 2t^2 \mathcal{A}^2 \quad \text{for } t \in \mathbb{R}.$$

Therefore, this lemma is proved. $\qquad\square$

At this point, we are ready to present tensor martingale deviation bounds. In probability theory, the Azuma inequality gives a concentration result for the values of martingales that have bounded differences. Below is a tensor extension.

Theorem 7.3 (Tensor Azuma inequality). *Given a finite adapted sequence of Hermitian tensors $\{\mathcal{X}_i \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}\}$ and a fixed sequence of Hermitian tensors $\{\mathcal{A}_i\}$ that satisfy*

$$\mathbb{E}_{i-1}\mathcal{X}_i = 0 \quad \text{and} \quad \mathcal{X}_i^2 \preceq \mathcal{A}_i^2 \quad \text{almost surely,}$$

where $i = 1, 2, 3, \ldots$. Define the total variance σ^2 as $\sigma^2 \overset{\text{def}}{=} \big\| \sum_i^n \mathcal{A}_i^2 \big\|$. Then we have the following inequality:

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}.$$

Proof. Define the filtration $\mathfrak{F}_i \overset{\text{def}}{=} \mathfrak{F}(\mathcal{X}_1, \ldots, \mathcal{X}_i)$ for the process $\{\mathcal{X}_i\}$. Then we have

$$\mathbb{E}\,\mathrm{Tr}\exp\left(\sum_{i=1}^n t\mathcal{X}_i \right) = \mathbb{E}\left(\mathbb{E}\left(\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + t\mathcal{X}_n \right) \Big| \mathfrak{F}_n \right) \Big| \mathfrak{F}_{n-1} \right)$$

$$(7.1) \qquad\qquad \leq \mathbb{E}\left(\mathbb{E}\left(\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + 2\beta t\mathcal{X}_n \right) \Big| \mathfrak{F}_n \right) \Big| \mathfrak{F}_n \right)$$

$$\leq \mathbb{E}\left(\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + \log \mathbb{E}\left(e^{2\beta t \mathcal{X}_n}\ \Big|\ \mathfrak{F}_n\right)\right)\ \Big|\ \mathfrak{F}_n\right)$$

$$\leq \mathbb{E}\,\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + 2t^2\mathcal{A}_n^2\right),$$

where the first equality comes from the tower property of conditional expectation; the first inequality comes from Lemma 7.1; the second inequality comes from Corollary 3.6 and the relaxation of the condition to the larger algebra set $\mathfrak{F}_n$; finally, the last inequality requires Lemma 7.2.

If we continue the iteration procedure based on (7.1), we have

$$(7.2) \qquad \mathbb{E}\,\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) \leq \mathrm{Tr}\exp\left(2t^2\sum_{i=1}^{n}\mathcal{A}_i^2\right).$$

Applying (7.2) into Lemma 3.1, we obtain

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right)$$

$$\leq \inf_{t>0}\left\{e^{-t\theta}\mathbb{E}\,\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right)\right\} \leq \inf_{t>0}\left\{e^{-t\theta}\mathbb{E}\,\mathrm{Tr}\exp\left(2t^2\sum_{i=1}^{n}\mathcal{A}_i^2\right)\right\}$$

$$\leq \inf_{t>0}\left\{e^{-t\theta}\mathbb{I}_1^M \lambda_{\max}\left(\exp\left(2t^2\sum_{i=1}^{n}\mathcal{A}_i^2\right)\right)\right\} = \inf_{t>0}\left\{e^{-t\theta}\mathbb{I}_1^M \exp\left(2t^2\sigma^2\right)\right\} \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}},$$

where the third inequality utilizes $\lambda_{\max}$ to bound trace, the equality applies the definition of σ^2 and spectral mapping theorem, finally, we select $t = \frac{\theta}{4\sigma^2}$ to minimize the upper bound to obtain this theorem. $\qquad\square$

If we add extra assumption that the summands are independent, Theorem 7.3 gives Hoeffding's inequality for random tensors. If we apply Theorem 7.3 to a Hermitian tensor martingale, we will have the following corollary.

Corollary 7.4. *Let $\{\mathcal{Y}_i : i = 1, 2, \ldots, n\} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ be a Hermitian tensor martingale, and let $\mathcal{X}_i$ be the difference sequence of $\{\mathcal{Y}_i\}$, i.e., $\mathcal{X}_i \overset{\text{def}}{=} \mathcal{Y}_i - \mathcal{Y}_{i-1}$ for $i = 1, 2, 3, \ldots$. If the difference sequence satisfies*

$$\mathbb{E}_{i-1}\mathcal{X}_i = 0 \quad and \quad \mathcal{X}_i^2 \preceq \mathcal{A}_i \quad almost\ surely,$$

where $i = 1, 2, 3, \ldots$ and the total variance σ^2 is defined as $\sigma^2 \overset{\text{def}}{=} \left\|\sum_i^n \mathcal{A}_i^2\right\|$. Then we have

$$\mathrm{Pr}\left(\lambda_{\max}(\mathcal{Y}_n - \mathbb{E}\mathcal{Y}_n) \geq \theta\right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}.$$

In the scalar setting, McDiarmid inequality can be treated as a corollary of Azuma's inequality. McDiarmid inequality states that a function of independent random variables exhibits normal concentration about its mean, and the variance depends on the function value sensitivity with respect to the input. A version of the bounded differences inequality holds in the tensor context.

Theorem 7.5 (Tensor McDiarmid inequality)**.** *Given a set of n independent random variables, i.e., $\{X_i : i = 1, 2, \ldots, n\}$, and let F be a Hermitian tensor-valued function that maps these n random variables to a Hermitian tensor of dimension within $\mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$. Consider a sequence of Hermitian tensors $\{\mathcal{A}_i\}$ that satisfy*

$$(7.3) \qquad \big(F(x_1, \ldots, x_i, \ldots, x_n) - F(x_1, \ldots, x_i', \ldots, x_n)\big)^2 \preceq \mathcal{A}_i^2,$$

where $x_i, x_i' \in X_i$ and $1 \leq i \leq n$. Define the total variance σ^2 as $\sigma^2 \overset{\text{def}}{=} \big\| \sum_i^n \mathcal{A}_i^2 \big\|$. Then we have the following inequality:

$$\Pr\big(\lambda_{\max}\big(F(x_1, \ldots, x_n) - \mathbb{E}F(x_1, \ldots, x_n)\big) \geq \theta\big) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}.$$

Proof. We define the following random tensors $\mathcal{Y}_i$ for $0 \leq i \leq n$ as

$$\mathcal{Y}_i \overset{\text{def}}{=} \mathbb{E}\big(F(x_1, \ldots, x_n) \mid X_1, \ldots, X_i\big) = \mathbb{E}_{X_{i+1}}\mathbb{E}_{X_{i+2}} \cdots \mathbb{E}_{X_n} F(x_1, \ldots, x_n),$$

where $\mathbb{E}_{X_{i+1}}$ is the expectation with respect to the random variable X_{i+1}. The constructed sequence $\mathcal{Y}_i$ forms a martingale. The associated difference sequence with respect to $\mathcal{Y}$, denoted as $\{\mathcal{Z}_i\}$, can be stated as

$$\mathcal{Z}_i \overset{\text{def}}{=} \mathcal{Y}_i - \mathcal{Y}_{i-1} = \mathbb{E}_{X_{i+1}}\mathbb{E}_{X_{i+2}} \cdots \mathbb{E}_{X_n}\big(F(x_1, \ldots, x_n) - \mathbb{E}_{X_i} F(x_1, \ldots, x_n)\big).$$

Because $(x_1, \ldots, x_i)$ forms a filtration with respect to i, we have

$$\mathbb{E}_{X_{i-1}}\mathcal{Y}_i = \mathbb{E}_{X_{i-1}}\big(\mathbb{E}_{X_{i+1}}\mathbb{E}_{X_{i+2}} \cdots \mathbb{E}_{X_n} F(x_1, \ldots, x_n) \mid X_{i-1}\big)$$
$$= \mathbb{E}_{X_{i-1}}\big(\mathbb{E}_{X_i}\mathbb{E}_{X_{i+1}} \cdots \mathbb{E}_{X_n} F(x_1, \ldots, x_n) \mid X_{i-1}\big) = \mathbb{E}_{X_{i-1}}\mathcal{Y}_{i-1},$$

then

$$(7.4) \qquad \mathbb{E}_{X_{i-1}}\mathcal{Z}_i = \mathbb{E}_{X_{i-1}}\mathcal{Y}_i - \mathbb{E}_{X_{i-1}}\mathcal{Y}_{i-1} = \mathcal{O}.$$

Let X_i' be an independent copy of X_i, and construct the random vector $\mathbf{x}' = (X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)$ and the random vector $\mathbf{x} = (X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n)$. Since $\mathbb{E}_{X_i} F(\mathbf{x}) = \mathbb{E}_{X_i'} F(\mathbf{x}')$, we have

$$\mathcal{Z}_i = \mathbb{E}_{X_{i+1}}\mathbb{E}_{X_{i+2}} \cdots \mathbb{E}_{X_n}\mathbb{E}_{X_i'}(F(\mathbf{x}) - F(\mathbf{x}')).$$

Then $(F(\mathbf{x}) - F(\mathbf{x}'))^2 \preceq \mathcal{A}_i^2$ from requirement provided by (7.3). We have the following upper bound

(7.5) $$\mathbb{E}_{X_{i+1}}\mathbb{E}_{X_{i+2}} \cdots \mathbb{E}_{X_n}\mathbb{E}_{X_i'}(F(\mathbf{x}) - F(\mathbf{x}'))^2 \preceq \mathcal{A}_i^2.$$

Therefore, from conditions provided by (7.4) and (7.5), this theorem is proved by applying Corollary 7.4 to the martingale $\{\mathcal{Y}_i\}$. $\qquad\square$

8. Conclusion

In this paper, we generalize Laplace transform method and Lieb's concavity theorem from matrices to tensors, and apply these techniques to extend the following classical bounds from the scaler to the tensor situation: Chernoff, Azuma, Hoeffding, Bennett, Bernstein, and McDiarmid. The purpose of these probability inequalities tries to identify large-deviation behavior of the extreme eigenvalue of the sums of random tensors. Tail bounds for the norm of a sum of random rectangular tensors follow as an immediate corollary. Finally, we also apply the proof techniques invented at this work to study tensor-valued martingales.

References

[1] R. Ahlswede and A. Winter, *Addendum to: "Strong converse for identification via quantum channels" [IEEE Trans. Inform. Theory **48** (2002), no. 3, 569–579; MR1889969]*, IEEE Trans. Inform. Theory **49** (2003), no. 1, 346.

[2] A. L. F. de Almeida, G. Favier and J. C. M. Mota, *Constrained tensor modeling approach to blind multiple-antenna CDMA schemes*, IEEE Trans. Signal Process. **56** (2008), no. 6, 2417–2428.

[3] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade and M. Telgarsky, *Tensor decompositions for learning latent variable models (a survey for ALT)*, in: *Algorithmic Learning Theory, 19–38*, Lecture Notes in Comput. Sci. **9355**, Lecture Notes in Artificial Intelligence, Springer, Cham, 2015.

[4] O. Bohigas, M.-J. Giannoni and C. Schmit, *Characterization of chaotic quantum spectra and universality of level fluctuation laws*, Phys. Rev. Lett. **52** (1984), no. 1, 1–4.

[5] M. G. Bulmer, *Principles of Statistics*, Corrected reprint of the 1967 second edition, Dover Publications, New York, 1979.

[6] S. Y. Chang, *Tensor multivariate trace inequalities and their applications*, available at arXiv:2010.02152.

[7] P. Chansangiam, *A survey on operator monotonicity, operator convexity, and operator means*, Int. J. Anal. **2015**, Art. ID 649839, 8 pp.

[8] Y. Chen, D. Han and L. Qi, *New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions*, IEEE Trans. Signal Process. **59** (2011), no. 12, 5888–5898.

[9] L.-B. Cui, C. Chen, W. Li and M. K. Ng, *An eigenvalue problem for even order tensors with its applications*, Linear Multilinear Algebra **64** (2016), no. 4, 602–621.

[10] W. Ding, L. Qi and Y. Wei, *Fast Hankel tensor-vector product and its application to exponential data fitting*, Numer. Linear Algebra Appl. **22** (2015), no. 5, 814–832.

[11] E. G. Effros, *A matrix convexity approach to some celebrated quantum inequalities*, Proc. Natl. Acad. Sci. USA **106** (2009), no. 4, 1006–1008.

[12] X. Fu, K. Huang, W.-K. Ma, N. D. Sidiropoulos and R. Bro, *Joint tensor factorization and outlying slab suppression with applications*, IEEE Trans. Signal Process. **63** (2015), no. 23, 6315–6328.

[13] H. H. Goldstine and J. von Neumann, *Numerical inverting of matrices of high order II*, Proc. Amer. Math. Soc. **2** (1951), 188–202.

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

[15] F. Hansen and G. K. Pedersen, *Jensen's inequality for operators and Löwner's theorem*, Math. Ann. **258** (1981/82), no. 3, 229–241.

[16] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.

[17] T.-X. Jiang, M. K. Ng, X.-L. Zhao and T.-Z. Huang, *Framelet representation of tensor nuclear norm for third-order tensor completion*, IEEE Trans. Image Process. **29** (2020), 7233–7244.

[18] T. Jiang and J. Xie, *Limiting behavior of largest entry of random tensor constructed by high-dimensional data*, J. Theoret. Probab. **33** (2020), no. 4, 2380–2400.

[19] I. R. Klebanov and G. Tarnopolsky, *Uncolored random tensors, melon diagrams, and the Sachdev–Ye–Kitaev models*, Phys. Rev. D **95** (2017), no. 4, 046004, 13 pp.

[20] C.-Y. Ko, K. Batselier, L Daniel, W. Yu and N. Wong, *Fast and accurate tensor completion with total variation regularized tensor trains*, IEEE Trans. Image Process. **29** (2020), 6918–6931.

[21] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev. **51** (2009), no. 3, 455–500.

[22] V. Lahoche and D. O. Samary, *Reliability of the local truncations for the random tensor models renormalization group flow*, Phys. Rev. D **102** (2020), no. 5, 056002, 30 pp.

[23] M. Liang and B. Zheng, *Further results on Moore–Penrose inverses of tensors with application to tensor nearness problems*, Comput. Math. Appl. **77** (2019), no. 5, 1282–1293.

[24] E. H. Lieb, *Convex trace functions and the Wigner–Yanase–Dyson conjecture*, Advances in Math. **11** (1973), no. 3, 267–288.

[25] S. Mirsamadi and J. H. L. Hansen, *A generalized nonnegative tensor factorization approach for distant speech recognition with distributed microphones*, IEEE/ACM Trans. Audio, Speech, Language Process. **24** (2016), no. 10, 1721–1731.

[26] D. Muti and S. Bourennane, *Survey on tensor signal algebraic filtering*, Signal Process. **87** (2007), no. 2, 237–249.

[27] D. Nion and N. D. Sidiropoulos, *Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar*, IEEE Trans. Signal Process. **58** (2010), no. 11, 5693–5705.

[28] L. Qi, H. Chen and Y. Chen, *Tensor eigenvalues and their applications*, Advances in Mechanics and Mathematics **39**, Springer, Singapore, 2018.

[29] L. Qi and Z. Luo, *Tensor Analysis: Spectral theory and special tensors*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

[30] Y. Shen, B. Baingana and G. B. Giannakis, *Tensor decompositions for identifying directed graph topologies and tracking dynamic networks*, IEEE Trans. Signal Process. **65** (2017), no. 14, 3675–3687.

[31] Y. Shen, X. Fu, G. B. Giannakis and N. D. Sidiropoulos, *Topology identification of directed graphs via joint diagonalization of correlation matrices*, IEEE Trans. Signal Inform. Process. Netw. **6** (2020), 271–283.

[32] N. D. Sidiropoulos, R. Bro and G. B. Giannakis, *Parallel factor analysis in sensor array processing*, IEEE Trans. Signal Process. **48** (2000), no. 8, 2377–2388.

[33] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Trans. Signal Process. **65** (2017), no. 13, 3551–3582.

[34] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. **12** (2012), no. 4, 389–434.

[35] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Foundations and Trends, Boston, 2004.

[36] S. J. Turnovsky, *The stability properties of optimal economic policies*, Am. Econ. Rev. **64** (1974), no. 1, 136–148.

[37] R. Vershynin, *Concentration inequalities for random tensors*, Bernoulli **26** (2020), no. 4, 3139–3162.

[38] G. Wainrib and J. Touboul, *Topological and dynamical complexity of random neural networks*, Phys. Rev. Lett. **110** (2013), no. 11, 118101, 4 pp.

[39] X. Wang, M. Che and Y. Wei, *Neural networks based approach solving multi-linear systems with $\mathcal{M}$-tensors*, Neurocomputing **351** (2019), 33–42.

[40] Q. Wu, L. Zhang and G. Shi, *Robust multifactor speech feature extraction based on Gabor analysis*, IEEE Trans. Audio, Speech, Lang. Process. **19** (2011), no. 4, 927–936.

[41] H.-R. Xu, D.-H. Li and S.-L. Xie, *An equivalent tensor equation to the tensor complementarity problem with positive semi-definite Z-tensor*, Optim. Lett. **13** (2019), no. 4, 685–694.

Shih Yu Chang
Department of Applied Data Science, San Jose State University, San Jose, CA, USA
E-mail address: `shihyu.chang@sjsu.edu`

Wen-Wei Lin
Department of Applied Mathematics, National Yang Ming Chiao Tung University,
Hsinchu, Taiwan
E-mail address: `wwlin@math.nctu.edu.tw`