Tail Bounds for Random Tensors and Their Applications

Shih Yu Chang

San Jose State University

Mar. 15, 2022

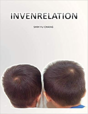
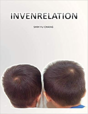


Table of Contents



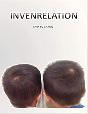
Tail Bounds for Random Variables

In probability theory, tail bounds (concentration inequalities) provide bounds on how a random variable deviates from some value, e.g., its expected value. The law of large numbers of classical probability theory states that sums of independent random variables are, close to their expectation with a large probability. Common tail bounds are:

Theorem 1 (Bernstein Bound for RVs)

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables taking values $+1$ and -1 with probability 1/2 (this distribution is also known as the Rademacher distribution), and θ be a given positive real number, then we have

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \theta\right) \leq 2\exp\left(-\frac{n\theta^2}{2 + 2\theta/3}\right). \tag{1}$$



Theorem 2 (Chernoff Bound for RVs)

Let $X_1, X_2, \ldots, X_n$ *be independent random variables taking values* $\{0, 1\}$ *with* $X = \sum_{i=1}^{n} X_i$, *and* $\mu = \mathbb{E}X$, *then we have*

$$\Pr\left(X > (1+\delta)\mu\right) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}, \tag{2}$$

where $\delta > 0$.

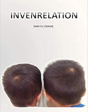
How about tail bounds for random matrices? We have to summarize a random matrix by a scaler value, e.g., the maximum eigenvalue, matrix norm, etc, before comparing to a real number. For example, we have following theorem about Matrix Chernoff bound, it is [Tro12]

Theorem 3 (Chernoff Bound for Random Matrices)

Consider a finite sequence $\boldsymbol{X}_i$ *of independent, random, Hermitian matrices with dimension* m. *Suppose we have* $\boldsymbol{X}_i \geq \boldsymbol{0}$ *and* $\lambda_{\max}(\boldsymbol{X}_i) \leq R$ *almost surely, we then have*

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \boldsymbol{X}_i\right) \geq (1+\delta)\mu_{\max}\right) \leq m\left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\frac{\mu_{\max}}{R}}, \tag{3}$$

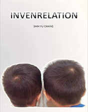
where $\mu_{\max} = \lambda_{\max}\left(\sum_{i=1}^{n} \mathbb{E}\boldsymbol{X}_i\right)$.



Applications of Tail Bounds

Tail bounds for random variables or random matrices have already found a place in many areas of the mathematical sciences, including [Tro19]

- numerical linear algebra
- combinatorics
- algorithms analysis
- optimization
- quantum information
- statistics
- signal processing
- machine learning
- uncertainty quantification



Why Tail Bounds for Random Tensors

- In recent years, tensors have been applied to deal with multirelational data in science and engineering which is crucial in current Big Data era.
- Very few works about tail bounds for random tensors.
- Unlike scalers or matrices, there are different ways to define the product between two tensors. We will discuss tensors under Einstein product and T-product.

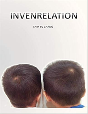
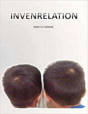


Table of Contents



Tensor Basic Facts Under Einstein Product

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

$$(\mathcal{X} \star_N \mathcal{Y})_{i_1,\cdots,i_M,k_1,\cdots,k_L} \stackrel{\text{def}}{=} \sum_{j_1,\cdots,j_N} a_{i_1,\cdots,i_M,j_1,\cdots,j_N} b_{j_1,\cdots,j_N,k_1,\cdots,k_L}. \quad (4)$$

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

$$\mathrm{Tr}(\mathcal{X}) \stackrel{\text{def}}{=} \sum_{1 \leq i_j \leq I_j,\, j \in [M]} \mathcal{X}_{i_1,\cdots,i_M,i_1,\cdots,i_M}. \quad (5)$$

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

$$\langle \mathcal{X}, \mathcal{Y} \rangle \stackrel{\text{def}}{=} \mathrm{Tr}\left(\mathcal{X}^H \star_M \mathcal{Y}\right). \quad (6)$$

Tensor Functions

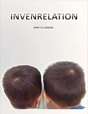
Given a function $g : \mathbb{R} \rightarrow \mathbb{R}$, the mapping result of a diagonal tensor by the function g is to obtain another same size diagonal tensor with diagonal entry mapped by the function g. Then, the function g can be extended to allow a Hermitian tensor $\mathcal{X} \in \mathbb{C}^{I_1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

$$g(\mathcal{X}) \stackrel{\text{def}}{=} \mathcal{U} \star_M g(\Lambda) \star_M \mathcal{U}^H, \quad \text{where } \mathcal{X} = \mathcal{U} \star_M \Lambda \star_M \mathcal{U}^H. \tag{7}$$

The *spectral mapping theorem* asserts that each eigenvale 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

$$f(x) \geq g(x), \text{for } x \in [a, b] \Rightarrow f(\mathcal{X}) \geq g(\mathcal{X}), \text{for eigenvalues of } \mathcal{X} \in [a, b]; \tag{8}$$

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



Tensor Exponential Function

Given a function $g : \mathbb{R} \rightarrow \mathbb{R}$, 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

$$g(\mathcal{X}) \stackrel{\text{def}}{=} \mathcal{U} \star_M g(\Lambda) \star_M \mathcal{U}^H, \quad \text{where } \mathcal{X} = \mathcal{U} \star_M \Lambda \star_M \mathcal{U}^H. \tag{9}$$

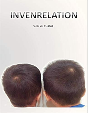
Definition 4

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

$$e^{\mathcal{X}} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{\mathcal{X}^k}{k!}, \tag{10}$$

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 $e^{\mathcal{X}} = \mathcal{Y}$



Tensor Moments and Cumulant

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

$$\mathbb{M}_{\mathcal{X}}(t) \stackrel{\text{def}}{=} \mathbb{E}e^{t\mathcal{X}}, \quad \text{and} \quad \mathbb{K}_{\mathcal{X}}(t) \stackrel{\text{def}}{=} \log \mathbb{E}e^{t\mathcal{X}}, \tag{11}$$

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

$$\mathbb{M}_{\mathcal{X}}(t) = \mathcal{I} + \sum_{n=1}^{\infty} \frac{t^n}{n!}\mathbb{E}(\mathcal{X}^n), \quad \text{and} \quad \mathbb{K}_{\mathcal{X}}(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!}\psi_n, \tag{12}$$

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

$$\psi_1 = \mathbb{E}(\mathcal{X}), \quad \text{and} \quad \psi_2 = \mathbb{E}(\mathcal{X}^2) - (\mathbb{E}(\mathcal{X}))^2. \tag{13}$$

T-product Tensors, I

For a third order tensor $\mathcal{C} \in \mathbb{C}^{m \times n \times p}$, we define bcirc operation to the tensor $\mathcal{C}$ as:

$$\mathrm{bcirc}(\mathcal{C}) \stackrel{\mathrm{def}}{=} \begin{pmatrix} \mathbf{C}^{(1)} & \mathbf{C}^{(p)} & \mathbf{C}^{(p-1)} & \cdots & \mathbf{C}^{(2)} \\ \mathbf{C}^{(2)} & \mathbf{C}^{(1)} & \mathbf{C}^{(p)} & \cdots & \mathbf{C}^{(3)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{C}^{(p)} & \mathbf{C}^{(p-1)} & \mathbf{C}^{(p-2)} & \cdots & \mathbf{C}^{(1)} \end{pmatrix}, \quad (14)$$

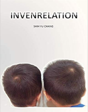
where $\mathbf{C}^{(1)}, \cdots, \mathbf{C}^{(p)} \in \mathbb{C}^{m \times n}$ are frontal slices of tensor $\mathcal{C}$. The inverse operation of bcirc is denoted as bcirc^{-1} with relation $\mathrm{bcirc}^{-1}(\mathrm{bcirc}(\mathcal{C})) \stackrel{\mathrm{def}}{=} \mathcal{C}$. For a third order tensor $\mathcal{C} \in \mathbb{C}^{m \times m \times p}$, we define Hermitian transpose of $\mathcal{C}$, denoted by $\mathcal{C}^{\mathrm{H}}$, as

$$\mathcal{C}^{\mathrm{H}} = \mathrm{bcirc}^{-1}((\mathrm{bcirc}(\mathcal{C}))^{\mathrm{H}}). \quad (15)$$

And a tensor $\mathcal{D} \in \mathbb{C}^{m \times m \times p}$ is called a Hermitian T-product tensor if $\mathcal{D}^{\mathrm{H}} = \mathcal{D}$. Similarly, we also define transpose of $\mathcal{C}$, denoted by $\mathcal{C}^{\mathrm{T}}$, as

$$\mathcal{C}^{\mathrm{T}} = \mathrm{bcirc}^{-1}((\mathrm{bcirc}(\mathcal{C}))^{\mathrm{T}}). \quad (16)$$

And a tensor $\mathcal{D} \in \mathbb{C}^{m \times m \times p}$ is called a Symmetric T-product tensor if $\mathcal{D}^{\mathrm{T}} = \mathcal{D}$.



T-product Tensors, II

The identity tensor $\mathcal{I}_{mmp} \in \mathbb{C}^{m \times m \times p}$ can be defined as:

$$\mathcal{I}_{mmp} = \text{bcirc}^{-1}(\mathbf{I}_{mp}), \tag{17}$$

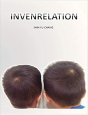
where $\mathbf{I}_{mp}$ is the identity matrix in $\mathbb{R}^{mp \times mp}$. A zero tensor, denoted as $\mathcal{O}_{mnp} \in \mathbb{C}^{m \times n \times p}$, is a tensor that all elements inside the tensor as 0.
In order to define the T-product operation, we need to define another kind of operation over a third order tensor. For a third order tensor $\mathcal{C} \in \mathbb{C}^{m \times n \times p}$, we define unfold operation to the tensor $\mathcal{C}$ as:

$$\text{unfold}(\mathcal{C}) \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{C}^{(1)} \\ \mathbf{C}^{(2)} \\ \vdots \\ \mathbf{C}^{(p)} \end{pmatrix}, \tag{18}$$

where $\text{unfold}(\mathcal{C}) \in \mathbb{C}^{mp \times n}$, and the inverse operation of unfold is fold with the relation $\text{fold}(\text{unfold}(\mathcal{C})) \stackrel{\text{def}}{=} \mathcal{C}$. Given $\mathcal{C} \in \mathbb{C}^{m \times n \times p}$ and $\mathcal{D} \in \mathbb{C}^{n \times k \times p}$, we define the T-product between $\mathcal{C}$ and $\mathcal{D}$ as

$$\mathcal{C} \star \mathcal{D} \stackrel{\text{def}}{=} \text{fold}(\text{bcirc}(\mathcal{C})\text{unfold}(\mathcal{D})), \tag{19}$$

where $\mathcal{C} \star \mathcal{D} \in \mathbb{C}^{m \times k \times p}$.



T-product Tensors, III

We define the T-product tensor *trace* for a tensor $\mathcal{C} = (c_{ijk}) \in \mathbb{C}^{m \times m \times p}$, denoted by $\mathrm{Tr}(\mathcal{C})$, as following

$$\mathrm{Tr}(\mathcal{C}) \overset{\text{def}}{=} \sum_{i=1}^{m} \sum_{k=1}^{p} c_{iik}, \tag{20}$$

which is the summation of all entries in f-diagonal components.

Given a tensor $\mathcal{C} \in \mathbb{R}^{m \times n \times p}$, Theorem 4.1 in [KM11] proposed a T-singular value decomposition (T-SVD) for $\mathcal{C}$ as:

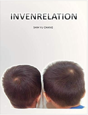
$$\mathcal{C} = \mathcal{U} \star \mathcal{S} \star \mathcal{V}^{\mathrm{T}}, \tag{21}$$

where $\mathcal{U} \in \mathbb{C}^{m \times m \times p}$ and $\mathcal{V} \in \mathbb{C}^{n \times n \times p}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{C}^{m \times n \times p}$ is a f-diagonal tensor. We also have $\mathcal{U}^{\mathrm{T}} \star \mathcal{U} = \mathcal{I}_{mmp}$ and $\mathcal{V}^{\mathrm{T}} \star \mathcal{V} = \mathcal{I}_{nnp}$. We define $\sigma(\mathcal{C})$ be the spectrum of $\mathcal{C}$, i.e., the set of $s \in \mathbb{C}$, where s are nonzero entries from tensor $\mathcal{S}$. We use $\|\cdot\|$ for the spectral norm, which is the largest singular value of a T-product tensor.

Given any integer k and $\mathcal{B} \in \mathbb{C}^{m \times m \times p}$, we define $\mathcal{B}^k$ as

$$\mathcal{B}^k \overset{\text{def}}{=} \overbrace{\mathcal{B} \star \mathcal{B} \star \mathcal{B} \star \cdots \star \mathcal{B}}^{k \text{ terms of } \mathcal{B} \text{ under T-square}} \tag{22}$$

where $\mathcal{B}^k \in \mathbb{C}^{m \times m \times p}$.



T-product Tensors, IV

Definition 5

Given a tensor $\mathcal{X} \in \mathbb{C}^{m \times m \times p}$, the *tensor exponential* of the tensor $\mathcal{X}$ is defined as

$$e^{\mathcal{X}} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{\mathcal{X}^k}{k!}, \tag{23}$$

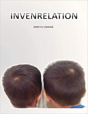
where $\mathcal{X}^0$ is defined as the identity tensor $\mathcal{I}_{mmp}$. Given a tensor $\mathcal{Y}$, the tensor $\mathcal{X}$ is said to be a *tensor logarithm* of $\mathcal{Y}$ if $e^{\mathcal{X}} = \mathcal{Y}$.

From T-SVD in Eq. (21), we can express a Hermitian T-product tensor $\mathcal{C} \in \mathbb{R}^{m \times m \times p}$ as

$$\mathcal{C} = \sum_{i=1}^{m} \sum_{k=0}^{p-1} s_{iik} \mathbf{U}_i^{[k]} \star \left(\mathbf{U}_i^{[k]} \right)^{\mathrm{T}}, \tag{24}$$

where s_{iik} are *eigenvalues* of the tensor $\mathcal{C}$, and $\mathbf{U}_i^{[k]} \in \mathbb{C}^{m \times 1 \times p}$ is the i-th lateral slice (matrix) of the tensor $\mathcal{U}$ after k cyclic permutations. The matrix $\mathbf{U}_i^{[0]}$ is obtained from the i-th lateral slice (matrix) of the tensor $\mathcal{U}$ with column vectors as $\mathbf{u}_i^{(1)}, \cdots, \mathbf{u}_i^{(p)}$, then we have

$$\mathbf{U}_i^{[k]} = \left(\mathbf{u}_i^{(p+1-k) \bmod p}, \mathbf{u}_i^{(p+2-k) \bmod p}, \cdots, \mathbf{u}_i^{(p)}, \mathbf{u}_i^{(1)}, \cdots \mathbf{u}_i^{(p-k)} \right). \tag{25}$$



T-product Tensors, V

Note that we have $\left(\mathbf{U}_i^{[k]}\right)^{\mathrm{H}} \star \mathbf{U}_i^{[k]} = \mathcal{I}_{11p}$ and $\left(\mathbf{U}_i^{[k]}\right)^{\mathrm{H}} \star \mathbf{U}_{i'}^{[k']} = \mathcal{O}_{11p}$ for $i \neq i'$ or $k \neq k'$. All values of s_{iik} are real and we define $\lambda_{\max} \stackrel{\text{def}}{=} \max_{\substack{1 \leq i \leq m \\ 0 \leq k \leq p-1}} \{s_{iik}\}$, and $\lambda_{\min} \stackrel{\text{def}}{=} \min_{\substack{1 \leq i \leq m \\ 0 \leq k \leq p-1}} \{s_{iik}\}$.

Given a Hermitian T-product tensor $\mathcal{C} \in \mathbb{C}^{m \times m \times p}$, and a tensor $\mathcal{X} \in \mathbb{C}^{m \times 1 \times p}$ obtained from treating the matrix $\mathbf{X} \in \mathbb{C}^{m \times p}$ as a tensor with dimensions $\mathbb{R}^{m \times 1 \times p}$. We define following quadratic form with respect to the matrix $\mathbf{X}$ as

$$F_{\mathcal{C}}(\mathbf{X}) \stackrel{\text{def}}{=} \mathcal{X}^{\mathrm{T}} \star \mathcal{C} \star \mathcal{X}, \tag{26}$$

and we say that a tensor $\mathcal{C}$ is T-positive definite (TPD) (or T-positive semi-definite (TPSD)) if $F_{\mathcal{C}}(\mathbf{X}) > \mathbf{0}$ (or $F_{\mathcal{C}}(\mathbf{X}) \geq \mathbf{0}$) for any $\mathbf{X} \in \mathbb{C}^{m \times p}$, where $\mathbf{0}$ is a zero vector with size p.

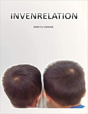
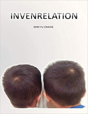


Table of Contents



Laplace Transform Method for Tensors

Lemma 6 (Laplace Transform Method for Tensors)

Let $\mathcal{X}$ be a random Hermitian tensor. For $\theta \in \mathbb{R}$, we have

$$\mathbb{P}(\lambda_{\max}(\mathcal{X}) \geq \theta) \leq \inf_{t>0} \left\{ e^{-\theta t} \mathbb{E} \mathrm{Tr} e^{t\mathcal{X}} \right\} \tag{27}$$

Proof: Given a fix value t, we have

$$\begin{aligned} \mathbb{P}(\lambda_{\max}(\mathcal{X}) \geq \theta) = \mathbb{P}(\lambda_{\max}(t\mathcal{X}) \geq t\theta) \\ = \mathbb{P}(e^{\lambda_{\max}(t\mathcal{X})} \geq e^{t\theta}) \leq e^{-t\theta} \mathbb{E} e^{\lambda_{\max}(t\mathcal{X})}. \end{aligned} \tag{28}$$

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

$$e^{\lambda_{\max}(t\mathcal{Y})} = \lambda_{\max}(e^{t\mathcal{Y}}) \leq \mathrm{Tr} e^{t\mathcal{Y}}, \tag{29}$$

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 [LZ19]. From Eqs (28) and (29), this lemma is established. □

Relative Entropy Between Tensors

Definition 7

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

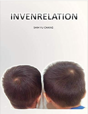
$$D(\mathcal{A} \parallel \mathcal{B}) \stackrel{\text{def}}{=} \mathrm{Tr}\mathcal{A} \star_M (\log \mathcal{A} - \log \mathcal{B}). \tag{30}$$

Lemma 8 (Joint Convexity of Relative Entropy for Tensors)

The relative entropy function of two positive-definite tensors is a jointly convex function. That is

$$\begin{aligned} \mathbb{D}(t\mathcal{A}_1 + (1-t)\mathcal{A}_2 \parallel t\mathcal{B}_1 + (1-t)\mathcal{B}_2) \leq \\ t\mathbb{D}(\mathcal{A}_1 \parallel \mathcal{B}_1) + (1-t)\mathbb{D}(\mathcal{A}_2 \parallel \mathcal{B}_2), \end{aligned} \tag{31}$$

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.*



Theorem 9 (Lieb's concavity theorem for tensors)

Let $\mathcal{H}$ be a Hermitian tensor. Following map

$$\mathcal{A} \to \mathrm{Tr} e^{\mathcal{H}+\log \mathcal{A}} \tag{32}$$

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

$$\mathrm{Tr}\mathcal{Y} \geq \mathrm{Tr}\mathcal{X} - \mathrm{Tr}\mathcal{X} \log \mathcal{X} + \mathrm{Tr}\mathcal{X} \log \mathcal{Y}. \tag{33}$$

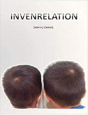
If we replace $\mathcal{Y}$ by $e^{\mathcal{H}+\log \mathcal{A}}$, we then have

$$\mathrm{Tr} e^{\mathcal{H}+\log \mathcal{A}} = \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\mathcal{X} \star \mathcal{H} - \mathbb{D}(\mathcal{X} \parallel \mathcal{A}) + \mathrm{Tr}\mathcal{X} \right\} \tag{34}$$

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

$$\begin{aligned}
\mathrm{Tr} e^{\mathcal{H}+\log(t\mathcal{A}_1+(1-t)\mathcal{A}_2)} &= \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\mathcal{X}\mathcal{H} - \mathbb{D}(\mathcal{X} \parallel t\mathcal{A}_1 + (1-t)\mathcal{A}_2) + \mathrm{Tr}\mathcal{X} \right\} \\
&\geq t \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\mathcal{X}\mathcal{H} - \mathbb{D}(\mathcal{X} \parallel t\mathcal{A}_1) + \mathrm{Tr}\mathcal{X} \right\} \\
&\quad +(1-t) \max_{\mathcal{X} \succ \mathcal{O}} \left\{ \mathrm{Tr}\mathcal{X}\mathcal{H} - \mathbb{D}(\mathcal{X} \parallel (1-t)\mathcal{A}_2) + \mathrm{Tr}\mathcal{X} \right\} \\
&= t\mathrm{Tr} e^{\mathcal{H}+\log \mathcal{A}_1} + (1-t)\mathrm{Tr} e^{\mathcal{H}+\log \mathcal{A}_2},
\end{aligned} \tag{35}$$

where the first and last equalities are obtained based on the variational formula provided by Eq. (34), and the inequality is due to the joint convexity property of the relative entropy from Leamm 8.



Lemma 10 (Subadditivity of Tensor CGFs)

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

$$\mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) \leq \mathrm{Tr}\exp\left(\sum_{i}^{n} \log \mathbb{E}e^{t\mathcal{X}_i}\right), \quad \textit{for } t \in \mathbb{R}. \tag{36}$$

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

$$\mathbb{K}_i(t) \stackrel{\text{def}}{=} \log(\mathbb{E}e^{t\mathcal{X}_i}). \tag{37}$$

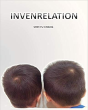
Then, we define the Hermitian tensor $\mathcal{H}_k$ as

$$\mathcal{H}_k(t) = \sum_{i=1}^{k-1} t\mathcal{X}_k + \sum_{i=k+1}^{n} \mathbb{K}_i(t). \tag{38}$$

By applying Eq. (38) to Theorem 9 repeatedly for $k = 1, 2, \cdots, n$, we have

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) &=_1 \mathbb{E}_0 \cdots \mathbb{E}_{n-1}\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + t\mathcal{X}_n\right) \\
&\leq \mathbb{E}_0 \cdots \mathbb{E}_{n-2}\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + \log\left(\mathbb{E}_{n-1}e^{t\mathcal{X}_n}\right)\right) \\
&= \mathbb{E}_0 \cdots \mathbb{E}_{n-2}\mathrm{Tr}\exp\left(\sum_{i=1}^{n-2} t\mathcal{X}_i + t\mathcal{X}_{n-1} + \mathbb{K}_n(t).\right) \\
&\leq \mathbb{E}_0 \cdots \mathbb{E}_{n-3}\mathrm{Tr}\exp\left(\sum_{i=1}^{n-2} t\mathcal{X}_i + \mathbb{K}_{n-1}(t) + \mathbb{K}_n(t)\right) \\
\cdots &\leq \mathrm{Tr}\exp\left(\sum_{i=1}^{n} \mathbb{K}_i(t)\right)
\end{aligned} \tag{39}$$

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

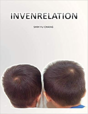


Theorem 11 (Master Tail Bound for Independent Sum of Random Tensors)

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

$$\Pr\left(\lambda_{\max}\Big(\sum_{i=1}^{n}\mathcal{X}_i\Big) \geq \theta\right) \leq \inf_{t>0}\left\{e^{-t\theta}\mathrm{Tr}\exp\left(\sum_{i=1}^{n}\log \mathbb{E}e^{t\mathcal{X}_i}\right)\right\}. \quad (40)$$

Proof: By substituing the Lemma 10 into the Laplace transform bound provided by the Lemma 6, this theorem is established. □



Corollaries for Master Tail Bound for Independent Sum of Random Tensors

Corollary 12

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 $f : (0, \infty) \rightarrow [0, \infty]$ and a sequence of non-random Hermitian tensors $\{\mathcal{A}_i\}$ with following condition:

$$f(t)\mathcal{A}_i \succeq \log \mathbb{E} e^{t\mathcal{X}_i}, \quad \text{for } t > 0. \tag{41}$$

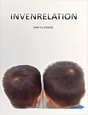
Then, for all $\theta \in \mathbb{R}$, we have

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \geq \theta\right) \quad \leq \quad \mathbb{I}_1^M \inf_{t>0} \left\{ \exp\left[-t\theta + f(\theta)\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{A}_i\right)\right]\right\} \tag{42}$$

Corollary 13

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

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \geq \theta\right) \quad \leq \quad \mathbb{I}_1^M \inf_{t>0} \left\{ \exp\left[-t\theta + n \log \lambda_{\max}\left(\frac{\sum_{i=1}^{n} \mathbb{E} e^{t\mathcal{X}_i}}{n}\right)\right]\right\} \tag{43}$$



Tensor with Gaussian and Rademacher Random Series, Square Tensor

Lemma 14

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

$$\mathbb{E}e^{\alpha t \mathcal{A}} = e^{t^2 \mathcal{A}^2/2} \text{ and } e^{t^2 \mathcal{A}^2/2} \succeq \mathbb{E}e^{\beta t \mathcal{A}}, \tag{44}$$

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

Theorem 15 (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 indepedent normal variables. We define

$$\sigma^2 \stackrel{\text{def}}{=} \left\| \sum_i^n \mathcal{A}_i^2 \right\|, \tag{45}$$

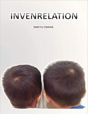
then, for all $\theta \geq 0$, we have

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^n \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}, \tag{46}$$

and

$$\mathrm{Pr}\left(\left\| \sum_{i=1}^n \alpha_i \mathcal{A}_i \right\| \geq \theta \right) \leq 2\mathbb{I}_1^M e^{-\frac{\theta^2}{2\sigma^2}}. \tag{47}$$

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



Corollary 16 (Rectangular Tensor with Gaussian and Rademacher Series)

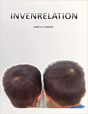
Given a finite sequence $\mathcal{A}_i$ *of fixed Hermitian tensors with dimensions as* $\mathbb{C}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_M}$*, and let* $\{\alpha_i\}$ *be a finite sequence of indepedent normal variables. We define*

$$\sigma^2 \stackrel{def}{=} \max \left\{ \left\| \sum_{i=1}^{n} \mathcal{A}_i \star_M \mathcal{A}_i^H \right\|, \left\| \sum_{i=1}^{n} \mathcal{A}_i^H \star_M \mathcal{A}_i \right\| \right\}. \tag{48}$$

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

$$\Pr \left(\left\| \sum_{i=1}^{n} \alpha_i \mathcal{A}_i \right\| \geq \theta \right) \leq \prod_{m=1}^{M} (I_m + J_m) e^{-\frac{\theta^2}{2\sigma^2}}. \tag{49}$$

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



Tensor Chernoff Bounds

Lemma 17

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

$$\mathcal{I} + (e^t - 1)\mathbb{E}\mathcal{X} \succeq \mathbb{E}e^{t\mathcal{X}}. \quad (50)$$

Theorem 18 (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

$$\mathcal{X}_i \succeq \mathcal{O} \quad \text{and} \quad \lambda_{\max}(\mathcal{X}_i) \leq 1 \quad \text{almost surely.} \quad (51)$$

Define following two quantaties:

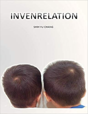
$$\overline{\mu}_{\max} \stackrel{def}{=} \lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i\right) \quad \text{and} \quad \overline{\mu}_{\min} \stackrel{def}{=} \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i\right), \quad (52)$$

then, we have following two inequalities:

$$\Pr\left(\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M e^{-n\mathfrak{D}(\theta || \overline{\mu}_{\max})}, \quad \text{for } \overline{\mu}_{\max} \leq \theta \leq 1; \quad (53)$$

and

$$\Pr\left(\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i\right) \leq \theta\right) \leq \mathbb{I}_1^M e^{-n\mathfrak{D}(\theta || \overline{\mu}_{\min})}, \quad \text{for } 0 \leq \theta \leq \overline{\mu}_{\min}. \quad (54)$$



Proof of Tensor Chernoff Bound I

From Lemma 17, we have

$$\mathcal{I} + f(t)\mathbb{E}\mathcal{X}_i \succeq \mathbb{E}e^{t\mathcal{X}_i}, \tag{55}$$

where $f(t) \stackrel{\text{def}}{=} e^t - 1$ for $t > 0$. By applying Corollary 13, we obtain

$$\begin{aligned}
\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \alpha\right) &\leq \mathbb{I}_1^M \exp\left(-t\alpha + n\log\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}(\mathcal{I} + f(t)\mathbb{E}\mathcal{X}_i)\right)\right) \\
&= \mathbb{I}_1^M \exp\left(-t\alpha + n\log\lambda_{\max}\left(\mathcal{I} + f(t)\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i\right)\right) \\
&= \mathbb{I}_1^M \exp\left(-t\alpha + n\log\left(1 + f(t)\overline{\mu}_{\max}\right)\right). \qquad (56)
\end{aligned}$$

The last equality follows from the definition of $\overline{\mu}_{\max}$ and the eigenvalue map properties. When the value t at the right-hand side of Eq. (56) is

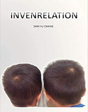
$$t = \log\frac{\alpha}{1-\alpha} - \log\frac{\overline{\mu}_{\max}}{1-\overline{\mu}_{\max}}, \tag{57}$$

we can achieve the tighest upper bound at Eq. (56). By substituing the value t in Eq. (57) into Eq. (56) and change the varaible $\alpha \rightarrow n\theta$, Eq. (53) is proved. The next goal is to prove Eq. (54).

If we apply Lemma 17 to the sequence $\{-\mathcal{X}_i\}$, we have

$$\mathcal{I} - g(t)\mathbb{E}\mathcal{X}_i \succeq \mathbb{E}e^{t(-\mathcal{X}_i)}, \tag{58}$$

where $g(t) \stackrel{\text{def}}{=} 1 - e^t$ for $t > 0$.



Proof of Tensor Chernoff Bound I, cont.

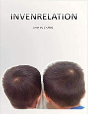
By applying Corollary 13 again, we obtain

$$
\begin{aligned}
\Pr\left(\lambda_{\min}\left(\sum_{i=1}^{n}\mathcal{X}_i\right)\leq\alpha\right) &= \Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}(-\mathcal{X}_i)\right)\geq\alpha\right) \\
&\leq \mathbb{I}_1^M\exp\left(t\alpha+n\log\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}(\mathcal{I}-g(t)\mathbb{E}\mathcal{X}_i)\right)\right) \\
&=_1 \mathbb{I}_1^M\exp\left(t\alpha+n\log\left(1-f(t)\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i\right)\right)\right) \\
&= \mathbb{I}_1^M\exp\left(t\alpha+n\log\left(1-g(t)\overline{\mu}_{\min}\right)\right), \qquad (59)
\end{aligned}
$$

where we apply the relation $\lambda_{\min}(-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i)=-\lambda_{\max}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i)$ at the equality $=_1$. When the value t at the right-hand side of Eq. (59) is

$$
t=log\frac{\overline{\mu}_{\max}}{1-\overline{\mu}_{\max}}-\log\frac{\alpha}{1-\alpha}, \qquad (60)
$$

we can achieve the tighest upper bound at Eq. (59). By substituing the value t in Eq. (60) into Eq. (59) and change the varaible $\alpha\rightarrow n\theta$, Eq. (54) is proved also.



Tensor Bernstein Bounds

Lemma 19

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:*

$$\mathbb{E}\mathcal{X} = 0 \text{ and } \lambda_{\max}(\mathcal{X}) \leq 1 \text{ almost surely.} \tag{61}$$

Then, we have

$$e^{(e^t - t - 1)\mathbb{E}(\mathcal{X}^2)} \succeq \mathbb{E}e^{t\mathcal{X}} \tag{62}$$

where $t > 0$*.*

Theorem 20 (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*

$$\mathbb{E}\mathcal{X}_i = 0 \text{ and } \lambda_{\max}(\mathcal{X}_i) \leq T \text{ almost surely.} \tag{63}$$

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

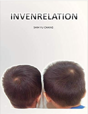
$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-\theta^2/2}{\sigma^2 + T\theta/3}\right); \tag{64}$$

and

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-3\theta^2}{8\sigma^2}\right) \quad \text{for } \theta \leq \sigma^2/T; \tag{65}$$

and

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i\right) \geq \theta\right) \leq \mathbb{I}_1^M \exp\left(\frac{-3\theta}{8T}\right) \quad \text{for } \theta \geq \sigma^2/T. \tag{66}$$



Proof for Bounded $\lambda_{\max}$ Tensor Bernstein Bounds

Without loss of generality, we can assume that $T = 1$ since the summands are 1-homogeneous and the variance is 2-homogeneous. From Lemma 19, we have

$$\mathbb{E}e^{t\mathcal{X}_i} \preceq e^{(e^t - t - 1)\mathbb{E}(\mathcal{X}_i^2)} \quad \text{for } t > 0. \tag{67}$$

By applying Corollary 12, we then have

$$\begin{aligned}
\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) &\leq \mathbb{I}_1^M \exp\left(-t\theta + (e^t - t - 1)\lambda_{\max}\left(\sum_{i=1}^{n}\mathbb{E}\left(\mathcal{X}_i^2\right)\right)\right) \\
&= \mathbb{I}_1^M \exp\left(-t\theta + \sigma^2(e^t - t - 1)\right). \qquad (68)
\end{aligned}$$

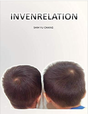
The right-hand side of Eq. (68) can be minimized by setting $t = \log(1 + \theta/\sigma^2)$. Substitute such t and simplify the right-hand side of Eq. (68), we obtain Eq. (135).
For $\theta \leq \sigma^2/T$, we have

$$\frac{1}{\sigma^2 + T\theta/3} \geq \frac{1}{\sigma^2 + T(\sigma^2/T)/3} = \frac{3}{4\sigma^2}, \tag{69}$$

then, we obtain Eq. (136). Correspondingly, for $\theta \geq \sigma^2/T$, we have

$$\frac{\theta}{\sigma^2 + T\theta/3} \geq \frac{\sigma^2/T}{\sigma^2 + T(\sigma^2/T)/3} = \frac{3}{4T}, \tag{70}$$

then, we obtain Eq. (137) also.



Tensor Martingales

Necessary definitions about tensor martingales will be provied 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:

$$\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \cdots \subset \mathfrak{F}_\infty \subset \mathfrak{F}. \tag{71}$$

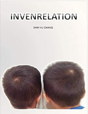
Given such a filtration, we define the conditional expectation $\mathbb{E}_i[\cdot] \stackrel{\text{def}}{=} \mathbb{E}_i[\cdot|\mathfrak{F}_i]$. A sequence $\{\mathcal{Y}_i\}$ of random tensors is called *adapted* to the filtration when each tensor $\mathcal{Y}_i$ is measuable with respec to $\mathfrak{F}_i$. 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

$$\mathbb{E}_{i-1}\mathcal{X}_i = \mathcal{X}_{i-1} \quad \text{and} \quad \mathbb{E}\,\|\mathcal{X}_i\| < \infty, \tag{72}$$

where $i = 1, 2, 3, \cdots$. 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

$$\mathcal{Y}_i \stackrel{\text{def}}{=} \mathcal{X}_i - \mathcal{X}_{i-1} \;\; \text{for } i = 1, 2, 3, \cdots \tag{73}$$

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



Two Lemmas About Tensor Martingales

Lemma 21 (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

$$\mathbb{E}\mathrm{Tr}e^{\mathcal{A}+\mathcal{X}} \leq \mathbb{E}\mathrm{Tr}e^{\mathcal{A}+2\beta\mathcal{X}}, \tag{74}$$

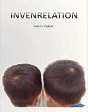
where β is Rademacher random variable.

The other Lemma is to provide the tensor cumulant-generating function of a symetrized random tensor.

Lemma 22 (Cumulant-Generating Function of Symetrized 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

$$\log \mathbb{E}\left[e^{2\beta t\mathcal{X}} | \mathcal{X}\right] \preceq 2t^2\mathcal{A}^2. \tag{75}$$



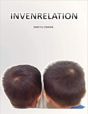
Theorem 23 (Tensor Azuma Inequality)

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*

$$\mathbb{E}_{i-1}\mathcal{X}_i = 0 \;\; \textit{and} \;\; \mathcal{X}_i^2 \preceq \mathcal{A}_i^2 \textit{almost surely}, \tag{76}$$

where $i = 1, 2, 3, \cdots$.
Define the total varaince σ^2 *as:* $\sigma^2 \stackrel{def}{=} \left\| \sum_i^n \mathcal{A}_i^2 \right\|$. *Then, we have following inequalities:*

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^n \mathcal{X}_i \right) \geq \theta \right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}. \tag{77}$$

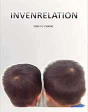


Proof of Tensor Azuma Inequality

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

$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) &= \mathbb{E}\left(\mathbb{E}\left(\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + t\mathcal{X}_n\right)|\mathfrak{F}_n\right)|\mathfrak{F}_{n-1}\right) \\
&\leq \mathbb{E}\left(\mathbb{E}\left(\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + 2\beta t\mathcal{X}_n\right)|\mathfrak{F}_n\right)|\mathfrak{F}_n\right) \\
&\leq \mathbb{E}\left(\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + \log\mathbb{E}\left(e^{2\beta t\mathcal{X}_n}|\mathfrak{F}_n\right)\right)|\mathfrak{F}_n\right) \\
&\leq \mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^{n-1} t\mathcal{X}_i + 2t^2\mathcal{A}_n^2\right), \qquad (78)
\end{aligned}
$$

where the first equality comes from the tower property of conditional expectation; the first inequality comes from Lemma 21, and the relaxation the conition to the larger algebra set $\mathfrak{F}_n$; finally, the last inequality requires Lemma 22.



Proof of Tensor Azuma Inequality, cont.

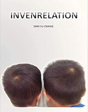
If we continue the iteration procedure based on Eq. (78), we have

$$\mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right) \leq \mathrm{Tr}\exp\left(2t^2\sum_{i=1}^{n}\mathcal{A}_i^2\right), \tag{79}$$

then apply Eq. (79) into Lemma 6, we obtain

$$\begin{aligned}
\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) &\leq \inf_{t>0}\left\{e^{-t\theta}\mathbb{E}\mathrm{Tr}\exp\left(\sum_{i=1}^{n} t\mathcal{X}_i\right)\right\} \\
&\leq \inf_{t>0}\left\{e^{-t\theta}\mathbb{E}\mathrm{Tr}\exp\left(2t^2\sum_{i=1}^{n}\mathcal{A}_i^2\right)\right\} \\
&\leq \inf_{t>0}\left\{e^{-t\theta}\mathbb{I}_1^M\lambda_{\max}\left(\exp\left(2t^2\sum_{i=1}^{n}\mathcal{A}_i^2\right)\right)\right\} \\
&= \inf_{t>0}\left\{e^{-t\theta}\mathbb{I}_1^M\exp\left(2t^2\sigma^2\right)\right\} \\
&\leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}},
\end{aligned} \tag{80}$$

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.



Theorem 24 (Tensor McDiarmid Inequality)

Given a set of n independent random variables, i.e. $\{X_i : i = 1, 2, \cdots n\}$*, 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*

$$\left(F(x_1, \cdots, x_i, \cdots, x_n) - F(x_1, \cdots, x_i', \cdots, x_n)\right)^2 \preceq \mathcal{A}_i^2, \tag{81}$$

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

$$\Pr\left(\lambda_{\max}\left(F(x_1, \cdots, x_n) - \mathbb{E}F(x_1, \cdots, x_n)\right) \geq \theta\right) \leq \mathbb{I}_1^M e^{-\frac{\theta^2}{8\sigma^2}}. \tag{82}$$

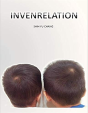
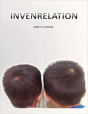


Table of Contents



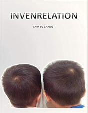
Motivation

Previous theory is based on the summation of independent random tensors, how about the tail bouns for the function of random tensors sum? We wish to consider the following problem:

$$\Pr\left(\left\|g\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right\| \geq \theta\right) \leq \text{some bounds ?} \tag{83}$$

where $\| \ \|$ is a tensor norm function. The answer is: Yes!.
But we need more tools:

- Unitarily Invariant Tensor Norms.
- Antisymmetric Tensor Product.
- Marorization.

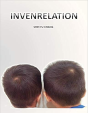


Unitarily Invariant Tensor Norms, I

Let us represent the Hermitian eigenvalues of a Hermitian tensor $\mathcal{H} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ in decreasing order by the vector $\vec{\lambda}(\mathcal{H}) = (\lambda_1(\mathcal{H}), \cdots, \lambda_r(\mathcal{H}))$, where r is the Hermitian rank of the tensor $\mathcal{H}$. We use $\mathbb{R}_{\geq 0}(\mathbb{R}_{>0})$ to represent a set of nonnegative (positive) real numbers. Let $\|\cdot\|_\rho$ be a unitarily invariant tensor norm, i.e., $\|\mathcal{H} \star_N \mathcal{U}\|_\rho = \|\mathcal{U} \star_N \mathcal{H}\|_\rho = \|\mathcal{H}\|_\rho$, where $\mathcal{U}$ is any unitary tensor. Let $\rho : \mathbb{R}^r_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ be the corresponding gauge function that satisfies Hölder's inequality so that

$$\|\mathcal{H}\|_\rho = \||\mathcal{H}|\|_\rho = \rho(\vec{\lambda}(|\mathcal{H}|)), \tag{84}$$

where $|\mathcal{H}| \stackrel{\text{def}}{=} \sqrt{\mathcal{H}^H \star_N \mathcal{H}}$. The bijective correspondence between symmetric gauge functions on $\mathbb{R}^r_{\geq 0}$.



Unitarily Invariant Tensor Norms, II

Several popular norms can be treated as special cases of unitarily invariant tensor norm. The first one is Ky Fan like k-norm [FH55] for tensors. For $k \in \{1, 2, \cdots, r\}$, the Ky Fan k-norm [FH55] for tensors $\mathcal{H}\mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$, denoted as $\|\mathcal{H}\|_{(k)}$, is defined as:

$$\|\mathcal{H}\|_{(k)} \stackrel{\text{def}}{=} \sum_{i=1}^{k} \lambda_i(|\mathcal{H}|). \tag{85}$$

If $k = 1$, the Ky Fan k-norm for tensors is the tensor operator norm, denoted as $\|\mathcal{H}\|$. The second one is Schatten p-norm for tensors, denoted as $\|\mathcal{H}\|_p$, is defined as:

$$\|\mathcal{H}\|_p \stackrel{\text{def}}{=} \left(\mathrm{Tr}|\mathcal{H}|^p\right)^{\frac{1}{p}}, \tag{86}$$

where $p \geq 1$. If $p = 1$, it is the trace norm. The third one is k-trace norm, denoted as $\mathrm{Tr}_k[\mathcal{H}]$, defined by [Hua20]. It is

$$\mathrm{Tr}_k[\mathcal{H}] \stackrel{\text{def}}{=} \sum_{1 \leq i_1 < i_2 < \cdots i_k \leq r} \lambda_{i_1}\lambda_{i_1}\cdots\lambda_{i_k} \tag{87}$$

where $1 \leq k \leq r$. If $k = 1$, $\mathrm{Tr}_k[\mathcal{H}]$ is reduced as trace norm.

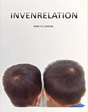
Unitarily Invariant Tensor Norms, III

Following inequality is the extension of Hölder inequality to gauge function ρ which will be used by later to prove majorization relations.

Lemma 25

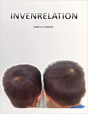
For n nonnegative real vectors with the dimension r, i.e., $\mathbf{b}_i = (b_{i_1}, \cdots, b_{i_r}) \in \mathbb{R}^r_{\geq 0}$, *and* $\alpha > 0$ *with* $\sum_{i=1}^{n} \alpha_i = 1$, *we have*

$$\rho\left(\prod_{i=1}^{n} b_{i_1}^{\alpha_i}, \prod_{i=1}^{n} b_{i_2}^{\alpha_i}, \cdots, \prod_{i=1}^{n} b_{i_r}^{\alpha_i}\right) \leq \prod_{i=1}^{n} \rho(\mathbf{b}_i)^{\alpha_i} \qquad (88)$$



Antisymmetric Tensor Product, I

Let $\mathfrak{H}$ be a Hilbert space of dimension r, $\mathfrak{L}(\mathfrak{H})$ be the set of tensors (linear operators) on $\mathfrak{H}$. Two tensors $\mathcal{A}, \mathcal{B} \in \mathfrak{L}(\mathfrak{H})$ is said $\mathcal{A} \geq \mathcal{B}$ if $\mathcal{A} - \mathcal{B}$ is a nonnegative Hermitian tensor. For any $k \in \{1, 2, \cdots, r\}$, let $\mathfrak{H}^{\otimes k}$ be the k-th tensor power of the space $\mathfrak{H}$ and let $\mathfrak{H}^{\wedge k}$ be the antisymmetric subspace of $\mathfrak{H}^{\otimes k}$. We define function $\wedge^k : \mathfrak{L}(\mathfrak{H}) \rightarrow \mathfrak{L}(\mathfrak{H}^{\wedge k})$ as mapping any tensor $\mathcal{A}$ to the restriction of $\mathcal{A}^{\otimes k} \in \mathfrak{L}(\mathfrak{H}^{\otimes k})$ to the antisymmetric subspace $\mathfrak{H}^{\wedge k}$ of $\mathfrak{H}^{\otimes k}$. Following lemma summarizes several useful properties of such antisymmetric tensor products.

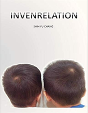


Antisymmetric Tensor Product, II

Lemma 26

Let $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ be tensors in $\mathfrak{L}(\mathfrak{H})$, and $\mathcal{D} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ be Hermitian tensors from $\mathfrak{H}$ with Hermitian rank r. For any $k \in \{1, 2, \cdots, r\}$, we have

1. $(\mathcal{A}^{\wedge k})^H = (\mathcal{A}^H)^{\wedge k}$,.
2. $(\mathcal{A}^{\wedge k}) \star_N (\mathcal{B}^{\wedge k}) = (\mathcal{A} \star_N \mathcal{B})^{\wedge k}$.
3. *If* $\lim\limits_{i \to \infty} \|\mathcal{A}_i - \mathcal{A}\| \to 0$, *then* $\lim\limits_{i \to \infty} \left\|\mathcal{A}_i^{\wedge k} - \mathcal{A}^{\wedge k}\right\| \to 0$.
4. *If* $\mathcal{C} \geq \mathcal{O}$ *(zero tensor), then* $\mathcal{C}^{\wedge k} \geq \mathcal{O}$ *and* $(\mathcal{C}^p)^{\wedge k} = (\mathcal{C}^{\wedge k})^p$ *for all* $p \in \mathbb{R}_{>0}$.
5. $|\mathcal{A}|^{\wedge k} = |\mathcal{A}^{\wedge k}|$.
6. *If* $\mathcal{D} \geq \mathcal{O}$ *and* $\mathcal{D}$ *is invertibale,* $(\mathcal{D}^z)^{\wedge k} = (\mathcal{D}^{\wedge k})^z$ *for all* $z \in \mathbb{D}$.
7. $\left\|\mathcal{A}^{\wedge k}\right\| = \prod_{i=1}^{k} \lambda_i(|\mathcal{A}|)$.



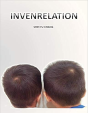
Marorization, I

Let $\mathbf{x} = [x_1, \cdots, x_r] \in \mathbb{R}^r, \mathbf{y} = [y_1, \cdots, y_r] \in \mathbb{R}^r$ be two vectors with following orders among entries $x_1 \geq \cdots \geq x_r$ and $y_1 \geq \cdots \geq y_r$, *weak majorization* between vectors $\mathbf{x}, \mathbf{y}$, represented by $\mathbf{x} \prec_w \mathbf{y}$, requires following relation for vectors $\mathbf{x}, \mathbf{y}$:

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i, \tag{89}$$

where $k \in \{1, 2, \cdots, r\}$. *Majorization* between vectors $\mathbf{x}, \mathbf{y}$, indicated by $\mathbf{x} \prec \mathbf{y}$, requires following relation for vectors $\mathbf{x}, \mathbf{y}$:

$$\begin{aligned} \sum_{i=1}^{k} x_i &\leq \sum_{i=1}^{k} y_i, \quad \text{for } 1 \leq k < r; \\ \sum_{i=1}^{r} x_i &= \sum_{i=1}^{r} y_i, \quad \text{for } k = r. \end{aligned} \tag{90}$$



Marorization, II

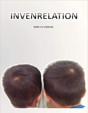
For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^r_{\geq 0}$ such that $x_1 \geq \cdots \geq x_r$ and $y_1 \geq \cdots \geq y_r$, *weak log majorization* between vectors $\mathbf{x}, \mathbf{y}$, represented by $\mathbf{x} \prec_{w \log} \mathbf{y}$, requires following relation for vectors $\mathbf{x}, \mathbf{y}$:

$$\prod_{i=1}^{k} x_i \leq \prod_{i=1}^{k} y_i, \tag{91}$$

where $k \in \{1, 2, \cdots, r\}$, and *log majorization* between vectors $\mathbf{x}, \mathbf{y}$, represented by $\mathbf{x} \prec_{\log} \mathbf{y}$, requires equality for $k = r$ in Eq. (91). If f is a single variable function, $f(\mathbf{x})$ represents a vector of $[f(x_1), \cdots, f(x_r)]$ From Lemma 1 in [HKT17], we have

Lemma 27

(1) For any convex function $f : [0, \infty) \rightarrow [0, \infty)$, if we have $\mathbf{x} \prec \mathbf{y}$, then $f(\mathbf{x}) \prec_w f(\mathbf{y})$.
(2) For any convex function and non-decreasing $f : [0, \infty) \rightarrow [0, \infty)$, if we have $\mathbf{x} \prec_w \mathbf{y}$, then $f(\mathbf{x}) \prec_w f(\mathbf{y})$.



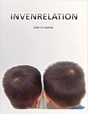
Marorization, III

Another lemma is from Lemma 12 in [HKT17], we have

Lemma 28

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^r_{\geq 0}$ *such that* $x_1 \geq \cdots \geq x_r$ *and* $y_1 \geq \cdots \geq y_r$ *with* $\mathbf{x} \prec_{\log} \mathbf{y}$*. Also let* $\mathbf{y}_i = [y_{i;1}, \cdots, y_{i;r}] \in \mathbb{R}^r_{\geq 0}$ *be a sequence of vectors such that* $y_{i;1} \geq \cdots \geq y_{i;r} > 0$ *and* $\mathbf{y}_i \to \mathbf{y}$ *as* $i \to \infty$*. Then, there exists* $i_0 \in \mathbb{N}$ *and* $\mathbf{x}_i = [x_{i;1}, \cdots, x_{i;r}] \in \mathbb{R}^r_{\geq 0}$ *for* $i \geq i_0$ *such that* $x_{i;1} \geq \cdots \geq x_{i;r} > 0$*,* $\mathbf{x}_i \to \mathbf{x}$ *as* $i \to \infty$*, and*

$$\mathbf{x}_i \prec_{\log} \mathbf{y}_i \quad \text{for } i \geq i_0. \tag{92}$$



Majorization wtih Integral Average

Let Ω be a σ-compact metric space and ν a probability measure on the Boreal σ-field of Ω. Let $\mathcal{C}, \mathcal{D}_\tau \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ be Hermitian tensors with Hermitian rank r. We further assume that tensors $\mathcal{C}, \mathcal{D}_\tau$ are uniformly bounded in their norm for $\tau \in \Omega$. Let $\tau \in \Omega \rightarrow \mathcal{D}_\tau$ be a continuous function such that $\sup\{\|D_\tau\| : \tau \in \Omega\} < \infty$. For notational convenience, we define the following relation:

$$\left[\int_\Omega \lambda_1(\mathcal{D}_\tau) d\nu(\tau), \cdots, \int_\Omega \lambda_r(\mathcal{D}_\tau) d\nu(\tau)\right] \stackrel{\text{def}}{=} \int_{\Omega^r} \vec{\lambda}(\mathcal{D}_\tau) d\nu^r(\tau). \qquad (93)$$

If f is a single variable function, the notation $f(\mathcal{C})$ represents a tensor function with respect to the tensor $\mathcal{C}$.

Theorem 29

Let $f : \mathbb{R} \rightarrow [0, \infty)$ be a non-decreasing convex function, we have following two equivalent statements:

$$\vec{\lambda}(\mathcal{C}) \prec_w \int_{\Omega^r} \vec{\lambda}(\mathcal{D}_\tau) d\nu^r(\tau) \Longleftrightarrow \|f(\mathcal{C})\|_\rho \leq \int_\Omega \|f(\mathcal{D}_\tau)\|_\rho \, d\nu(\tau), \qquad (94)$$

where $\|\cdot\|_\rho$ is the unitarily invariant norm defined in Eq. (84).

Majorization wtih Integral Average, cont.

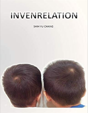
Next theorem will provide a stronger version of Theorem 29 by removing weak majorization conditions.

Theorem 30

Let $f : \mathbb{R} \rightarrow [0, \infty)$ *be a convex function, we have following two equivalent statements:*

$$\vec{\lambda}(\mathcal{C}) \prec \int_{\Omega^r} \vec{\lambda}(\mathcal{D}_\tau) d\nu^r(\tau) \Longleftrightarrow \left\| f(\mathcal{C}) \right\|_\rho \leq \int_{\Omega} \left\| f(\mathcal{D}_\tau) \right\|_\rho d\nu(\tau), \qquad (95)$$

where $\left\| \cdot \right\|_\rho$ *is the unitarily invariant norm defined in Eq.* (84)*.*



Log-Majorization wtih Integral Average

Theorem 31

Let $\mathcal{C}, \mathcal{D}_\tau$ be nonnegative Hermitian tensors, $f : (0, \infty) \rightarrow [0, \infty)$ be a continous function such that the mapping $x \rightarrow \log f(e^x)$ is convex on $\mathbb{R}$, and $g : (0, \infty) \rightarrow [0, \infty)$ be a continous function such that the mapping $x \rightarrow g(e^x)$ is convex on $\mathbb{R}$, then we have following three equivalent statements:

$$\vec{\lambda}(\mathcal{C}) \quad \prec_{w \log} \quad \exp \int_{\Omega^r} \log \vec{\lambda}(\mathcal{D}_\tau) d\nu^r(\tau); \tag{96}$$

$$\left\| f(\mathcal{C}) \right\|_\rho \quad \leq \quad \exp \int_{\Omega} \log \left\| f(\mathcal{D}_\tau) \right\|_\rho d\nu(\tau); \tag{97}$$

$$\left\| g(\mathcal{C}) \right\|_\rho \quad \leq \quad \int_{\Omega} \left\| g(\mathcal{D}_\tau) \right\|_\rho d\nu(\tau). \tag{98}$$

Log-Majorization wtih Integral Average, cont

Next theorem will extend Theorem 31 to non-weak version.

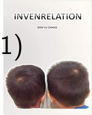
Theorem 32

Let $\mathcal{C}, \mathcal{D}_\tau$ be nonnegative Hermitian tensors with $\int_\Omega \left\| \mathcal{D}_\tau^{-p} \right\|_\rho d\nu(\tau) < \infty$ for any $p > 0$, $f : (0, \infty) \rightarrow [0, \infty)$ be a continous function such that the mapping $x \rightarrow \log f(e^x)$ is convex on $\mathbb{R}$, and $g : (0, \infty) \rightarrow [0, \infty)$ be a continous function such that the mapping $x \rightarrow g(e^x)$ is convex on $\mathbb{R}$, then we have following three equivalent statements:

$$\vec{\lambda}(\mathcal{C}) \quad \prec_{\log} \quad \exp \int_{\Omega^r} \log \vec{\lambda}(\mathcal{D}_\tau) d\nu^r(\tau); \tag{99}$$

$$\left\| f(\mathcal{C}) \right\|_\rho \quad \leq \quad \exp \int_{\Omega} \log \left\| f(\mathcal{D}_\tau) \right\|_\rho d\nu(\tau); \tag{100}$$

$$\left\| g(\mathcal{C}) \right\|_\rho \quad \leq \quad \int_{\Omega} \left\| g(\mathcal{D}_\tau) \right\|_\rho d\nu(\tau). \tag{101}$$

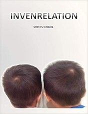


Multivaraite Tensor Norm Inequalities

Lemma 33 (Lie-Trotter product formula for tensors)

Let $m \in \mathbb{N}$ *and* $(\mathcal{L}_k)_{k=1}^{m}$ *be a finite sequence of bounded tensors with dimensions* $\mathcal{L}_k \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, *then we have*

$$\lim_{n \to \infty} \left(\prod_{k=1}^{m} \exp\left(\frac{\mathcal{L}_k}{n}\right) \right)^{n} = \exp\left(\sum_{k=1}^{m} \mathcal{L}_k \right) \quad (102)$$



Multivaraite Tensor Norm Inequalities, cont

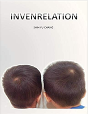
Theorem 34

Let $\mathcal{C}_i \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ *be positive Hermitian tensors for* $1 \leq i \leq n$ *with Hermitian rank* r, $\left\|\cdot\right\|_\rho$ *be a unitarily invaraint norm with corresponding gauge function* ρ. *For any continous function* $f : (0, \infty) \rightarrow [0, \infty)$ *such that* $x \rightarrow \log f(e^x)$ *is convex on* $\mathbb{R}$, *we have*

$$\left\| f\left(\exp\left(\sum_{i=1}^{n} \log \mathcal{C}_i\right)\right)\right\|_\rho \leq \exp \int_{-\infty}^{\infty} \log \left\| f\left(\left|\prod_{i=1}^{n} \mathcal{C}_i^{1+\iota t}\right|\right)\right\|_\rho \beta_0(t) dt, \quad (103)$$

where $\beta_0(t) = \frac{\pi}{2(\cosh(\pi t)+1)}$.
For any continous function $g(0, \infty) \rightarrow [0, \infty)$ *such that* $x \rightarrow g(e^x)$ *is convex on* $\mathbb{R}$, *we have*

$$\left\| g\left(\exp\left(\sum_{i=1}^{n} \log \mathcal{C}_i\right)\right)\right\|_\rho \leq \int_{-\infty}^{\infty} \left\| g\left(\left|\prod_{i=1}^{n} \mathcal{C}_i^{1+\iota t}\right|\right)\right\|_\rho \beta_0(t) dt. \quad (104)$$



Ky Fan k-norm Tail Bound

Lemma 35

Let $\mathcal{C}_i \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ with Hermitian rank r and let p_i be positive real numbers satisfying $\sum\limits_{i=1}^{m} \frac{1}{p_i} = 1$. Then, we have

$$\left\| \left| \prod_{i=1}^{m} \mathcal{C}_i \right|^s \right\|_{(k)} \leq \prod_{i=1}^{m} \left(\left\| |\mathcal{C}_i|^{sp_i} \right\|_{(k)} \right)^{\frac{1}{p_i}} \leq \sum_{i=1}^{m} \frac{\left\| |\mathcal{C}_i|^{sp_i} \right\|_{(k)}}{p_i} \quad (105)$$

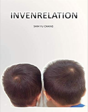
where $s \geq 1$ and $k \in \{1, 2, \cdots, r\}$.

Lemma 36

Let $\mathcal{C}_i \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ with Hermitian rank r, then we have

$$\left\| \left| \sum_{i=1}^{m} \mathcal{C}_i \right|^s \right\|_{(k)} \leq m^{s-1} \sum_{i=1}^{m} \left\| |\mathcal{C}_i|^s \right\|_{(k)} \quad (106)$$

where $s \geq 1$ and $k \in \{1, 2, \cdots, r\}$.



Ky Fan k-norm Tail Bound, cont

Theorem 37

Consider a sequence $\{\mathcal{X}_j \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}\}$ *of independent, random, Hermitian tensors. Let* g *be a polynomial function with degree* n *and nonnegative coeffecients* $a_0, a_1, \cdots, a_n$ *raised by power* $s \geq 1$*, i.e.,* $g(x) = (a_0 + a_1 x + \cdots + a_n x^n)^s$*. Suppose following condition is satisfied:*

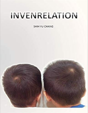
$$g\left(\exp\left(t\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \succeq \exp\left(tg\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \quad \textit{almost surely}, \tag{107}$$

where $t > 0$*. Then, we have*

$$\Pr\left(\left\|g\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right\|_{(k)} \geq \theta\right) \leq$$

$$(n+1)^{s-1}\inf_{t,p_j} e^{-\theta t}\left(ka_0^s + \sum_{l=1}^{n}\sum_{j=1}^{m}\frac{a_l^{ls}\mathbb{E}\left\|\exp\left(p_j l s t \mathcal{X}_j\right)\right\|_{(k)}}{p_j}\right). \tag{108}$$

where $\sum_{j=1}^{m}\frac{1}{p_j} = 1$ *and* $p_j > 0$*.*



Ky Fan k-norm Tail Bound, proof

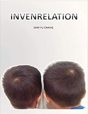
Let $t > 0$ be a parameter to be chosen later. Then

$$\begin{aligned}
\mathrm{Pr}\left(\left\|g\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right\|_{(k)} \geq \theta\right) &= \mathrm{Pr}\left(\left\|\exp\left(tg\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right)\right\|_{(k)} \geq \exp(\theta t)\right) \\
&\leq_1 \exp(-\theta t)\,\mathbb{E}\left(\left\|\exp\left(tg\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right)\right\|_{(k)}\right) \\
&\leq_2 \exp(-\theta t)\,\mathbb{E}\left(\left\|g\left(\exp\left(t\sum_{j=1}^{m}\mathcal{X}_j\right)\right)\right\|_{(k)}\right) \qquad (109)
\end{aligned}$$

where $\leq_1$ uses Markov's inequality, $\leq_2$ requires condition provided by Eq. (107).
We can further bound the expectation term in Eq. (108) as

$$\begin{aligned}
\mathbb{E}\left(\left\|g\left(\exp\left(t\sum_{j=1}^{m}\mathcal{X}_j\right)\right)\right\|_{(k)}\right) &\leq_3 \mathbb{E}\int_{-\infty}^{\infty}\left\|g\left(\left|\prod_{j=1}^{m}e^{(1+\iota\tau)t\mathcal{X}_j}\right|\right)\right\|_{(k)} \beta_0(\tau)d\tau \\
&\leq_4 (n+1)^{s-1}\left(ka_0^s + \sum_{l=1}^{n}a_l^{ls}\mathbb{E}\int_{-\infty}^{\infty}\left\|\left|\prod_{j=1}^{m}e^{(1+\iota\tau)t\mathcal{X}_j}\right|^{ls}\right\|_{(k)} \beta_0(\tau)d\tau\right), \quad (110)
\end{aligned}$$

where $\leq_3$ from Eq. (104) in Theorem 34, $\leq_4$ is obtained from function g definition and Lemma 36.

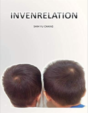


Ky Fan k-norm Tail Bound, proof, cont.

Again, the expectation term in Eq. (110) can be further bounded by Lemma 35 as

$$
\begin{aligned}
\mathbb{E}\int_{-\infty}^{\infty}\left\|\left|\prod_{j=1}^{m} e^{(1+\iota\tau)t\mathcal{X}_j}\right|^{ls}\right\|_{(k)}\beta_0(\tau)d\tau &\leq \mathbb{E}\int_{-\infty}^{\infty}\sum_{j=1}^{m}\frac{\left\|\left|e^{t\mathcal{X}_j}\right|^{p_j ls}\right\|_{(k)}}{p_j}\beta_0(\tau)d\tau \\
&= \sum_{j=1}^{m}\frac{\mathbb{E}\left\|e^{p_j lst\mathcal{X}_j}\right\|_{(k)}}{p_j}. \qquad (111)
\end{aligned}
$$

Note that the final equality is obtained due to that the integrand is indepedent of the variable τ and $\int_{-\infty}^{\infty}\beta_0(\tau)d\tau = 1$.
Finally, this theorem is established from Eqs. (109), (110), and (111). $\square$



Theorem 38 (Generalized Tensor Chernoff Bound)

Consider a sequence $\{\mathcal{X}_j \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}\}$ *of independent, random, Hermitian tensors. Let* g *be a polynomial function with degree* n *and nonnegative coeffecients* $a_0, a_1, \cdots, a_n$ *raised by power* $s \geq 1$*, i.e.,* $g(x) = (a_0 + a_1 x + \cdots + a_n x^n)^s$ *with* $s \geq 1$*. Suppose following condition is satisfied:*

$$g\left(\exp\left(t\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \succeq \exp\left(tg\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \quad \textit{almost surely}, \qquad (112)$$

where $t > 0$*. Moreover, we require*

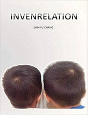
$$\mathcal{X}_i \succeq \mathcal{O} \quad \textit{and} \quad \lambda_{\max}(\mathcal{X}_i) \leq \mathrm{R} \quad \textit{almost surely.} \qquad (113)$$

Then we have following inequality:

$$\Pr\left(\left\|g\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right\|_{(k)} \geq \theta\right) \leq (n+1)^{s-1}\inf_{t>0} e^{-\theta t} \cdot$$

$$\left\{ka_0^s + \sum_{l=1}^{n}\sum_{j=1}^{m}\frac{ka_l^{ls}}{m}\left[1 + \left(e^{mlsRt} - 1\right)\overline{\sigma_1(\mathcal{X}_j)} + C\left(e^{mlsRt} - 1\right)\Xi(\mathcal{X}_j)\right]\right\} (114)$$

where $\overline{\sigma_1(\mathcal{X}_j)} \stackrel{\text{def}}{=} \left[\sigma_1\left(\overline{\frac{\mathcal{X}_j + \mathcal{X}_j^*}{2}}\right) + \sigma_1\left(\overline{\frac{\mathcal{X}_j - \mathcal{X}_j^*}{2}}\right)\right]$.



Theorem 39 (Generalized Tensor Bernstein Bound)

Consider a sequence $\{\mathcal{X}_j \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}\}$ *of independent, random, Hermitian tensors. Let* g *be a polynomial function with degree* n *and nonnegative coeffecients* $a_0, a_1, \cdots, a_n$ *raised by power* $s \geq 1$*, i.e.,* $g(x) = (a_0 + a_1 x + \cdots + a_n x^n)^s$ *with* $s \geq 1$*. Suppose following condition is satisfied:*

$$g\left(\exp\left(t\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \succeq \exp\left(tg\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \quad \textit{almost surely,} \tag{115}$$

where $t > 0$*, and we also have*

$$\mathbb{E}\mathcal{X}_j = \mathcal{O} \quad \textit{and} \quad \mathcal{X}_j^p \preceq \frac{p!\mathcal{A}_j^2}{2} \quad \textit{almost surely for } p = 2, 3, 4, \cdots. \tag{116}$$

Then we have following inequality:

$$\Pr\left(\left\|g\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right\|_{(k)} \geq \theta\right) \leq (n+1)^{s-1}\inf_{t>0} e^{-\theta t}k \cdot$$
$$\left\{a_0^s + \sum_{l=1}^{n}\sum_{j=1}^{m} a_l^{ls}\left[\frac{1}{m} + \frac{m(lst)^2\sigma_1(\mathcal{A}_j^2)}{2(1-mlst)} + lstC\Upsilon(\mathcal{X}_j)\right]\right\}. \tag{117}$$

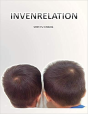
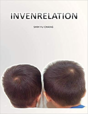


Table of Contents

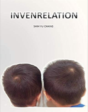


Another Direction of Concentration Bounds

Another direction to extend from the basic Chernoff bound is to consider non-independent assumptions for random variables. By Gillman [Gil98], they changed the independence assumption to Markov dependence and we summarize their works as follows. We are given $\mathfrak{G}$ as a regular λ-expander graph with vertex set $\mathfrak{V}$, and $g : \mathfrak{V} \rightarrow \mathbb{C}$ as a bounded function. Suppose $v_1, v_2 \cdots, v_\kappa$ is a stationary random walk of length κ on $\mathfrak{G}$, it is shown that:

$$\Pr\left(\left\|\frac{1}{\kappa}\sum_{j=1}^{\kappa} g(v_i) - \mathbb{E}[g]\right\| \geq \vartheta\right) \leq 2\exp(-\Omega(1-\lambda)\kappa\vartheta^2). \tag{118}$$

The value of λ is also the second largest eigenvalue of the transition matrix of the underlying graph $\mathfrak{G}$. The bound given in Eq. (118) is named as "Expander Chernoff Bound".



Expectation Estimation for Product of Tensors, I

Let $\mathbf{A}$ be the normalized adjacency matrix of the underlying graph $\mathfrak{G}$ and let $\tilde{\mathbf{A}} = \mathbf{A} \otimes \mathcal{I}_{(\mathbb{I}_1^M)^2}$, where the identity tensor $\mathcal{I}_{(\mathbb{I}_1^M)^2}$ has dimensions as $I_1^2 \times \cdots \times I_M^2 \times I_1^2 \times \cdots \times I_M^2$. We use $\mathcal{F} \in \mathbb{C}^{\left(n \times I_1^2 \times \cdots \times I_M^2\right) \times \left(n \times I_1^2 \times \cdots \times I_M^2\right)}$ to represent block diagonal tensor valued matrix where the v-th diagonal block is the tensor

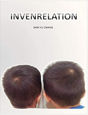
$$\mathcal{T}_v = \exp\left(\frac{t\mathrm{g}(v)(a + \iota b)}{2}\right) \otimes \exp\left(\frac{t\mathrm{g}(v)(a - \iota b)}{2}\right). \tag{119}$$

The tensor $\mathcal{F}$ can also be expressed as

$$\mathcal{F} = \begin{bmatrix} \mathcal{T}_{v_1} & \mathcal{O} & \cdots & \mathcal{O} \\ \mathcal{O} & \mathcal{T}_{v_2} & \cdots & \mathcal{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{O} & \mathcal{O} & \cdots & \mathcal{T}_{v_n} \end{bmatrix}. \tag{120}$$

Then the tensor $\left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}}\right)^{\kappa}$ is a block tensor valued matrix whose (u, v)-block is a tensor with dimensions as $I_1^2 \times \cdots \times I_M^2 \times I_1^2 \times \cdots \times I_M^2$ expressed as :

$$\sum_{v_1, \cdots, v_{\kappa-1} \in \mathfrak{V}} \mathbf{A}_{u,v_1} \left(\prod_{j=1}^{\kappa-2} \mathbf{A}_{v_j, v_{j+1}}\right) \mathbf{A}_{v_{\kappa-1}, v} \left(\mathcal{T}_u \star_{2M} \mathcal{T}_{v_1} \star_{2M} \cdots \star_{2M} \mathcal{T}_{v_{\kappa-1}}\right) \tag{121}$$



Expectation Estimation for Product of Tensors, II

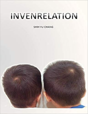
Let $\mathbf{u}_0 \in \mathbb{C}^{n \times \mathbb{I}_1^2 \times \cdots \times \mathbb{I}_M^2}$ be the tensor obtained by $\frac{\mathbf{1}}{\sqrt{n}} \otimes \mathbf{col}(\mathcal{I}_{\mathbb{I}_1^M})$, where $\mathbf{1}$ is the all ones vector with size n and $\mathbf{col}(\mathcal{I}_{\mathbb{I}_1^M}) \in \mathbb{C}^{I_1^2 \times \cdots \times I_M^2 \times 1}$ is the column tensor of the identity tensor $\mathcal{I}_{\mathbb{I}_1^M} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$. By applying the following relation:

$$\left\langle \mathbf{col}(\mathcal{I}_{\mathbb{I}_1^M}), \mathcal{C} \otimes \mathcal{B} \star_M \mathbf{col}(\mathcal{I}_{\mathbb{I}_1^M}) \right\rangle = \mathrm{Tr}\left(\mathcal{C} \star_M \mathcal{B}^T \right), \tag{122}$$

where $\mathcal{C}, \mathcal{B} \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$; we will have following expectation of κ steps transition of Hermitian tensors from the vertex v_1 to the vertex v_κ,

$$\mathbb{E}\left[\mathrm{Tr}\left(\prod_{i=1}^{\kappa} \exp\left(\frac{\mathrm{tg}(v_i)(a + \iota b)}{2} \right) \star_M \prod_{i=\kappa}^{1} \exp\left(\frac{\mathrm{tg}(v_i)(a - \iota b)}{2} \right) \right) \right] =$$

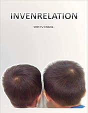
$$= \mathbb{E}\left[\left\langle \mathbf{col}(\mathcal{I}_{\mathbb{I}_1^M}), \prod_{i=1}^{\kappa} \mathcal{T}_{v_i} \star_M \mathbf{col}(\mathcal{I}_{\mathbb{I}_1^M}) \right\rangle \right] = \left\langle \mathbf{u}_0, \left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}} \right)^{\kappa} \star_{M+1} \mathbf{u}_0 \right\rangle \tag{123}$$



Expectation Estimation for Product of Tensors, III

If we define $\left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}}\right)^{\kappa} \star_{M+1} \mathbf{u}_0$ as $\mathbf{u}_{\kappa}$, the goal of this section is to estimate $\langle \mathbf{u}_0, \mathbf{u}_{\kappa} \rangle$.

The trick is to separate the space of $\mathbf{u}$ as the subspace spanned by the $(\mathbb{I}_1^M)^2$ tensors $\mathbf{1} \otimes e_i$ denoted by $\mathbf{u}^{\parallel}$, where $1 \leq i \leq (\mathbb{I}_1^M)^2$ and $e_i \in \mathbb{C}^{I_1^2 \times \cdots \times I_M^2 \times 1}$ is the column tensor of size $(\mathbb{I}_1^M)^2$ with 1 in position i and 0 elsewhere, and its orthogonal complement space, denoted by $\mathbf{u}^{\perp}$. Following lemma is required to bound how the tensor norm is changed in terms of aforementioned subspace and its orthogonal space after acting by the tensor $\mathcal{F} \star_{2M+1} \tilde{\mathbf{A}}$. We require two lemmas.

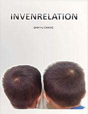


Expectation Estimation for Product of Tensors, IV

Lemma 40

Given paramters $\lambda \in (0,1)$, $a \geq 0$, $r > 0$, *and* $t > 0$. *Let* $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E})$ *be a regular* λ*-expander graph on the vetices set* $\mathfrak{V}$ *and* $\|\mathrm{g}(v_i)\| \leq r$ *for all* $v_i \in \mathfrak{V}$. *Each vertex* $v \in \mathfrak{V}$ *will be assigned a tensor* $\acute{\mathcal{T}}_v$, *where* $\acute{\mathcal{T}}_v \stackrel{def}{=} \frac{\mathrm{g}(v)(a+\iota b)}{2} \otimes \mathcal{I}_{\mathbb{I}_1^M} + \mathcal{I}_{\mathbb{I}_1^M} \otimes \frac{\mathrm{g}(v)(a-\iota b)}{2} \in \mathbb{C}^{I_1^2 \times \cdots \times I_M^2 \times I_1^2 \times \cdots \times I_M^2}$. *Let* $\mathcal{F} \in \mathbb{C}^{(n \times I_1^2 \times \cdots \times I_M^2) \times (n \times I_1^2 \times \cdots \times I_M^2)}$ *to represent block diagonal tensor valued matrix where the* v*-th diagonal block is the tensor* $\exp(t\acute{\mathcal{T}}_v) = \mathcal{T}_v$. *For any tensor* $\mathbf{u} \in \mathbb{C}^{n \times \mathbb{I}_1^2 \times \cdots \times \mathbb{I}_M^2}$, *we have*

1. $\left\| \left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}} \star_{M+1} \mathbf{u}^{\parallel} \right)^{\parallel} \right\| \leq \gamma_1 \left\| \mathbf{u}^{\parallel} \right\|$, *where* $\gamma_1 = \exp(tr\sqrt{a^2 + b^2})$;
2. $\left\| \left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}} \star_{M+1} \mathbf{u}^{\perp} \right)^{\parallel} \right\| \leq \gamma_2 \left\| \mathbf{u}^{\perp} \right\|$, *where* $\gamma_2 = \lambda(\exp(tr\sqrt{a^2 + b^2}) - 1)$;
3. $\left\| \left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}} \star_{M+1} \mathbf{u}^{\parallel} \right)^{\perp} \right\| \leq \gamma_3 \left\| \mathbf{u}^{\parallel} \right\|$, *where* $\gamma_3 = \exp(tr\sqrt{a^2 + b^2}) - 1$;
4. $\left\| \left(\mathcal{F} \star_{M+1} \tilde{\mathbf{A}} \star_{M+1} \mathbf{u}^{\perp} \right)^{\perp} \right\| \leq \gamma_4 \left\| \mathbf{u}^{\perp} \right\|$, *where* $\gamma_4 = \lambda \exp(tr\sqrt{a^2 + b^2})$.



Expectation Estimation for Product of Tensors, V

In the following, we will apply Lemma 40 to bound the following term provided by Eq. (123)

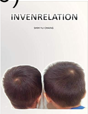
$$\left\langle \mathbf{u}_0, \left(\mathbf{F} \star_{M+1} \tilde{\mathbf{A}}\right)^{\kappa} \star_{M+1} \mathbf{u}_0 \right\rangle \tag{124}$$

This bound is formulated by the following Lemma 41

Lemma 41

Let $\mathfrak{G}$ be a regular λ-expander graph on the vertex set $\mathfrak{V}$, $g : \mathfrak{V} \rightarrow \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$, and let $v_1, \cdots, v_\kappa$ be a stationary random walk on $\mathfrak{G}$. If $tr\sqrt{a^2+b^2} < 1$ and $\lambda(2\exp(tr\sqrt{a^2+b^2}) - 1) \leq 1$, we have:

$$\mathbb{E}\left[\mathrm{Tr}\left(\prod_{i=1}^{\kappa} \exp\left(\frac{t g(v_i)(a+\iota b)}{2}\right) \star_M \prod_{i=\kappa}^{1} \exp\left(\frac{t g(v_i)(a-\iota b)}{2}\right)\right)\right] \leq$$
$$\mathbb{I}_1^M \exp\left[\kappa\left(2tr\sqrt{a^2+b^2} + \frac{8}{1-\lambda} + \frac{16tr\sqrt{a^2+b^2}}{1-\lambda}\right)\right]. \tag{125}$$



Theorem 42 (Tensor Expander Chernoff Bound)

Let $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E})$ be a regular undirected graph whose transition matrix has second eigenvalue λ, and let $g : \mathfrak{V} \rightarrow \in \mathbb{C}^{I_1 \times \cdots \times I_M \times I_1 \times \cdots \times I_M}$ be a function. We assume following:

1. *For each $v \in \mathfrak{V}$, $g(v)$ is a Hermitian tensor;*
2. $\|g(v)\| \leq r$;
3. *A nonnegative coefficients polynomial raised by the power $s \geq 1$ as $f : x \rightarrow (a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n)^s$ satisfying $f\left(\exp\left(t \sum_{j=1}^{\kappa} g(v_j)\right)\right) \geq \exp\left(tf\left(\sum_{j=1}^{\kappa} g(v_j)\right)\right)$ almost surely;*
4. *For $\tau \in [\infty, \infty]$, we have constants C and σ such that $\beta_0(\tau) \leq \frac{C \exp(\frac{-\tau^2}{2\sigma^2})}{\sigma\sqrt{2\pi}}$.*

Then, we have

$$\mathrm{Pr}\left(\left\| f\left(\sum_{j=1}^{\kappa} g(v_j)\right)\right\|_{(k)} \geq \vartheta\right) \leq$$
$$\min_{t>0}\left[(n+1)^{(s-1)} e^{-\vartheta t}\left(a_0 k + C\left(k + \sqrt{\frac{\mathbb{I}_1^M - k}{k}}\right)\cdot\right.\right.$$
$$\left.\left.\sum_{l=1}^{n} a_l \exp(8\kappa\overline{\lambda} + 2(\kappa + 8\overline{\lambda}) lsrt + 2(\sigma(\kappa + 8\overline{\lambda}) lsr)^2 t^2)\right)\right], \qquad (126)$$

where $\overline{\lambda} = 1 - \lambda$.

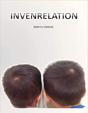
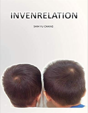
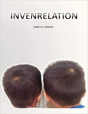


Table of Contents



Tail Bounds for T-product Tensors

- Aforementioned methods or non-indepedent for tensors under Einstein product can also be considered again for tensors under T-product tensors. Works details about tail bounds for tensors under Einstein product can be found [CL, Cha21a, Cha21c].
- The T-product has been shown as a powerful tool in many fields: signal processing, machine learning, computer vision, image processing, low-rank tensor approximation,etc, see [CW21b, CW21c] and references therein. We will show some results about those bounds for T-product tensors, more details can be found at [CW21b, CW21c, Cha21b, CW21a]



T-product Tensor with Concavity Approach, I

Theorem 43 (Hermitian T-product Tensor with Gaussian and Rademacher Series Eigenvalue Version)

Given a finite sequence of fixed T-product tensors $\mathcal{A}_i \in \mathbb{C}^{m \times m \times p}$, *and let* $\{\alpha_i\}$ *be a finite sequence of independent standard normal variables. We define*

$$\sigma_{GR}^2 \stackrel{def}{=} \left\| \sum_{i}^{n} \mathcal{A}_i^2 \right\|, \tag{127}$$

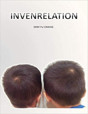
then, for all $\theta \geq 0$, *we have*

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \alpha_i \mathcal{A}_i \right) \geq \theta \right) \leq mpe^{-\frac{\theta^2}{2\sigma_{GR}^2}}. \tag{128}$$

We use $\|\mathcal{X}\|$ *for the spectral norm, which is the largest singular value for the T-product tensor* $\mathcal{X}$. *Then, we have*

$$\Pr\left(\left\| \sum_{i=1}^{n} \alpha_i \mathcal{A}_i \right\| \geq \theta \right) \leq 2mpe^{-\frac{\theta^2}{2\sigma_{GR}^2}}. \tag{129}$$

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



T-product Tensor with Concavity Approach, II

Theorem 44 (T-product Tensor Chernoff Bound I)

Consider a sequence $\{\mathcal{X}_i \in \mathbb{C}^{m \times m \times p}\}$ *of independent, random, Hermitian T-product tensors that satisfy*

$$\mathcal{X}_i \succeq \mathcal{O} \quad \text{and} \quad \lambda_{\max}(\mathcal{X}_i) \leq 1 \quad \text{almost surely.} \tag{130}$$

Define following two quantaties:

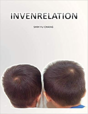
$$\overline{\mu}_{\max} \stackrel{def}{=} \lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i\right) \quad \text{and} \quad \overline{\mu}_{\min} \stackrel{def}{=} \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\mathcal{X}_i\right), \tag{131}$$

then, we have following two inequalities:

$$\Pr\left(\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i\right) \geq \theta\right) \leq mpe^{-n\mathfrak{D}(\theta||\overline{\mu}_{\max})}, \quad \text{for } \overline{\mu}_{\max} \leq \theta \leq 1, \tag{132}$$

and

$$\Pr\left(\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i\right) \leq \theta\right) \leq mpe^{-n\mathfrak{D}(\theta||\overline{\mu}_{\min})}, \quad \text{for } 0 \leq \theta \leq \overline{\mu}_{\min}. \tag{133}$$



T-product Tensor with Concavity Approach, III

Theorem 45 (T-product Tensor Bernstein Bounds with Bounded $\lambda_{\max}$)

Given a finite sequence of independent Hermitian T-product tensors $\{\mathcal{X}_i \in \mathbb{C}^{m \times m \times p}\}$ *that satisfy*

$$\mathbb{E}\mathcal{X}_i = 0 \quad \textit{and} \quad \lambda_{\max}(\mathcal{X}_i) \leq T \quad \textit{almost surely.} \tag{134}$$

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

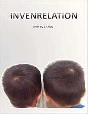
$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \geq \theta \right) \leq mp \exp\left(\frac{-\theta^2/2}{\sigma^2 + T\theta/3} \right); \tag{135}$$

and

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \geq \theta \right) \leq mp \exp\left(\frac{-3\theta^2}{8\sigma^2} \right) \quad \textit{for } \theta \leq \sigma^2/T; \tag{136}$$

and

$$\Pr\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \geq \theta \right) \leq mp \exp\left(\frac{-3\theta}{8T} \right) \quad \textit{for } \theta \geq \sigma^2/T. \tag{137}$$



T-product Tensor with Concavity Approach, IV

Theorem 46 (T-product Tensor Azuma Inequality for Eigenvalue)

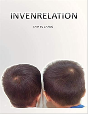
Given a finite adapted sequence of Hermitian tensors $\{\mathcal{X}_i \in \mathbb{C}^{m \times m \times p}\}$ *and a fixed sequence of Hermitian T-product tensors* $\{\mathcal{A}_i\}$ *that satisfy*

$$\mathbb{E}_{i-1}\mathcal{X}_i = 0 \quad \text{and} \quad \mathcal{X}_i^2 \preceq \mathcal{A}_i^2 \quad \text{almost surely}, \tag{138}$$

where $i = 1, 2, 3, \cdots$.

Define the total varaince σ^2 *as:* $\sigma^2 \stackrel{def}{=} \left\| \sum_i^n \mathcal{A}_i^2 \right\|$. *Then, we have following inequalities:*

$$\mathrm{Pr}\left(\lambda_{\max}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \geq \theta \right) \leq mpe^{-\frac{\theta^2}{8\sigma^2}}. \tag{139}$$



Generalized T-product Tensor Chernoff Bounds

Theorem 47 (Generalized T-product Tensor Chernoff Bound)

Consider a sequence $\{\mathcal{X}_j \in \mathbb{C}^{m' \times m' \times p}\}$ *of independent, random, Hermitian tensors. Let* g *be a polynomial function with degree* n *and nonnegative coefficients* $a_0, a_1, \ldots, a_n$ *raised by power* $s \geq 1$*, i.e.,* $g(x) = (a_0 + a_1 x + \cdots + a_n x^n)^s$ *with* $s \geq 1$*. Suppose the following condition is satisfied:*

$$g\left(\exp\left(t \sum_{j=1}^{m} \mathcal{X}_j\right)\right) \geq \exp\left(tg\left(\sum_{j=1}^{m} \mathcal{X}_j\right)\right) \quad \textit{almost surely}, \tag{140}$$

where $t > 0$*. Moreover, we require*

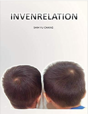
$$\mathcal{X}_i \geq \mathcal{O} \quad \textit{and} \quad \lambda_{\max}(\mathcal{X}_i) \leq \mathrm{R} \quad \textit{almost surely.} \tag{141}$$

Then we have the following inequality:

$$\Pr\left(\left\| g\left(\sum_{j=1}^{m} \mathcal{X}_j\right)\right\|_{(k)} \geq \theta\right) \leq (n+1)^{s-1} \inf_{t>0} e^{-\theta t} \cdot$$

$$\left\{ k a_0^s + \sum_{\ell=1}^{n} \sum_{j=1}^{m} \frac{k a_\ell^{\ell s}}{m} \left[1 + \left(e^{m\ell sRt} - 1\right) \overline{\sigma_1(\mathcal{X}_j)} + C\left(e^{m\ell sRt} - 1\right) \Xi(\mathcal{X}_j)\right] \right\} \tag{142}$$

where $\overline{\sigma_1(\mathcal{X}_j)} \stackrel{\text{def}}{=} \left[\sigma_1\left(\frac{\mathcal{X}_j + \mathcal{X}_j^*}{2}\right) + \sigma_1\left(\frac{\mathcal{X}_j - \mathcal{X}_j^*}{2}\right)\right]$.



Generalized T-product Tensor Bernstein Bounds

Theorem 48 (Generalized T-product Tensor Bernstein Bound)

Consider a sequence $\{\mathcal{X}_j \in \mathbb{C}^{m' \times m' \times p}\}$ *of independent, random, Hermitian tensors. Let* g *be a polynomial function with degree* n *and nonnegative coefficients* $a_0, a_1, \ldots, a_n$ *raised by power* $s \geq 1$*, i.e.,* $g(x) = (a_0 + a_1 x + \cdots + a_n x^n)^s$ *with* $s \geq 1$*. Suppose the following condition is satisfied:*

$$g\left(\exp\left(t\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \geq \exp\left(tg\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right) \quad \text{almost surely}, \qquad (143)$$

where $t > 0$*, and we also have*

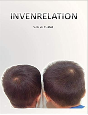
$$\mathbb{E}\mathcal{X}_j = \mathcal{O} \quad \text{and} \quad \mathcal{X}_j^p \leq \frac{p!\mathcal{A}_j^2}{2} \quad \text{almost surely for } p = 2, 3, 4, \ldots. \qquad (144)$$

Then we have the following inequality:

$$\Pr\left(\left\|g\left(\sum_{j=1}^{m}\mathcal{X}_j\right)\right\|_{(k)} \geq \theta\right) \leq (n+1)^{s-1}\inf_{t>0} e^{-\theta t} k \cdot$$

$$\left\{a_0^s + \sum_{\ell=1}^{n}\sum_{j=1}^{m} a_\ell^{\ell s}\left[\frac{1}{m} + \frac{m(lst)^2\sigma_1(\mathcal{A}_j^2)}{2(1-mlst)} + lstC\Upsilon(\mathcal{X}_j)\right]\right\}, \qquad (145)$$

where C *is a constant and* $\Upsilon(\mathcal{X}_j)$ *is determined from the expectation of entries from the tensor* $\mathcal{X}_j$*.*



T-product Tensor Expander Chernoff Bound

Theorem 49 (T-product Tensor Expander Chernoff Bound)

Let $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E})$ be a regular undirected graph whose transition matrix has second eigenvalue λ, and let $g : \mathfrak{V} \rightarrow \mathbb{R}^{m \times m \times p}$ be a function. We assume following:

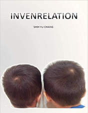
1. *A nonnegative coefficients polynomial raised by the power $s \geq 1$ as $f : x \rightarrow (a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n)^s$ satisfying $f\left(\exp\left(t \sum_{j=1}^{\kappa} g(v_j)\right)\right) \succeq \exp\left(tf\left(\sum_{j=1}^{\kappa} g(v_j)\right)\right)$ almost surely;*
2. *For each $v \in \mathfrak{V}$, $g(v)$ is a symmetric T-product tensor with $f\left(\sum_{j=1}^{\kappa} g(v_j)\right)$ as TPD T-product tensor;*
3. *$\|g(v)\| \leq r$;*
4. *For $\tau \in [\infty, \infty]$, we have constants C and σ such that $\beta_0(\tau) \leq \frac{C}{\sigma\sqrt{2\pi}} \exp\left(\frac{-\tau^2}{2\sigma^2}\right)$.*

Then, we have

$$\Pr\left(\left\| f\left(\sum_{j=1}^{\kappa} g(v_j)\right)\right\|_{(k)} \geq \vartheta\right) \leq$$

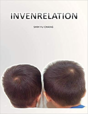
$$\min_{t>0}\left\{(n+1)^{(s-1)} e^{-\vartheta t}\left[a_0 k + C\left(mp + \sqrt{\frac{(mp-k)mp}{k}}\right)\cdot \sum_{l=1}^{n} a_l \exp\left(8\kappa\overline{\lambda} + 2(\kappa + 8\overline{\lambda})lsrt + 2(\sigma(\kappa+8\overline{\lambda})lsr)^2 t^2\right)\right]\right\}, \quad (146)$$

where $\overline{\lambda} = 1 - \lambda$.

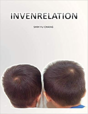


Conclusions and future works

- We established tail bounds for random tensors under situations with independent sum, dependent sum, and function of sum (Einstein product and T-product).
- Three main techniques used here are trace concavity method, majorization approach, and Markov chain embedding.
- Tightness of these bounds.
- Non-linear form.
- Applications to numerical computations, data science, etc.

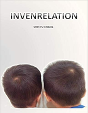


Questions?



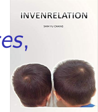
References I

- Shih Yu Chang, *General tail bounds for random tensors summation: Majorization approach*, 2021.
- ———, *T-product tensor expander chernoff bound*, 2021.
- ———, *Tensor expander chernoff bounds*, 2021.
- Shih Yu Chang and Wei Wen Lin, *Convenient tail bounds for sums of random tensors*, accepted by *Taiwanese J. Math.*, 2021.
- Shih Yu Chang and Yimin Wei, *Generalized t-product tensor bernstein bounds*, 2021.
- ———, *T product tensors part i: Inequalities*, 2021.
- ———, *T product tensors part ii: Tail bounds for sums of random t product tensors*, 2021.
- Ky Fan and Alan J Hoffman, *Some metric inequalities in the space of matrices*, Proceedings of the American Mathematical Society **6** (1955), no. 1, 111–116.



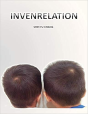
References II

- David Gillman, *A chernoff bound for random walks on expander graphs*, SIAM Journal on Computing **27** (1998), no. 4, 1203–1220.
- Fumio Hiai, Robert König, and Marco Tomamichel, *Generalized log-majorization and multivariate trace inequalities*, Annales Henri Poincaré, vol. 18, Springer, 2017, pp. 2499–2521.
- De Huang, *Generalizing lieb's concavity theorem via operator interpolation*, Advances in Mathematics **369** (2020), 107208.
- Misha E Kilmer and Carla D Martin, *Factorization strategies for third-order tensors*, Linear Algebra and its Applications **435** (2011), no. 3, 641–658.
- Maolin Liang and Bing Zheng, *Further results on Moore-Penrose inverses of tensors with application to tensor nearness problems*, Comput. Math. Appl. **77** (2019), no. 5, 1282–1293. MR 3913666
- Joel A. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. **12** (2012), no. 4, 389–434. MR 2946459



References III

Joel A Tropp, *Matrix concentration & computational linear algebra*.



Thank you very much!

Please email shihyu.chang@sjsu.edu for any further question.

Shih Yu Chang
Department of Applied Data Science
San Jose State University, USA