Sharp Concentration of Simple Random Tensors

Jiaheng Chen

Computational and Applied Mathematics, UChicago

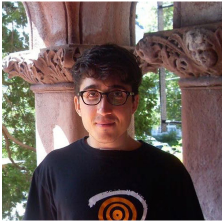
May 22, 2025

Papers and Collaborators

Papers:

1. Sharp concentration of simple random tensors
   https://arxiv.org/abs/2502.16916
2. Sharp concentration of simple random tensors II: asymmetry
   To appear soon

Omar Al-Ghattas
UChicago $\rightarrow$ Broad Institute of MIT and Harvard



Daniel Sanz-Alonso
UChicago

Outline

Setup

Simple (rank-one) tensor:

$$(v_1 \otimes \cdots \otimes v_p)_{i_1,\ldots,i_p} = (v_1)_{i_1} \cdots (v_p)_{i_p}, \quad v^{\otimes p} = \underbrace{v \otimes \cdots \otimes v}_{p}$$

Tensor space:

$$(\mathbb{R}^d)^{\otimes p} = \underbrace{\mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d}_{p} = \mathrm{span}\left\{v_1 \otimes \cdots \otimes v_p : v_1, \ldots, v_p \in \mathbb{R}^d\right\}$$

Frobenius inner product: for $T, W \in (\mathbb{R}^d)^{\otimes p}$,

$$\langle T, W \rangle := \sum_{i_1,\ldots,i_p \in [d]} T_{i_1,\ldots,i_p} W_{i_1,\ldots,i_p}$$

Injective norm: for $T \in (\mathbb{R}^d)^{\otimes p}$,

$$\|T\| := \sup_{v_1,\ldots,v_p \in B_2^d} \langle T, v_1 \otimes \cdots \otimes v_p \rangle, \quad B_2^d = \{v \in \mathbb{R}^d : \|v\|_2 \leq 1\}$$

Remark: $p = 2$: matrix operator norm

Main problems

Problem 1 (sample moment tensor)

Let $X, X_1, \ldots, X_N$ be i.i.d. centered sub-Gaussian random vectors in $\mathbb{R}^d$ with covariance Σ.

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim ?$$

Problem 2 (simple random tensor)

For $p \geq 2$ and $1 \leq k \leq p$, let $X^{(k)}, X_1^{(k)}, \ldots, X_N^{(k)}$ be i.i.d. centered sub-Gaussian random vectors in $\mathbb{R}^{d_k}$ with covariance $\Sigma^{(k)}$.

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{(1)} \otimes \cdots \otimes X_i^{(p)} - \mathbb{E} X^{(1)} \otimes \cdots \otimes X^{(p)} \right\| \lesssim ?$$

Remark: $p = 2$: covariance matrix, cross-covariance matrix

Main results: random tensors

Main result 1: sample moment tensor

Theorem 1 (Al-Ghattas, C., Sanz-Alonso, 2025)

Let $X, X_1, \ldots, X_N$ be i.i.d. centered sub-Gaussian random vectors in $\mathbb{R}^d$ with covariance Σ. For any integer $p \geq 2$ and any $u \geq 1$, it holds with probability at least $1 - \exp(-u)$ that

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim_p \|\Sigma\|^{p/2} \left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N} + \sqrt{\frac{u}{N}} + \frac{u^{p/2}}{N} \right),$$

where the *effective rank* $r(\Sigma) := \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|}$. As a corollary,

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim_p \|\Sigma\|^{p/2} \left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N} \right).$$

If X is Gaussian, then

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \asymp_p \|\Sigma\|^{p/2} \left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N} \right).$$

Remarks

- **Sharpness:** Achieves optimal tail bound and expectation bound without extra logarithmic factors
- **Dimension free:** Bound is independent of ambient dimension d; effective rank $r(\Sigma) = \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|} \in [1, d]$ quantifies the rate of eigenvalue decay of Σ
- **Generality:** Holds for random variables in general separable Hilbert spaces
- Sub-Gaussian random vector: $\|\langle X, v\rangle\|_{\psi_2} \leq \boldsymbol{K}\|\langle X, v\rangle\|_{L_2}, \forall\ v \in \mathbb{R}^d$
  Dependence on K:

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E}X^{\otimes p}\right\| \lesssim_p \boldsymbol{K^p}\|\Sigma\|^{p/2}\left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N}\right)$$

Related literature

$p = 2$ (covariance) [Kannan, Lovaśz, Simonovits 1997], [Bourgain 1998], [Rudelson 1999], [Giannopoulos, Milman 2000], [Paouris 2006], [Guédon, Rudelson 2007], [Mendelson 2008], [Vershynin 2010], [Adamczak, Litvak, Pajor, Tomczak-Jaegermann 2010], **[Koltchinskii, Lounici 2014]**, [Tropp 2015], [Minsker 2017], **[van Handel 2017]**, [Liaw, Mehrabian, Plan, Vershynin 2017], [Tikhomirov 2018], **[Han 2022]**, etc

$p \geq 2$ [Giannopoulos, Milman 2000], **[Guédon, Rudelson 2007]**, [Mendelson 2008], [Adamczak, Litvak, Pajor, Tomczak-Jaegermann 2010], [Vershynin 2011], [Mendelson 2021], **[Even, Massoulié 2021]**, **[Zhivotovskiy 2021]**, [Bartl, Mendelson 2025], etc

Related work on random tensors

[Vershynin 2020], [Zhou, Zhu 2021], [Bamberger, Krahmer, Ward 2022], [Jiang 2022], [Bandeira, Boedihardjo, van Handel 2023], [Bandeira, Gopi, Jiang, Lucca, Rothvoss 2024], [Boedihardjo 2024], etc

Related literature: $p = 2$ (covariance)

- [Koltchinskii, Lounici 2014]: for centered Gaussian data

$$\mathbb{E}\left\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \mathbb{E}\, X \otimes X \right\| \asymp \|\Sigma\| \left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)}{N} \right)$$

method: quadratic empirical processes [Klartag, Mendelson 2005], [Mendelson 2010], [Dirksen 2013], [Bednorz 2014]

- [van Handel 2017]: decoupling and Gaussian comparison inequalities
- [Han 2022]: Gaussian min-max theorem, for centered Gaussian data

$$\mathbb{E}\left\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \mathbb{E}\, X \otimes X \right\| \le \|\Sigma\| \left(1 + \frac{C}{\sqrt{r(\Sigma)}} \right) \left(2\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)}{N} \right)$$

Remark: When $p = 2$, our result matches the optimal bounds (both in probability and in expectation) known for sample covariance matrices.

Related literature: $p \geq 2$

- [Guédon, Rudelson 2007]: use majorizing measures

$$\mathbb{E}\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim_p \|\Sigma\|^{p/2} \left(\sqrt{\varepsilon} + \varepsilon \right),$$

where $\varepsilon = \frac{\log N}{N} \cdot \frac{\mathbb{E}\max_{1 \leq i \leq N} \|X_i\|^p}{\|\Sigma\|^{p/2}} \asymp_p \frac{(\log N)(r(\Sigma)^{p/2} + (\log N)^{p/2})}{N}$.

- [Even, Massoulié 2021]: estimate the metric entropy of ellipsoids

$$\mathbb{E}\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim_p \|\Sigma\|^{p/2} \sqrt{\frac{(\log N)^p (r(\Sigma) + \log d)^{p+1}}{N}}.$$

- [Zhivotovskiy 2021]: PAC-Bayesian methods, if $N \geq r(\Sigma)^{p-1}$, then w.p. $\geq 1 - N\exp(-\sqrt{r(\Sigma)})$,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim_p \|\Sigma\|^{p/2} \sqrt{\frac{r(\Sigma)}{N}}.$$

Remark: Our result improves upon existing bounds and is optimal for sub-Gaussian distributions.

Main result 2: simple random tensor

Theorem 2 (C., Sanz-Alonso, 2025)

For any integer $p \geq 2$ and $1 \leq k \leq p$, let $X^{(k)}, X_1^{(k)}, \ldots, X_N^{(k)}$ be i.i.d. centered sub-Gaussian random vectors in $\mathbb{R}^{d_k}$ with covariance $\Sigma^{(k)}$. Then,

$$\mathbb{E}\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{(1)} \otimes \cdots \otimes X_i^{(p)} - \mathbb{E} X^{(1)} \otimes \cdots \otimes X^{(p)} \right\| \lesssim_p \left(\prod_{k=1}^{p} \|\Sigma^{(k)}\|^{1/2} \right) \mathscr{E}_N\big((\Sigma^{(k)})_{k=1}^p \big),$$

where

$$\mathscr{E}_N\big((\Sigma^{(k)})_{k=1}^p \big) := \left(\frac{\sum_{k=1}^p r(\Sigma^{(k)})}{N} \right)^{1/2} + \frac{1}{N} \prod_{k=1}^{p} \Big(r(\Sigma^{(k)}) + \log N \Big)^{1/2}.$$

Moreover, the upper bound can be reversed (up to a constant) if $(X_i^{(1)})_{i=1}^N, \ldots, (X_i^{(p)})_{i=1}^N, X^{(1)}, \ldots, X^{(p)}$ are independent Gaussian.

Remarks

- If $X^{(1)} = \cdots = X^{(p)}$, $\Sigma^{(1)} = \cdots = \Sigma^{(p)}$, and $X_i^{(1)} = \cdots = X_i^{(p)}$ for all $1 \le i \le N$, then Thm 2 recovers the upper bound in Thm 1:

$$\mathcal{E}_N\big((\Sigma^{(k)})_{k=1}^p\big) \asymp_p \sqrt{\frac{r(\Sigma)}{N}} + \frac{(r(\Sigma)+\log N)^{p/2}}{N} \asymp_p \sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N}.$$

- Assume $r(\Sigma^{(1)}) \ge \cdots \ge r(\Sigma^{(p)})$ and $r(\Sigma^{(t)}) \ge \log N \ge r(\Sigma^{(t+1)})$.

$$\mathbb{E}\left\| \frac{1}{N}\sum_{i=1}^N X_i^{(1)} \otimes \cdots \otimes X_i^{(p)} - \mathbb{E}X^{(1)} \otimes \cdots \otimes X^{(p)} \right\|$$

$$\lesssim_p \left(\prod_{k=1}^p \|\Sigma^{(k)}\|^{1/2}\right)\left(\left(\frac{r(\Sigma^{(1)})}{N}\right)^{1/2} + \frac{(\log N)^{(p-t)/2}}{N}\prod_{k=1}^t r(\Sigma^{(k)})^{1/2}\right).$$

- **Example:** $p = 3$, $r(\Sigma^{(1)}) = r(\Sigma^{(2)}) = d \gg \log N$, $r(\Sigma^{(3)}) = O(1)$ $t = 2$, the error is $\left(\frac{d}{N}\right)^{1/2} + \frac{\sqrt{\log N}}{N} d$, consistency: $N \gg d\sqrt{\log N}$

- **New phenomenon arises only when $p \ge 3$!**

Main results: empirical processes

Motivation

Facts:

1. $\left\| \frac{1}{N}\sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E}X^{\otimes p} \right\| = \sup_{\|v\|=1} \left| \frac{1}{N}\sum_{i=1}^{N} \langle X_i, v\rangle^p - \mathbb{E}\langle X, v\rangle^p \right|$

2. $\left\| \frac{1}{N}\sum_{i=1}^{N} X_i^{(1)} \otimes \cdots \otimes X_i^{(p)} - \mathbb{E}\, X^{(1)} \otimes \cdots \otimes X^{(p)} \right\|$
   $= \sup_{\|v_1\|=\cdots=\|v_p\|=1} \left| \frac{1}{N}\sum_{i=1}^{N} \prod_{k=1}^{p} \langle X_i^{(k)}, v_k\rangle - \mathbb{E}\prod_{k=1}^{p} \langle X^{(k)}, v_k\rangle \right|$

Problem 3 (L_p empirical process)

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{N}\sum_{i=1}^{N} f^p(X_i) - \mathbb{E}f^p(X) \right| \lesssim ?$$

Problem 4 (multi-product empirical process)

$$\sup_{f^{(k)}\in\mathcal{F}^{(k)}, 1\le k\le p} \left| \frac{1}{N}\sum_{i=1}^{N} \prod_{k=1}^{p} f^{(k)}(X_i^{(k)}) - \mathbb{E}\prod_{k=1}^{p} f^{(k)}(X^{(k)}) \right| \lesssim ?$$

Complexity parameters of function class

Talagrand's γ functional: Let $(\mathcal{F}, d)$ be a metric space. An admissible sequence is an increasing sequence $(\mathcal{F}_s)_{s\geq 0} \subset \mathcal{F}$ which satisfies $\mathcal{F}_s \subset \mathcal{F}_{s+1}$, $|\mathcal{F}_0| = 1, |\mathcal{F}_s| \leq 2^{2^s}$ for $s \geq 1$, and $\bigcup_{s=0}^{\infty} \mathcal{F}_s$ is dense in $\mathcal{F}$. Let

$$\gamma(\mathcal{F}, d) := \inf \sup_{f\in\mathcal{F}} \sum_{s\geq 0} 2^{s/2} d\left(f, \mathcal{F}_s\right),$$

where the infimum is taken over all admissible sequences. We write $\gamma(\mathcal{F}, \psi_2)$ when the distance on $\mathcal{F}$ is induced by the ψ_2-norm.

ψ_2-diameter: $d_{\psi_2}(\mathcal{F}) := \sup_{f\in\mathcal{F}} \|f\|_{\psi_2}$

L_p empirical process

Theorem 3 (Al-Ghattas, C., Sanz-Alonso, 2025)

For any $p \geq 2$ and $u \geq 1$, it holds with probability at least $1 - \exp(-u^2(\gamma(\mathcal{F}, \psi_2)/d_{\psi_2}(\mathcal{F}))^2)$ that

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{i=1}^{N} f^p(X_i) - \mathbb{E} f^p(X) \right| \lesssim_p u \frac{\gamma(\mathcal{F}, \psi_2) d_{\psi_2}^{p-1}(\mathcal{F})}{\sqrt{N}} + u^p \frac{\gamma^p(\mathcal{F}, \psi_2)}{N}.$$

As a corollary,

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{i=1}^{N} f^p(X_i) - \mathbb{E} f^p(X) \right| \lesssim_p \frac{\gamma(\mathcal{F}, \psi_2) d_{\psi_2}^{p-1}(\mathcal{F})}{\sqrt{N}} + \frac{\gamma^p(\mathcal{F}, \psi_2)}{N}.$$

Remark: $p = 2$: quadratic empirical processes [Kannan, Lovaśz, Simonovits 1997], [Bourgain 1998], [Rudelson 1999], [Giannopoulos, Milman 2000], [Klartag, Mendelson 2005], [Mendelson, Pajor, Tomczak, Jaegermann 2007], [Mendelson 2010], [Mendelson, Paouris 2011], [Dirksen 2013], [Bednorz 2014], [Mendelson 2014]

Multi-product empirical process

Theorem 4 (C., Sanz-Alonso, 2025)

For $1 \le k \le p$, let $X^{(k)}, X_1^{(k)}, \ldots, X_N^{(k)} \overset{\text{i.i.d.}}{\sim} \mu^{(k)}$ be a sequence of random variables on the probability space $(\Omega^{(k)}, \mu^{(k)})$, and let $\mathcal{F}^{(k)}$ be a class of functions defined on $(\Omega^{(k)}, \mu^{(k)})$. Then,

$$\mathbb{E} \sup_{f^{(k)} \in \mathcal{F}^{(k)}, 1 \le k \le p} \left| \frac{1}{N} \sum_{i=1}^{N} \prod_{k=1}^{p} f^{(k)}(X_i^{(k)}) - \mathbb{E} \prod_{k=1}^{p} f^{(k)}(X^{(k)}) \right| \lesssim_p \left(\prod_{k=1}^{p} d_{\psi_2}(\mathcal{F}^{(k)}) \right) \mathscr{E}_N((\mathcal{F}^{(k)})_{k=1}^{p}),$$

where

$$\mathscr{E}_N((\mathcal{F}^{(k)})_{k=1}^{p}) := \frac{\sum_{k=1}^{p} \bar{\gamma}(\mathcal{F}^{(k)}, \psi_2)}{\sqrt{N}} + \frac{\prod_{k=1}^{p} (\bar{\gamma}(\mathcal{F}^{(k)}, \psi_2) + (\log N)^{1/2})}{N}$$

$$\bar{\gamma}(\mathcal{F}^{(k)}, \psi_2) := \frac{\gamma(\mathcal{F}^{(k)}, \psi_2)}{d_{\psi_2}(\mathcal{F}^{(k)})}.$$

Remark: $p = 2$: product empirical processes [Mendelson 2014]

Proof sketch

Empirical process $\Rightarrow$ Tensor concentration

Let the function class $\mathcal{F} := \{\langle \cdot, v\rangle : \|v\|_2 = 1\}$.

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E}X^{\otimes p}\right\| &= \mathbb{E}\sup_{\|v\|_2=1}\left|\frac{1}{N}\sum_{i=1}^{N}\langle X_i, v\rangle^p - \mathbb{E}\langle X, v\rangle^p\right| \\
&= \mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{i=1}^{N} f^p(X_i) - \mathbb{E}f^p(X)\right| \\
&\lesssim_p \frac{\gamma(\mathcal{F},\psi_2)\, d_{\psi_2}^{p-1}(\mathcal{F})}{\sqrt{N}} + \frac{\gamma^p(\mathcal{F},\psi_2)}{N} \\
&\overset{(\star)}{\asymp}_p \|\Sigma\|^{p/2}\left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N}\right)
\end{aligned}$$

$(\star)$: $d_{\psi_2}(\mathcal{F}) \asymp \|\Sigma\|^{1/2}$

$\gamma(\mathcal{F},\psi_2) \asymp \mathrm{Tr}(\Sigma)^{1/2}$ (Talagrand's majorizing-measure theorem)

Proof outline: L_p empirical process

High-level summary: Based on a chaining framework developed by [Mendelson 2014] for analyzing multiplier and product empirical processes, extend it to L_p and multi-product empirical processes

- Symmetrization (Giné-Zinn inequality):

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{i=1}^{N} f^p(X_i) - \mathbb{E}f^p(X)\right| \lesssim \sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{N}\varepsilon_i f^p(X_i)\right|$$

- Hoeffding's inequality: for any $z=(z_i)_{i=1}^N$ and any set I, w.p. $\geq 1-2\exp(-t^2/2)$

$$\left|\sum_{i=1}^{N}\varepsilon_i z_i\right| \leq \sum_{i\in I}|z_i| + t\left(\sum_{i\in I^c} z_i^2\right)^{1/2}.$$

In particular,

$$\left|\sum_{i=1}^{N}\varepsilon_i z_i\right| \leq \sum_{i=1}^{k}|z_i^*| + t\left(\sum_{i=k+1}^{N} (z_i^*)^2\right)^{1/2},$$

where $(z_i^*)_{i=1}^N$ denotes the non-increasing rearrangement of $(|z_i|)_{i=1}^N$. This estimate is optimal when $k \asymp t^2$.

Proof outline: L_p empirical process

1. Let $(\mathcal{F}_s)_{s\geq 0} \subset \mathcal{F}$ be an admissible sequence, $|\mathcal{F}_s| \leq 2^{2^s}$, let $\pi_s f$ denote the projection of f to $\mathcal{F}_s$.

2. Chaining: $f^p = \sum_{s\geq s_0}((\pi_{s+1}f)^p - (\pi_s f)^p) + (\pi_{s_0} f)^p$

$$\left|\sum_{i=1}^N \varepsilon_i f^p(X_i)\right| \leq \sum_{s\geq s_0}\left|\sum_{i=1}^N \varepsilon_i((\pi_{s+1}f)^p - (\pi_s f)^p)(X_i)\right| + \left|\sum_{i=1}^N \varepsilon_i(\pi_{s_0}f)^p(X_i)\right|$$

3. Fix $f \in \mathcal{F}$ and $(X_i)_{i=1}^N$, with $(\varepsilon_i)_{i=1}^N$ probability $\geq 1 - 2\exp(-u^2 2^s)$:

$$\left|\sum_{i=1}^N \varepsilon_i((\pi_{s+1}f)^p - (\pi_s f)^p)_i\right|$$

Hoeffding : $t = u2^{s/2}$

$$\leq \sum_{i\in I_s}|((\pi_{s+1}f)^p - (\pi_s f)^p)_i| + u2^{\frac{s}{2}}\left(\sum_{i\in I_s^c}((\pi_{s+1}f)^p - (\pi_s f)^p)_i^2\right)^{\frac{1}{2}}$$

$\Delta_s f := \pi_{s+1}f - \pi_s f$

$$\lesssim \sum_{i\in I_s}|(\Delta_s f)_i||(\pi_s f)_i|^{p-1} + u2^{\frac{s}{2}}\left(\sum_{i\in I_s^c}(\Delta_s f)_i^2(\pi_s f)_i^{2(p-1)}\right)^{\frac{1}{2}}$$

Cauchy-Schwarz

$$\leq \left(\sum_{i\in I_s}(\Delta_s f)_i^2\right)^{\frac{1}{2}}\left(\sum_{i\in I_s}(\pi_s f)_i^{2(p-1)}\right)^{\frac{1}{2}} + u2^{\frac{s}{2}}\left(\sum_{i\in I_s^c}(\Delta_s f)_i^4\right)^{\frac{1}{4}}\left(\sum_{i\in I_s^c}(\pi_s f)_i^{4(p-1)}\right)^{\frac{1}{4}}$$

$$\leq \left(\sum_{i\in I_s}(\Delta_s f)_i^2\right)^{\frac{1}{2}}\sup_{f\in\mathcal{F}}\left(\sum_{i=1}^N f_i^{2(p-1)}\right)^{\frac{1}{2}} + u2^{\frac{s}{2}}\left(\sum_{i\in I_s^c}(\Delta_s f)_i^4\right)^{\frac{1}{4}}\left(\sum_{i\in I_s^c}(\pi_s f)_i^{4(p-1)}\right)^{\frac{1}{4}}$$

Proof outline: L_p empirical process

4. For a non-decreasing sequence $(j_s)_{s\geq s_0}$ (to be defined later), choose I_s to be the union of the j_s-1 largest coordinates of $(|(\Delta_s f)_i|)_{i=1}^N$ and the j_s-1 largest coordinates of $(|(\pi_s f)_i|)_{i=1}^N$.

5. Fix $f\in\mathcal{F}$ and $(X_i)_{i=1}^N$, with $(\varepsilon_i)_{i=1}^N$ probability $\geq 1-2\exp(-u^2 2^s)$:

$$\Big|\sum_{i=1}^N \varepsilon_i((\pi_{s+1}f)^p-(\pi_s f)^p)_i\Big|$$

$$\lesssim \Big(\sum_{i\in I_s}(\Delta_s f)_i^2\Big)^{\frac12}\sup_{f\in\mathcal{F}}\Big(\sum_{i=1}^N f_i^{2(p-1)}\Big)^{\frac12}+u2^{\frac{s}{2}}\Big(\sum_{i\in I_s^c}(\Delta_s f)_i^4\Big)^{\frac14}\Big(\sum_{i\in I_s^c}(\pi_s f)_i^{4(p-1)}\Big)^{\frac14}$$

$$\lesssim \Big(\sum_{i<j_s}\Big[(\Delta_s f)^2\Big]_i^*\Big)^{1/2}\sup_{f\in\mathcal{F}}\Big(\sum_{i=1}^N f_i^{2(p-1)}\Big)^{1/2}$$

$$+u2^{\frac{s}{2}}\Big(\sum_{i\geq j_s}\big[(\Delta_s f)^4\big]_i^*\Big)^{1/4}\Big(\sum_{i\geq j_s}\Big[(\pi_s f)^{4(p-1)}\Big]_i^*\Big)^{1/4}.$$

Remark: The $(\varepsilon_i)_{i=1}^N$ probability estimate (apply a union bound for all $f\in\mathcal{F}$ and all $s\geq s_0$):

$$1-\sum_{s\geq s_0}2\exp(-u^2 2^s)\cdot 2^{2^{s+2}}\geq 1-2\exp(-u^2 2^{s_0}/8)$$

Lemma (order statistics, adapted from [Mendelson 2014]): Define

$$j_s := \min\left\{ \left\lceil \frac{c_0 u^2 2^s}{\log(4 + eN/u^2 2^s)} \right\rceil, N+1 \right\}.$$

$\left(\frac{eN}{j}\right)^{-j} \leq \exp(-u^2 2^s)$ if $j \geq j_s$

With $(X_i)_{i=1}^N$ probability at least $1 - 2\exp(-cu^2 2^{s_0})$, for every $f \in \mathcal{F}$ and $s \geq s_0$,

$$\left(\sum_{i<j_s} \left[(\Delta_s f)^2 (X_i) \right]^* \right)^{\frac{1}{2}} \lesssim u 2^{\frac{s}{2}} \|\Delta_s f\|_{(u^2 2^s)}, \left(\sum_{i \geq j_s} \left[(\Delta_s f)^4 (X_i) \right]^* \right)^{\frac{1}{4}} \lesssim N^{\frac{1}{4}} \|\Delta_s f\|_{L_8},$$

$$\left(\sum_{i<j_s} \left[(\pi_s f)^2 (X_i) \right]^* \right)^{\frac{1}{2}} \lesssim u 2^{\frac{s}{2}} \|\pi_s f\|_{(u^2 2^s)}, \left(\sum_{i \geq j_s} \left[(\pi_s f)^{4(p-1)} (X_i) \right]^* \right)^{\frac{1}{4(p-1)}} \lesssim N^{\frac{1}{4(p-1)}} \|\pi_s f\|_{L_{8(p-1)}}$$

The graded L_q norm: $\|f\|_{(q)} := \sup_{1 \leq p \leq q} \frac{\|f\|_{L_p}}{\sqrt{p}} \lesssim \|f\|_{\psi_2}$

Proposition (Al-Ghattas, C., Sanz-Alonso, 2025)

For any $m \geq 2$ and $u \geq 1$, it holds with probability at least $1 - 2\exp(-u^2)$ that

$$\sup_{f \in \mathcal{F}} \left(\sum_{i=1}^N |f|^m (X_i) \right)^{1/m} \lesssim_m \gamma(\mathcal{F}, \psi_2) + N^{1/m} d_{\psi_2}(\mathcal{F}) + d_{\psi_2}(\mathcal{F}) u.$$

Proof: generic chaining + ideas from [Bednorz 2014] for analyzing quadratic processes, along with α-sub-exponential inequalities [Götze, Sambale, Sinulis 2019]

Proof outline: multi-product empirical process

High-level summary:

Mendelson's framework (order statistics) + a new proposition

Proposition (C., Sanz-Alonso, 2025)

$$\mathbb{E} \sup_{f^{(k)} \in \mathcal{F}^{(k)}, 1 \le k \le p} \left(\sum_{i=1}^{N} \prod_{k=1}^{p} \left(f^{(k)}(X_i^{(k)}) \right)^2 \right)^{1/2}$$

$$\lesssim_p \left(\prod_{k=1}^{p} d_{\psi_2}(\mathcal{F}^{(k)}) \right) \left(\sqrt{N} + \prod_{k=1}^{p} \left(\bar{\gamma}(\mathcal{F}^{(k)}, \psi_2) + (\log N)^{1/2} \right) \right).$$

Remarks:

1. Sharp bound, $\log N$ factors cannot be removed
2. $\log N$ factors arise from the ℓ_∞-norm bound: for any $u > 0$, it holds w.p. $\ge 1 - \exp(-u^2)$ that

$$\sup_{f \in \mathcal{F}} \max_{1 \le i \le N} |f(X_i)| \lesssim \gamma(\mathcal{F}, \psi_2) + d_{\psi_2}(\mathcal{F})(\sqrt{\log N} + u).$$

Lower bound in the asymmetric case

Example: $p = 3$: $X \sim \mathcal{N}(0, I_d), Y \sim \mathcal{N}(0, I_d), Z \sim \mathcal{N}(0, 1)$.

A lower bound:

$$\mathbb{E}\left\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes Y_i \otimes Z_i - \mathbb{E}X \otimes Y \otimes Z \right\| = \mathbb{E}\sup_{u,v,w}\left| \frac{1}{N}\sum_{i=1}^{N}\langle X_i, u\rangle\langle Y_i, v\rangle\langle Z_i, w\rangle \right|$$

$$\geq \mathbb{E}_{X_1,Y_1,Z_1}\sup_{u,v,w}\left| \mathbb{E}_{\{X_i,Y_i,Z_i\}_{i\geq 2}} \frac{1}{N}\sum_{i=1}^{N}\langle X_i, u\rangle\langle Y_i, v\rangle\langle Z_i, w\rangle \right|$$

$$= \mathbb{E}\sup_{u,v,w}\frac{1}{N}|\langle X_1, u\rangle\langle Y_1, v\rangle\langle Z_1, w\rangle| = \frac{1}{N}\mathbb{E}\left[\|X_1\|\|Y_1\|\|Z_1\|\right] \sim \frac{d}{N}$$

Improved lower bound: there exists some index i_* such that $|Z_{i_*}| \asymp \sqrt{\log N}$

$$\mathbb{E}\sup_{u,v,w}\left| \frac{1}{N}\sum_{i=1}^{N}\langle X_i, u\rangle\langle Y_i, v\rangle\langle Z_i, w\rangle \right| \geq \mathbb{E}_{i_*}\sup_{u,v,w}\left| \mathbb{E}_{\sim i_*}\frac{1}{N}\sum_{i=1}^{N}\langle X_i, u\rangle\langle Y_i, v\rangle\langle Z_i, w\rangle \right|$$

$$= \frac{1}{N}\mathbb{E}\sup_{u,v,w}|\langle X_{i_*}, u\rangle\langle Y_{i_*}, v\rangle\langle Z_{i_*}, w\rangle| \sim \frac{\sqrt{\log N}}{N}\mathbb{E}\left[\|X_{i_*}\|\|Y_{i_*}\|\right] \sim \frac{d\sqrt{\log N}}{N}$$

Main message: "short" vectors contribute at least a $\sqrt{\log N}$ factor

Summary

Summary

- Sharp dimension-free concentration inequalities for simple random tensors under sub-Gaussian distribution

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E}X^{\otimes p}\right\| \lesssim_p \|\Sigma\|^{p/2}\left(\sqrt{\frac{r(\Sigma)}{N}} + \frac{r(\Sigma)^{p/2}}{N}\right)$$

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} X_i^{(1)} \otimes \cdots \otimes X_i^{(p)} - \mathbb{E}X^{(1)} \otimes \cdots \otimes X^{(p)}\right\|$$
$$\lesssim_p \left(\prod_{k=1}^{p} \|\Sigma^{(k)}\|^{1/2}\right)\left(\left(\frac{\sum_{k=1}^{p} r(\Sigma^{(k)})}{N}\right)^{1/2} + \frac{1}{N}\prod_{k=1}^{p}\left(r(\Sigma^{(k)}) + \log N\right)^{1/2}\right)$$

- New upper bounds on L_p empirical processes and multi-product empirical processes, extend [Mendelson 2014] to higher-order settings
- **Future directions and potential applications:** Tensor data analysis, structured tensor estimation, exact constants, heavy-tailed distributions, cumulant tensors, method of moments, independent component analysis...

Thank you!

Papers:

1. Sharp concentration of simple random tensors.
   https://arxiv.org/abs/2502.16916
2. Sharp concentration of simple random tensors II: asymmetry.
   To appear soon

Back-up slides

Empirical process ⇒ Tensor concentration

Fact 1:

$$d_{\psi_2}(\mathcal{F}) = \sup_{f \in \mathcal{F}} \|f\|_{\psi_2} \asymp \sup_{f \in \mathcal{F}} \|f\|_{L_2} = \sup_{\|v\|_2 = 1} \left(\mathbb{E}\langle X, v\rangle^2\right)^{1/2} = \|\Sigma\|^{1/2}.$$

Fact 2: Let $Y \sim \mathcal{N}(0, \Sigma)$.

$$\gamma(\mathcal{F}, \psi_2) \asymp \gamma(\mathcal{F}, L_2) \overset{(\star)}{=} \gamma(\mathbb{S}^{d-1}, d_Y)$$
$$\overset{(\star\star)}{\asymp} \mathbb{E} \sup_{u \in \mathbb{S}^{d-1}} \langle Y, u\rangle = \mathbb{E}\|Y\| \asymp (\mathbb{E}\|Y\|^2)^{1/2} = \mathrm{Tr}(\Sigma)^{1/2}.$$

$(\star)$: Let μ be the law of X. For any $u, v \in \mathbb{S}^{d-1}$,

$$\begin{aligned}\|\langle \cdot, u\rangle - \langle \cdot, v\rangle\|_{L_2(\mu)} &= \langle u - v, \Sigma(u - v)\rangle^{1/2} \\ &= \left(\mathbb{E}(\langle Y, u\rangle - \langle Y, v\rangle)^2\right)^{1/2} =: d_Y(u, v).\end{aligned}$$

$(\star\star)$: Talagrand's majorizing-measure theorem