THE UNIVERSITY OF CHICAGO

Sharp Concentration of Simple Random Tensors

Omar Al-Ghattas [1]   *Jiaheng Chen* [2]   Daniel Sanz-Alonso [1]

[1]Department of Statistics, University of Chicago   [2]Committee on Computational and Applied Mathematics, University of Chicago

Theorem 1: sample moment tensor

Data: $X, X_1, X_2, \dots, X_N$ i.i.d. mean-zero sub-Gaussian random vectors in $\mathbb{R}^d$ with covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$

Question: What is the difference between the sample moment tensor $\frac{1}{N}\sum_{i=1}^N X_i^{\otimes p}$ and the true moment tensor $\mathbb{E}X^{\otimes p}$?

Spectral norm (ℓ_2 injective norm): For a d-dimensional, p-th order tensor $T \in (\mathbb{R}^d)^{\otimes p}$,

$$\|T\| := \sup_{\|v_1\|_2 = \cdots = \|v_p\|_2 = 1} \langle T, v_1 \otimes \cdots \otimes v_p \rangle.$$

Effective rank: $r(\Sigma) := \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|} \in [1, d]$, quantifies the rate of eigenvalue decay

Theorem 1: For any integer $p \ge 2$ and any $u \ge 1$, it holds with probability at least $1 - \exp(-u)$ that

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

As a corollary,

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

Moreover, 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:

- **Dimension-free:** Bound independent of ambient dimension d
- **Generality:** Holds for random variables in general separable Banach or Hilbert spaces
- **Sharpness:** Achieves optimal tail bound without extra logarithmic factors
- The case $p = 2$ (sample covariance) was obtained by Koltchinskii and Lounici (2014)

Theorem 2: p-th order empirical process

Goal: Study the p-th order empirical process: $\sup_{f\in\mathcal{F}} \left| \frac{1}{N}\sum_{i=1}^N f^p(X_i) - \mathbb{E}f^p(X) \right|$

Motivation:

$$\begin{aligned} \left\| \frac{1}{N}\sum_{i=1}^N X_i^{\otimes p} - \mathbb{E}X^{\otimes p} \right\| &= \sup_{\|v\|_2=1} \left| \left\langle \frac{1}{N}\sum_{i=1}^N X_i^{\otimes p} - \mathbb{E}X^{\otimes p}, v^{\otimes p} \right\rangle \right| \\ &= \sup_{\|v\|_2=1} \left| \frac{1}{N}\sum_{i=1}^N \langle X_i, v\rangle^p - \mathbb{E}\langle X, v\rangle^p \right| \\ &= \sup_{f\in\mathcal{F}} \left| \frac{1}{N}\sum_{i=1}^N f^p(X_i) - \mathbb{E}f^p(X) \right|, \quad \mathcal{F} = \{\langle\cdot, v\rangle : \|v\|_2 = 1\} \end{aligned}$$

Definition (Orlicz ψ_2-norm): For any function f on a probability space (Ω, μ),

$$\|f\|_{\psi_2} := \inf\left\{ c > 0 : \mathbb{E}_{X\sim\mu}\left[\exp\left(\frac{|f(X)|^2}{c^2} \right) \right] \le 2 \right\}.$$

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

Definition (Talagrand's γ functional): Let $(\mathcal{F}, d)$ be a metric space. An admissible sequence is an increasing sequence $(\mathcal{F}_s)_{s\ge0} \subset \mathcal{F}$ which satisfies $\mathcal{F}_s \subset \mathcal{F}_{s+1}$, $|\mathcal{F}_0| = 1$, $|\mathcal{F}_s| \le 2^{2^s}$ for $s \ge 1$, and $\bigcup_{s=0}^\infty \mathcal{F}_s$ is dense in $\mathcal{F}$. Let

$$\gamma(\mathcal{F}, d) := \inf \sup_{f\in\mathcal{F}} \sum_{s\ge0} 2^{s/2} d(f, \mathcal{F}_s),$$

where the infimum is taken over all admissible sequences.

Theorem 2: p-th order empirical process (continued)

Theorem 2: Assume $0 \in \mathcal{F}$ or $\mathcal{F}$ is symmetric (i.e., $f \in \mathcal{F}$ implies $-f \in \mathcal{F}$). For any $p \ge 2$ and $u \ge 1$, it holds with probability at least $1 - \exp(-u)$ that

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{N}\sum_{i=1}^N f^p(X_i) - \mathbb{E}f^p(X) \right| \lesssim_p \frac{\gamma(\mathcal{F}, \psi_2) d_{\psi_2}^{p-1}(\mathcal{F})}{\sqrt{N}} + \frac{\gamma^p(\mathcal{F}, \psi_2)}{N} + d_{\psi_2}^p(\mathcal{F}) \left(\sqrt{\frac{u}{N}} + \frac{u^{p/2}}{N} \right).$$

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: For the case $p = 2$, the expectation bound was established by Klartag and Mendelson (2005) and by Mendelson, Pajor, and Tomczak-Jaegermann (2007); improved tail bounds were later obtained by Dirksen (2013) and Bednorz (2014).

Proof:

- By symmetrization, reduce to analyzing the supremum of the Bernoulli process:

$$\sup_{f\in\mathcal{F}} \left| \sum_{i=1}^N \varepsilon_i f^p(X_i) \right|$$

- Apply chaining methods carefully within the powerful framework developed by Mendelson (2016) for analyzing multiplier and product empirical processes.
- Key lemma: For any $m \ge 2$ and $u \ge 1$, with probability at least $1 - 2\exp(-u^2)$, it holds that

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

- The key lemma itself is proved via chaining arguments, incorporating ideas from Bednorz (2014).

Proof of Theorem 1

- Apply Theorem 2 to the function class $\mathcal{F} = \{\langle\cdot, v\rangle : \|v\|_2 = 1\}$.
- Since X is sub-Gaussian, the ψ_2-norm of linear functionals is equivalent to the L_2-norm:

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

- There exists a centered Gaussian random vector Y in $\mathbb{R}^d$ with the same covariance Σ. For any $u, v \in \mathbb{S}^{d-1}$,

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

where μ is the law of X.

- Talagrand's majorizing-measure theorem:

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

- Consequently,

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

- The lower bound for the Gaussian case builds upon ideas from Koltchinskii and Lounici (2014).

Theorem 3: simple random tensor

Theorem 3: For any integer $p \ge 2$ and $1 \le k \le p$, let $X^{(k)}, X_1^{(k)}, \dots, X_N^{(k)}$ be i.i.d. mean-zero sub-Gaussian random vectors in $\mathbb{R}^{d_k}$ with covariance matrix $\Sigma^{(k)}$. Then,

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

where

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

Moreover, if $(X_i^{(1)})_{i=1}^N, \dots, (X_i^{(p)})_{i=1}^N, X^{(1)}, \dots, X^{(p)}$ are independent Gaussian, then

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

Remarks:

- The sequence of random variables $(X_i^{(k)})_{i=1}^N$ and $(X_i^{(k')})_{i=1}^N$ are not necessarily independent for $k \neq k'$.
- Achieves the optimal expectation bound; the logarithmic factors cannot be removed, as the bound can be reversed in the independent case.
- If $r(\Sigma^{(1)}) \ge \cdots \ge r(\Sigma^{(t)}) \ge \log N \ge r(\Sigma^{(t+1)}) \cdots \ge r(\Sigma^{(p)})$, then

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

- 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 Theorem 3 recovers the upper bound in Theorem 1:

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

Theorem 4: multi-product empirical process

Theorem 4: For $1 \le k \le p$, let $X^{(k)}, X_1^{(k)}, \dots, X_N^{(k)} \overset{\text{i.i.d.}}{\sim} \mu^{(k)}$ 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 $(\Omega^{(k)}, \mu^{(k)})$. Assume $0 \in \mathcal{F}^{(k)}$ or $\mathcal{F}^{(k)}$ is symmetric (i.e., $f^{(k)} \in \mathcal{F}^{(k)}$ implies $-f^{(k)} \in \mathcal{F}^{(k)}$). Then,

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

where

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

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

References

[1] Omar Al-Ghattas, Jiaheng Chen, and Daniel Sanz-Alonso. Sharp concentration of simple random tensors. *arXiv preprint arXiv:2502.16916*, 2025.

[2] Witold Bednorz. Concentration via chaining method and its applications. *arXiv preprint arXiv:1405.0676*, 2014.

[3] Jiaheng Chen and Daniel Sanz-Alonso. Sharp concentration of simple random tensors II: asymmetry. *arXiv preprint*, 2025. To appear.

[4] Vladimir Koltchinskii and Karim Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133, 2017.

[5] Shahar Mendelson. Upper bounds on product and multiplier empirical processes. *Stochastic Processes and their Applications*, 126(12):3652–3680, 2016.