Matrix Concentration Inequalities via the Method of Exchangeable Pairs

Original Paper by Mackey et al.

Presenters: Sehun Kim, Bright Liu, Anthony Rodriguez-Miranda, Dewey To

Harvard University

Stat 212: Probability II
July 15, 2025

Notation and preliminaries

- Let $\mathbb{M}^d$ represent the space of all $d \times d$ complex matrices.
- We use the following constructions of trace and normalized trace of a square matrix,

$$\operatorname{tr} \mathbf{B} := \sum_{j=1}^{d} b_{jj} \text{ and } \overline{\operatorname{tr}}\, \mathbf{B} := \frac{1}{d} \sum_{j=1}^{d} b_{jj} \text{ for } \mathbf{B} \in \mathbb{M}^d$$

- Let $\mathbb{H}^d$ represent the space of Hermitian $d \times d$ matrices, i.e the space of complex square matrices that are equal to their own conjugate transpose. Equivalently, for $\mathbf{A} \in \mathbb{M}^d$, if $\mathbf{A} = \mathbf{A}^*$, then $\mathbf{A} \in \mathbb{H}^d$ as well.

Notation and preliminaries (continued)

- The symbols $\lambda_{\text{max}}(\mathbf{A})$ and $\lambda_{\text{min}}(\mathbf{A})$ refer to the algebraic maximum and minimum eigenvalues of a matrix $\mathbf{A} \in \mathbb{H}^d$.
- For each interval $I \subset \mathbb{R}$, we define the set of Hermitian matrices whose eigenvalues fall in that interval,

$$\mathbb{H}^d(I) := \{\mathbf{A} \in \mathbb{H}^d : [\lambda_{\text{min}}(\mathbf{A}), \lambda_{\text{max}}(\mathbf{A})] \subset I\}.$$

- The set $\mathbb{H}^d_+$ consists of all positive-semidefinite (psd) $d \times d$ matrices, i.e all $\mathbf{B} \in \mathbb{H}^d$ such that $\mathbf{x}^*\mathbf{B}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{C}^d$
  - Note that this necessarily means all eigenvalues are non-negative

Notation and preliminaries (continued)

- Curly inequalities refer to the semidefinite partial order on Hermitian matrices. For example, we write $\mathbf{A} \preccurlyeq \mathbf{B}$ to signify that the matrix $\mathbf{B} - \mathbf{A}$ is positive semidefinite.
- More generally, we have the operator Jensen inequality

$$(\mathbb{E}[\mathbf{X}])^2 \preccurlyeq \mathbb{E}[\mathbf{X}^2],$$

  valid for any random Hermitian matrix, provided that $\mathbb{E}\|\mathbf{X}\|^2 < \infty$.
- We also note that $||\mathbf{B}||$ represents the maximum singular value of $\mathbf{B}$, or the maximum absolute value of its eigenvalues.

Exchangeable pairs of random matrices

Definition (Exchangeable pair)

Let Z and Z' be random variables. We say that (Z, Z') is an exchangeable pair if $(Z, Z') \stackrel{d}{=} (Z', Z)$.

Definition (Matrix Stein pair)

For an exchangeable pair of random variables (Z, Z'), let $\Psi : \mathcal{Z} \to \mathbb{H}^d$ be a measurable function. Define the random Hermitian matrices $\mathbf{X} := \Psi(Z)$ and $\mathbf{X}' := \Psi(Z')$.

We say $(\mathbf{X}, \mathbf{X}')$ is a *matrix Stein pair* if there is a constant scale factor $\alpha \in (0, 1]$ for which

$$\mathbb{E}[\mathbf{X} - \mathbf{X}' | Z] = \alpha \mathbf{X} \quad \text{almost surely.}$$

Exchangeable pairs of random matrices (continued)

When discussing a matrix Stein pair $(\mathbf{X}, \mathbf{X}')$, we always assume that $\mathbb{E}||\mathbf{X}||^2 < \infty$.

We also note two useful properties. First, $(\mathbf{X}, \mathbf{X}')$ always forms an exchangeable pair. Second, it must be the case that $\mathbb{E}[\mathbf{X}] = 0$. Indeed,

$$\mathbb{E}[\mathbf{X}] = \frac{1}{\alpha}\mathbb{E}\left[\mathbb{E}[\mathbf{X} - \mathbf{X}'|Z]\right] = \frac{1}{\alpha}\mathbb{E}[\mathbf{X} - \mathbf{X}'] = 0$$

by definition of matrix Stein pair, the tower property of conditional expectation, and the exchangeability of $(\mathbf{X}, \mathbf{X}')$.

Conditional Variance

Definition (Conditional Variance)

Suppose that $(\mathbf{X}, \mathbf{X}')$ is a matrix Stein pair with scale factor α, constructed from the exchangeable pair (Z, Z'). The conditional variance is the random matrix

$$\Delta_{\mathbf{X}} := \Delta_{\mathbf{X}}(Z) := \frac{1}{2\alpha}\mathbb{E}\left((\mathbf{X} - \mathbf{X}')^2 \,|\, Z\right).$$

- $\Delta_{\mathbf{X}}$ is a stochastic estimate for the variance, $\mathbb{E}\mathbf{X}^2$.
- Control over $\Delta_{\mathbf{X}}$ yields control over $\lambda_{\max}(\mathbf{X})$.

Example of Exchangeable Pair

- $Z := (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ of random Hermitian matrices, $\mathbb{E}\mathbf{Y}_k = \mathbf{0}$ and $\mathbb{E}\left\|\mathbf{Y}_k\right\|^2 < \infty$ for each k. With independent hermitian matrices, we define the sequence and Stein pair as the sequence and the total sum.
- $\mathbf{X} := \mathbf{Y}_1 + \cdots + \mathbf{Y}_n$
- $\mathbf{Y}'_k$ be an independent copy of $\mathbf{Y}_k$, draw a random index K uniformly from $\{1, \ldots, n\}$ $Z' := \left(\mathbf{Y}_1, \ldots, \mathbf{Y}_{K-1}, \mathbf{Y}'_K, \mathbf{Y}_{K+1}, \ldots, \mathbf{Y}_n\right)$
- $\mathbb{E}\left[\mathbf{X} - \mathbf{X}' \mid Z\right] = \mathbb{E}\left[\mathbf{Y}_K - \mathbf{Y}'_K \mid Z\right] = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[\mathbf{Y}_j - \mathbf{Y}'_j \mid Z\right] = \frac{1}{n}\sum_{j=1}^{n} \mathbf{Y}_j = \frac{1}{n}\mathbf{X}$

Conditional Variance Calculation Example

$$
\begin{aligned}
\Delta_{\mathbf{X}} &= \frac{n}{2} \cdot \mathbb{E}\left[\left(\mathbf{X} - \mathbf{X}' \right)^2 \mid Z \right] \\
&= \frac{n}{2} \cdot \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[\left(\mathbf{Y}_k - \mathbf{Y}'_k \right)^2 \mid Z \right] \\
&= \frac{1}{2} \sum_{k=1}^{n} \left[\mathbf{Y}_k^2 - \mathbf{Y}_k \left(\mathbb{E}\mathbf{Y}'_k \right) - \left(\mathbb{E}\mathbf{Y}'_k \right) \mathbf{Y}_k + \mathbb{E} \left(\mathbf{Y}'_k \right)^2 \right] \\
&= \frac{1}{2} \sum_{k=1}^{n} \left(\mathbf{Y}_k^2 + \mathbb{E}\mathbf{Y}_k^2 \right)
\end{aligned}
$$

- Recap: Exploit symmetries of distribution to construct matrix Stein pair
- We utilize independent copies, similar to that of McDiarmid's concentration inequality for the univariate case

Matrix Combinatorial Stein Pair

- Deterministic array $(A_{jk})_{j,k=1}^{n}$ of Hermitian matrices, and let π be uniformly random permutation, we let $\mathbf{Y} := \sum_{j=1}^{n} A_{j\pi(j)}, \mathbb{E}\mathbf{Y} = \frac{1}{n}\sum_{j=1}^{n} A_{jk}, \mathbf{X} := \mathbf{Y} - \mathbb{E}\mathbf{Y}$
- We draw a pair (J, K) of indices uniformly at random from $\{1, \ldots, n\}^2$ and define second random permutation $\pi' := \pi \circ (J, K)$. The pair (π, π') is exchangeable.
- $\mathbf{X}' := \sum_{j=1}^{n} \mathbf{A}_{j\pi'(j)} - \mathbb{E}\mathbf{Y}$ is exchangeable with $\mathbf{X}$.

Matrix Stein Conditional Variance

$$\begin{aligned}\mathbb{E}\left[\mathbf{X}-\mathbf{X}' \mid \pi\right] &= \mathbb{E}\left[\mathbf{A}_{J\pi(J)}+\mathbf{A}_{K\pi(K)}-\mathbf{A}_{J\pi(K)}-\mathbf{A}_{K\pi(J)} \mid \pi\right] \\ &= \frac{1}{n^2}\sum_{j,k=1}^{n}\left[\mathbf{A}_{j\pi(j)}+\mathbf{A}_{k\pi(k)}-\mathbf{A}_{j\pi(k)}-\mathbf{A}_{k\pi(j)}\right] \\ &= \frac{2}{n}(\mathbf{Y}-\mathbb{E}\mathbf{Y}) = \frac{2}{n}\mathbf{X}\end{aligned}$$

$$\Delta_X(\pi) = \frac{n}{4}E[(X-X')^2|\pi] = \frac{1}{4n}\sum_{j,k=1}^{n}[A_{j\pi(j)}+A_{k\pi(k)}-A_{j\pi(k)}-A_{k\pi(j)}]^2$$

- We note the Δ_X is controlled when A_{jk} are bounded.

Method of exchangeable pairs

Theorem

Let $(\mathbf{X}, \mathbf{X}')$ be a matrix Stein pair with scale factor α (where we take $\mathbf{X}, \mathbf{X}'$ to be d-dimensional Hermitian matrices). Given measurable $\boldsymbol{F} : \mathbb{H}^d \to \mathbb{H}^d$ where $\mathbb{E}\|(\mathbf{X} - \mathbf{X}') \cdot \mathbf{F}(\mathbf{X})\|$ is finite, we have

$$\mathbb{E}[\mathbf{X} \cdot \boldsymbol{F}(\mathbf{X})] = \frac{1}{2\alpha}\mathbb{E}[(\mathbf{X} - \mathbf{X}')(\boldsymbol{F}(\mathbf{X}) - \boldsymbol{F}(\mathbf{X}'))].$$

Method of exchangeable pairs

Proof.

By the definition of a matrix Stein pair and the tower rule, we have

$$\alpha\mathbb{E}[\mathbf{X}\cdot\mathbf{F}(\mathbf{X})] = \mathbb{E}[\mathbb{E}[(\mathbf{X}-\mathbf{X}')|Z]\cdot\mathbf{F}(\mathbf{X})] = \mathbb{E}[(\mathbf{X}-\mathbf{X}')\cdot\mathbf{F}(\mathbf{X})].$$

$(\mathbf{X}, \mathbf{X}')$ being exchangeable then gives

$$\mathbb{E}[(\mathbf{X}-\mathbf{X}')\mathbf{F}(\mathbf{X})] = \mathbb{E}[(\mathbf{X}'-\mathbf{X})\mathbf{F}(\mathbf{X}')] = -\mathbb{E}[(\mathbf{X}-\mathbf{X}')\mathbf{F}(\mathbf{X})].$$

Taking the average of the LHS and the RHS above then shows that

$$\mathbb{E}[\mathbf{X}\cdot\mathbf{F}(\mathbf{X})] = \frac{1}{2\alpha}\mathbb{E}[(\mathbf{X}-\mathbf{X}')(\mathbf{F}(\mathbf{X})-\mathbf{F}(\mathbf{X}'))].$$

□

Standard Matrix Functions

Definition

Let I be an interval on the real line. By the Spectral Theorem, a d-dimensional Hermitian matrix $\mathbf{A}$ can be written $\mathbf{A} = \mathbf{Q} \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_d) \cdot \mathbf{Q}^*$ for some unitary $\mathbf{Q}$. For any $f : I \to \mathbb{R}$, we can extend f to take in matrices as input by defining

$$f(\mathbf{A}) = \mathbf{Q} \cdot \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_d)) \cdot \mathbf{Q}^*.$$

Corollary

If λ is an eigenvalue of $\mathbf{A}$, then $f(\lambda)$ is an eigenvalue of $f(\mathbf{A})$.

Normalized Trace MGF

Now that we know how standard scalar functions can be extended to Hermitian matrices, we can define the (normalized) trace MGF.

Definition

Given a random Hermitian matrix $\mathbf{X}$, the (normalized) trace MGF of $\mathbf{X}$ is

$$m(\theta) := \mathbb{E}[\overline{\mathrm{tr}}(e^{\theta \mathbf{X}})].$$

Why is this useful? It gives us concentration inequalities via the matrix Laplace transform method!

Matrix Laplace Transform Method

Theorem

Let $\mathbf{X}$ be a random d-dimensional Hermitian matrix with normalized trace MGF $m(\theta)$. For all real t, we have

$$\mathbb{P}(\lambda_{max}(\mathbf{X}) \geq t) \leq d \cdot \inf_{\theta > 0} \exp(-\theta t + \log m(\theta))$$

$$\mathbb{P}(\lambda_{min}(\mathbf{X}) \leq t) \leq d \cdot \inf_{\theta < 0} \exp(-\theta t + \log m(\theta))$$

$$\mathbb{E}(\lambda_{max}(\mathbf{X})) \leq \inf_{\theta > 0} \frac{1}{\theta}[\log d + \log m(\theta)]$$

$$\mathbb{E}(\lambda_{min}(\mathbf{X})) \geq \sup_{\theta < 0} \frac{1}{\theta}[\log d + \log m(\theta)].$$

Thus, we merely need to bound $m(\theta)$ to get new concentration inequalities!

Matrix Laplace Transform Method

Proof.

(We'll only prove the first inequality.) Applying Markov's inequality and our corollary concerning the eigenvalues of a standard matrix function gives us

$$
\begin{aligned}
\mathbb{P}(\lambda_{\max}(\mathbf{X}) \geq t) &= \mathbb{P}(e^{\lambda_{\max}(\theta \mathbf{X})} \geq e^{\theta t}) \\
&\leq e^{-\theta t} \cdot \mathbb{E}(e^{\lambda_{\max}(\theta \mathbf{X})}) \\
&= e^{-\theta t} \cdot \mathbb{E}(\lambda_{\max}(e^{\theta \mathbf{X}})) \\
&\leq e^{-\theta t} \mathbb{E}(\operatorname{tr}(e^{\theta \mathbf{X}})),
\end{aligned}
$$

where the last line arises from $e^{\theta \mathbf{X}}$ having only positive eigenvalues, implying that $\lambda_{\max}(e^{\theta \mathbf{X}}) \leq \operatorname{tr}(e^{\theta \mathbf{X}})$. Taking the infimum of the trace MGF here and rewriting it using the normalized trace MGF then proves our desired result. □

Alternative approach to bounding $m(\theta)$

We bound $m(\theta)$ by bounding the derivative $m'(\theta)$

$$m'(\theta) = \mathbb{E}\bar{\operatorname{tr}}[\mathbf{X}e^{\theta\mathbf{X}}] = \frac{1}{2\alpha}\mathbb{E}\bar{\operatorname{tr}}[(\mathbf{X}-\mathbf{X}')(e^{\theta\mathbf{X}} - e^{\theta\mathbf{X}'})]$$

We use the smoothness of $e^{\theta\mathbf{X}}$

Lemma: Mean Value Trace Inequality

Lemma

Interval I on the real line. Functions $g : I \to \mathbb{R}$, weakly increasing, $h : I \to \mathbb{R}$, h' convex, is given. For all matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{H}^d(I)$,

$$\bar{tr}[(g(\boldsymbol{A}) - g(\boldsymbol{B}))(h(\boldsymbol{A}) - h(\boldsymbol{B}))] \leq \frac{1}{2}\bar{tr}[(g(\boldsymbol{A}) - g(\boldsymbol{B}))(A - B)(h'(\boldsymbol{A}) + h'(\boldsymbol{B}))]$$

(If h' is concave, the inequality is reversed.)

Proof.

Sketch: Prove the inequality pointwise for $a, b \in I$. Then, use Generalized Klein Inequality on traces to change pointwise inequality to hold for $\mathbf{A}, \mathbf{B} \in \mathbb{H}^d(I)$. $\square$

Direct bound of $m'(\theta)$

Theorem

1. $m'(\theta) \leq \theta \cdot \mathbb{E}\bar{tr}[\triangle_{\mathbf{X}} e^{\theta \mathbf{X}}]$ *for* $\theta \geq 0$.
2. $m'(\theta) \geq \theta \cdot \mathbb{E}\bar{tr}[\triangle_{\mathbf{X}} e^{\theta \mathbf{X}}]$ *for* $\theta \leq 0$.

Proof.

Suppose $\theta \geq 0$. (Proof is similar when $\theta \leq 0$.) $g : x \mapsto x, h : s \mapsto e^{\theta s}$

$$
\begin{aligned}
m'(\theta) &= \frac{1}{2\alpha} \mathbb{E}\bar{\mathrm{tr}}[(\mathbf{X} - \mathbf{X}')(e^{\theta \mathbf{X}} - e^{\theta \mathbf{X}'})] \\
&\leq \frac{\theta}{4\alpha} \mathbb{E}\bar{\mathrm{tr}}[(\mathbf{X} - \mathbf{X}')^2 (e^{\theta \mathbf{X}} + e^{\theta \mathbf{X}'})] \\
&= \frac{\theta}{2\alpha} \mathbb{E}\bar{\mathrm{tr}}[(\mathbf{X} - \mathbf{X}')^2 e^{\theta \mathbf{X}}] \\
&= \theta \cdot \mathbb{E}\bar{\mathrm{tr}}[\triangle_{\mathbf{X}} e^{\theta \mathbf{X}}]
\end{aligned}
$$

Conditional variance bound

Definition

Matrix stein pair $(\mathbf{X}, \mathbf{X}') \in \mathbb{H}^d \times \mathbb{H}^d$. Assume there exists nonnegative constants c, v such that

$$\triangle_{\mathbf{X}} \preccurlyeq c\mathbf{X} + v\mathbf{I}$$

Corollary

1. $m'(\theta) \leq c\theta \cdot m'(\theta) + v\theta \cdot m(\theta)$ *for* $\theta \geq 0$.
2. $m'(\theta) \geq c\theta \cdot m'(\theta) + v\theta \cdot m(\theta)$ *for* $\theta \leq 0$.

Concentration Inequality for bounded Random Matrices

Theorem

Matrix stein pair $(\mathbf{X}, \mathbf{X}') \in \mathbb{H}^d \times \mathbb{H}^d$ *satisfying conditional variance bound* $\triangle_{\mathbf{X}} \preccurlyeq c\mathbf{X} + v\mathbf{I}$.

1. $\mathbb{P}(\lambda_{min}(\mathbf{X}) \leq -t) \leq d \exp\left(\frac{-t^2}{2v}\right)$
2. $\mathbb{P}(\lambda_{max}(\mathbf{X}) \geq t) \leq d \exp\left(-\frac{t}{c} + \frac{v}{c^2}\log(1 + \frac{ct}{v})\right) \leq d \exp\left(\frac{-t^2}{2v+2ct}\right)$
3. $\mathbb{E}[\lambda_{min}(\mathbf{X})] \geq -\sqrt{2v \log d}$
4. $\mathbb{E}[\lambda_{max}(\mathbf{X})] \leq \sqrt{2v \log d} + c \log d$

Proof.

Sketch: Solve the differential inequality on $\log m(\theta)$, integrate it, and use the laplace transformation method. $\square$

Matrix Hoeffding Inequqality

Theorem

Finite sequence of independent random matrices $(\boldsymbol{Y}_k)_{k\geq 1}$ *in* $\mathbb{H}^d$*. Finite, deterministic sequence of matrices* $(\boldsymbol{A}_k)_{k\geq 1}$ *in* $\mathbb{H}^d$*. Suppose* $\mathbb{E}[\mathbf{Y}_k] = 0$ *and* $\boldsymbol{Y}_k^2 \preccurlyeq \boldsymbol{A}_k^2$ *almost surely for all index k. Then, for all* $t \geq 0$*,*

$$\mathbb{P}\left(\lambda_{max}(\sum_k \boldsymbol{Y}_k) \geq t\right) \leq d \exp\left(\frac{-t^2}{2\sigma^2}\right), \sigma^2 := \frac{1}{2}\left\|\sum_k (\boldsymbol{A}_k^2 + \mathbb{E}[\boldsymbol{Y}_k^2])\right\|$$

and

$$\mathbb{E}[\lambda_{max}(\sum_k \boldsymbol{Y}_k)] \leq \sigma\sqrt{2\log d}$$

Proof.

Observe that $\triangle_X = \frac{1}{2}\sum_k(\mathbf{Y}_k^2 + \mathbb{E}[\mathbf{Y}_k^2]) \preccurlyeq \sigma^2\mathbf{I}$ and the theorem directly follows. $\square$

Recap: Azuma Concentration Inequality

Theorem (Azuma)

Finite sequence of Martingale Difference $(Y_i, \mathcal{F}_i)$ *satisfies* $|Y_i| \leq c_i$ *almost surely for some constants* c_i*. Then, for all* $t \geq 0$,

$$\mathbb{P}(\sum_{i=1}^{n} Y_i \geq t) \leq \exp\left(\frac{-t^2}{2\sum_{i=1}^{n} c_i^2}\right)$$