Deck 7: Matrix concentration: Matrix Chernoff

Math 7870: Topics in Randomized Numerical Linear Algebra

Spring 2026

Akil Narayan

Matrix Chernoff setup

We'll begin with one of the main results: The matrix Chernoff bound.

Setup: Let $\{\boldsymbol{X}_j\}_{j\in[n]}$ be a sequence of independent, positive-definite (Hermitian) matrices. We are interested in studying the extremal eigenvalues of,

$$\boldsymbol{S}_n := \sum_{j\in[n]} \boldsymbol{X}_j$$

(As before, an empirical iid sum is a specialization by building in a $1/n$ factor in the $\boldsymbol{X}_j$ summands.)

Our question of concentration involves understanding, e.g.,

$$\lambda_{\max}(\mathbb{E}\boldsymbol{S}_n) \quad \text{vs} \quad \mathbb{E}\lambda_{\max}(\boldsymbol{S}_n) \quad \text{vs} \quad \Pr\left(\lambda_{\max}(\boldsymbol{S}_n) > (1+\epsilon)\lambda_{\max}(\mathbb{E}\boldsymbol{S}_n)\right)$$

We expect that how "large" the $\boldsymbol{X}_j$ summands are plays a role. To quantify this, let's assume that

$$\|\boldsymbol{X}_j\|_2 \leqslant B \quad \text{with probability 1.}$$

Matrix Chernoff

Theorem 1 (Matrix Chernoff inequality). *With the previous setup, let $\mu_{\min}$ and $\mu_{\max}$ denote the extremal eigenvalues of $\mathbb{E}\boldsymbol{S}_n$:*

$$\mu_{\min} = \lambda_{\min}(\mathbb{E}\boldsymbol{S}_n) \geqslant 0 \qquad\qquad \mu_{\max} = \lambda_{\max}(\mathbb{E}\boldsymbol{S}_n).$$

If $\boldsymbol{X}_j$ are $m \times m$ matrices, then:

$$\begin{aligned}
\Pr\left(\lambda_{\min}(\boldsymbol{S}_n) \leqslant (1-\epsilon)\mu_{\min}\right) &\leqslant m \exp\left(-\frac{\mu_{\min}}{B}\left[\epsilon + (1-\epsilon)\log(1-\epsilon)\right]\right), && 0 \leqslant \epsilon < 1 \\
\Pr\left(\lambda_{\max}(\boldsymbol{S}_n) \leqslant (1+\epsilon)\mu_{\max}\right) &\leqslant m \exp\left(-\frac{\mu_{\max}}{B}\left[-\epsilon + (1+\epsilon)\log(1+\epsilon)\right]\right), && \epsilon \geqslant 0 \\
\mathbb{E}\lambda_{\min}(\boldsymbol{S}_n) &\geqslant \mu_{\min}\left(\frac{1-e^{-s}}{s}\right) - \frac{B}{s}\log d, && s > 0 \\
\mathbb{E}\lambda_{\max}(\boldsymbol{S}_n) &\leqslant \mu_{\max}\left(\frac{e^{s}-1}{s}\right) + \frac{B}{s}\log d, && s > 0.
\end{aligned}$$

NB some observations: large B makes these bounds less attractive.

If this is an empirical sum, then $B \sim 1/n$.

Some proof preliminaries: Matrix functions

Recall, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive-definite, then

$$\boldsymbol{A} \preceq \boldsymbol{B} \quad \text{means} \quad \boldsymbol{B} - \boldsymbol{A} \text{ is positive-definite.}$$

In particular, this definition is equivalent to the condition:

$$\boldsymbol{A} \preceq \boldsymbol{B} \quad \text{means} \quad \boldsymbol{v}^* \boldsymbol{B} \boldsymbol{v} \geqslant \boldsymbol{v}^* \boldsymbol{A} \boldsymbol{v} \text{ for any nontrivial } \boldsymbol{v}.$$

Recall: if $f : (0, \infty) \to (0, \infty)$, then we define $f(\boldsymbol{A})$ for a positive-definite $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \boldsymbol{U} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \boldsymbol{U}^* \quad \Longrightarrow \quad f(\boldsymbol{A}) = \boldsymbol{U} \begin{pmatrix} f(\lambda_1) & & & \\ & f(\lambda_2) & & \\ & & \ddots & \\ & & & f(\lambda_n) \end{pmatrix} \boldsymbol{U}^*$$

This definition does imply that if $f(x) \leqslant g(x)$ for all $x \in (0, \infty)$, then $f(\boldsymbol{A}) \preceq g(\boldsymbol{A})$.

More preliminaries: Some matrix function facts

There are a few observations we can make from the definition of the partial order and matrix functions.

- If $\boldsymbol{A} \preceq \boldsymbol{B}$, then for every $z > 0$, $\boldsymbol{A} + z\boldsymbol{I} \preceq \boldsymbol{B} + z\boldsymbol{I}$.
- (Conjugation rule) If $\boldsymbol{A} \preceq \boldsymbol{B}$, then for any matrix $\boldsymbol{M}$, we have that $\boldsymbol{M}\boldsymbol{A}\boldsymbol{M}^* \preceq \boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^*$.
- (The negative inverse is operator monotone) If $\boldsymbol{0} \prec \boldsymbol{A} \preceq \boldsymbol{B}$, then $\boldsymbol{B}^{-1} \preceq \boldsymbol{A}^{-1}$.
- (The logarithm is operator monotone) If $\boldsymbol{A} \preceq \boldsymbol{B}$, then $\log \boldsymbol{A} \preceq \log \boldsymbol{B}$.
- (Eigenvalue monotonicity) Let $\lambda_k(\boldsymbol{M})$ denote the kth eigenvalue (ordered smallest-to-largest) of a positive-definite matrix $\boldsymbol{M}$.
  If $\boldsymbol{A} \preceq \boldsymbol{B}$, then $\lambda_k(\boldsymbol{A}) \leqslant \lambda_k(\boldsymbol{B})$ for all $k \in [n]$.
- (The trace-exponential is a monotone function) If $\boldsymbol{A} \preceq \boldsymbol{B}$, then $\operatorname{tr} e^{\boldsymbol{A}} \leqslant \operatorname{tr} e^{\boldsymbol{B}}$.

An initial lemma, I

We will need to estimate the matrix CGF. Here's an intermediate result we'll need:

Lemma 1 (Chernoff CGF bound)**.** *Suppose $\boldsymbol{X}$ is a random spd matrix, with $\lambda_{\max}(\boldsymbol{X}) = \|\boldsymbol{X}\|_2 \leqslant B$ with probability 1. Then for any $s \in \mathbb{R}$:*

$$\boldsymbol{K}_{\boldsymbol{X}}(s) \preceq h(s)\mathbb{E}\boldsymbol{X}, \qquad\qquad h(s) := \frac{e^{sB} - 1}{B}$$

Proof idea: The exponential function is convex, so it's dominated by any secant line:

$$e^{sx} \leqslant \underbrace{1 + x\frac{e^{sB} - 1}{B}}_{\text{linear function connects } (0,1) \text{ to } (B, e^B)}$$

This implies that for SPD matrices:

$$e^{s\boldsymbol{X}} \preceq \boldsymbol{I} + \boldsymbol{X}\frac{e^{sB} - 1}{B}.$$

An initial lemma, II

$$e^{s\boldsymbol{X}} \preceq \boldsymbol{I} + \boldsymbol{X}\frac{e^{sB}-1}{B}.$$

Since $\boldsymbol{Y} \succeq \boldsymbol{0}$ with probability 1 implies $\mathbb{E}\boldsymbol{Y} \succeq \boldsymbol{0}$, then

$$\boldsymbol{M}_{\boldsymbol{X}}(s) \preceq \boldsymbol{I} + \frac{e^{sB}-1}{B}\mathbb{E}\boldsymbol{X} \preceq \exp\left(\frac{e^{sB}-1}{B}\mathbb{E}\boldsymbol{X}\right),$$

where the last inequality uses $1 + x \leqslant e^x$.

Now define $h(s) = (e^{sB}-1)/B$, and take logarithms. (log is operator monotone!)

Chernoff proof: probabilities

We can now complete the proof. From the master bound:

$$\Pr\left(\lambda_{\max}(\boldsymbol{S}_n) \geqslant t\right) \leqslant \inf_{s>0} e^{-st} \operatorname{tr} \exp\left(\sum_{j\in[n]} \boldsymbol{K}_{\boldsymbol{X}_j}(s)\right)$$

Recall that $\boldsymbol{K}_{\boldsymbol{X}_j}(s) \preceq h(s)\mathbb{E}\boldsymbol{X}_j$ and that the trace-exponential is operator monotone:

$$\begin{aligned}
\Pr\left(\lambda_{\max}(\boldsymbol{S}_n) \geqslant t\right) &\leqslant \inf_{s>0} e^{-st} \operatorname{tr} \exp\left(\sum_{j\in[n]} h(s)\mathbb{E}\boldsymbol{X}_j\right) \\
&= \inf_{s>0} e^{-st} \operatorname{tr} \exp\left(h(s)\mathbb{E}\boldsymbol{S}_n\right) \\
&\leqslant \inf_{s>0} e^{-st} m \lambda_{\max} \exp\left(h(s)\mathbb{E}\boldsymbol{S}_n\right) \\
&\leqslant \inf_{s>0} e^{-st} m \exp h(s) \lambda_{\max} \boldsymbol{S}_n \\
&= \inf_{s>0} \exp\left(-st + \log m + h(s)\mu_{\max}\right)
\end{aligned}$$

(Infimum achieved at $s = \frac{1}{B}\log\left(t\mu_{\max}\right)$.)

A similar computation can be done to yield the lower bound, and the expectation bounds.

An Application: Random column selection, I

Suppose we have a $m \times n$ matrix $\boldsymbol{A}$ with columns $\{\boldsymbol{a}_j\}_{j\in[n]}$:

$$\boldsymbol{A} = \sum_{j\in[n]} \boldsymbol{e}_j^* \boldsymbol{a}_j.$$

We will randomly subsample columns of this matrix: let $\{\gamma_j\}_{j\in[n]} \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$.

Our randomly sampled matrix is:

$$\boldsymbol{S} = \sum_{j\in[n]} \gamma_j \boldsymbol{e}_j^* \boldsymbol{a}_j.$$

If $\gamma_j = 1$, we keep the jth column. If $\gamma_j = 0$, we zero out that column.

(Note that the dominant singular values of $\boldsymbol{S}$ match those of the version of $\boldsymbol{S}$ where we remove the zeroed-out columns.)

Our question: are the singular values of $\boldsymbol{S}$ close to those of $\boldsymbol{A}$?

An application: Random column selection, II

$$\boldsymbol{S} = \sum_{j \in [n]} \gamma_j \boldsymbol{e}_j^* \boldsymbol{a}_j.$$

We consider the matrix,

$$\boldsymbol{M} := \boldsymbol{S}\boldsymbol{S}^* = \sum_{j \in [n]} \gamma_j \boldsymbol{a}_k \boldsymbol{a}_k^*$$

Note that $\lambda_{\max}(\boldsymbol{M}) = \sigma_1^2(\boldsymbol{S})$, and $\lambda_{\min}(\boldsymbol{M}) = \sigma_m^2(\boldsymbol{S})$.

Clearly $\boldsymbol{M}$ is a sum of independent, semi-positive-definite Hermitian matrices $(\gamma_j \boldsymbol{a}_k \boldsymbol{a}_k^*)$.

Using $\mu_{\min} = \lambda_{\min}(\mathbb{E}\boldsymbol{M}) = p\lambda_{\min}(\boldsymbol{A}\boldsymbol{A}^*)$, and $\mu_{\max} = \lambda_{\max}(\mathbb{E}\boldsymbol{M}) = p\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^*)$, and $B = \max_j \|\boldsymbol{a}_j\|_2^2$, the Chernoff bounds, after some simplification, yield,

$$\mathbb{E}\sigma_1^2(\boldsymbol{S}) \leqslant c_1 p \sigma_1^2(\boldsymbol{A}) + B(\log d)$$
$$\mathbb{E}\sigma_m^2(\boldsymbol{S}) \leqslant c_2 p \sigma_m^2(\boldsymbol{A}) - B(\log d),$$

where c_1, c_2 are $\mathcal{O}(1)$ absolute constants.