Deck 6: Matrix concentration: Initial results

Math 7870: Topics in Randomized Numerical Linear Algebra

Spring 2026

Akil Narayan

$S_n = \sum_{i=1}^{n} X_i$, X_i : independent, n fixed.

Q: $\Pr(S_n \geq t) \leq$? , t : tolerance, assume S_n is centered.

$$\Pr(S_n \geq t) \underset{\substack{\uparrow \\ s>0}}{=} \Pr(\exp(s S_n) \geq \exp(st)) \leq \frac{\mathbb{E} e^{s S_n}}{e^{st}} = e^{-st} M_{S_n}(s)$$

$$\Pr(S_n \geq t) \leq \inf_s \exp(-st + \log M_{S_n}(s)) = \inf_s \exp(-st + K_{S_n}(s))$$

$$= \exp\left(-\sup_s (st - K_{S_n}(s))\right)$$

$$= \exp(-K^*_{S_n}(t))$$

Specialize: $X_i \sim \frac{1}{n} X$

$$\Pr(S_n \geq t) \leq \exp\left(-\sup_s (st - K_{S_n}(s))\right)$$

$\log \mathbb{E} e^{s(X+Y)} = \log \mathbb{E} e^{sX} + \log \mathbb{E} e^{sY}$ if $X \perp Y$

$$= \exp\left(-\sup_s (st - n K_{\frac{1}{n}X}(s))\right)$$

$$K_{\frac{1}{n}X}(s) = \log \mathbb{E} e^{sX/n} = K_X(s/n)$$

$$= \exp\left(-\sup_s \left(st - n K_X(s/n)\right)\right)$$

$$= \exp\left(-\sup_u \left(nut - n K_X(u)\right)\right)$$

$$= \exp\left(-n K_X^*(t)\right)$$

Estimation: if X is "nice"

$$\Rightarrow K_X^*(t) \sim \begin{cases} t^2, & t \text{ small} \\ t, & t \text{ large} \end{cases}$$

Beyond scalars

We now have tools to understand sums of independent (scalar) random variables:

$$S_n := \sum_{i\in[n]} X_i \xrightarrow{\text{with mild assumptions}} \Pr(|S_n - \mathbb{E}S_n| \geqslant t) \lesssim \exp\left(-\frac{C_1 t^2}{C_2 + C_3 t}\right)$$
$$\sim \exp\left(-\min\left\{K_1 t, K_2 t^2\right\}\right).$$

This allows us to guarantee that

- If $\epsilon > 0$ is "small", then $n \gtrsim \epsilon^{-2}\log(1/\delta)$ implies $\Pr(|S_n - \mathbb{E}S_n| \geqslant \epsilon) \leqslant \delta$.
- If $\epsilon > 0$ is "large", then $n \gtrsim \epsilon^{-1}\log(1/\delta)$ implies $\Pr(|S_n - \mathbb{E}S_n| \geqslant \epsilon) \leqslant \delta$.

Beyond scalars

We now have tools to understand sums of independent (scalar) random variables:

$$S_n := \sum_{i\in[n]} X_i \quad \xrightarrow{\text{with mild assumptions}} \quad \Pr(|S_n - \mathbb{E}S_n| \geqslant t) \lesssim \exp\left(-\frac{C_1 t^2}{C_2 + C_3 t}\right)$$

$$\sim \exp\left(-\min\left\{K_1 t, K_2 t^2\right\}\right).$$

This allows us to guarantee that

- If $\epsilon > 0$ is "small", then $n \gtrsim \epsilon^{-2}\log(1/\delta)$ implies $\Pr(|S_n - \mathbb{E}S_n| \geqslant \epsilon) \leqslant \delta$.
- If $\epsilon > 0$ is "large", then $n \gtrsim \epsilon^{-1}\log(1/\delta)$ implies $\Pr(|S_n - \mathbb{E}S_n| \geqslant \epsilon) \leqslant \delta$.

Unfortunately, one cannot immediately conclude similar results for matrices because many of the tools we require cannot be directly applied to matrices.
(E.g., orderings on general matrices are more painful.)

Fortunately, the essential (Bernstein) result above still holds. (!!)

Proving it requires more sophisticated arguments.

Some caveats with matrices

If $\boldsymbol{X}_i$, $i \in [n]$ are iid, centered random matrices, we'd like to bound something like:

$$\Pr\left(\|\boldsymbol{S}_n\|_2 \geqslant t\right).$$

We cannot immediately use our scalar strategy with Markov's inequality.

Even if we could proceed, the matrix exponential is not monotone:

$$\mathbf{0} \preceq \boldsymbol{A} \preceq \boldsymbol{B} \quad \not\Longrightarrow \quad e^{\boldsymbol{A}} \preceq e^{\boldsymbol{B}}.$$

(Recall: concave functions on $(0, \infty)$ give rise to operator monotone functions.)

In particular, we will first understand concentration with (symmetric) positive definite matrices.

This allows us to sensibly consider matrix functions exp, log, etc.

The matrix Laplace transform, I

Despite challenges with generalizing to the matrix case, there is a somewhat straightforward way to derive some bounds on the extremal eigenvalues of a matrix.

Let $\boldsymbol{A}$ be a random Hermitian matrix.

We can play essentially the same trick as we did for scalars: For any $s > 0$:

$$\begin{aligned}
\Pr(\lambda_{\max}(\boldsymbol{A}) \geqslant t) &= \Pr\left(e^{s\lambda_{\max}(\boldsymbol{A})} \leqslant e^{st}\right) \\
&\leqslant e^{-st}\mathbb{E}e^{s\lambda_{\max}(\boldsymbol{A})} \\
&= e^{-st}\mathbb{E}e^{\lambda_{\max}(s\boldsymbol{A})} \\
&= e^{-st}\mathbb{E}\lambda_{\max}(e^{s\boldsymbol{A}}) \\
&\leqslant e^{-st}\mathbb{E}\,\mathrm{tr}\, e^{s\boldsymbol{A}}.
\end{aligned}$$

$$R_A(x) = \frac{x^* A x}{\|x\|^2}$$

$$\lambda_{max}(A) = \max_{x \neq 0} R_A(x)$$

$$e^A = U e^D U^*, \quad A = UDU^*$$

The matrix Laplace transform, II

Hence, we've shown the following:

Lemma (Matrix Laplace transform)**.** *Suppose* $\boldsymbol{A}$ *is a random Hermitian matrix. Then:*

$$\Pr(\lambda_{\max}(\boldsymbol{A}) \geqslant t) \leqslant \inf_{s>0} e^{-st}\mathbb{E}\,\mathrm{tr}\, e^{s\boldsymbol{A}}$$

$$\Pr(\lambda_{\min}(\boldsymbol{A}) \leqslant t) \leqslant \inf_{s<0} e^{-st}\mathbb{E}\,\mathrm{tr}\, e^{s\boldsymbol{A}}.$$

The matrix Laplace transform, II

Hence, we've shown the following:

Lemma (Matrix Laplace transform)**.** *Suppose* $\boldsymbol{A}$ *is a random Hermitian matrix. Then:*

$$\Pr(\lambda_{\max}(\boldsymbol{A}) \geqslant t) \leqslant \inf_{s>0} e^{-st}\mathbb{E}\operatorname{tr} e^{s\boldsymbol{A}}$$
$$\Pr(\lambda_{\min}(\boldsymbol{A}) \leqslant t) \leqslant \inf_{s<0} e^{-st}\mathbb{E}\operatorname{tr} e^{s\boldsymbol{A}}.$$

An application of Jensen's inequality yields the following ~~corollary:~~

Lemma. *Suppose A is a random Hermitian matrix. Then:*

Idea:

$$\lambda_{max}(A) = \frac{1}{s}\log e^{s\lambda_{max}(A)}$$

$$\mathbb{E}\,\lambda_{\max}(\boldsymbol{A}) \leqslant \inf_{s>0} \frac{1}{s}\log \mathbb{E}\operatorname{tr} e^{s\boldsymbol{A}}$$
$$\mathbb{E}\,\lambda_{\min}(\boldsymbol{A}) \geqslant \sup_{s<0} \frac{1}{s}\log \mathbb{E}\operatorname{tr} e^{s\boldsymbol{A}}$$

Matrix MGF properties

The matrix Laplace transform allows us to use Markov's inequality to bound extremal eigenvalues.

In the scalar setting, we continued the argument by using product/additive structure of MGF's/CGF's:

$$S_n = \sum_{i \in [n]} X_i, \qquad M_{S_n}(s) = \prod_{i \in [n]} M_{X_i}(s), \qquad K_{S_n}(s) = \sum_{i \in [n]} K_{X_k}(s).$$

Matrix MGF properties

The matrix Laplace transform allows us to use Markov's inequality to bound extremal eigenvalues.

In the scalar setting, we continued the argument by using product/additive structure of MGF's/CGF's:

$$S_n = \sum_{i \in [n]} X_i, \qquad M_{S_n}(s) = \prod_{i \in [n]} M_{X_i}(s), \qquad K_{S_n}(s) = \sum_{i \in [n]} K_{X_k}(s).$$

Matrix MGF's and CGF's still exist: If $\boldsymbol{A}$ is a random Hermitian matrix, its moment generating function (MGF) and cumulant generating function (CGF) are given by,

$$\boldsymbol{M}_{\boldsymbol{A}}(s) = \mathbb{E} e^{s\boldsymbol{A}}, \qquad \boldsymbol{K}_{\boldsymbol{A}}(s) = \log \mathbb{E} e^{s\boldsymbol{A}}.$$

Like before, when $\boldsymbol{M}_{\boldsymbol{A}}(s)$ exists, its Taylor coefficients around $s = 0$ are matrix moments $\mathbb{E}\boldsymbol{A}^k$.

Matrix MGF properties

The matrix Laplace transform allows us to use Markov's inequality to bound extremal eigenvalues.

In the scalar setting, we continued the argument by using product/additive structure of MGF's/CGF's:

$$S_n = \sum_{i \in [n]} X_i, \qquad M_{S_n}(s) = \prod_{i \in [n]} M_{X_i}(s), \qquad K_{S_n}(s) = \sum_{i \in [n]} K_{X_k}(s).$$

Matrix MGF's and CGF's still exist: If $\boldsymbol{A}$ is a random Hermitian matrix, its moment generating function (MGF) and cumulant generating function (CGF) are given by,

$$\boldsymbol{M}_{\boldsymbol{A}}(s) = \mathbb{E} e^{s\boldsymbol{A}}, \qquad \boldsymbol{K}_{\boldsymbol{A}}(s) = \log \mathbb{E} e^{s\boldsymbol{A}}.$$

Like before, when $\boldsymbol{M}_{\boldsymbol{A}}(s)$ exists, its Taylor coefficients around $s = 0$ are matrix moments $\mathbb{E}\boldsymbol{A}^k$.

Unfortunately, the MGF/CGF decomposition property simply fails in the matrix setting, largely because for general matrices $\boldsymbol{A}, \boldsymbol{B}$:

$$\boldsymbol{A}\boldsymbol{B} \neq \boldsymbol{B}\boldsymbol{A}, \quad \text{and} \quad e^{\boldsymbol{A}+\boldsymbol{B}} \neq e^{\boldsymbol{A}} e^{\boldsymbol{B}}.$$

In particular, for general Hermitian $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\boldsymbol{M}_{\boldsymbol{A}+\boldsymbol{B}}(s) \neq \boldsymbol{M}_{\boldsymbol{A}}(s)\boldsymbol{M}_{\boldsymbol{B}}(s), \qquad \text{and} \qquad \boldsymbol{K}_{\boldsymbol{A}+\boldsymbol{B}}(s) \neq \boldsymbol{K}_{\boldsymbol{A}}(s) + \boldsymbol{K}_{\boldsymbol{B}}(s).$$

The resolution

Recall that the matrix Laplace transform yields a formula of the form, $\mathbb{E}\,\mathrm{tr}\, e^{s\boldsymbol{A}}$.

If $\boldsymbol{A}$ is an iid, then we've seen that the MGF's of the summands will not decompose.

The *Golden-Thompson* inequality gives us a glimmer of hope by adding a trace operator:

$$\mathrm{tr}\, e^{\boldsymbol{A}+\boldsymbol{B}} \leqslant \mathrm{tr}\left(e^{\boldsymbol{A}} e^{\boldsymbol{B}}\right).$$

The resolution

Recall that the matrix Laplace transform yields a formula of the form, $\mathbb{E}\,\mathrm{tr}\, e^{s\boldsymbol{A}}$.

If $\boldsymbol{A}$ is an iid, then we've seen that the MGF's of the summands will not decompose.

The *Golden-Thompson* inequality gives us a glimmer of hope by adding a trace operator:

$$\mathrm{tr}\, e^{\boldsymbol{A}+\boldsymbol{B}} \leqslant \mathrm{tr}\left(e^{\boldsymbol{A}} e^{\boldsymbol{B}}\right).$$

Unfortunately, this inequality is false for three (or more) summands. However, including the expectation *does* yield a subadditive result:

Lemma 1 (Matrix CGF subadditivity). *If $\{\boldsymbol{X}_i\}_{i\in[n]}$ are random, independent, Hermitian matrices:*

$$\mathbb{E}\,\mathrm{tr}\exp\left(s\sum_{i\in[n]}\boldsymbol{X}_i\right) \leqslant \mathrm{tr}\exp\left(\sum_{i\in[n]}\boldsymbol{K}_{\boldsymbol{X}_i}(s)\right), \quad \textit{i.e.,}$$

$$\mathrm{tr}\exp \boldsymbol{K}_{\sum_i \boldsymbol{X}_i}(s) \leqslant \mathrm{tr}\exp\left(\sum_i \boldsymbol{K}_{\boldsymbol{X}_i}(s)\right)$$

$\mathrm{tr}\, M_{S_n}(s)$

$= \mathrm{tr}\exp \log M_{S_n}(s)$

$= \mathrm{tr}\exp K_{S_n}(s)$

The resolution

Recall that the matrix Laplace transform yields a formula of the form, $\mathbb{E} \operatorname{tr} e^{s\boldsymbol{A}}$.

If $\boldsymbol{A}$ is an iid, then we've seen that the MGF's of the summands will not decompose.

The *Golden-Thompson* inequality gives us a glimmer of hope by adding a trace operator:

$$\operatorname{tr} e^{\boldsymbol{A}+\boldsymbol{B}} \leqslant \operatorname{tr} \left(e^{\boldsymbol{A}} e^{\boldsymbol{B}} \right).$$

Unfortunately, this inequality is false for three (or more) summands. However, including the expectation *does* yield a subadditive result:

Lemma 1 (Matrix CGF subadditivity). *If $\{\boldsymbol{X}_i\}_{i \in [n]}$ are random, independent, Hermitian matrices:*

$$\mathbb{E} \operatorname{tr} \exp\left(s \sum_{i \in [n]} \boldsymbol{X}_i \right) \leqslant \operatorname{tr} \exp\left(\sum_{i \in [n]} \boldsymbol{K}_{\boldsymbol{X}_i}(s) \right), \quad \textit{i.e.,}$$

$$\operatorname{tr} \exp \boldsymbol{K}_{\sum_i \boldsymbol{X}_i}(s) \leqslant \operatorname{tr} \exp \left(\sum_i \boldsymbol{K}_{\boldsymbol{X}_i}(s) \right)$$

Proving this requires some rather heavy machinery: for fixed Hermitian $\boldsymbol{B}$, the function $\boldsymbol{A} \mapsto \operatorname{tr} \exp(\boldsymbol{B} + \log \boldsymbol{A})$ is concave on positive-definite matrices.

"Master" bounds/inequalities

With the matrix CGF subadditivity, we can essentially establish an important intermediate result:

Theorem. *Let* $\{\boldsymbol{X}_i\}_{i \in [n]}$ *be independent, Hermitian, random matrices, and let* $\boldsymbol{S}_n$ *denote their sum. Then:*

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

$$\mathbb{E}\, \lambda_{\max}(\boldsymbol{S}_n) \leqslant \inf_{s>0} \frac{1}{s} \log \operatorname{tr} \exp\left(\sum_{i \in [n]} \boldsymbol{K}_{\boldsymbol{X}_i}(s)\right)$$

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

$$\mathbb{E}\, \lambda_{\min}(\boldsymbol{S}_n) \leqslant \sup_{s<0} \frac{1}{s} \log \operatorname{tr} \exp\left(\sum_{i \in [n]} \boldsymbol{K}_{\boldsymbol{X}_i}(s)\right)$$

Using the master bounds: A warmup, I

The master bounds are powerful, but before applying them in general let's consider a specialized case.

Suppose the matrix sequence $\boldsymbol{X}_i$ is given by:

$$\boldsymbol{X}_i = x_i \boldsymbol{A}_i, \qquad \{\boldsymbol{A}_i\}_{i \in [n]} \text{ Hermitian, deterministic}, \qquad \{x_i\}_{i \in [n]} \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

In this special case, the randomness in the $\boldsymbol{X}_i$ is controlled by a multiplicative scalar.

Using the master bounds: A warmup, I

The master bounds are powerful, but before applying them in general let's consider a specialized case.

Suppose the matrix sequence $\boldsymbol{X}_i$ is given by:

$$\boldsymbol{X}_i = x_i \boldsymbol{A}_i, \qquad \{\boldsymbol{A}_i\}_{i\in[n]} \text{ Hermitian, deterministic}, \qquad \{x_i\}_{i\in[n]} \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

In this special case, the randomness in the $\boldsymbol{X}_i$ is controlled by a multiplicative scalar.

To use the master bounds, we must compute the CGF of $\boldsymbol{X}_i$. As an intermediate step, we identify the moments of standard normal random variables:

k odd: $\int_{\mathbb{R}} x^k e^{-x^2/2}\, dx = 0$

$$\mathbb{E} x_i^k = \begin{cases} \frac{k!}{2^{k/2}(k/2)!}, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

$t = x^2/2$

k even: $\int_{\mathbb{R}} x^k e^{-x^2/2}\, dx \sim \int_0^\infty x^k e^{-x^2/2}\, dx \sim \int_0^\infty x^{k-1} e^{-t}\, dt \sim \int_0^\infty t^{(k-1)/2} e^{-t}\, dt$

NB: $\Gamma(z)\Gamma(z+\frac{1}{2}) \sim \Gamma(2z)$

$$= \Gamma\left(\frac{k}{2}+\frac{1}{2}\right)$$

$$\sim \frac{\Gamma(k)}{\Gamma(k/2)}$$

Using the master bounds: A warmup, II

We can now compute the CGF of $\boldsymbol{X}_i$:

$$
\begin{aligned}
\boldsymbol{K}_{\boldsymbol{X}_i}(s) &= \log \mathbb{E} \exp(s\boldsymbol{X}_i) \\
&= \log \mathbb{E} \exp(s x_i \boldsymbol{A}_i) \\
&= \log \sum_{k \in \mathbb{N}_0} \frac{\mathbb{E} x_i^k}{k!} (s\boldsymbol{A}_i)^k
\end{aligned}
$$

$(\underline{X}_i = x_i \underline{A}_i)$

$\exp(s x_i \underline{A}_i)$

$= \sum_k \frac{1}{k!} (s x_i)^k (\underline{A}_i)^k$

$\mathbb{E} x_i^k = \frac{k!}{(k/2)!\, 2^k}$ (k even)

Using the master bounds: A warmup, II

We can now compute the CGF of $\boldsymbol{X}_i$:

$$\begin{aligned} \boldsymbol{K}_{\boldsymbol{X}_i}(s) &= \log \mathbb{E} \exp(s\boldsymbol{X}_i) \\ &= \log \mathbb{E} \exp(s x_i \boldsymbol{A}_i) \\ &= \log \sum_{k \in \mathbb{N}_0} \frac{\mathbb{E} x_i^k}{k!} (s\boldsymbol{A}_i)^k \end{aligned}$$

Using the moments of a standard normal random variable:

$$\begin{aligned} \boldsymbol{K}_{\boldsymbol{X}_i}(s) &= \log \sum_{\ell \in \mathbb{N}_0} \frac{1}{2^\ell \ell!} (s\boldsymbol{A}_i)^{2\ell} \\ &= \log \sum_{\ell \in \mathbb{N}_0} \frac{1}{\ell!} \left(\frac{s^2}{2} \boldsymbol{A}_i^2 \right)^\ell \\ &= \log \exp\left(\frac{1}{2} s^2 \boldsymbol{A}_i^2 \right) = \frac{1}{2} s^2 \boldsymbol{A}_i^2. \end{aligned}$$

$$\Pr\left(\cancel{|S_n - \mathbb{E} S_n| \geq t} \right) \leq \inf_{s>0} e^{-st} \operatorname{tr} \cancel{\mathbb{E}} \exp\left(\sum k_{X_i}(s)\right)$$

$$\lambda(S_n - \mathbb{E} S_n) \geq t$$

$$= \inf_{s>0} e^{-st} \operatorname{tr} \cancel{\mathbb{E}} \exp\left(\tfrac{1}{2} s^2 \sum_i \underline{\underline{A}}_i^2\right)$$

$$\leq \inf_{s>0} e^{-st}\, d\, \lambda_{max}\left(\exp\left(\tfrac{1}{2} s^2 \sum_i \underline{\underline{A}}_i^2\right)\right)$$

$$\underline{\underline{A}}_i \in \mathbb{C}^{d \times d}$$

$$= \inf_{s>0} d\, e^{-st} \exp\left(\tfrac{1}{2} s^2 \lambda_{max}\left(\sum_i \underline{\underline{A}}_i^2\right)\right)$$

... optimize over s ...

Using the master bounds: A warmup, III

Given the previous, the following is the result of this estimation.

Theorem (Hermitian Matrix Gaussian Series)**.** *Let* $\boldsymbol{X}_i = x_i \boldsymbol{A}_i$*,* $i \in [n]$*, be a sequence of random* $m \times m$ *Hermitian matrices where* $\{x_i\}_i$ *are iid standard normal random variables, and* $\{\boldsymbol{A}_i\}_i$ *are Hermitian, deterministic matrices. Define*

$$\boldsymbol{S}_n := \sum_{i \in [n]} \boldsymbol{X}_i = \sum_{i \in [n]} x_i \boldsymbol{A}_i, \qquad V(\boldsymbol{S}_n) = \|\mathbb{E}\boldsymbol{S}_n^2\|_2 = \left\| \sum_{i \in [n]} \boldsymbol{A}_i^2 \right\|_2 .$$

Then:

$$\Pr\left(\lambda_{\max}(\boldsymbol{S}_n) \geqslant t\right) \leqslant m \exp\left(-\frac{t^2}{2V(\boldsymbol{S}_n)}\right)$$
$$\mathbb{E}\,\lambda_{\max}(\boldsymbol{S}_n) \leqslant \sqrt{2V(\boldsymbol{S}_n) \log m}.$$

($m = d$ from before)

$V(\boldsymbol{S}_n)$ is called the *matrix variance statistic* of $\boldsymbol{S}_n$.

To rectangular matrices: Matrix dilation

We can essentially port the result for Hermitian matrices directly to rectangular ones. The appropriate tool is a matrix dilation.

Definition (Matrix dilation)**.** *Let $\boldsymbol{A}$ be a $m_1 \times m_2$ matrix. Then $\mathcal{D}(\boldsymbol{A})$ is the size-$(m_1 + m_2)$ Hermitian matrix defined by:*

$$\mathcal{D}(\boldsymbol{A}) = \begin{pmatrix} \mathbf{0} & \boldsymbol{A} \\ \boldsymbol{A}^* & \mathbf{0} \end{pmatrix}$$

$$D(A)^2 = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} = \begin{pmatrix} AA^* & 0 \\ 0 & A^*A \end{pmatrix}$$

$$\lambda(D^2(A)) = \sigma^2(A)$$

$$\Rightarrow |\lambda(D(A))| = \sigma(A) \Longrightarrow \|D(A)\|_2 = \sigma_{max}(A) = \|A\|_2$$

To rectangular matrices: Matrix dilation

We can essentially port the result for Hermitian matrices directly to rectangular ones. The appropriate tool is a matrix dilation.

Definition (Matrix dilation). *Let $\boldsymbol{A}$ be a $m_1 \times m_2$ matrix. Then $\mathcal{D}(\boldsymbol{A})$ is the size-$(m_1 + m_2)$ Hermitian matrix defined by:*

$$\mathcal{D}(\boldsymbol{A}) = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}^* & \boldsymbol{0} \end{pmatrix}$$

To see why this is a useful concept, notice that:

$$\mathcal{D}(\boldsymbol{A})^2 = \begin{pmatrix} \boldsymbol{A}\boldsymbol{A}^* & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^*\boldsymbol{A} \end{pmatrix}.$$

modulus

Therefore: the spectrum of $\mathcal{D}(\boldsymbol{A})$ is a duplication of the singular values of $\boldsymbol{A}$ (plus some zeros if $\boldsymbol{A}$ is rectangular).

Moreover, matrix dilation is a linear operation, and

$$\lambda_{\max}(\mathcal{D}(\boldsymbol{A})) = \left\|\mathcal{D}(\boldsymbol{A})\right\|_2 = \left\|\boldsymbol{A}\right\|_2.$$

why is $\lambda_{max}(D(A)) = \|D(A)\|_2$?

$$\lambda_{max}(D(A)) = \max_i \lambda_i(D(A)) \leq \max_i |\lambda_i(D(A))| = \sigma_1(A)$$
$$= \|A\|_2$$

First, observe: take $\underline{u} \in \mathbb{C}^{m_1}$, $\underline{v} \in \mathbb{C}^{m_2}$ to be unit norm:

$$R_{D(A)}\left(\begin{pmatrix} \underline{u} \\ \underline{v} \end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix} u^* & v^* \end{pmatrix}\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix} u^* & v^* \end{pmatrix}\begin{pmatrix} Av \\ A^*u \end{pmatrix} = \frac{1}{2}(u^*Av + v^*A^*u)$$
$$= \mathrm{Re}(u^*Av)$$

Take: $u = u_1$ (maximal left-sing. vector of A)
$v = v_1$ (maximal right-sing vector of A)

$$u^*Av = u_1^*Av_1 = u_1^*\underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^*\underline{v_1} = \sigma_1 = \|A\|_2$$

$$\|A\|_2 = \sigma_1 = u_1^*Av_1 = \mathrm{Re}(u_1^*Av_1) \leq \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \mathrm{Re}(u^*Av)$$

$$= \max R_{D(A)}\left(\begin{pmatrix} u \\ v \end{pmatrix}\right)$$

$$= \lambda_{max}(D(A))$$

$$\Rightarrow \lambda_{max}(D(A)) \leq \|A\|_2 \leq \lambda_{max}(D(A))$$

Rectangular Gaussian Matrix Series

Then consider a similar setup: $\boldsymbol{X}_i = x_i \boldsymbol{A}_i$, with $\{x_i\}_{i \in [n]}$ iid standard normal random variables.

But now: let $\boldsymbol{A}_i$ be deterministic $m_1 \times m_2$ *rectangular* matrices.

- Consider $\mathcal{D}(\boldsymbol{S}_n)$, with $\boldsymbol{S}_n = \sum_{i \in [n]} \boldsymbol{X}_i$.
- Note that norm bounds on $\boldsymbol{S}_n$ are equivalent to eigenvalue bounds on $\mathcal{D}(\boldsymbol{S}_n)$.
- Relate $V(\mathcal{D}(\boldsymbol{A}))$ to an appropriate $\boldsymbol{S}_n$-dependent quantity.

Rectangular Gaussian Matrix Series

Then consider a similar setup: $\boldsymbol{X}_i = x_i \boldsymbol{A}_i$, with $\{x_i\}_{i \in [n]}$ iid standard normal random variables.

But now: let $\boldsymbol{A}_i$ be deterministic $m_1 \times m_2$ *rectangular* matrices.

- Consider $\mathcal{D}(\boldsymbol{S}_n)$, with $\boldsymbol{S}_n = \sum_{i \in [n]} \boldsymbol{X}_i$.
- Note that norm bounds on $\boldsymbol{S}_n$ are equivalent to eigenvalue bounds on $\mathcal{D}(\boldsymbol{S}_n)$.
- Relate $V(\mathcal{D}(\boldsymbol{A}))$ to an appropriate $\boldsymbol{S}_n$-dependent quantity.

Theorem. *With the above setup, then:*

$$\Pr(\|\boldsymbol{S}_n\|_2 \geqslant t) \leqslant (m_1 + m_2) \exp\left(-\frac{t^2}{2V(\boldsymbol{S}_n)}\right)$$

$$\mathbb{E}\,\|\boldsymbol{S}_n\|_2 \leqslant \sqrt{2V(\boldsymbol{S}_n)(m_1 + m_2)},$$

log

where $V(\boldsymbol{S}_n)$ is the matrix variance statistic of $\boldsymbol{S}_n$:

$$V(\boldsymbol{S}_n) = \max\{\|\mathbb{E}\,\boldsymbol{S}_n \boldsymbol{S}_n^*\|_2, \|\mathbb{E}\,\boldsymbol{S}_n^* \boldsymbol{S}_n\|_2\} = \max\left\{\left\|\sum_{i \in [n]} \boldsymbol{A}_i \boldsymbol{A}_i^*\right\|_2, \left\|\sum_{i \in [n]} \boldsymbol{A}_i^* \boldsymbol{A}_i\right\|_2\right\}.$$