Spectral norm of sum of independent random matrices

Set up

Let $n \in \mathbf{Z}_+$ and $S_1, S_2, ..., S_n$ be an independent family of random $d_1 \times d_2$ complex-valued matrices with $\mathbf{E}(S_i) = 0$ and bounded spectral norm for every $1 \leq i \leq n$. Define:

$$X := \sum_{i=1}^{n} S_i.$$

We are interested in results concerning $\|X\|$, both in expectation and large deviation.

1D case

Let us assume that $n = 1$, then X is just a some of independent centered random variables.

Theorem (Bernstein's inequality)

Let $S_1, S_2, ..., S_n$ be independent zero-mean random variables such that $|S_i| \leq R$ almost surely for all i and let $X = \sum S_i$. Then, for any $t > 0$ we have:

$$\mathbf{P}(X > t) \leq \exp\left(-\frac{t^2/2}{\sum_j \mathbf{E}(S_j^2) + Rt/3}\right).$$

Proof of Bernstein's inequality

Lemma

Let h be a random variable with $\mathbf{E}(h) = 0$ and $|h| \leq R$ almost surely. Then, for $0 < \theta < 3/R$,

$$\mathbf{E}\left(e^{\theta h}\right) \leq \exp\left(\frac{\theta^2/2}{1-\theta R/3} \cdot \mathbf{E}(h^2)\right)$$

and

$$\log\left(\mathbf{E}\left(e^{\theta h}\right)\right) \leq \frac{\theta^2/2}{1-\theta R/3} \cdot \mathbf{E}(h^2).$$

Proof of Lemma

Note that if we prove the first relation, the second one follows by talking logarithm and using the fact that the log is a monotone function.

Proof of Lemma

Note that if we prove the first relation, the second one follows by talking logarithm and using the fact that the log is a monotone function.

Fix parameter $\theta > 0$. Write

$$e^{\theta h} = I + \theta h + \left(e^{\theta h} - \theta h - 1\right) = 1 + \theta h + h^2 f(h)$$

where f is defined by:

$$f(x) = \begin{cases} \frac{e^{\theta x} - \theta x - 1}{x^2} & \text{if } x \neq 0 \\ f(x) = 0 & \text{if } x = 0. \end{cases}$$

Proof of Lemma

Note that f is increasing as its derivatives is positive, hence

$$f(x) \leq f(R) \text{ for } x \leq R.$$

Proof of Lemma

Note that f is increasing as its derivatives is positive, hence

$$f(x) \leq f(R) \text{ for } x \leq R.$$

Since $h \leq R$, we have that

$$f(h) \leq f(R).$$

Proof of Lemma

Note that f is increasing as its derivatives is positive, hence

$$f(x) \leq f(R) \text{ for } x \leq R.$$

Since $h \leq R$, we have that

$$f(h) \leq f(R).$$

It follows that

$$e^{\theta h} \leq 1 + \theta h + h^2 f(R).$$

By Taylor, we can estimate $f(R)$:

$$f(R) = \frac{e^{\theta R} - \theta R - 1}{R^2} = \frac{1}{R^2} \sum_{q=2}^{\infty} \frac{(\theta R)^q}{q!} \leq \frac{\theta^2}{2} \sum_{q=2}^{\infty} \frac{(\theta R)^{q-2}}{3^{q-2}} = \frac{\theta^2/2}{1 - \theta R/3},$$

where we used that $q! \geq 2 \cdot 3^{q-2}$, for $q \geq 2$.

Proof of Lemma

Let

$$g(\theta, R) = \frac{\theta^2/2}{1 - \theta R/3}.$$

This translates as

$$e^{\theta h} \leq 1 + \theta h + g(\theta, R)h^2,$$

which implies, by the linearity of expectation and the fact that $\mathbf{E}(h) = 0$,

$$\mathbf{E}\left(e^{\theta h}\right) \leq 1 + g(\theta, R)\mathbf{E}(h^2) \leq \exp\left(g(\theta, R)\mathbf{E}(h^2)\right),$$

where in the last step we used that $1 + a \leq e^a$, which completes the proof of lemma.

Proof of Bernstein's inequality

Let $0 < \theta < 3/R$ be a real number to be chosen later. By Markov inequality we have:

$$\begin{aligned} \mathbf{P}(X > t) &= \mathbf{P}\left(e^{\theta X} > e^{\theta t}\right) \\ &\leq e^{-\theta t}\mathbf{E}\left(e^{\theta X}\right) \end{aligned}$$

Note that since $S_1, S_2, ..., S_n$ are independent we have:

$$\mathbf{E}\left(e^{\theta(\sum_{i=1}^{n} S_i)}\right) = e^{\sum_{i=1}^{n} \log \mathbf{E} e^{\theta S_i}},$$

which further implies:

$$\mathbf{P}(X > t) \leq e^{-\theta t} e^{\sum_{i=1}^{n} \log \mathbf{E} e^{\theta S_i}}.$$

Proof of Bernstein's inequality

We can apply our lemma to bound the logarithmic factors

$$\begin{aligned} \mathbf{P}(X > t) &\leq e^{-\theta t} e^{\sum_{i=1}^n \log \mathbf{E} e^{\theta S_i}} \\ &\leq e^{-\theta t} e^{\sum_{i=1}^n g(\theta) \mathbf{E}(S_i^2)} \\ &\leq e^{-\theta t} e^{g(\theta) \cdot \sum_{i=1}^n \mathbf{E}(S_i^2)}, \end{aligned}$$

Pick $\theta = t / \left(\sum_{i=1}^n \mathbf{E}(S_i^2) + Rt/3 \right)$ to conclude the proof of Bernstein's inequality.

Matrix parameters

Matrix parameters

Definition

- The matrix variance parameter is defined by:

$$\nu(X) := \max\left\{\|\mathbf{E}[XX^*]\|, \|\mathbf{E}[X^*X]\|\right\}$$
$$= \max\left\{\left\|\sum_{i=1}^{n}\mathbf{E}[S_iS_i^*]\right\|, \left\|\sum_{i=1}^{n}\mathbf{E}[S_i^*S_i]\right\|\right\}$$

Matrix parameters

Definition

- The matrix variance parameter is defined by:

$$\nu(X) := \max\left\{\|\mathbf{E}[XX^*]\|, \|\mathbf{E}[X^*X]\|\right\}$$
$$= \max\left\{\left\|\sum_{i=1}^{n}\mathbf{E}[S_iS_i^*]\right\|, \left\|\sum_{i=1}^{n}\mathbf{E}[S_i^*S_i]\right\|\right\}$$

- The large deviation parameter is defined by:

$$L := \left(\mathbf{E}\left[\max_{i=1,\ldots,n}\|S_i\|^2\right]\right)^{1/2}$$

Matrix parameters

Definition

- The matrix variance parameter is defined by:

$$\nu(X) := \max\left\{\|\mathbf{E}[XX^*]\|, \|\mathbf{E}[X^*X]\|\right\}$$
$$= \max\left\{\left\|\sum_{i=1}^{n}\mathbf{E}[S_iS_i^*]\right\|, \left\|\sum_{i=1}^{n}\mathbf{E}[S_i^*S_i]\right\|\right\}$$

- The large deviation parameter is defined by:

$$L := \left(\mathbf{E}\left[\max_{i=1,\ldots,n}\|S_i\|^2\right]\right)^{1/2}$$

- The dimensional constant is defined by:

$$C_d := C(d_1, d_2) := 4 \cdot (1 + 2\log(d_1 + d_2))$$

Theorem 1

Theorem (The norm of an independent sum of matrices)

Let $S_1, S_2, ..., S_n$ be independent $d_1 \times d_2$ random matrices with $\mathbf{E}(S_i) = \mathbf{0}$ for each i. Let $X := S_1 + ... + S_n$ and $\nu(X), C_d$ and L defined previously. Then the following is true:

1. $\sqrt{\frac{1}{4} \cdot \nu(X)} + \frac{1}{4} \cdot L \leq \left(\mathbf{E}\left(\|X\|^2\right)\right)^{1/2} \leq \sqrt{C_d \cdot \nu(X)} + C_d \cdot L.$

Theorem 1

Theorem (The norm of an independent sum of matrices)

Let $S_1, S_2, ..., S_n$ be independent $d_1 \times d_2$ random matrices with $\mathbf{E}(S_i) = \mathbf{0}$ for each i. Let $X := S_1 + ... + S_n$ and $\nu(X)$, C_d and L defined previously. Then the following is true:

1. $\sqrt{\frac{1}{4} \cdot \nu(X)} + \frac{1}{4} \cdot L \leq \left(\mathbf{E}\left(\|X\|^2\right)\right)^{1/2} \leq \sqrt{C_d \cdot \nu(X)} + C_d \cdot L.$
2. *Moreover, if there exists $R > 0$ such that $\|S_i\|$'s are uniformly bounded by R then*
$$\mathbf{P}(\|X\| \geq t) \leq (d_1 + d_2) \cdot \exp\left(\frac{-t^2/2}{\nu(X) + Rt/3}\right).$$

Observations

Observation

- *In the case where S_i's are Hermitians, Theorem 1 can be used to get bounds for $\lambda_{\min}(X)$, by replacing S_i with $-S_i$ and X with $-X$.*

Observations

Observation

- *In the case where S_i's are Hermitians, Theorem 1 can be used to get bounds for $\lambda_{\min}(X)$, by replacing S_i with $-S_i$ and X with $-X$.*
- *Theorem 1 can be extended to non-centered matrices too, by replacing S_i with $S_i - \mathbf{E}(S_i)$.*

Observations

Observation

- *In the case where S_i's are Hermitians, Theorem 1 can be used to get bounds for $\lambda_{\min}(X)$, by replacing S_i with $-S_i$ and X with $-X$.*
- *Theorem 1 can be extended to non-centered matrices too, by replacing S_i with $S_i - \mathbf{E}(S_i)$.*
- *There exists a strong conection between $\left(\mathbf{E}(\|X\|^2)\right)^{1/2}$ and $\left(\mathbf{E}(\|X\|^p)\right)^{1/p}$ due to Jensen and Khintchine inequalities, so there are equivalents of Theorem 1.1 for other norms too.*

Observations

Observation

- *In the case where S_i's are Hermitians, Theorem 1 can be used to get bounds for $\lambda_{\min}(X)$, by replacing S_i with $-S_i$ and X with $-X$.*
- *Theorem 1 can be extended to non-centered matrices too, by replacing S_i with $S_i - \mathbf{E}(S_i)$.*
- *There exists a strong conection between $\left(\mathbf{E}(\|X\|^2)\right)^{1/2}$ and $(\mathbf{E}(\|X\|^p))^{1/p}$ due to Jensen and Khintchine inequalities, so there are equivalents of Theorem 1.1 for other norms too.*
- *The large deviation bound in Theorem 1.2 is an extension of the well-known Bernstein inequality for random matrices.*

The optimality of Theorem 1.1

The optimality of Theorem 1.1

The lower and the upper bounds in Theorem 1.1 match, except for the dimensional factor C_d ($\approx 8 \log d$). We will show by four examples that neither the lower bound nor the upper bound can be sharpened substantially without further assumptions.

The optimality of Theorem 1.1

The lower and the upper bounds in Theorem 1.1 match, except for the dimensional factor C_d ($\approx 8 \log d$). We will show by four examples that neither the lower bound nor the upper bound can be sharpened substantially without further assumptions.

In what follows, let $E_{i,j}$ denote the matrix with all entries 0 except the $(i,j)^{th}$ entry which is 1.

Optimality of the upper bound: variance term

Optimality of the upper bound: variance term

Example

Let

$$Z := \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{\sqrt{n}} \xi_{ij} E_{ii},$$

where ξ_{ij}'s are independently Rademacher random variables talking values ± 1 each with probability $1/2$.

Optimality of the upper bound: variance term

Example

Let

$$Z := \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{\sqrt{n}} \xi_{ij} E_{ii},$$

where ξ_{ij}'s are independently Rademacher random variables talking values ± 1 each with probability $1/2$.

It is easy to estimate directly

$$\mathbf{E}(\|Z\|^2) \approx \mathbf{E}\left(\left\|\sum_{i=1}^{j} \gamma_i E_{ii}\right\|^2\right) = \mathbf{E} \max_i |\gamma_i|^2 \approx 2 \log d,$$

where γ_i's are independent standard gaussian random variables.

Optimality of the upper bound: variance term

Optimality of the upper bound: variance term

The variance parameter satisfies:

$$\nu(Z) := \left\| \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{n} E_{ii} \right\| = \|I_d\| = 1.$$

Optimality of the upper bound: variance term

The variance parameter satisfies:

$$\nu(Z) := \left\| \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{n} E_{ii} \right\| = \|I_d\| = 1.$$

The large deviation parameter satisfies:

$$L^2 = \mathbf{E} \max_{i,j} \left\| \frac{1}{\sqrt{n}} \xi_{ij} \mathbf{E}_{ii} \right\|^2 = \frac{1}{n}.$$

Optimality of the upper bound: variance term

The variance parameter satisfies:

$$\nu(Z) := \left\| \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{n} E_{ii} \right\| = \|I_d\| = 1.$$

The large deviation parameter satisfies:

$$L^2 = \mathbf{E} \max_{i,j} \left\| \frac{1}{\sqrt{n}} \xi_{ij} \mathbf{E}_{ii} \right\|^2 = \frac{1}{n}.$$

It follows that

$$\left(\mathbf{E}\|Z\|^2\right)^{1/2} \approx \sqrt{2 \log d \, \nu(Z)},$$

so the logarithm factor in the variance term in the upper bound is needed.

Optimality of the upper bound: large-deviation term

Optimality of the upper bound: large-deviation term

Example

Let

$$Z := \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ij} - n^{-1}) \cdot E_{ii},$$

where δ_{ij} is an independent family of Bernoulli $(1/n)$ random variables.

Using the properties of the Bernoulli random variables, we have

$$\left(\mathbf{E}(\|Z\|^2)\right)^{1/2} \approx \text{constant} \cdot \frac{\log d}{\log \log d},$$

Optimality of the upper bound: large-deviation term

Optimality of the upper bound: large-deviation term

The variance parameter satisfies:

$$\nu(Z) = \left\| \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{E}(\delta_{ij} - n^{-1})^2 \cdot E_{ii} \right\| \approx 1.$$

Optimality of the upper bound: large-deviation term

The variance parameter satisfies:

$$\nu(Z) = \left\| \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{E}(\delta_{ij} - n^{-1})^2 \cdot E_{ii} \right\| \approx 1.$$

The large-deviation parameter is

$$L^2 = \mathbf{E} \left(\max_{i,j} \|(\delta_{ij} - n^{-1}) \cdot E_{ii}\|^2 \right) \approx 1.$$

Optimality of the upper bound: large-deviation term

The variance parameter satisfies:

$$\nu(Z) = \left\| \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{E}(\delta_{ij} - n^{-1})^2 \cdot E_{ii} \right\| \approx 1.$$

The large-deviation parameter is

$$L^2 = \mathbf{E} \left(\max_{i,j} \|(\delta_{ij} - n^{-1}) \cdot E_{ii}\|^2 \right) \approx 1.$$

This implies that the large-deviation parameter in the upper bound can not be improved, except by an iterated logarithm factor.

Optimality of the lower bound: variance term

Optimality of the lower bound: variance term

Example

Let

$$Z := \sum_{i,j=1}^{d} \xi_{ij} E_{ij},$$

where ξ_{ij}'s are independent Rademacher variables.

Optimality of the lower bound: variance term

Example

Let

$$Z := \sum_{i,j=1}^{d} \xi_{ij} E_{ij},$$

where ξ_{ij}'s are independent Rademacher variables.

It is known that

$$(\mathbf{E}\|Z\|^2)^{1/2} \approx \sqrt{2d}.$$

Optimality of the lower bound: variance term

Optimality of the lower bound: variance term

The variance parameter satisfies:

$$\nu(Z) = \max\{\|d \cdot I_d\|, \|d \cdot I_d\|\} = d$$

Optimality of the lower bound: variance term

The variance parameter satisfies:

$$\nu(Z) = \max\{\|d \cdot I_d\|, \|d \cdot I_d\|\} = d$$

and the large deviation parameter is:

$$L^2 = \mathbf{E} \max_{i,j} \|\xi_{ij} E_{ij}\|^2 = 1.$$

Optimality of the lower bound: variance term

The variance parameter satisfies:

$$\nu(Z) = \max\{\|d \cdot I_d\|, \|d \cdot I_d\|\} = d$$

and the large deviation parameter is:

$$L^2 = \mathbf{E} \max_{i,j} \|\xi_{ij} E_{ij}\|^2 = 1.$$

We conclude that the variance term in the lower bound can not have a logarithmic factor.

Optimality of the lower bound: large-deviation term

Optimality of the lower bound: large-deviation term

Example

Let

$$Z := \sum_{i=1}^{d} P_i E_{i,i},$$

where $\{P_i\}$ is an independent family of symmetric random variables whose tails satisfy:

$$\mathbf{P}\left(|P_i| \geq t\right) = \begin{cases} t^{-4} & \text{if } t \geq 1 \\ 1 & \text{if } t \leq 1. \end{cases}$$

Optimality of the lower bound: large-deviation term

Example

Let

$$Z := \sum_{i=1}^{d} P_i E_{i,i},$$

where $\{P_i\}$ is an independent family of symmetric random variables whose tails satisfy:

$$\mathbf{P}\left(|P_i| \geq t\right) = \begin{cases} t^{-4} & \text{if } t \geq 1 \\ 1 & \text{if } t \leq 1. \end{cases}$$

The key properties of these variables are that:

$$\mathbf{E}(P_i^2) = 2 \text{ and } \mathbf{E} \max_i P_i^2 \approx \text{constant} \cdot d^2.$$

Optimality of the lower bound: large-deviation term

Optimality of the lower bound: large-deviation term

The variance parameter is:

$$\nu(Z) = \left\| \sum_{i=1}^{d} (\mathbf{E} P_i^2) E_i \right\| = 2,$$

Optimality of the lower bound: large-deviation term

The variance parameter is:

$$\nu(Z) = \left\| \sum_{i=1}^{d} (\mathbf{E} P_i^2) E_i \right\| = 2,$$

and the large deviation parameter satisfy:

$$L^2 = \mathbf{E} \max_i \|P_i E_i\|^2 = \mathbf{E} \max_i |P_i|^2 \approx \text{constant} \cdot d^2..$$

Optimality of the lower bound: large-deviation term

The variance parameter is:

$$\nu(Z) = \left\| \sum_{i=1}^{d} (\mathbf{E} P_i^2) E_i \right\| = 2,$$

and the large deviation parameter satisfy:

$$L^2 = \mathbf{E} \max_i \|P_i E_i\|^2 = \mathbf{E} \max_i |P_i|^2 \approx \text{constant} \cdot d^2..$$

By direct calculation, we have:

$$\left(\mathbf{E}(\|Z\|^2)\right)^{1/2} \approx \text{constant} \cdot d.$$

Optimality of the lower bound: large-deviation term

The variance parameter is:

$$\nu(Z) = \left\| \sum_{i=1}^{d} (\mathbf{E} P_i^2) E_i \right\| = 2,$$

and the large deviation parameter satisfy:

$$L^2 = \mathbf{E} \max_i \|P_i E_i\|^2 = \mathbf{E} \max_i |P_i|^2 \approx \text{constant} \cdot d^2..$$

By direct calculation, we have:

$$\left(\mathbf{E}(\|Z\|^2)\right)^{1/2} \approx \text{constant} \cdot d.$$

We conclude that the large-deviation term in the lower bound can not carry a logarithmic factor.

Other results

Theorem (Matrix Chernoff Bound part 1)

Let $\{S_1, S_2, ..., S_n\}$ be a finite sequence of independent $d \times d$ Hermitian random matrices such that for each i, S_i is positive semi-definite and $\lambda_{\max}(S_i) \leq L$. Define $X = \sum_{i=1}^{n} S_i$ and let $\mu_{\min} := \lambda_{\min}(\mathbf{E}(X))$ and $\mu_{\max} = \lambda_{\max}(\mathbf{E}(X))$. For any $\theta > 0$ we have:

$$\mathbf{E}\left(\lambda_{\min}\left(X\right)\right) \geq \frac{1 - e^{-\theta}}{\theta}\mu_{\min} - \frac{1}{\theta}L \log d \tag{1}$$

$$\mathbf{E}\left(\lambda_{\max}\left(X\right)\right) \leq \frac{e^{\theta} - 1}{\theta}\mu_{\max} + \frac{1}{\theta}L \log d \tag{2}$$

Theorem (Matrix Chenoff Bound part 2)

Also, for any $\epsilon > 0$ *we have:*

$$\mathbf{P}\left(\lambda_{\max}(X) \geq (1+\epsilon)\mu_{\max}\right) \leq d\left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\mu_{\max}/L}, \tag{3}$$

and for any $\epsilon \in [0,1]$ *we have:*

$$\mathbf{P}\left(\lambda_{\min}(X) \leq (1-\epsilon)\mu_{\min}\right) \leq d\left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{\mu_{\min}/L}. \tag{4}$$

Observations

Observation

- *If we pick θ to be 1, we get*

$$\mathbf{E}\lambda_{\min}(X) \geq 0.63\mu_{\min} - L\log d \text{ and}$$
$$\mathbf{E}\lambda_{\max}(X) \leq 1.72\mu_{\max} + L\log d.$$

- *If the matrices S_i are unbounded, we have:*

$$\mathbf{E}\lambda_{\max}(X) \leq 2\mu_{\max} + 8e\left(\mathbf{E}\left(\max_k \lambda_{\max}(S_k)\right)\right)\log d.$$

Other results

Theorem (Matrix Azuma Inequality)

Let $\{X_1, X_2, ..., X_k\}$ be a finite adapted sequence of self-adjoint $d \times d$ random matrices and let $\{A_1, A_2, ..., A_k\}$ be a fixed sequence of self-adjoint matrices. Assume that each random variables satisfies $\mathbf{E}_{i-1} X_i = \mathbf{0}$ and $X_i^2 \preceq A_i^2$ almost surely for any $1 \leq i \leq k$, where $\mathbf{0}$ is the zero $d \times d$ matrix. Let

$$\sigma^2 = \left\| \sum_k A_k^2 \right\|,$$

then for all $t \geq 0$ we have:

$$\mathbf{P}\left(\lambda_{\max}\left(\sum_k X_k \right) \geq t \right) \leq d \cdot e^{-t^2/(8\sigma^2)}.$$

Other results

Theorem (Matrix McDiarmid Inequality)

Let $\{Z_1, Z_2, ..., Z_n\}$ be independent random variables and $\mathbf{z} := (Z_1, Z_2, ..., Z_n)$. Let H be a function that maps n variables to a $d \times d$ self-adjoint matrix. Consider a sequence $\{A_1, A_2, ..., A_n\}$ of fixed self-adjoint matrices that satisfy:

$$(H(z_1, ..., z_k, ..., z_n) - H(z_1, ..., z'_k, ..., z_n))^2 \preceq A_k^2,$$

where z_i and z'_i range over all possible values of Z_i for each $1 \leq i \leq n$. Let

$$\sigma^2 := \left\| \sum_k A_k^2 \right\|,$$

then for any $t \geq 0$ we have:

$$\mathbf{P}\left(\lambda_{\max}\left(H(\mathbf{z}) - \mathbf{E}(H(\mathbf{z}))\right) \leq d \cdot e^{-t^2/8\sigma^2}.$$

Other results

Theorem (Matrix Hoeffding Inequality)

Let $(X_i)_{i \geq 0}$ be a sequence of independent, self-adjoint $d \times d$ random matrices and let $(A_i)_{i \geq 0}$ be a fixed sequence of self-adjoint matrices . Assume that each random variables satisfies $\mathbf{E} X_i = \mathbf{0}$ and $X_i \preceq A_i$ almost surely for any $i \geq 0$, where $\mathbf{0}$ is the zero $d \times d$ matrix. Let

$$\sigma^2 = \frac{1}{2} \left\| \sum_k A_k^2 + \mathbf{E} X_k^2 \right\| \leq \left\| \sum_k A_k^2 \right\|,$$

then for all $t \geq 0$ we have:

$$\mathbf{P}\left(\lambda_{\max}\left(\sum_k X_k \right) \geq t \right) \leq d \cdot e^{-t^2/(2\sigma^2)}.$$

Proof of Theorem 1

We direct our attention to the proof of Theorem 1 as the other theorems have similar proofs.

Proof of Theorem 1

We direct our attention to the proof of Theorem 1 as the other theorems have similar proofs. We start with the proof of Theorem 1.1. Recall

Theorem (Theorem 1.1)

Let $S_1, S_2, ..., S_n$ be independent $d_1 \times d_2$ random matrices with $\mathbf{E}(S_i) = \mathbf{0}$ for each i. Then the following is true:

$$\sqrt{\frac{1}{4} \cdot \nu(X)} + \frac{1}{4} \cdot L \leq \left(\mathbf{E}\left(\|X\|^2\right)\right)^{1/2} \leq \sqrt{C_d \cdot \nu(X)} + C_d \cdot L.$$

Hermitian dilatation

Hermitian dilatation

Definition

Let M be a $d_1 \times d_2$ matrix. We define the Hermitian dilatation $H(M)$ of M by:

$$H(M) := \begin{bmatrix} 0 & M \\ M^* & 0 \end{bmatrix}.$$

Note that $H(M)$ is symmetric and satisfies:

$$\|H(M)\| = \|M\|$$

and

$$\|\mathbf{E}H(M)^2\| = \max\left\{\|\mathbf{E}(MM^*)\|, \|\mathbf{E}(M^*M)\|\right\}. \tag{5}$$

Hermitian dilatation

As the Hermitian dilation is a linear map, we have:

$$H(X) = \sum_{i=1}^{n} H(S_i)$$

and so, we can assume without loss of generality that X and S_i's are centered Hermitian for any $1 \leq i \leq n$.

Main idea of the proof

The main idea behind the proof is that if we let $\xi_1, \xi_2, ..., \xi_n$ be n Rademacher random variables talking values ± 1 each with probability $1/2$ independent of the S_i's, then

$$X' := \sum_i \xi_i S_i,$$

has the same distribution as X. The advantage of working with X' is that we can condition on the values of S_i's and still get good bounds for $\|X'\|$.

Main Lemma

Lemma (Lemma 1)

Let $H_1, H_2, ..., H_n$ be fixed $d \times d$ Hermitian matrices and let $\xi_1, ..., \xi_n$ be independent Rademacher random variables. Then the following holds:

$$\left(\mathbf{E} \left\| \sum_{i=1}^{n} \xi_i H_i \right\|^2 \right)^{1/2} \leq \sqrt{1 + 2 \log d} \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^{1/2} .$$

Other version of the lemma

The same result holds if we replace the Rademacher random variables with standard normal ones. The proofs are almost identical.

Other version of the lemma

The same result holds if we replace the Rademacher random variables with standard normal ones. The proofs are almost identical.

Lemma (Lemma for Gaussian random variables)

Let $H_1, H_2, ..., H_n$ be fixed $d \times d$ Hermitian matrices and let $\gamma_1, ..., \gamma_n$ be independent $\mathcal{N}(0, 1)$ random variables. Then the following holds:

$$\left(\mathbf{E} \left\| \sum_{i=1}^{n} \gamma_i H_i \right\|^2 \right)^{1/2} \leq \sqrt{1 + 2 \log d} \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^{1/2}.$$

Proof of Lemma 1

The proof of Lemma 1 is based on the moment method. Define

$$Y := \sum_{i=1}^{n} \xi_i H_i.$$

Proof of Lemma 1

The proof of Lemma 1 is based on the moment method. Define

$$Y := \sum_{i=1}^{n} \xi_i H_i.$$

Let p be a fixed integer that we will choose it later. By Jensen we have:

$$\left(\mathbf{E}(\|Y\|^2\right)^{1/2} \leq \left(\mathbf{E}(\|Y\|^{2p}\right)^{1/2p}.$$

Proof of Lemma 1

The proof of Lemma 1 is based on the moment method. Define

$$Y := \sum_{i=1}^{n} \xi_i H_i.$$

Let p be a fixed integer that we will choose it later. By Jensen we have:

$$\left(\mathbf{E}(\|Y\|^2\right)^{1/2} \leq \left(\mathbf{E}(\|Y\|^{2p}\right)^{1/2p}.$$

Since all the eigenvalues of a Hermitian matrix are real, we have:

$$\left(\mathbf{E}(\|Y\|^2\right)^{1/2} \leq \left(\mathbf{E}(\|Y\|^{2p}\right)^{1/2p} \leq \left(\mathbf{E}(\mathrm{Trace}(Y^{2p}))\right)^{1/2p}.$$

Proof of Lemma

Proof of Lemma

Let Y_{+i} be the value of Y conditioned on the event that $Y_i = 1$ and define Y_{-i} similarly.

Precisely, we have

$$Y_{+i} := H_i + \sum_{j \neq i} \xi_j H_j \text{ and } Y_{-i} := -H_i + \sum_{j \neq i} \xi_j H_j.$$

Proof of Lemma

Let Y_{+i} be the value of Y conditioned on the event that $Y_i = 1$ and define Y_{-i} similarly.

Precisely, we have

$$Y_{+i} := H_i + \sum_{j \neq i} \xi_j H_j \text{ and } Y_{-i} := -H_i + \sum_{j \neq i} \xi_j H_j.$$

$$\begin{aligned}
\mathbf{E}\left(\mathrm{Trace}(Y^{2p})\right) &= \mathbf{E}\,\mathrm{Trace}(Y \cdot Y^{2p-1}) \\
&= \sum_{i=1}^{n} \mathbf{E}\left(\mathbf{E}_{\xi_i}\,\mathrm{Trace}(\xi_i H_i \cdot Y^{2p-1})\right) \\
&= \frac{1}{2}\sum_{i=1}^{n} \mathbf{E}\,\mathrm{Trace}\left(H_i \cdot \left(Y_{+i}^{2p-1} - Y_{-i}^{2p-1}\right)\right)
\end{aligned}$$

Proof of Lemma

Proof of Lemma

We can write

$$Y_{+i}^{2p-1} - Y_{-i}^{2p-1} = \sum_{q=0}^{2p-2} Y_{+i}^{q}(Y_{+i} - Y_{-i})Y_{-i}^{2p-q-2},$$

Proof of Lemma

We can write

$$Y_{+i}^{2p-1} - Y_{-i}^{2p-1} = \sum_{q=0}^{2p-2} Y_{+i}^{q}(Y_{+i} - Y_{-i})Y_{-i}^{2p-q-2},$$

It follows that

$$\mathbf{E}\left(\mathrm{Trace}(Y^{2p})\right) = \frac{1}{2}\sum_{i=1}^{n} \mathbf{E}\,\mathrm{Trace}\left(H_i \cdot \left(\sum_{j=0}^{2p-2} Y_{+i}^{j}(Y_{+i} - Y_{-i})Y_{-i}^{2p-2-j}\right)\right)$$
$$= \sum_{i=1}^{n}\sum_{j=0}^{2p-2} \mathbf{E}\,\mathrm{Trace}\left(H_i^2 \cdot \left(Y_{+i}^{j} Y_{-i}^{2p-2-j}\right)\right)$$

since $Y_{+i} - Y_{-i} = 2H_i$.

Proof of Lemma

Proof of Lemma

For real numbers a and b we have by AM-GM that:

$$a^j b^{2p-2-j} + a^{2p-2-j} b^j \leq a^{2p-2} + b^{2p-2}.$$

Proof of Lemma

For real numbers a and b we have by AM-GM that:

$$a^j b^{2p-2-j} + a^{2p-2-j} b^j \leq a^{2p-2} + b^{2p-2}.$$

The equivalent version for the trace of matrices is the following fact

Fact (The trace fomula)

$$\text{Trace}(Y_{+i}^j Y_{-i}^{2p-2-j} + Y_{+i}^{2p-2-j} Y_{-i}^j) \leq \text{Trace}(Y^{2p-2} + Y^{2p-2}).$$

Proof of the Trace formula

Proof of the Trace formula

To see this, let $Y_{+i} = \sum_k \lambda_k u_k u_k^*$ and $Y_{-i} = \sum_k \mu_k v_k v_k^*$ be the SVD decompositions of Y_{+i} and Y_{-i} respectively. Then:

$$\begin{aligned}
\text{Trace}(Y_{+i}^j Y_{-i}^{2p-2-j}) &= \text{Trace}\left(\left(\sum_{k=1}^{d} \lambda_k^j u_k u_k^*\right)\left(\sum_{k=1}^{d} \mu_k^{2p-2-j} v_k v_k^*\right)\right) \\
&= \sum_{k_1,k_2=1}^{d} \lambda_{k_1}^j \mu_{k_2}^{2p-2-j} \text{Trace}(u_{k_1} u_{k_1}^* v_{k_2} v_{k_2}^*) \\
&\leq \sum_{k_1,k_2=1}^{d} |\lambda_{k_1}|^j |\mu_{k_2}|^{2p-2-j} \left|u_{k_1}^* v_{k_2}\right|^2
\end{aligned}$$

Proof of the Trace formula

Proof of the Trace formula

It follows that:

$$
\begin{aligned}
\mathrm{Trace}(Y_{+i}^{j} Y_{-i}^{2p-2-j} + Y_{+i}^{2p-2-j} Y_{-i}^{j}) \\
&\leq \sum_{k_1,k_2=1}^{d} \left(\lambda_{k_1}^{2p-2} + \mu_{k_2}^{2p-2}\right) \left|u_{k_1}^{*} v_{k_2}\right|^2 \\
&= \sum_{k_1,k_2=1}^{d} \left(\lambda_{k_1}^{2p-2} + \mu_{k_2}^{2p-2}\right) \mathrm{Trace}(u_{k_1} u_{k_1}^{*} v_{k_2} v_{k_2}^{*}) \\
&= \mathrm{Trace}\left(\left(\sum_{k=1}^{d} \lambda_k^{2p-2} u_k u_k^{*}\right)\left(\sum_{k=1}^{d} v_k v_k^{*}\right)\right) + \\
&+ \mathrm{Trace}\left(\left(\sum_{k=1}^{d} \mu_k^{2p-2} v_k v_k^{*}\right)\left(\sum_{k=1}^{d} u_k u_k^{*}\right)\right) \\
&= \mathrm{Trace}\left(Y_{+i}^{2p-2} + Y_{-i}^{2p-2}\right)
\end{aligned}
$$

Back to the proof of Lemma

Back to the proof of Lemma

We established that

$$\mathrm{Trace}(Y_{+i}^{j} Y_{-i}^{2p-2-j} + Y_{+i}^{2p-2-j} Y_{-i}^{j}) \leq \mathrm{Trace}(Y_{+i}^{2p-2} + Y_{-i}^{2p-2}).$$

Back to the proof of Lemma

We established that

$$\mathrm{Trace}(Y_{+i}^{j} Y_{-i}^{2p-2-j} + Y_{+i}^{2p-2-j} Y_{-i}^{j}) \le \mathrm{Trace}(Y_{+i}^{2p-2} + Y_{-i}^{2p-2}).$$

The same proof is valid for

$$\mathrm{Trace}\left(H_i^2\left(Y_{+i}^{j} Y_{-i}^{2p-2-j} + Y_{+i}^{2p-2-j} Y_{-i}^{j}\right)\right) \le \mathrm{Trace}\left(H_i^2\left(Y_{+i}^{2p-2} + Y_{-i}^{2p-2}\right)\right).$$

as the trace is linear.

Proof of Lemma

We have

Proof of Lemma

We have

$$
\begin{aligned}
\mathbf{E}\left(\mathrm{Trace}(Y^{2p})\right) &= \sum_{i=1}^{n}\sum_{j=0}^{2p-2} \mathbf{E}\,\mathrm{Trace}\left(H_i^2 \cdot \left(Y_{+i}^{j} Y_{-i}^{2p-2-j}\right)\right) \\
&\leq \sum_{i=1}^{n} \frac{2p-1}{2} \mathbf{E}\,\mathrm{Trace}\left(H_i^2 \cdot \left(Y_{+i}^{2p-2} + Y_{-i}^{2p-2-j}\right)\right) \\
&= (2p-1)\sum_{i=1}^{n} \mathbf{E}\,\mathrm{Trace}\left(H_i^2\left(\mathbf{E}_{\xi_i} Y^{2p-2}\right)\right) \\
&= (2p-1)\mathbf{E}\left(\mathrm{Trace}\left(\sum_{i=1}^{n} H_i^2\right) Y^{2p-2}\right) \\
&\leq (2p-1)\left\|\sum_{i=1}^{n} H_i^2\right\| \mathbf{E}\left(\mathrm{Trace}(Y^{2p-2})\right)
\end{aligned}
$$

Proof of Lemma

Proof of Lemma

Recursively it follows that:

$$\begin{aligned}\mathbf{E}\left(\text{Trace } Y^{2p}\right) &\le (2p-1)!! \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^p \cdot \text{Trace } Y^0 \\ &= d \cdot (2p-1)!! \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^p\end{aligned}$$

where $(2p-1)!! = 1 \cdot 3 \cdot \ldots \cdot (2p-1)$.

Proof of Lemma

Recursively it follows that:

$$\mathbf{E}\left(\text{Trace } Y^{2p}\right) \le (2p-1)!! \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^p \cdot \text{Trace } Y^0$$
$$= d \cdot (2p-1)!! \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^p$$

where $(2p-1)!! = 1 \cdot 3 \cdot \ldots \cdot (2p-1)$.

$$\mathbf{E}(\|Y\|^2)^{1/2} \le \left(\mathbf{E}(\text{Trace}(Y^{2p}))\right)^{1/2p} \le (d \cdot (2p-1)!!)^{1/2p} \cdot \left\| \sum_{i=1}^{n} H_i^2 \right\|^{1/2}.$$

Proof of Lemma 1

Proof of Lemma 1

Note that:

$$(2p-1)!! \le \left(\frac{2p+1}{e}\right)^p,$$

pick $p = \lceil \log d \rceil$ to get

$$\mathbf{E}\left\|Y^2\right\|^{1/2} \le \sqrt{1+2\log d} \cdot \left\|\sum_{i=1}^{n} H_i^2\right\|^{1/2},$$

which completes the proof of Lemma 1.

Proposition

Proposition (Symmetrization)

Let $W_1, W_2, \ldots, W_n$ be $d_1 \times d_2$ independent random matrices. Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent Rademacher variables that are also independent of the W's. The following is true:

$$\frac{1}{2}\left(\mathbf{E}\left\|\sum_{i=1}^{n}\xi_i W_i\right\|^r\right)^{1/r} \leq \left(\mathbf{E}\left\|\sum_{i=1}^{n}(W_i - \mathbf{E}(W_i))\right\|^r\right)^{1/r} \leq 2\left(\mathbf{E}\left\|\sum_{i=1}^{n}\xi_i W_i\right\|^r\right)^{1/r}.$$

Proof of Proposition

Assume $r = 1$ (the proof for the general case is similar and it uses the convexity of $\|\cdot\|^r$).

Proof of Proposition

Assume $r = 1$ (the proof for the general case is similar and it uses the convexity of $\|\cdot\|^r$).

Let $W_1', W_2', \ldots W_n'$ be an identical copies of W_i's and let $\mathbf{E}'$ be the expectation with respect to those. Since $\|\cdot\|$ is convex, by Jensen we have

$$\begin{aligned}
\mathbf{E}\left\|\sum_{i=1}^{n}(W_i - \mathbf{E}W_i)\right\| &= \mathbf{E}\left\|\sum_{i=1}^{n}[(W_i - \mathbf{E}W_i) - \mathbf{E}'(W_i' - \mathbf{E}'W_i')]\right\| \\
&\leq \mathbf{E}\left[\mathbf{E}'\left\|\sum_{i=1}^{n}(W_i - \mathbf{E}W_i) - (W_i' - \mathbf{E}(W_i))\right\|\right] \\
&= \mathbf{E}\left\|\sum_{i=1}^{n}(W_i - W_i')\right\|.
\end{aligned}$$

Proof of Proposition

Proof of Proposition

Recall that $\xi_1, .., \xi_n$ are independent Rademacher random variables.

$$\begin{aligned}
\mathbf{E}\left\|\sum_{i=1}^{n}(W_i - \mathbf{E}W_i)\right\| &= \mathbf{E}\left\|\sum_{i=1}^{n}(W_i - W_i')\right\| \\
&= \mathbf{E}\left\|\sum_{i=1}^{n}\xi_i(W_i - W_i')\right\| \\
&\le \mathbf{E}\left\|\sum_{i=1}^{n}\xi_i W_i\right\| + \mathbf{E}\left\|\sum_{i=1}^{n}-\xi_i W_i'\right\| \\
&= 2\,\mathbf{E}\left\|\sum_{i=1}^{n}\xi_i W_i\right\|
\end{aligned}$$

Proof of Proposition

Recall that $\xi_1, .., \xi_n$ are independent Rademacher random variables.

$$\begin{aligned}\mathbf{E}\left\|\sum_{i=1}^{n}(W_i - \mathbf{E}W_i)\right\| &= \mathbf{E}\left\|\sum_{i=1}^{n}(W_i - W_i')\right\| \\ &= \mathbf{E}\left\|\sum_{i=1}^{n}\xi_i(W_i - W_i')\right\| \\ &\leq \mathbf{E}\left\|\sum_{i=1}^{n}\xi_i W_i\right\| + \mathbf{E}\left\|\sum_{i=1}^{n}-\xi_i W_i'\right\| \\ &= 2\,\mathbf{E}\left\|\sum_{i=1}^{n}\xi_i W_i\right\|\end{aligned}$$

The lower bound uses similar techniques.

Simple fact about positive-definite matrices

Simple fact about positive-definite matrices

Fact (Fact 1)

Let $A_1, A_2, \ldots, A_n$ be $d \times d$ positive-semidefinite matrices. Then:

$$\left\| \sum_{i=1}^{n} A_i^2 \right\| \leq \max_i \|A_i\| \cdot \left\| \sum_{i=1}^{n} A_i \right\|.$$

Simple fact about positive-definite matrices

Fact (Fact 1)

Let $A_1, A_2, \ldots, A_n$ be $d \times d$ positive-semidefinite matrices. Then:

$$\left\| \sum_{i=1}^{n} A_i^2 \right\| \leq \max_i \|A_i\| \cdot \left\| \sum_{i=1}^{n} A_i \right\| .$$

Proof.

Let $m \geq \lambda_{\max}(A)$, by writing the eigenvalue decomposition of A we have $A^2 \preceq mA$. Pick $m = \max_i \lambda_{\max}(A_i)$, then

$$\sum_{i=1}^{n} A_i^2 \preceq m \sum_{i=1}^{n} A_i .$$

The conclusion follows by talking the spectral norm of both sides. □

Proof structure

Now we have all the ingredients to finish our proof. We will first proof the equivalent for the positive-semidefinite case and then for the centered Hermitian case which implies Theorem 1. Note that in the positive-semidefinite case we do not require the matrices to be centered, so, the bounds are slightly different then the ones in Theorem 1.1.

Proof structure

Now we have all the ingredients to finish our proof. We will first proof the equivalent for the positive-semidefinite case and then for the centered Hermitian case which implies Theorem 1. Note that in the positive-semidefinite case we do not require the matrices to be centered, so, the bounds are slightly different then the ones in Theorem 1.1.

Theorem (Theorem 1.1 for positive-semidefinite matrices)

Assume that S_i's are $d \times d$ independent positive-semidefinite random matrices.

$$\frac{1}{4}\left(\|\mathbf{E}X\|^{1/2} + \left(\mathbf{E}\max_i \|S_i\|\right)^{1/2}\right)^2 \leq$$

$$\leq \mathbf{E}\|X\| \leq \left(\|\mathbf{E}X\|^{1/2} + \sqrt{C_d} \cdot \left(\mathbf{E}\max_i \|S_i\|\right)^{1/2}\right)^2.$$

Proof of the upper bound for the positive-semidefinite case

Proof of the upper bound for the positive-semidefinite case

We want to proof that when S_i's are positive-semidefinite we have

$$\mathbf{E}\|X\| \leq \left(\|\mathbf{E}X\|^{1/2} + \sqrt{C_d} \cdot \left(\mathbf{E} \max_i \|S_i\| \right)^{1/2} \right)^2 .$$

Proof of the upper bound for the positive-semidefinite case

We want to proof that when S_i's are positive-semidefinite we have

$$\mathbf{E}\|X\| \leq \left(\|\mathbf{E}X\|^{1/2} + \sqrt{C_d} \cdot \left(\mathbf{E} \max_i \|S_i\| \right)^{1/2} \right)^2 .$$

By the triangle inequality and the Proposition 1, we have:

$$\begin{aligned} \mathbf{E}\|X\| = \mathbf{E} \left\| \sum_{i=1}^n S_i \right\| &\leq \left\| \sum_{i=1}^n \mathbf{E} S_i \right\| + \mathbf{E} \left\| \sum_{i=1}^n (S_i - \mathbf{E} S_i) \right\| \\ &\leq \left\| \sum_{i=1}^n \mathbf{E} S_i \right\| + 2\mathbf{E} \left\| \sum_{i=1}^n \xi_i S_i \right\| , \end{aligned}$$

where ξ_i's are independent Rademacher random variables.

Proof of upper bound for the positive-semidefinite case

Condition on the matrices S_i's and use Lemma 1:

$$\mathbf{E}\left\|\sum_{i=1}^{n}\xi_i S_i\right\| = \mathbf{E}\left(\mathbf{E}_{\xi}\left\|\sum_{i=1}^{n}\xi_i S_i\right\|\right) \leq \sqrt{1+2\log d}\cdot\mathbf{E}\left(\left\|\sum_{i=1}^{n}S_i^2\right\|^{1/2}\right)$$

Proof of upper bound for the positive-semidefinite case

Condition on the matrices S_i's and use Lemma 1:

$$\mathbf{E}\left\|\sum_{i=1}^{n}\xi_i S_i\right\| = \mathbf{E}\left(\mathbf{E}_\xi\left\|\sum_{i=1}^{n}\xi_i S_i\right\|\right) \leq \sqrt{1+2\log d}\cdot\mathbf{E}\left(\left\|\sum_{i=1}^{n}S_i^2\right\|^{1/2}\right)$$

By Fact 1 and Cauchy inequality we have:

$$\begin{aligned}\mathbf{E}\left(\left\|\sum_{i=1}^{n}S_i^2\right\|^{1/2}\right) &\leq \mathbf{E}\left((\max_i\|S_i\|)^{1/2}\cdot\left\|\sum_{i=1}^{n}S_i\right\|^{1/2}\right)\\ &\leq \left(\mathbf{E}\max_i\|S_i\|\right)^{1/2}\cdot\left(\mathbf{E}\left\|\sum_{i=1}^{n}S_i\right\|\right)^{1/2}\\ &= \left(\mathbf{E}\max_i\|S_i\|\right)^{1/2}\cdot(\mathbf{E}\|X\|)^{1/2}\end{aligned}$$

Proof of the upper bound for the positive-semidefinite case

Proof of the upper bound for the positive-semidefinite case

It follows that:

$$\mathbf{E}\|X\| \le \left\|\sum_{i=1}^{n} \mathbf{E} S_i\right\| + \sqrt{4 + 8\log d} \cdot \left(\mathbf{E}\max_i \|S_i\|\right)^{1/2} \cdot (\mathbf{E}\|X\|)^{1/2}.$$

Proof of the upper bound for the positive-semidefinite case

It follows that:

$$\mathbf{E}\|X\| \leq \left\|\sum_{i=1}^{n} \mathbf{E} S_i\right\| + \sqrt{4 + 8\log d} \cdot \left(\mathbf{E}\max_i \|S_i\|\right)^{1/2} \cdot (\mathbf{E}\|X\|)^{1/2}.$$

This implies:

$$\mathbf{E}\|X\|^{1/2} \leq \left\|\sum_{i=1}^{n} \mathbf{E} S_i\right\|^{1/2} + \sqrt{4 + 8\log d} \cdot \left(\mathbf{E}\max_i \|S_i\|\right)^{1/2},$$

which completes the proof for the upper bound.

Proof of the lower bound for the positive-semidefinite case

Proof of the lower bound for the positive-semidefinite case

Note that since $S_1, \dots, S_n$ are positive-definite we have:

$$\mathbf{E}\|X\| \geq \mathbf{E}\max_i \|S_i\|.$$

Proof of the lower bound for the positive-semidefinite case

Note that since $S_1, \ldots, S_n$ are positive-definite we have:

$$\mathbf{E}\|X\| \geq \mathbf{E}\max_i \|S_i\|.$$

By Jensen, we also have:

$$\mathbf{E}\|X\| \geq \|\mathbf{E}X\|.$$

Proof of the lower bound for the positive-semidefinite case

Note that since $S_1, \ldots, S_n$ are positive-definite we have:

$$\mathbf{E}\|X\| \geq \mathbf{E}\max_i \|S_i\|.$$

By Jensen, we also have:

$$\mathbf{E}\|X\| \geq \|\mathbf{E}X\|.$$

This implies that:

$$\mathbf{E}\|X\| \geq \frac{1}{4}\left(\|\mathbf{E}X\|^{1/2} + (\mathbf{E}\max_i \|S_i\|)^{1/2}\right)^2,$$

which completes the proof for positive-semidefinite case.

Centered-Hermitian case

Theorem (Theorem 1.1 for centered Hermitian matrices)

Assume that S_i's are $d \times d$ independent centered Hermitian random matrices.

$$\frac{1}{2}\|\mathbf{E}X^2\|^{1/2}+\frac{1}{4}\left(\mathbf{E}\max_i\|S_i\|^2\right)^{1/2}\leq$$

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2}\leq\sqrt{C_d}\cdot\|\mathbf{E}(X^2)\|^{1/2}+C_d\cdot\left(\mathbf{E}\max_i\|S_i\|^2\right)^{1/2}$$

Proof of the upper bound for the centered Hermitian case

Proof of the upper bound for the centered Hermitian case

Assume that S_i's are centered Hermitian matrices. We want to prove that:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} \leq \sqrt{C_d} \cdot \|\mathbf{E}(X^2)\|^{1/2} + C_d \cdot \left(\mathbf{E} \max_i \|S_i\|^2\right)^{1/2}.$$

Proof of the upper bound for the centered Hermitian case

Assume that S_i's are centered Hermitian matrices. We want to prove that:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} \leq \sqrt{C_d} \cdot \|\mathbf{E}(X^2)\|^{1/2} + C_d \cdot \left(\mathbf{E} \max_i \|S_i\|^2\right)^{1/2}.$$

Condition on the values of S_i's and apply Lemma 1:

$$(\mathbf{E}\|X\|^2)^{1/2} = \left(\mathbf{E}\left\|\sum_{i=1}^n S_i\right\|^2\right)^{1/2} \leq 2\left(\mathbf{E}\left[\mathbf{E}_\xi \left\|\sum_{i=1}^n \xi_i S_i\right\|^2\right]\right)^{1/2}$$

$$\leq \sqrt{4 + 8\log d} \cdot \left(\mathbf{E}\left\|\sum_{i=1}^n S_i^2\right\|\right)^{1/2}$$

Proof of the upper bound for the centered Hermitian case

Note that S_i^2's are positive definite matrices, we have just proved that

$$\mathbf{E}\left\|\sum_{i=1}^{n} S_i^2\right\| \le \left(\left\|\mathbf{E}\sum_{i=1}^{n} S_i^2\right\|^{1/2} + \sqrt{C_d}\cdot\left(\mathbf{E}\max_i \|S_i^2\|\right)^{1/2}\right)^2.$$

Proof of the upper bound for the centered Hermitian case

Note that S_i^2's are positive definite matrices, we have just proved that

$$\mathbf{E}\left\|\sum_{i=1}^n S_i^2\right\| \leq \left(\left\|\mathbf{E}\sum_{i=1}^n S_i^2\right\|^{1/2} + \sqrt{C_d}\cdot\left(\mathbf{E}\max_i \|S_i^2\|\right)^{1/2}\right)^2.$$

This implies

$$(\mathbf{E}\|X\|^2)^{1/2} \leq \sqrt{C_d}\left(\left\|\mathbf{E}\sum_{i=1}^n S_i^2\right\|^{1/2} + \sqrt{C_d}\cdot\left(\mathbf{E}\max_i \|S_i^2\|\right)^{1/2}\right),$$

which completes the proof for the upper bound.

Proof of the lower bound for the centered Hermitian case

Proof of the lower bound for the centered Hermitian case

Let S_i's be centered Hermitian matrices. We want to prove that:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} \geq \frac{1}{2}\|\mathbf{E}(X^2)\|^{1/2} + \frac{1}{4}\left(\mathbf{E}\left(\max_i \|S_i\|^2\right)\right)^{1/2}.$$

Proof of the lower bound for the centered Hermitian case

Let S_i's be centered Hermitian matrices. We want to prove that:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} \geq \frac{1}{2}\|\mathbf{E}(X^2)\|^{1/2} + \frac{1}{4}\left(\mathbf{E}\left(\max_i \|S_i\|^2\right)\right)^{1/2}.$$

Using Fact 1 we have:

$$\mathbf{E}(\|X\|^2) = \mathbf{E}\left(\left\|\sum_{i=1}^n S_i\right\|^2\right) \geq \frac{1}{4}\mathbf{E}\left(\left\|\sum_{i=1}^n \xi_i S_i\right\|^2\right),$$

where ξ_i's are Rademacher independent random variables.

Proof of the lower bound for the centered Hermitian case

Condition on the values of S_i. Without loss of generality we assume that $\|S_1\| = \max_i \|S_i\|$. Condition further on the value of ξ_1 and so, by Jensen we have:

$$\mathbf{E}_{\xi}\left(\left\|\sum_{i=1}^{n}\xi_i S_i\right\|^2\right) \geq \mathbf{E}_{\xi_1}\left(\left\|\mathbf{E}\left(\sum_{i=1}^{n}\xi_i S_i|\xi_1\right)\right\|^2\right)$$

$$= \mathbf{E}_{\xi_1}\left(\|\xi_1 S_1\|^2\right) = \|S_1\|$$

$$= \max_i \|S_i\|$$

Proof of the lower bound for the centered Hermitian case

Combining the last two inequalities and take square root we have:

$$\mathbf{E}(\|X\|^2)^{1/2} \geq \frac{1}{2}\left(\mathbf{E}\max_i \|S_i\|^2\right)^{1/2}.$$

Proof of the lower bound for the centered Hermitian case

Combining the last two inequalities and take square root we have:

$$\mathbf{E}(\|X\|^2)^{1/2} \geq \frac{1}{2}\left(\mathbf{E}\max_i \|S_i\|^2\right)^{1/2}.$$

Using that S_i's are centered and Jensen's inequality we have:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} = \left(\mathbf{E}\|X^2\|\right)^{1/2} \geq \|\mathbf{E}X^2\|^{1/2}.$$

Proof of the lower bound for the centered Hermitian case

Combining the last two inequalities and take square root we have:

$$\mathbf{E}(\|X\|^2)^{1/2} \geq \frac{1}{2}\left(\mathbf{E}\max_i \|S_i\|^2\right)^{1/2}.$$

Using that S_i's are centered and Jensen's inequality we have:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} = \left(\mathbf{E}\|X^2\|\right)^{1/2} \geq \|\mathbf{E}X^2\|^{1/2}.$$

Averaging the last two inequalities leads to:

$$\left(\mathbf{E}(\|X\|^2)\right)^{1/2} \geq \frac{1}{2}\|\mathbf{E}(X^2)\|^{1/2} + \frac{1}{4}\left(\mathbf{E}\left(\max_i \|S_i\|^2\right)\right)^{1/2},$$

which completes the proof of the Hermitian case and hence Theorem 1.1.

Proof of Theorem 1.2

In order to proof Theorem 1.2 we need to define the exponential and the logarithm function of a matrix and discuss some properties.

Definition

- Let A be a fixed Hermitian matrix, define:

$$e^A := I + \sum_{q=1}^{\infty} \frac{A^q}{q!}.$$

A more rigurose definition can be done using the SVD decomposition.

Proof of Theorem 1.2

In order to proof Theorem 1.2 we need to define the exponential and the logarithm function of a matrix and discuss some properties.

Definition

- Let A be a fixed Hermitian matrix, define:

$$e^A := I + \sum_{q=1}^{\infty} \frac{A^q}{q!}.$$

  A more rigurose definition can be done using the SVD decomposition.
- Let A be a fixed Hermitian matrix, define:

$$\log\left(e^A\right) = A.$$

Proof of Theorem 1.2

We recall some properties from the Linear Algebra that we will use in the proof.

Properties

- *Let A and B be Hermitian matrices, then:*

$$A \preceq B \text{ implies } \lambda_i(A) \leq \lambda_i(B) \text{ for each } i.$$

Proof of Theorem 1.2

We recall some properties from the Linear Algebra that we will use in the proof.

Properties

- *Let A and B be Hermitian matrices, then:*

$$A \preceq B \text{ implies } \lambda_i(A) \leq \lambda_i(B) \text{ for each } i.$$

- *Let A and B be Hermitian matrices, then:*

$$A \preceq B \text{ implies } \operatorname{Trace} e^A \leq \operatorname{Trace} e^B.$$

Proof of Theorem 1.2

We recall some properties from the Linear Algebra that we will use in the proof.

Properties

- *Let A and B be Hermitian matrices, then:*

$$A \preceq B \text{ implies } \lambda_i(A) \leq \lambda_i(B) \text{ for each } i.$$

- *Let A and B be Hermitian matrices, then:*

$$A \preceq B \text{ implies } \operatorname{Trace} e^A \leq \operatorname{Trace} e^B.$$

- *Let A and B be positive-semidefinite matrices, then:*

$$A \preceq B \text{ implies } \log(A) \leq \log(B).$$

Two important facts

Proposition (Conjugation Rule)

Let A and B be two Hermitian matrices of the same dimension, and let H be a general matrix with compatible dimensions. Then

$$A \preceq B \text{ implies } HAH^* \preceq HBH^*.$$

Two important facts

Proposition (Conjugation Rule)

Let A and B be two Hermitian matrices of the same dimension, and let H be a general matrix with compatible dimensions. Then

$$A \preceq B \text{ implies } HAH^* \preceq HBH^*.$$

Proposition (Transfer Rule)

Let f and g be real-valued functions defined on the interval I of the real line, and let A be an Hermitian matrix whose eigenvalues are contained in I. Then

$$f(a) \leq g(a) \text{ for each } a \in I \text{ implies } f(A) \preceq g(A).$$

Lieb's Theorem

Theorem (Lieb's Theorem)

Let H be a fixed Hermitian matrix. The function:

$$A \longrightarrow \operatorname{Trace} e^{H+\log A},$$

is a concave map on the convex cone of positive-definite matrices. As a consequence, we have:

$$\mathbf{E} \operatorname{Trace} e^{H+X} \leq \operatorname{Trace} e^{H+\log(\mathbf{E} e^{X})},$$

where X is a random Hermitian matrix.

Main lemma

Lemma (Lemma 2)

Let H be a random centered Hermitian matrix such that $\lambda_{\max}(H) \leq R$. Then, for $0 < \theta < 3/R$,

$$\mathbf{E}\left(e^{\theta H}\right) \preceq \exp\left(\frac{\theta^2/2}{1-\theta R/3} \cdot \mathbf{E}(H^2)\right)$$

and

$$\log\left(\mathbf{E}\left(e^{\theta H}\right)\right) \preceq \frac{\theta^2/2}{1-\theta R/3} \cdot \mathbf{E}(H^2).$$

Proof of Lemma 2

Note that if we prove the first relation, the second one follows by talking logarithm and using the fact that the log is a monotone function.

Proof of Lemma 2

Note that if we prove the first relation, the second one follows by talking logarithm and using the fact that the log is a monotone function.
Fix parameter $\theta > 0$. Write

$$e^{\theta H} = I + \theta H + \left(e^{\theta H} - \theta H - I\right) = I + \theta H + H \cdot f(H) \cdot H,$$

where f is defined by:

$$f(x) = \begin{cases} \frac{e^{\theta x} - \theta x - 1}{x^2} & \text{if } x \neq 0 \\ f(x) = 0 & \text{if } x = 0. \end{cases}$$

Proof of Lemma 2

Note that f is increasing as its derivatives is positive, hence

$$f(x) \leq f(R) \text{ for } x \leq R.$$

Proof of Lemma 2

Note that f is increasing as its derivatives is positive, hence

$$f(x) \leq f(R) \text{ for } x \leq R.$$

Since $\|H\| \leq R$, we have, by Transfer Rule

$$f(H) \preceq f(R)I.$$

Proof of Lemma 2

Note that f is increasing as its derivatives is positive, hence

$$f(x) \leq f(R) \text{ for } x \leq R.$$

Since $\|H\| \leq R$, we have, by Transfer Rule

$$f(H) \preceq f(R)I.$$

By Conjugation Rule we have:

$$e^{\theta H} \preceq I + \theta H + H\left(f(R) \cdot I\right)H = I + \theta H + f(R) \cdot H^2$$

Proof of Lemma 2

By Taylor, we can estimate $f(R)$:

$$f(R) = \frac{e^{\theta R} - \theta R - 1}{R^2} = \frac{1}{R^2}\sum_{q=2}^{\infty}\frac{(\theta R)^q}{q!} \leq \frac{\theta^2}{2}\sum_{q=2}^{\infty}\frac{(\theta R)^{q-2}}{3^{q-2}} = \frac{\theta^2/2}{1-\theta R/3},$$

where we used that $q! \geq 2 \cdot 3^{q-2}$, for $q \geq 2$.

Proof of Lemma 2

By Taylor, we can estimate $f(R)$:

$$f(R) = \frac{e^{\theta R} - \theta R - 1}{R^2} = \frac{1}{R^2} \sum_{q=2}^{\infty} \frac{(\theta R)^q}{q!} \leq \frac{\theta^2}{2} \sum_{q=2}^{\infty} \frac{(\theta R)^{q-2}}{3^{q-2}} = \frac{\theta^2/2}{1 - \theta R/3},$$

where we used that $q! \geq 2 \cdot 3^{q-2}$, for $q \geq 2$. As H^2 is positive-semidefinite, this implies,

$$e^{\theta H} \preceq I + \theta H + \frac{\theta^2/2}{1 - \theta R/3} H^2 := I + \theta H + g(\theta) H^2.$$

Proof of Lemma 2

By Taylor, we can estimate $f(R)$:

$$f(R) = \frac{e^{\theta R} - \theta R - 1}{R^2} = \frac{1}{R^2} \sum_{q=2}^{\infty} \frac{(\theta R)^q}{q!} \leq \frac{\theta^2}{2} \sum_{q=2}^{\infty} \frac{(\theta R)^{q-2}}{3^{q-2}} = \frac{\theta^2/2}{1 - \theta R/3},$$

where we used that $q! \geq 2 \cdot 3^{q-2}$, for $q \geq 2$. As H^2 is positive-semidefinite, this implies,

$$e^{\theta H} \preceq I + \theta H + \frac{\theta^2/2}{1 - \theta R/3} H^2 := I + \theta H + g(\theta) H^2.$$

The expectation preserves the semidefinite order:

$$\mathbf{E}\left(e^{\theta H}\right) \preceq I + g(\theta) \cdot \mathbf{E}(H^2) \preceq \exp\left(g(\theta) \cdot \mathbf{E}(H^2)\right),$$

where in the last step we used that $1 + a \leq e^a$.

Proof of Theorem 1.2

Let $0 < \theta < 3/R$ be a real number to be chosen later. Recall that we are working under the assumptions that X is a Hermitian matrix. By Markov's inequality we have:

$$\begin{aligned} \mathbf{P}(\lambda_{\max}(X) \geq t) &= \mathbf{P}\left(e^{\theta\lambda_{\max}(X)} \geq e^{\theta t}\right) \\ &\leq e^{-\theta t}\mathbf{E}\left(e^{\theta\lambda_{\max}(X)}\right) \\ &= e^{-\theta t}\mathbf{E}\left(e^{\lambda_{\max}(\theta X)}\right) \\ &= e^{-\theta t}\mathbf{E}\left(\lambda_{\max}(e^{\theta X})\right) \\ &= e^{-\theta t}\mathbf{E}\left(\operatorname{Trace} e^{\theta X}\right) \end{aligned}$$

Proof of Theorem 1.2

If we apply the Lieb's theorem recursively, for each S_i we have:

$$\mathbf{E}\left(\mathrm{Trace}\, e^{\theta(\sum_{i=1}^{n} S_i)}\right) \leq \mathrm{Trace}\left(e^{\sum_{i=1}^{n} \log \mathbf{E} e^{\theta S_i}}\right).$$

By Lemma 2 we have:

$$\begin{aligned}\mathbf{P}(\lambda_{\max}(X) \geq t) &\leq e^{-\theta t}\, \mathrm{Trace}\left(e^{\sum_{i=1}^{n} \log \mathbf{E} e^{\theta S_i}}\right)\\ &\leq e^{-\theta t}\, \mathrm{Trace}\left(e^{\sum_{i=1}^{n} g(\theta)\mathbf{E}(S_i^2)}\right)\\ &\leq d e^{-\theta t} e^{g(\theta)\cdot\nu(X)},\end{aligned}$$

where in the last step we bounded the trace of a hermitian matrix by d times its largest eigenvalue.

Proof of Theorem 1.2

If we apply the Lieb's theorem recursively, for each S_i we have:

$$\mathbf{E}\left(\operatorname{Trace} e^{\theta(\sum_{i=1}^n S_i)}\right) \le \operatorname{Trace}\left(e^{\sum_{i=1}^n \log \mathbf{E} e^{\theta S_i}}\right).$$

By Lemma 2 we have:

$$\begin{aligned} \mathbf{P}(\lambda_{\max}(X) \ge t) &\le e^{-\theta t} \operatorname{Trace}\left(e^{\sum_{i=1}^n \log \mathbf{E} e^{\theta S_i}}\right) \\ &\le e^{-\theta t} \operatorname{Trace}\left(e^{\sum_{i=1}^n g(\theta) \mathbf{E}(S_i^2)}\right) \\ &\le d e^{-\theta t} e^{g(\theta)\cdot \nu(X)}, \end{aligned}$$

where in the last step we bounded the trace of a hermitian matrix by d times its largest eigenvalue.
Pick $\theta = t/(\nu(X) + Rt/3)$ to conclude the proof of Theorem 1.2.

Application: Randomized Sparsification of a Matrix

Sparse matrices has several potential advantages. Firstly, it is considerably less expansive to store than a dense one. Secondly, many algorithms run more efficient and faster on sparse matrices.

Our task is that given a dense $d_1 \times d_2$ matrix A, find a sparse matrix R, which approximate A with respect to the spectral norm, that is we want $\|A - R\|_2$ to be as small as possible.

Application: Randomized Sparsification of a Matrix

We start by expressing the matrix A as a sum of its entries,

$$A = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} a_{ij} E_{ij},$$

where E_{ij} is the matrix with all zero entries, but its $(i,j)^{th}$ which is 1.

Application: Randomized Sparsification of a Matrix

Define the following sampling probabilities:

$$p_{ij} = \frac{1}{2}\left(\frac{|a_{ij}|^2}{\|A\|_F^2} + \frac{|a_{ij}|}{\|A\|_1}\right),$$

where $\|A\|_1 := \sum_{i,j} |a_{ij}|$. Note that:

$$\sum_{i=1}^{d_1}\sum_{j=1}^{d_2} p_{ij} = 1.$$

Define R to be the random matrix that has exactly one entry:

$$R = \frac{1}{p_{ij}} a_{ij} E_{ij} \text{ with probability } p_{ij}.$$

Application: Randomized Sparsification of a Matrix

Note that

$$\mathbf{E}(R) = \sum_{i,j} \left(\frac{1}{p_{ij}} a_{ij} E_{ij} \right) p_{Ij} = A.$$

The problem with R to be an approximation for A is that it has huge variance. This can be overcome if we take the average of n of them, where n is big. Fix n big and define:

$$R_n = \frac{1}{n} \sum_{i=1}^{n} R(i),$$

where $R(i)$'s are independent copies of R.

Application: Randomized Sparsification of a Matrix

Proposition

$$\mathbf{E}\|R_n - A\| \leq \sqrt{\frac{4\|A\|_F^2 \cdot \max(d_1, d_2) \log(d_1 + d_2)}{n}} + \frac{4\|A\|_1 \log(d_1 + d_2)}{3n}.$$

Application: Randomized Sparsification of a Matrix

Let $D := \max(d_1, d_2) \log(d_1 + d_2)$. Note that $\|A\|_1 \leq \max(d_1, d_2)\|A\|_F$ so the bound can be interpreted also as:

$$\frac{\mathbf{E}\|R_n - A\|}{\|A\|} \leq \frac{\|A\|_F}{\|A\|} \cdot \left(\sqrt{\frac{4D}{n}} + \frac{4D}{3n} \right) = \mathrm{srank}(A) \cdot \left(\sqrt{\frac{4D}{n}} + \frac{4D}{3n} \right),$$

where $\mathrm{srank}(A) := \|A\|_F / \|A\|$ is the stable rank.

Application: Randomized Sparsification of a Matrix

The proof of the proposition will follow by Theorem 1. Note that by definition

$$p_{ij} \geq \frac{1}{2}\frac{|a_{ij}|}{\|A\|_1} \text{ and } p_{ij} \geq \frac{1}{2} \cdot \frac{|a_{ij}|^2}{\|A\|_F^2}.$$

This implies that:

$$\|R\| \leq \max_{i,j} \|p_{ij}^{-1} a_{ij} E_{ij}\| = \max_{i,j} \frac{|a_i j|}{p_{ij}} \leq 2\|A\|_1$$

and

$$\mathbf{E}(RR^*) = \sum_{i,j} \frac{|a_{ij}|^2}{p_{ij}} E_{ii} \preceq 2d_2\|A\|_F^2 \cdot I_{d_1},$$

$$\mathbf{E}(RR^*) = \sum_{i,j} \frac{|a_{ij}|^2}{p_{ij}} E_{ii} \preceq 2d_1\|A\|_F^2 \cdot I_{d_2},$$

which implies

$$\nu(R_n) \leq 2\max(d_1, d_2).$$

Questions?