Chapter 6: Random Matrices and Covariance Estimation:

6.1. Eigenvalues shift a function toward the roots with regularly, Zero equalities.

$$X^T A X = X^T B X + X^T (A-B) X$$

$$\lambda_{min}(A) = \lambda_{min}(B) + \|A-B\| \quad \text{or} \quad \lambda_{max}(A) = \lambda_{max}(B) + \|A-B\|$$

$$\lambda_{min}(A) - \lambda_{min}(B) = \|A-B\| \qquad \lambda_{max}(A) - \lambda_{max}(B) = \|A-B\|$$

6.2. $A_{n\times m} \in \mathbb{R}$; $q \in [1, \infty]$; $|||A|||_q = \sup\limits_{\|x\|=1} \|Ax\|_q$

a) $|||A|||_2 = \sup\limits_{\|x\|_2=1} \sqrt{X^T A^T A X} = \sigma_{max}(A)$

$|||A|||_1 = \sup\limits_{\|x\|=1} |A_{ij}| \leq \max |A_{ij}|$

$|||A|||_\infty = \sup\limits_{\|x\|=1} \max\limits_{i\in[m]} |\sum_j A_{ij}| \leq \max\limits_{i\in[m]} \|A_{ij}\|$

b) $|||AB|||_q = \left\| A \frac{B}{\|B\|} \right\| \|B\| = \|A\| \|B\|$

c) $|||A|||_2^2 = \|A^T A\| = \|A\|_1 \|A\|_1 = \|A\|_\infty \|A\|_1$

6.3. $A_{d\times d}$, $B_{d\times d}$ where $0 \leq A_{ij} \leq B_{ij}$ for all $i,j = 1 \ldots d$

a) Proof by Induction:

Base case: $0 \leq A^1 \leq B^1$

Next step: $0 \leq A^2 = \sum\limits_{k=1}^{d} A_{ik} A_{kj} \leq \sum\limits_{k=1}^{d} B_{ik} B_{kj} = B^2$

Induction Step: $0 \leq A^m \leq B^m$

b) If $0 \leq A^m \leq B^m$ then $0 \leq (A^m)^{1/2} = |A^m| \leq (B^m)^{1/2} = |B^m|$

and $|||A|||_2 \leq |||B|||_2$

c) Proof by Induction:

Base case: $C \le |C|$

Next step: $C^2 = \sum_{i=1}^{d} C_{ik} C_{kj} \le \left| \sum_{i=1}^{d} C_{ik} C_{kj} \right| \le |C^2|$

Induction Step: $C^m \le |C^m|$

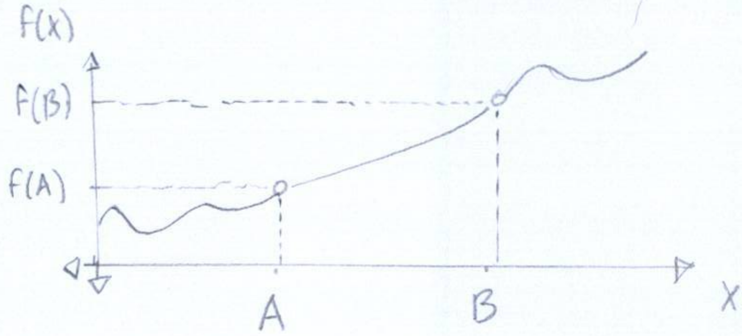
$|||C|||_2 \le \| \, |C| \, \|_2$

6.4. $A \in S^{d \times d}$

$I_d + A = Q^T Q + Q^T \Lambda Q$ where Q is involuntary and $A = Q^T \Lambda Q$

$\le Q e^{T \Lambda} Q$

$\le e^{A}$

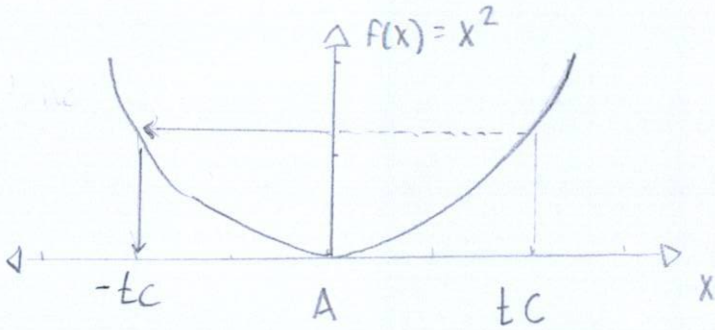


(Matrix Monotone)

$f(A) \le f(B)$ when $A < B$

6.5.

a) $f(A) = A^2$ is not matrix monotone because multiple solutions for specific matrix cases, such as:



$A = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix}$; $C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$f(A + tC) = (A + tC)^2 = A(A + tC) + tC(A + tC)$

$= A^2 + t(AC + CA) + (tC)^2$

$= \begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix} + t\left(\begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix}\right) + t^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$= f(A - tC)$

When $A - tC < A$, not $f(A - tC)$ and $f(A)$.

Matrix Rule:

$(a+b)^2 \neq a^2 + 2ab + b^2$

$(A+B)^2 = A^2 + (AB + BA) + B^2$

b) If $A < A + tC$, then $e^A < e^{A+tC}$, but not always true for matrices.

$$e^{(A+tC)} = 1 + \frac{(A+tC)}{1!} + \frac{(A+tC)^2}{2!} + \frac{(A+tC)^3}{3!} + \cdots$$

$$= 1 + A + tC + \frac{A^2 + t(AC+CA) + (tC)^2}{2!} + \cdots$$

$$= e^A + tC + \frac{t(AC+CA) + (tC)^2}{2!} + \cdots$$

$$= e^A + e^{tC} - 1 + \frac{t(AC+CA)}{2!} + \cdots$$

$e^A < e^{(A+tC)}$ looks true until an expansion with a positive or negative remainder.

Matrix monotone always increases.

(Löwner-Heinz Inequality)

$$f(z) = a + bt + \int_0^\infty \frac{ts}{t+s}\, dm(s) = a + bt + \int_0^\infty \left(s - \frac{s^2}{t+s}\right) dm(s)$$

This guy stated around 1930, $f(z) - f(Z) > 0$, in the case of matrix monotone functions and matrices. Where $a \in \mathbb{R}$, $b \geq 0$, and $dm(s)$ is a positive measure, such as the arithmetic, geometric, or harmonic mean.

c) If $A > B$, then $A^\beta > B^\beta$ and $\log(A^\beta) > \log(B^\beta)$, so $\log(A) > \log(B)$

6.6. $Var(Q) = \mathbb{E}[Q^2] - (\mathbb{E}[Q])^2$
$= \mathbb{E}[(Q - \mathbb{E}[Q])^2]$ is positive semidefinite

6.7. $Q = gB$

a) $\mathbb{E}[e^{\lambda Q}] = U^{\top} \circ \mathbb{E}[e^{\lambda g B}] \circ U$

$= U^{\top} \circ \mathbb{E}[e^{1 + \lambda \mathbb{E}[g] B + \frac{\lambda^2 \mathbb{E}[g^2]^2 B^2}{2}}] \circ U$

$\leq U^{\top} \circ \mathbb{E}[e^{\frac{\lambda^2 \mathbb{E}[g^2]^2 B^2}{2}}] \circ U$

$\leq e^{\frac{\lambda^2 \sigma^2 B^2}{2}}$

$\leq e^{\frac{1}{2} V}$ where $V = \lambda^2 \sigma^2 B^2$

b) $|||B||| \leq b$; $Q = gB \leq gb$

$\mathbb{E}[e^{\lambda Q}] \leq U^{\top} \circ \mathbb{E}[e^{\lambda g b}] \circ U$

$\leq U^{\top} \circ \mathbb{E}[e^{1 + \lambda \mathbb{E}[g] B + \frac{\lambda^2 \mathbb{E}[g^2]^2 B^2}{2}}] \circ U$

$\leq U^{\top} \circ \mathbb{E}[e^{\frac{\lambda^2 \mathbb{E}[g^2] b^2}{2}}] \circ U$

$\leq e^{\frac{\lambda^2 \sigma^2 b^2}{2}}$

$\leq e^{\frac{1}{2} V}$ where $V = \lambda^2 \sigma^2 b^2$

(Definition 6.6)

$\Psi_Q(\lambda) \cong e^{\frac{\lambda^2 V}{2}}$ "A random matrix tends Gaussian"

(Chernoff Bound and Technique)

$\mathbb{P}[\gamma_{max}(Q) \geq \delta] \leq \operatorname{trace}(\Psi_Q(\lambda)) e^{-\lambda \delta}$

$\mathbb{P}[|||Q||| \geq \delta] \leq 2 \circ \operatorname{trace}(\Psi_Q(\lambda)) e^{-\lambda \delta}$

(Hoeffding's Bound for random matrices)

$\mathbb{P}[||| \frac{1}{n} \sum_{i=1}^{n} Q_i ||| \geq \delta] \leq 2 \circ \operatorname{rank}(\sum_{i=1}^{n} V_i) e^{\frac{-n\delta^2}{2\sigma^2}}$

$\boxed{\text{One guy studied random matrix eigen values}}$

$\boxed{\text{Another literary studied random matrix rank.}}$

6.8.

a) $\mathbb{E}[\gamma_{max}(S_n)] :=\mathbb{E}\left[e^{\lambda \gamma_{max}(S_n)}\right]$

$$= tr\left(\mathbb{E}\left[e^{\lambda \gamma_{max}(S_n)}\right]\right)$$

$$= tr\left(\Psi_Q\left(\frac{\lambda}{n}\right)\right)$$

$$= tr\left(e^{\sum_{i=1}^{n} \log \Psi_Q\left(\frac{\lambda}{n}\right)}\right)$$

$$\leq tr\left(e^{\frac{\lambda^2 \sum V_i}{2n}}\right)$$

$$\leq d \cdot e^{\frac{\lambda^2 \sigma^2}{2n}}$$

Notes:
$tr(e^{R}) \leq d e^{\|R\|_2}$
where $R = \frac{\lambda^2}{2} \sum V_i$
and $\|R\| = \frac{\lambda^2}{2} n \sigma^2$

$$\mathbb{E}\left[e^{\lambda \gamma_{max}(S_n)}\right] \leq d \cdot e^{\frac{\lambda^2 \sigma^2}{2n}}$$

$$\mathbb{E}[\lambda \gamma_{max}(S_n)] \leq \log d e^{\frac{\lambda^2 \sigma^2}{2n}}$$

$$\mathbb{E}[\gamma_{max}(S_n)] \leq \frac{\log d + \frac{\lambda^2 \sigma^2}{2n}}{\lambda}$$

Wow Whoaaa

The expected function at an eigenvalue is zero.

$$\underset{\lambda}{\operatorname{argmin}} \left\{ \frac{\log d + \frac{\lambda^2 \sigma^2}{2n}}{\lambda} \right\} = 0$$

$$\lambda^* = \sqrt{\frac{2n}{\sigma^2} \log d}$$

$$\mathbb{E}[\gamma^*_{max}(S_n)] \leq \frac{\log d + \frac{\lambda^{*2} \sigma^2}{2n}}{\lambda^*}$$

$$\leq \sqrt{\frac{2\sigma^2}{n} \log d}$$

b) $\mathbb{E}\left[\left|\left|\left| \frac{1}{n} \sum_{i=1}^{n} Q_i \right|\right|\right|\right] = 2 \cdot \mathbb{E}\left[\left\| \frac{1}{n} \sum_{i=1}^{n} Q_i \right\|\right]$

$$:= 2 \cdot \mathbb{E}\left[e^{\lambda \| \frac{1}{n} \sum_{i}^{n} Q_i \|}\right]$$

$$= 2 \cdot \mathbb{E}\left[e^{1 + \lambda \mathbb{E}[\frac{1}{n} \sum Q_i] + \frac{\lambda^2 \mathbb{E}[\frac{1}{n} \sum Q_i^2]^2}{2} + \cdots}\right]$$

$$\leq 2 \cdot \mathrm{tr}\left(\Psi_Q\left(\frac{\lambda}{n}\right)\right)$$

$$\leq 2 \cdot \mathrm{tr}\left(e^{\sum_{i=1}^{n} \log \Psi_Q\left(\frac{\lambda}{n}\right)}\right)$$

$$\leq 2 \cdot \mathrm{tr}\left(e^{\frac{\lambda^2 \sum_i V_i}{2n}}\right)$$

$$\leq 2 \cdot d \cdot e^{\frac{\lambda^2 \sigma^2}{2n}}$$

$$\mathbb{E}\left[\left|\left|\left| \frac{1}{n}\sum_{i=1}^{n} Q_i \right|\right|\right|\right] \leq \frac{\log 2 \cdot d + \frac{\lambda^2 \sigma^2}{2n}}{\lambda}$$ FROM PART (a) OF THE PROBLEM.

$$\arg\min_{\lambda} \left\{ \frac{\log 2d + \frac{\lambda^2 \sigma^2}{2n}}{\lambda} \right\} = 0$$

$$\lambda^* = \sqrt{\frac{2n}{\sigma^2} \log 2d}$$

$$\mathbb{E}\left[\left|\left|\left| \frac{1}{n}\sum_{i=1}^{n} Q_i \right|\right|\right|\right] \leq \sqrt{\frac{2\sigma^2 \log(2d)}{n}}$$

6.9

a) $|||Q|||_2 \leq b$ implies $Q^{j-2} \leq b^{j-2}$ because

$Q^j \leq b^j$; $\rightarrow$ $\dfrac{Q^j}{Q^2} \leq \dfrac{b^j}{Q^2}$ $\rightarrow$ $Q^{j-2} \leq b^j \cdot Q^{-2}$ $\rightarrow$ $Q^{j-2} \leq b^{j-2}$

b)
$$\begin{array}{ccc}
A \cong B & \longrightarrow & A - B \cong 0 \\
\downarrow & & \downarrow \\
 & & Q(A-B) \cong 0 \\
 & & \downarrow \\
 & & Q(A-B)Q \cong 0 \\
\downarrow & & \downarrow \\
QA \cong QB & \longrightarrow & QAQ \cong QBQ
\end{array}$$

c) $\mathbb{E}[Q^j] \leq b^{j-2}$

$$\mathbb{E}[Q \circ Q^{j-2}_{\circ} Q] \leq b^{j-2}\,\mathbb{E}[QQ]$$

$$\leq b^{j-2}\, Var(Q).$$

(Bernsten bound for random matrices)

$$\mathbb{P}\left[\frac{1}{n}\left|\left|\left|\sum_{i=1}^{n} Q_i\right|\right|\right|_2 \geq \delta\right] \leq 2 \circ \operatorname{rank}\left(\sum_{i=1}^{n} var(Q_i)\right) \exp\left\{-\frac{n\delta^2}{2(\sigma^2+b\delta)}\right\}$$

when $\{Q_i\}_{i=1}^{n}$ is independent, zero mean, and symmetric

6.10

a) $A \in \mathbb{R}^{d \times d}$; $Q_i = \begin{bmatrix} 0_{d\times d} & A_i \\ A_i & 0_{d\times d} \end{bmatrix}$

$$Q_i^2 = \begin{bmatrix} A_i^T A_i & 0 \\ 0 & A_i^T A \end{bmatrix}$$

$$|Q_i^2| = \begin{bmatrix} |A_i^T A_i| & 0 \\ 0 & |A_i^T A_i| \end{bmatrix}$$

$$|||Q_i|||_2 = |||A_i|||_2$$

b) $\left|\left|\left|\frac{1}{n}\sum_{i=1}^{n} var(Q_i)\right|\right|\right|_2 = \left|\left|\left|\frac{1}{n}\sum_{i=1}^{n}\left[Q_i - \mathbb{E}[Q_i]\right]^2\right|\right|\right|$ but mean is zero

$$= \left|\left|\left|\frac{1}{n}\sum_{i=1}^{n} Q_i^2\right|\right|\right|$$

$$= \left|\left|\left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[A^T A]\right|\right|\right|$$

$$\leq \sigma^2$$

c) $\mathbb{P}\left[\left|\left|\left|\sum_{i=1}^{n} A_i\right|\right|\right| \geq n\delta\right] = 2 \circ \mathbb{P}\left[\left|\left|\sum_{i=1}^{n} A_i\right|\right| \geq n\delta\right]$

$$\overset{\circ}{=} 2 \circ \mathbb{P}\left[e^{\lambda \|\sum A_i\|} \geq e^{\lambda n \delta}\right]$$

$$= 2 \circ \mathbb{P}\left[e^{1 + \lambda \mathbb{E}[\|\sum A_i\|] + \frac{\lambda^2 \mathbb{E}[\|\sum A_i\|^2]^2}{2}} \geq e^{\lambda n \delta}\right]$$

$$\le 2\circ \mathbb{P}\left[e^{\frac{\lambda^2 \mathbb{E}[Q^2]^2}{2(1-b|\lambda|)}} \ge e^{-\lambda n\delta} \right] \quad \text{where } |\lambda| \in \frac{1}{b}$$

$$\le 2\circ \operatorname{rank}\left(\sum_i \operatorname{Var}(Q)\right)\circ \operatorname{tr}\left(e^{\frac{\lambda^2\sigma^2 \circ n}{2(1-b|\lambda|)}}\right)\circ e^{-\lambda n\delta}$$

$$\operatorname*{argmin}_\lambda \left\{ \frac{\lambda^2\sigma^2 n}{2(1-b|\lambda|)} - \delta n\lambda \right\} = 0$$

$$\lambda^* = \frac{\delta}{\sigma^2 + b\delta}$$

$$\mathbb{P}\left[\left|\left|\left| \sum_{i=1}^n A_i \right|\right|\right| \ge n\delta \right] \le 2\circ(d_1+d_2)\, e^{\frac{-n\delta^2}{2(\sigma^2+b\delta)}}$$

6.11

a) $\{A_i\}_{i=1}^n \in \mathbb{R}^{d\times d}$; $A_i = g_i B_0$; $\mathbb{E}[g_i^j] \le \frac{j!}{2} b_1^{j-2}\sigma^2$

$|||B_i|||_2 \le b_2$

Early, $\mathbb{E}[A_i] = \mathbb{E}[g_i B_i]$

$$= \mathbb{E}[g_i]\,\mathbb{E}[B_i]$$

$$= \frac{1}{2}\,\frac{\sigma^2}{b_1}$$

$$\mathbb{E}[A_i^2] = \mathbb{E}[g_i^2 B_i^2]$$

$$= \mathbb{E}[g_i^2]\,\mathbb{E}[B_i^2]$$

$$= b_1\sigma^2 b_2$$

$$\mathbb{P}\left[\left|\left|\left| \frac{1}{n}\sum_{i=1}^n A_i \right|\right|\right| \ge \delta \right] = 2\circ \mathbb{P}\left[\left\| \frac{1}{n}\sum_{i=1}^n A_i \right\| \ge \delta \right]$$

$$\overset{\circ}{=} 2\circ \mathbb{P}\left[e^{\lambda \|\sum A_i\|} \ge e^{\lambda n\delta} \right]$$

$$= 2\circ \mathbb{P}\left[e^{1+\lambda \mathbb{E}[\|\sum A_i\|] + \frac{\lambda^2 \mathbb{E}[\|\sum A_i\|^2]^2}{2}} \ge e^{\lambda n\delta} \right]$$

$$\le 2\circ \mathbb{P}\left[e^{\frac{\lambda^2 \mathbb{E}[\|\sum A_i\|^2]^2}{2}} \ge e^{\lambda n\delta} \right]$$

$$\le 2\circ \mathbb{P}\left[e^{\frac{\lambda^2 b_2^2\sigma^2 n}{2(1-b_1 b_2|\lambda|)}} \ge e^{\lambda n\delta} \right] \quad \text{for all } |\lambda| = \frac{1}{b}$$

$$\leq 2 \circ \mathrm{rank}\left(\|\sum A_i^2\|\right) e^{\frac{\lambda^2 b_2^2 \sigma^2 n}{2(1-b_1 b_2|\lambda|)} - \lambda n \delta}$$

$$\operatorname*{argmin}_{\lambda} \left\{ \frac{\lambda^2 b_2^2 \sigma^2 n}{2(1-b_1 b_2|\lambda|)} - \lambda n \delta \right\} = 0$$

$$\lambda^* = \frac{\delta}{b_2^2 \sigma^2 - b_1 b_2 \delta}$$

$$\mathbb{P}\left[\left|\left|\left| \frac{1}{n}\sum_{i=1}^{n} A_i \right|\right|\right| \geq \delta\right] \leq 2 \circ (d_1 + d_2) e^{\frac{-n\delta^2}{2(\sigma^2 b_2^2 + b_1 b_2 \delta)}}$$

b)

$$\mathbb{E}\left[\left|\left|\left| \frac{1}{n}\sum_{i=1}^{n} A_i \right|\right|\right|_2\right] = 2 \circ \mathbb{E}\left[\left\| \frac{1}{n}\sum_{i=1}^{n} A_i \right\|\right]$$

$$= 2 \circ \int P\left[\left\| \frac{1}{n}\sum_{i=1}^{n} A_i \right\|\right]$$

$$= 2 \circ \int e^{\frac{-n\delta^2}{2(\sigma^2 b_2^2 + b_1 b_2 \delta)}} d\delta \qquad \boxed{C = d_1 + d_2}$$

$$\leq 2\left[C_1 \int_1 e^{\frac{-n\delta^2}{2\sigma^2 b_2^2}} d\delta + C \int_2 e^{\frac{-n\delta^2}{2 b_1 b_2 \delta}} d\delta \right]$$

Integral #1:

$$C \int e^{\frac{-n\delta^2}{2\sigma^2 b_2^2}} d\delta = \int_0^{\sqrt{\sigma^2 b_2^2 \log C/n}} d\sigma + \int_{\sqrt{\sigma^2 b_2^2 \log C/n}}^{\infty} e^{\frac{-n\delta^2}{2\sigma^2 b_2^2}} d\delta$$

$$= \sigma b_2 \sqrt{\frac{\log C}{n}} + \int_{\sqrt{\sigma^2 b_2^2 \log C/n}}^{\infty} e^{\frac{-n u^2}{2}} \circ \sigma^2 \circ b_2^2 \, du \qquad \boxed{\frac{\delta^2}{\sigma^2 b_2^2} = u^2}$$

$$= \sigma b_2 \sqrt{\frac{\log C}{n}} + \sqrt{\pi}\left(\left|\frac{\sigma b_2}{\sqrt{n}}\right| - \frac{2\sigma b_2}{\sqrt{n}} \operatorname{erf}\left(\sqrt{\log C}\right)\right)$$

$$= \frac{\sigma b_2}{\sqrt{n}} \left\{ \sqrt{\log(d_1 + d_2)} + \sqrt{\pi} \right\}$$

Integral #2:

$$c\int e^{\frac{-n\delta^2}{2b_1b_2\delta}}d\delta = \int_0^{\frac{2b_1b_2}{n}\log c} d\delta + \int_{\frac{2b_1b_2\log c}{n}}^{\infty} e^{\frac{-n\delta}{2b_1b_2}} d\delta$$

$$= \frac{2b_1b_2\log c}{n} + \frac{2b_1b_2}{n}\left[\log c + \frac{1}{c}\right]$$

$$\le \frac{2b_1b_2}{n}\{\log(d_1+d_2) + 1\}$$

An integral combination:

$$\mathbb{E}\left[\left|\left|\left|\frac{1}{n}\sum_{i=1}^{n}A_i\right|\right|\right|_2\right] \le \frac{2\sigma b_2}{\sqrt{n}}\{\sqrt{\log(d_1+d_2)} + \sqrt{\pi}\} + \frac{4b_1b_2}{n}\{\log(d_1+d_2) + 1\}$$

6.12.

a) $\mathbb{P}[\gamma_{max}(S) \ge \delta] := \mathbb{P}[e^{\gamma_{max}(\lambda S)} \ge e^{\lambda\delta}]$

$$= \mathbb{P}[\gamma_{max}(e^{\lambda S}) \ge e^{\lambda\delta}]$$

$$= \mathbb{E}[\gamma_{max}(e^{\lambda S})]e^{-\lambda\delta}$$

$$\le \mathbb{E}[tr(e^{\lambda S})]e^{-\lambda\delta}$$

$$\le tr(\mathbb{E}[e^{\lambda S}])e^{-\lambda\delta}$$

$$\le \frac{tr(\mathbb{E}[\phi(\lambda S)])}{\phi(\lambda\delta)} \quad \text{where } \phi(x) = e^x$$

b) $\log \Psi_Q(\lambda) = \log e^{1+\lambda\mathbb{E}[\Psi_Q(\lambda)] + \frac{\lambda^2\mathbb{E}[\Psi_Q(\lambda)^2]^2}{2}}$

The mean is zero and $e^\lambda = 1+\lambda+\sum_{n=2}^{\infty}\frac{\lambda^n}{n!}$

$$e^\lambda - 1 - \lambda = \sum_{n=2}^{\infty}\frac{\lambda^n}{n!}$$

$$\log \mathcal{Z}_Q(\lambda) = \log e^{1+\sum_{n=2}^{\infty}\frac{\lambda^n}{n!}\mathbb{E}[\mathcal{Z}(\lambda)^n]^2}$$

$$\leq \log e^{\phi(\lambda)\,var(Q)} \qquad \text{Where } \phi(\lambda) = e^{\lambda}-\lambda-1$$

$$\leq \phi(\lambda)\,var(Q)$$

c) $$tr(\mathbb{E}[\phi(\lambda S_n)]) = tr(\mathbb{E}[e^{\lambda S}-1])$$

$$\leq tr(\exp\{\log \mathcal{Z}_Q(\lambda)\}-1)$$

$$\leq tr(\exp\{\log \phi(\lambda)\,Var(V(Q))\}-1)$$

$$\leq \frac{tr(V)}{\|V\|_2}\, e^{\phi(\lambda)|||V|||_2}$$

6.13

a) $Q = \begin{bmatrix} 0 & X_i^T \\ X_i & 0_d \end{bmatrix}$; $V_n = \sum_{i=1}^{n} var(Q)$

Bernstein's inequality for random matrices defines rank as a coefficient for sub-Gaussian description.

$$rank(V_n) = rank\left(\sum_{i=1}^{n} var(Q)\right) = (d+1) \quad \text{For a } d\times 1 \text{ matrix.}$$

b) The prefactor "2" appears from two-sided inequalities. Equation (6.48) has no absolute symbols in the probability, nor a "2"-prefactor in the Sub-Gaussian inequality.

6.14

a) Sphere packing history began in Virginia in 1588 about transport and goods perfourmed thrice a yeere in Middle English. Around 1611, Keplar examined topics with spheres inside cubes. Later Gauß in 1831 defined face-centered cubic structures. Clicks near '58, '61 and '66 about rigid points and hemispheres. By year 1993, Conway and Sloan modelled 3-D lattice packing units. Maryna Viazovska in 2016 studied lower and upper metrics for d-dimensional volumes $V_n = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$.

With her formula, a packing number at 1/2 requires one sphere in one unit cube.

b) Cardinality above an equipolar constant. Cantor-Bernstein's theorem (1887) judged metrics in disjoint fractals with a unique conclusion, a base-cardinality in real systems. $\mathbb{R}^d = e^{X}$.

c) $$|||\frac{1}{M}\sum_{j=1}^{M}\theta^j \otimes \theta^j|||_2 = 2 \cdot \|\frac{1}{d}\sum_{j=1}^{M}\theta^j \otimes \theta^j\||_2$$

$$= 2 \cdot \|\frac{1}{d}(\sum_{j=1}^{M}\theta^j \otimes \theta^j - 1 \geq + D\|_2$$

$$\therefore = \frac{2}{\sqrt{d}}\| e^{\lambda(\sum_{j=1}^{m}\theta^j \otimes \theta^j - 1)} \geq e^{\lambda}\|_2$$

$$= \frac{2}{\sqrt{d}}\| e^{1+\lambda\mathbb{E}[\sum_{j=1}^{m}\theta^j \otimes \theta^j] + \frac{\lambda^2\mathbb{E}[\sum_{j=1}^{m}\theta^j \otimes \theta^{j\,2}]^2}{2}} \geq e^{\lambda}\|$$

$$\leq \frac{2}{\sqrt{d}}\| e^{\frac{\lambda^2\mathbb{E}[\sum_{j=1}^{m}\theta^j \otimes \theta^{j\,2}]^2}{2} - \lambda} \geq e^{\lambda}\|$$

$$\arg\min_{\lambda}\{\lambda^2\mathbb{E}[\sum_{j=1}^{M}\theta^j \otimes \theta^{j\,2}]^2 - 2\lambda\} = 0$$

$$\lambda^* = \frac{4}{\mathbb{E}[\sum_{j=1}^{m} \theta^j \otimes \theta^j{}^2]}$$

$$|||\frac{1}{M}\sum_{j=1}^{M} \theta^j \otimes \theta^j|||_2 \le \frac{2}{\sqrt{d}} e^{\frac{\lambda^{*2}\mathbb{E}[\sum_{j=1}^{m}\theta^j\otimes\theta^j{}^2]^2}{2} - 2\lambda^*}$$

$$\le \frac{2}{\sqrt{d}} e^0$$

$$\le \frac{2}{d}$$

(Thresholding-based Covariance Estimation) "The Trace Threshold"

When $\{X\}_{i=1}^n = N(0,\Sigma)$ and $n > \log d$ for any $\delta > 0$

$$\mathbb{P}[|||T(\hat{\Sigma}) - \Sigma||| \ge 2|||A|||\lambda] \le 8e^{\frac{-n}{16}\min(\delta,\delta^2)}$$

where $\frac{\lambda}{\sigma^2} = 8\sqrt{\frac{\log d}{n}} + \delta$

6.15

a) $\{X_i\}_{i=1}^n$ be an i.i.d sequence $= N(0,\Sigma)$

when $\hat{D} = \text{diag}(\hat{\Sigma})$ where $\hat{\Sigma}$ is a sample covariance

$$\mathbb{P}\left[|||\hat{D}-D|||_2/\sigma^2 \ge c_0\sqrt{\frac{\log d}{n}} + \delta\right] = \mathbb{P}[||\hat{D}-D||/\sigma^2 \ge 2|||A|||\lambda_n]$$

in that $|||A|||$ = sparsity metric

$$2\lambda_n = c_0\sqrt{\frac{\log d}{n}} + \delta$$

$$\mathbb{P}\left[\frac{|||\hat{D}-D|||_2}{\sigma^2} \ge 2|||A|||\lambda_n\right] \le \mathbb{P}[|||\hat{D}-D|||/\sigma^2 \ge t]$$

where $t = 2|||A|||\lambda_n$

$$\le \mathbb{P}\left[e^{1+\lambda\mathbb{E}[|||\hat{D}-D|||] + \frac{\lambda^2\mathbb{E}[|||\hat{D}-D|||^2]^2}{2}} \ge e^{\lambda t}\right]$$

$$\le \mathbb{P}\left[e^{\frac{\lambda^2\mathbb{E}[|||\hat{D}-D|||^2]^2}{2}} \ge e^{\lambda t}\right]$$

$$\arg\min_\lambda \left\{\frac{\lambda^2\mathbb{E}[|||\hat{D}-D|||^2]^2}{2} - \lambda t\right\} = 0$$

$$\lambda^* = t/\mathbb{E}[|||\hat{D}-D|||]^2$$

$$\mathbb{P}\left[|||\hat{D}-D|||/\sigma^2 \geq c_0\sqrt{\frac{\log d}{n}}+\delta\right] \leq e^{-t^2/2\cdot\mathbb{E}[|||\hat{D}-D|||]^2}$$

$$\leq e^{-\frac{1}{2}\left(\sqrt{\frac{n}{s\log d}}\right)^2\left(c_0\sqrt{\frac{\log d}{n}}+\delta\right)^2}$$

$$\leq e^{-\frac{1}{2}\frac{n}{s\log d}\left[c_0^2\frac{\log d}{n}+2c_0\sqrt{\frac{\log d}{n}}\delta+\delta^2\right]}$$

$$\leq e^{-\frac{1}{2}[c_0^2/s]}\cdot e^{-\frac{n}{2s\log d}\left[2c_0\sqrt{\frac{\log d}{n}}\delta+\delta^2\right]}$$

$$\leq c_1\cdot e^{-c_2\cdot n\cdot\min(\delta,\delta^2)}$$

where $c_1 = e^{-\frac{1}{2}c_0^2/s}$

$$c_2 = \frac{1}{s\log d}$$

b) When $\underbrace{\mathbb{E}[(X_{ij}^2-\Sigma_{ij})^m]}_{\|X_{ij}^2-\Sigma_{jj}\|_m^m} \leq K_m$

$$\mathbb{P}\left[|||\hat{D}-D|||_2 \geq 4\delta\sqrt{\frac{d^{2/m}}{n}}\right] \leq \frac{1}{(2\delta)^m}\cdot\frac{m^{m/2}}{2^m}\|\hat{D}-D\|_m^m$$

$$\text{where } \|\hat{D}-D\| \leq \frac{C_m^m}{n^m}\left(\left[\Sigma\right]\mathbb{E}[(X_{ii}^2-D_{ii})^2]\right)^{1/2}$$

$$\mathbb{P}\left[|||\hat{D}-D|||_2 \geq 4\delta\sqrt{\frac{d^{2/m}}{n}}\right] \leq \frac{1}{(2\delta)^m}\cdot K_m'$$

$$\text{When } K_m' = \frac{m^{m/2}}{2^m}\cdot K = \frac{m^{m/2}}{2^m}\left(\left[\Sigma\right]\mathbb{E}[(X^2-D)^2]\right)^2$$

6.16. <u>Proof By Induction</u>:

Base case: $A = \sqrt{5}$

Next case: $|||A|||_2 = 5$

Induction Step: $|||A|||_m = 5^{m/2}$

Good thoughts from Chapter 6:

① A random matrix at larger and larger size has an upper bound similar with a Gaussian.

② Lucky number, 8 appears in packing problems, estimation of diagonal covariances and matrix rank.

③ The Holy Grail was the "Trace Threshold." A trace is an upper limit in a random matrix. Traces lower computational steps, as a diagonal analysis, but their performance is not as accurate as a covariance method. In a case when $n > \log d$ where n, d, and a trace are number of Gaussians, matrix dimensions and the operation, respectively. Per se a trace is a good guess.

④ A random matrix has a baseline amplitude or average amplitude dependent on the dimension.