# Random matrices and covariance estimation

Junyi FAN

23/03/2022

# Set up of covariance estimation

Let $x_1, \cdots, x_n$ be collection of $n$ independent and identically distributed samples from a distribution in $\mathbb{R}^d$ with zero mean and covariance matrix $\Sigma$. Generally we use *sample covariance matrix* $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ as a estimator.

- we want to bound the error $\|\hat{\Sigma} - \Sigma\|_2$.
- Let $X = [x_1, \cdots, x_n]^T$ with singular values denoted by $\{\sigma_j(X)\}_{j=1}^{\min(n,d)}$. Using $\hat{\Sigma} = \frac{1}{n}X^TX$ we know the eigenvalues of $\hat{\Sigma}$ is the squares of the singular values of $X/\sqrt{n}$

# Wishart matrices and their behavior

Each $x_i \sim^{i.i.d} \mathcal{N}(0, \Sigma)$, then associated matrix $X$ is called drawn from $\Sigma$-*Gaussian ensemble* and the sample covariance $\hat{\Sigma}$ is said to follow multivariate Wishart distribution

## Theorem (6.1)

*Let $X \in \mathbb{R}^{nxd}$ be drawn from $\Sigma$-Gaussian ensemble. Then for all $\delta > 0$, the maximum singular value $\sigma_{max}(X)$ satisfies following inequality*

$$\mathbb{P}[\frac{\sigma_{max}(X)}{\sqrt{n}} \geq \gamma_{max}(\sqrt{\Sigma})(1+\delta) + \sqrt{\frac{tr(\Sigma)}{n}}] \leq e^{-n\delta^2/2}$$

*and for $n \geq d$, we also have*

$$\mathbb{P}[\frac{\sigma_{min}(X)}{\sqrt{n}} \geq \gamma_{min}(\sqrt{\Sigma})(1-\delta) - \sqrt{\frac{tr(\Sigma)}{n}}] \leq e^{-n\delta^2/2}$$

# Proof of Theorem 6.1

Let $X = W\sqrt{\Sigma}$, where $W_{ij} \sim \mathcal{N}(0,1)$. And we consider function $f(W) = \frac{\sigma_{max}(W\sqrt{\Sigma})}{\sqrt{n}}$.

- $f$ is Lipschitz: $\|f(W) - f(W')\|_2 \leq \frac{\gamma_{max}(\sqrt{\Sigma})}{\sqrt{n}}\|W - W'\|_2$
- Using Concentration of Lipschitz function: $\mathbb{P}[\sigma_{max}(X) \geq \mathbb{E}[\sigma_{max}(X)] + \sqrt{n}\gamma_{max}(\sqrt{\Sigma})\delta] \leq e^{-n\delta^2/2}$
- Then it suffices to show that $\mathbb{E}[\sigma_{max}(X)] \leq \sqrt{n}\gamma_{max}(\sqrt{\Sigma}) + \sqrt{tr(\Sigma)}$

# Proof of Theorem 6.1

**Theorem (Sudakov-Fernique)**

*Given a pair of zero-mean N-dimension Gaussian vectors* $(X_1, \cdots, X_N)$ *and* $(Y_1, \cdots, Y_N)$*, suppose that*

$$\mathbb{E}[(X_i - X_j)^2] \leq \mathbb{E}[(Y_i - Y_j)^2] \quad \forall (i,j)$$

*Then* $\mathbb{E}[\max_{j=1,\cdots,N} X_j] \leq \mathbb{E}[\max_{j=1,\cdots,N} Y_j]$

- $\sigma_{max}(X) = \max\limits_{v' \in \mathbb{S}^{d-1}} \|Xv'\|_2 = \max\limits_{v \in \mathbb{S}^{d-1}(\Sigma^{-1})} \|Wv\|_2 = \max\limits_{u \in \mathbb{S}^{d-1}} \max\limits_{v' \in \mathbb{S}^{d-1}(\Sigma^{-1})} u^T W v$
- Consider Gaussian process $Z_{u,v} = u^T W v$, given $(u, v) and (\tilde{u}, \tilde{v})$,w.l.o.g we assume $\|v\|_2 \leq \|\tilde{v}\|_2$,then we have $\mathbb{E}[(Z_{u,v} - Z_{\tilde{u},\tilde{v}})^2] = \|uv^T - \tilde{u}\tilde{v}^T\|_F^2$

# Proof of Theorem 6.1

$$
\begin{aligned}
\|uv^T - \tilde{u}\tilde{v}^T\|_F^2 &= \|u(v-\tilde{v})^T + (u-\tilde{u})\tilde{v}^T\|_F^2 \\
&= \|u(v-\tilde{v})^T\|_F^2 + \|(u-\tilde{u})\tilde{v}^T\|_F^2 + 2 < u(v-\tilde{v})^T, (u-\tilde{u})\tilde{v} \\
&\leq \|\tilde{v}\|^2 \|(u-\tilde{u})\|^2 + \|(v-\tilde{v})\|^2 \\
&\leq \gamma_{max}(\sqrt{\Sigma})^2 \|(u-\tilde{u})\|^2 + \|(v-\tilde{v})\|^2
\end{aligned}
$$

- Define another Gaussian process $Y_{u,v} = \gamma_{max}(\sqrt{\Sigma}) < g, u > + < h, v >$ with $g, v$ have entries from $\mathcal{N}(0,1)$.
- $\mathbb{E}[(Y_{u,v} - Y_{\tilde{u},\tilde{v}})^2] = \gamma_{max}(\sqrt{\Sigma})^2 \|(u-\tilde{u})\|^2 + \|(v-\tilde{v})\|^2$
- Using Sudakov-Fernique theorem, $\mathbb{E}[\sigma_{max}(X)] \leq \mathbb{E}[\sup_{(u,v)} Y] = \gamma_{max}(\sqrt{\Sigma})\mathbb{E}[\|g\|_2] + \mathbb{E}[\|\sqrt{\Sigma}h\|_2]$. With that $\mathbb{E}[\|g\|_2] \leq \sqrt{n}$ and $\mathbb{E}[\|\sqrt{\Sigma}h\|_2] \leq \sqrt{\mathbb{E}[h^T\Sigma h]} \leq \sqrt{tr(\Sigma)}$.

# Some Example

- $\Sigma = I_d$ and $n \geq d$. We have $\frac{\sigma_{max(W)}}{\sqrt{n}} \leq 1 + \delta + \sqrt{\frac{d}{n}}$ and $\frac{\sigma_{min(W)}}{\sqrt{n}} \geq 1 - \delta - \sqrt{\frac{d}{n}}$. Implying that $\|\frac{1}{n}W^TW - I_d\|_2 \leq 2\epsilon + \epsilon^2$ where $\epsilon = \delta + \sqrt{\frac{d}{n}}$
- For general $\Sigma$, Using $X = W\sqrt{\sigma}$, we know $\|\frac{1}{n}X^TX - I_d\|_2 = \|\sqrt{\Sigma}(\frac{1}{n}W^TW - I_d)\sqrt{\Sigma}\|_2 \leq \|\Sigma\|_2\|\frac{1}{n}X^TX - I_d\|_2$. Then we have

$$\frac{\|\hat{\Sigma} - \Sigma\|_2}{\|\Sigma\|_2} \leq 2\sqrt{\frac{d}{n}} + 2\delta + (\sqrt{\frac{d}{n}} + \delta)^2$$

  w.p at least $1 - 2e^{-n\delta^2/2}$.
- *trace constraint*: $\frac{tr(\Sigma)}{\|\Sigma\|_2} \leq C$, we have $\frac{\sigma_{max(W)}}{\sqrt{n}} \leq 1 + \delta + \sqrt{\frac{C}{n}}$. The parameter $C$ is more like effective dimension

# Covariance matrices from sub-Gaussian ensembles

- *Definition*: A random vector $x_i \in \mathbb{R}^d$ is zero-mean, sub-Gaussian with parameter $\sigma$, i.e. for each fixed $v \in \mathbb{S}^{d-1}$, $\mathbb{E}[e^{\lambda < v, x_i >}] \leq e^{\frac{\lambda^2 \sigma^2}{2}}, \forall \lambda \in \mathbb{R}$
- Example 1: Matrix $X \in \mathbb{R}^{n \times d}$ with i.i.d zero-mean and sub-Gaussian $(\sigma = 1)$ entries. Like Gaussian $(x_{ij} \sim \mathcal{N}(0,1))$ or Rademacher ensemble $(x_{ij} \in -1, +1)$.
- Example 2: $x_i \sim \mathcal{N}(0, \Sigma)$. For any $v \in \mathbb{S}^{d-1}$, we have $< v, x_i > \sim \mathcal{N}(0, v^T \Sigma v)$ and $v^T \Sigma v \leq \|\Sigma\|_2$. $x_i$ is sub-Gaussian with parameter $\|\Sigma\|_2$
- If random matrix $X \in \mathbb{R}^{n \times d}$ with row $x_i \in \mathbb{R}^d$ follows $\sigma$-sub-Gaussian. We say $X$ is *row-wise $\sigma$-sub-Gaussian ensemble*.

# Sub-Gaussian ensemble covariance matrix estimation

## Theorem (6.5)

*There are universal constants* $c_{j}{}_{j=0}^{3}$ *such that, for any row-wise* $\sigma$*-sub-Gaussian ensemble* $X \in \mathbb{R}^{n\times d}$*, the sample covariance* $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ *satisfies*

$$\mathbb{E}[e^{\lambda\|\hat{\Sigma}-\Sigma\|_2}] \leq e^{c_0 \frac{\lambda^2\sigma^4}{n} + 4d} \quad \forall|\lambda| < \frac{n}{64e^2\sigma^2}.$$

*and hence* $\mathbb{P}[\frac{\|\hat{\Sigma}-\Sigma\|_2}{\sigma^2} \geq c_1(\sqrt{\frac{d}{n}} + \frac{d}{n}) + \delta] \leq c_2 e^{-c_3 n \min(\delta,\delta^2)} \quad \forall\delta > 0.$

# Proof of Theorem 6.5

- (Discretization) Let $Q = \hat{\Sigma} - \Sigma$ and $\|Q\|_2 = \max_{v \in \mathbb{S}^{d-1}} |< v, Qv >|$. Let $\{v^1, \cdots, v^N\}$ be $\frac{1}{8}$-covering of $\mathbb{S}^{d-1}$, where $N \leq 17^d$. Then $\forall v \in \mathbb{S}^{d-1}, v = v^j + \Delta$ and $\|\Delta\|_2 \leq \frac{1}{8}$.
- From $< v, Qv >=< v^j, Qv^j > +2 < \Delta, Qv^j > + < \Delta, Q\Delta >$, we have

$$
\begin{aligned}
|< v, Qv >| &\leq \left|< v^j, Qv^j >\right| + 2\|\Delta\|_2 \|Q\|_2 \|v^j\|_2 + \|Q\|_2 \|\Delta\|^2 \\
&\leq \left|< v^j, Qv^j >\right| + (\frac{1}{4} + \frac{1}{64}) \|Q\|_2 \\
&\leq \left|< v^j, Qv^j >\right| + (\frac{1}{2}) \|Q\|_2
\end{aligned}
$$

Hence, $\|Q\|_2 = \max_{v \in \mathbb{S}^{d-1}} |< v, Qv >| \leq 2 \max_{j=1,\cdots,N} \left|< v^j, Qv^j >\right|$

# Proof of Theorem 6.5

- $\mathbb{E}[e^{\lambda\|Q\|_2}] \leq \mathbb{E}[\exp\{2\lambda \max_{j=1,\cdots,N} |< v^j, Qv^j >|\}] \leq \sum_{j=1}^{N}(\mathbb{E}[e^{2\lambda< v^j, Qv^j>}] + \mathbb{E}[e^{-2\lambda< v^j, Qv^j>}])$
- We just need to establish bound for fixed $u \in \mathbb{S}^{d-1}$: $\mathbb{E}[e^{t<u,Qu>}] \leq e^{512\frac{t^2}{n}e^4\sigma^4} \quad \forall |t| \leq \frac{n}{32e^2\sigma^2}$.
- Let $t = 2\lambda$ and $t = -2\lambda$. We have: $\mathbb{E}[e^{\lambda\|Q\|_2}] \leq 2Ne^{2048\frac{\lambda^2}{n}e^4\sigma^4} \leq e^{c_0\frac{\lambda^2\sigma^4}{n}+4d} \quad \forall |\lambda| \leq \frac{n}{64e^2\sigma^2}$.
- Follow Ex1,$\Sigma = I_d$ and $n \geq d$, we have $1 - c'\sqrt{\frac{d}{n}} \leq \frac{\sigma_{min}(X)}{\sqrt{n}} \leq \frac{\sigma_{max}(X)}{\sqrt{n}} \leq 1 + c'\sqrt{\frac{d}{n}}$, where $c' > 1$.

# Bound for general matrices

To give similar bound for general matrices, we first define $\Psi_Q(\lambda) = \mathbb{E}[e^{\lambda Q}]$

- *sub-Gaussian*: A zero-mean symmetric random matrix $Q \in \mathbb{S}^{d\times d}$ is sub-Gaussian with matrix parameter $V \in \mathbb{S}_+^{d\times d}$, i.e.

$$\Psi_Q(\lambda) \leq e^{\frac{\lambda^2 V}{2}} \quad \forall \lambda \in \mathbb{R}$$

- Example: $Q = \epsilon B$ with $\epsilon \in -1, +1$ a Rademacher variable and $B$ a fixed matrix., $Q$ is sub-Gaussian with $V = B^2 = var(Q)$.
- Example: $Q = \epsilon C$ with $\epsilon \in -1, +1$ a Rademacher variable and $C$ a random matrix. $C$ is independent of $\epsilon$ and $\|C\|_2 \leq b$. Then $Q$ is sub-Gaussian with $V = b^2 I_d$

# Bound for general matrices

- *sub-exponential*: A zero-mean symmetric random matrix $Q \in \mathbb{S}^{d\times d}$ is sub-exponential with matrix parameter $(V, \alpha)$, i.e.

$$\Psi_Q(\lambda) \leq e^{\frac{\lambda^2 V}{2}} \quad \forall |\lambda| < \frac{1}{\alpha}$$

- *Bernstein's condition for matrices*: A zero-mean symmetric random matrix $Q$ satisfies a Bernstein condition with parameter $b > 0$, i.e.

$$\mathbb{E}[Q^j] \leq \frac{1}{2} j! b^{j-2} var(Q) \quad j = 3, 4, \cdots$$

**Lemma (Bernstein in matrix)**

*For any symmetric zero-mean random matrix satisfying the Bernstein condition, we have*

$$\Psi_Q(\lambda) \leq \exp\left(\frac{\lambda^2 var(Q)}{2(1 - b\|\lambda\|)}\right) \quad \forall |\lambda| < \frac{1}{b}.$$

# Proof

$$
\begin{aligned}
\mathbb{E}[e^{\lambda Q}] &= I_d + \frac{\lambda^2 var(Q)}{2} + \sum_{j=3} \infty \frac{\lambda^j \mathbb{E}[Q^j]}{j!} \\
&\leq I_d + \frac{\lambda^2 var(Q)}{2} (\sum_{j=0}^{\infty} ||\lambda|^j b^j) \\
&= I_d + \frac{\lambda^2 var(Q)}{2(1 - b\|\lambda\|)} \\
&\leq \exp(\frac{\lambda^2 var(Q)}{2(1 - b\|\lambda\|)}).
\end{aligned}
$$

# Matrix Chernoff approach

## Lemma (Matrix chernoff technique)

*Let $Q$ be a zero-mean symmetric random matrix with $\Psi_Q$ exists in an open interval $(-a, a)$. Then for any $\delta > 0$, we have*

$$\mathbb{P}[\gamma_{max}(Q) \geq \delta] \leq tr(\Psi_Q(\lambda))e^{-\lambda\delta} \quad \forall \lambda \in [0, a).$$

*Similarly,*

$$\mathbb{P}[\|Q\|_2 \geq \delta] \leq 2tr(\Psi_Q(\lambda))e^{-\lambda\delta} \quad \forall \lambda \in [0, a).$$

Proof:

- $\mathbb{P}[\gamma_{max}(Q) \geq \delta] = \mathbb{P}[e^{\gamma_{max}(\gamma Q)} \geq e^{\lambda\delta}] = \mathbb{P}[\gamma_{max}(e^{\lambda Q}) \geq e^{\lambda\delta}]$
- Using Markov inequality:
  $\mathbb{P}[\gamma_{max}(e^{\lambda Q}) \geq e^{\lambda\delta}] \leq \mathbb{E}[\gamma_{max}(e^{\lambda Q})]e^{\lambda\delta} \leq \mathbb{E}[tr(e^{\lambda Q})]e^{\lambda\delta}$
- Trace and expectation commute:
  $\mathbb{E}[tr(e^{\lambda Q})] = tr(\mathbb{E}[e^{\lambda Q}]) = tr(\Psi_Q(\lambda))$
- For $\|Q\|_2$, bound $\gamma_{max}(-Q) \geq \delta$ similarly and use
  $\|Q\|_2 = max\{\gamma_{max}(Q), |\gamma_{min}(Q)|\}$

# Summation of random matrices

**Lemma (Summation of random matrices)**

*Let $Q_1, \cdots, Q_n$ be independent zero-mean symmetric random matrix with $\Psi_{Q_i}$ exists for all $\lambda \in I$. Then for $S_n = \sum_{i=1}^n Q_i$, we have*

$$tr(\Psi_{S_n}(\lambda)) \leq tr(e^{\sum_{i=1}^n \log \Psi_{Q_i}(\lambda)}) \quad \forall \lambda \in I.$$

- From this result, we have,

$$\mathbb{P}[\|\frac{1}{n}\sum_{i=1}^n Q_i\|_2 \geq \delta] \leq 2tr(e^{\sum_{i=1}^n \log \Psi_{Q_i}(\lambda)})e^{-n\delta}.$$

- Proof: We need such result from Lieb (1973). For any fixed matrix $H \in \mathbb{S}^{d\times d}$, function $f(A) = tr(e^{H+\log(A)})$ is concave.

# Proof of Lemma

Denote $G(\lambda) = tr(\Psi_{S_n}(\lambda))$

$$
\begin{aligned}
G(\lambda) &= tr(\Psi_{S_n}(\lambda)) \\
&= tr(\mathbb{E}[e^{\lambda S_{n-1} + \log \exp(\lambda Q_n)}]) \\
&= \mathbb{E}_{S_{n-1}} \mathbb{E}_{Q_n} tr(e^{\lambda S_{n-1} + \log \exp(\lambda Q_n)}) \\
&\leq \mathbb{E}_{S_{n-1}} tr(e^{\lambda S_{n-1} + \log \mathbb{E}_{Q_n} \exp(\lambda Q_n)}) \\
&\leq \mathbb{E}_{S_{n-2}} \mathbb{E}_{Q_{n-1}} tr(e^{\lambda S_{n-1} + \log \exp(\lambda Q_{n-1}) + \Psi_{Q_n}(\lambda)}) \\
&\leq \cdots \\
&\leq tr(e^{\sum_{i=1}^{n} \log \Psi_{Q_i}(\lambda)})
\end{aligned}
$$

# Upper tail bounds for random matrices

Here we established tail bound for both sub-Gaussian and Bernstein random matrices.

## Theorem (Hoeffding bound for random matrices)

*Let $\{Q_i\}_{i=1}^n$ be a sequence of zero-mean independent symmetric random matrices that satisfy the sub-Gaussian condition with parameters $\{V_i\}_{i=1}^n$. Then for all $\delta > 0$, we have*

$$\mathbb{P}[\|\frac{1}{n}\sum_{i=1}^n Q_i\|_2 \geq \delta] \leq 2rank(\sum_{i=1}^n V_i)e^{\frac{n\delta^2}{2\sigma^2}} \leq 2de^{\frac{n\delta^2}{2\sigma^2}},$$

*where $\sigma^2 = \|\frac{1}{n}\sum_{i=1}^n V_i\|_2$*

# Proof

- First we prove for $V = \sum_{i=1}^{n} V_i$ is full rank.
- From $\sum_{i=1}^{n} \log \Psi_{Q_i}(\lambda) \leq \frac{\lambda^2}{2} \sum_{i=1}^{n} V_i$, we have $tr(e^{\sum_{i=1}^{n} \log \Psi_{Q_i}(\lambda)}) \leq tr(e^{\frac{\lambda^2}{2} \sum_{i=1}^{n} V_i})$.
- Then $\mathbb{P}[\|\frac{1}{n} \sum_{i=1}^{n} Q_i\|_2 \geq \delta] \leq 2tr(e^{\frac{\lambda^2}{2} \sum_{i=1}^{n} V_i}) e^{-\lambda n \delta}$.
- For any $R \in \mathbb{S}^{d \times d}$ we have $tr(e^R) \leq de^{\|R\|_2}$, and $\|\frac{\lambda^2}{2} \sum_{i=1}^{n} V_i\|_2 = \frac{\lambda^2}{2} n\sigma^2$, we have $\mathbb{P}[\|\frac{1}{n} \sum_{i=1}^{n} Q_i\|_2 \geq \delta] \leq 2de^{\frac{\lambda^2}{2} n\sigma^2 - \lambda n \delta}$. Let $\lambda = \delta/\sigma^2$we get the result.
- For rank deficient case (rank $= r$), Let $V = UDU^T$ be eigendecomposition. Let $\tilde{Q} = U^T QU$ and $Q = \sum_{i=1}^{n} Q_i$,we get $\|\tilde{Q}\|_2 = \|Q\|_2$. Then we just apply previous result on $\tilde{Q}$.

# Some remarks

- The previous result also works under non-symmetric $\{A_i\}_{i=1}^n \in \mathbb{R}^{d_1 \times d_2}$ with $d$ replaced by $d_1 + d_2$. We just have to consider $Q_i = \begin{pmatrix} 0 & A_i \\ A_i^T & 0 \end{pmatrix}$
- (Looseness/Sharpness of Hoeffding)
  Let $Q_i = y_i E_i$, where $E_i$ is diagonal matrix with $E_{ii} = 1$ and $y_i$ is 1-sub-Gaussian variables. Then $Q_i$ is sub-Gaussian with parameter $V_i = E_i$. $\sigma^2 = \|\frac{1}{d}\sum_{i=1}^n V_i\|_2 = \frac{1}{d}$. Then we get

$$\mathbb{P}[\|\frac{1}{d}\sum_{i=1}^n Q_i\|_2 \geq \delta] \leq 2de^{-\frac{d^2\sigma^2}{2}}$$

  This implies $\|\frac{1}{d}\sum_{i=1}^n Q_i\|_2 \lesssim \frac{\sqrt{2\log(2d)}}{d}$. But we know $\|\frac{1}{d}\sum_{i=1}^n Q_i\|_2 = \max_{i=1,\cdots,d}\frac{|y_i|}{d}$. When $y_i$ is Rademacher variables, $\|\frac{1}{d}\sum_{i=1}^n Q_i\|_2 = \frac{1}{d}$. When $y_i$ is standard Gaussian variables, $\|\frac{1}{d}\sum_{i=1}^n Q_i\|_2 \approx \frac{\sqrt{2\log d}}{d}$.

# Bernstein bound

## Theorem (Bernstein bound for random matrices)

*Let $\{Q_i\}_{i=1}^n$ be a sequence of zero-mean independent symmetric random matrices that satisfy the Bernstein condition with parameters $b > 0$. Then for all $\delta > 0$, we have*

$$\mathbb{P}[\|\frac{1}{n}\sum_{i=1}^{n} Q_i\|_2 \geq \delta] \leq 2rank(\sum_{i=1}^{n} Var(Q_i))e^{-\frac{n\delta^2}{2(\sigma^2+b\delta)}},$$

*where $\sigma^2 = \|\frac{1}{n}\sum_{i=1}^{n} Var(Q_i)\|_2$*

Proof: Using $tr(\Psi_{S_n}(\lambda)) \leq tr(e^{\sum_{i=1}^{n}\log \Psi_{Q_i}(\lambda)})$ andlog$(\Psi_{Q_i}(\lambda) \leq \frac{\lambda^2 Var(Q_i)}{1-b|\lambda|}) \quad \forall |\lambda| < \frac{1}{b}$. Then,

$$tr(e^{\sum_{i=1}^{n}\log \Psi_{Q_i}(\lambda)}) \leq tr(\exp(\frac{\lambda^2 \sum_{i=1}^{n} Var(Q_i)}{1-b|\lambda|})) \leq rank(\sum_{i=1}^{n} Var(Q_i))e^{\frac{n\lambda^2\sigma^2}{1-b|\lambda|}}$$

# Proof

we use Lemma of summation of random matrices, we get

$$\mathbb{P}[\|\frac{1}{n}\sum_{i=1}^{n} Q_i\|_2 \geq \delta] \leq 2rank(\sum_{i=1}^{n} Var(Q_i))e^{\frac{n\lambda^2\sigma^2}{1-b|\lambda|}-\lambda n\delta}.$$

Picking $\lambda = \frac{\delta}{\sigma^2+b\delta} \in (0, \frac{1}{b})$. We get the result.

- This type of bounds also can generalize to non-symmetric case.

# Consequences for covariance matrices

## Corollary (Bounded samples)

*Let $x_1, \cdots, x_n$ be .i.d. zero-mean random vectors with covariance $\Sigma$ and $\|x_j\|_2 \leq \sqrt{b}$ almost surely. Then for all $\delta > 0$, the sample covariance matrix $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ satisfies*

$$\mathbb{P}[\|\hat{\Sigma} - \Sigma\|_2 \geq \delta] \leq 2d \exp(-\frac{n\delta^2}{2b(\|\Sigma\|_2 + \delta)}).$$

Proof: Let $Q_i = x_i x_i^T - \Sigma$. we have $\|Q_i\|_2 \leq b + \|\Sigma\|_2$. Also $\Sigma = \mathbb{E}[x_i x_i^T]$, then $\|\Sigma\|_2 = \max_{v \in \mathbb{S}^{d-1}} \mathbb{E}[< v, x_i >^2] \leq b$. Hence $\|Q_i\|_2 \leq 2b$. We have

$$Var(Q_i) = \mathbb{E}[(x_i x_i^T)^2] - \Sigma^2 \leq \mathbb{E}[\|x_i\|_2^2 x_i x_i^T] \leq b\Sigma$$

And $\|Var(Q_i)\|_2 \leq b\|\Sigma\|_2$. Using Bernstein's theorem we can get the result.