Random matrices and covariance estimation

Tianhao Wang, Liangchen He

Department of Statistics and Finance, USTC

October 16, 2022

Notation and basic facts

Given a rectangular matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ with $n \geq m$, we write its ordered singular values as

$$\sigma_{\max}(\mathbf{A}) = \sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \cdots \geq \sigma_m(\mathbf{A}) = \sigma_{\min}(\mathbf{A}) \geq 0$$

Note that $\sigma_{\min}(\mathbf{A}) = \min_{v \in \mathbb{S}^{m-1}} \|\mathbf{A}v\|_2$, and

$$\sigma_{\max}(\mathbf{A}) = \max_{v \in \mathbb{S}^{m-1}} \|\mathbf{A}v\|_2 =: |||\mathbf{A}|||_2$$

Denote $\mathcal{S}^{d \times d} := \left\{\mathbf{Q} \in \mathbb{R}^{d \times d} \mid \mathbf{Q} = \mathbf{Q}^{\mathrm{T}}\right\}$, $\mathcal{S}_+^{d \times d} := \left\{\mathbf{Q} \in \mathcal{S}^{d \times d} \mid \mathbf{Q} \geq 0\right\}$. For a matrix $\mathbf{Q} \in \mathcal{S}^{d \times d}$, we write its eigenvalues as

$$\gamma_{\max}(\mathbf{Q}) = \gamma_1(\mathbf{Q}) \geq \gamma_2(\mathbf{Q}) \geq \cdots \geq \gamma_d(\mathbf{Q}) = \gamma_{\min}(\mathbf{Q})$$

Notation and basic facts

$$|||\mathbf{Q}|||_2 = \max_{v \in \mathbb{S}^{d-1}} \left| v^{\mathrm{T}} \mathbf{Q} v \right| = \max \left\{ \gamma_{\max}(\mathbf{Q}), \left| \gamma_{\min}(\mathbf{Q}) \right| \right\}$$

Then we have the relationship $\gamma_j(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = (\sigma_j(\mathbf{A}))^2$ for $j = 1, \ldots, m$.

Let $\{x_1, \ldots, x_n\}$ be a collection of n independent and identically distributed samples from a distribution in $\mathbb{R}^d$ with zero mean, and covariance matrix $\mathbf{\Sigma} = \operatorname{cov}(x_1) \in \mathcal{S}_+^{d \times d}$. A standard estimator of $\mathbf{\Sigma}$ is the sample covariance matrix

$$\widehat{\mathbf{\Sigma}} := \frac{1}{n} \sum_{i=1}^{n} x_i x_i^{\mathrm{T}} = \frac{1}{n} \mathbf{X}^{\mathrm{T}} \mathbf{X}$$

Theorem

Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be drawn according to the $\mathbf{\Sigma}$-Gaussian ensemble. Then for all $\delta > 0$, the maximum singular value $\sigma_{\max}(\mathbf{X})$ satisfies the upper deviation inequality

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

Moreover, for $n \geq d$, the minimum singular value $\sigma_{\min}(\mathbf{X})$ satisfies the analogous lower deviation inequality

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

Proof

We can write $\mathbf{X} = \mathbf{W}\sqrt{\mathbf{\Sigma}}$, where the random matrix $\mathbf{W} \in \mathbb{R}^{n \times d}$ has i.i.d. $\mathcal{N}(0,1)$ entries. Let $\bar{\sigma}_{\max} = \gamma_{\max}(\sqrt{\mathbf{\Sigma}})$, $\bar{\sigma}_{\min} = \gamma_{\min}(\sqrt{\mathbf{\Sigma}})$. Consider the mapping $f : \mathbf{W} \mapsto \sigma_{\max}(\mathbf{W}\sqrt{\mathbf{\Sigma}})$ as a real-valued function on $\mathbb{R}^{nd}$. Since we have

$$|||\mathbf{W}\sqrt{\mathbf{\Sigma}}|||_2 = \max_{v \in \mathbb{S}^{d-1}} ||\mathbf{W}\sqrt{\mathbf{\Sigma}}v||_2 \le |||\mathbf{W}|||_2 \max_{v \in \mathbb{S}^{d-1}} ||\sqrt{\mathbf{\Sigma}}v||_2 \le |||\mathbf{W}|||_F \bar{\sigma}_{\max}$$

Then f is a $\bar{\sigma}_{\max}$-Lipschitz function. According to Theorem 2.26, we conclude that

$$\mathbb{P}\left[\sigma_{\max}(\mathbf{X}) \ge \mathbb{E}\left[\sigma_{\max}(\mathbf{X})\right] + \sqrt{n}\bar{\sigma}_{\max}\delta\right] \le e^{-n\delta^2/2}.$$

Proof

Consequently, it suffices to show that

$$\mathbb{E}\left[\sigma_{\max}(\mathbf{X})\right] \leq \sqrt{n}\bar{\sigma}_{\max} + \sqrt{\operatorname{tr}(\mathbf{\Sigma})}$$

Since $\sigma_{\max}(\mathbf{X}) = \max_{v' \in \mathbb{S}^{d-1}} \|\mathbf{X}v'\|_2$, making the substitution $v = \sqrt{\mathbf{\Sigma}}v'$, we can write

$$\sigma_{\max}(\mathrm{X}) = \max_{v \in \mathbb{S}^{d-1}(\mathbf{\Sigma}^{-1})} \|\mathrm{W}v\|_2 = \max_{u \in \mathbb{S}^{n-1}} \max_{v \in \mathbb{S}^{d-1}(\mathbf{\Sigma}^{-1})} \underbrace{u^{\mathrm{T}}\mathbf{W}v}_{Z_{u,v}}$$

where $\mathbb{S}^{d-1}\left(\mathbf{\Sigma}^{-1}\right) := \left\{v \in \mathbb{R}^d \mid \left\|\mathbf{\Sigma}^{-\frac{1}{2}}v\right\|_2 = 1\right\}$ is an ellipse.

Proof

We try to apply the Sudakov–Fernique comparison (Theorem 5.27) to conclude that

$$\mathbb{E}\left[\sigma_{\max}(\mathbf{X})\right] = \mathbb{E}\left[\max_{(u,v)\in\mathbb{T}} Z_{u,v}\right] \leq \mathbb{E}\left[\max_{(u,v)\in\mathbb{T}} Y_{u,v}\right] \leq \sqrt{n}\bar{\sigma}_{\max} + \sqrt{\operatorname{tr}(\boldsymbol{\Sigma})}$$

which means we need to construct another Gaussian process $\{Y_{u,v}, (u,v) \in \mathbb{T}\}$ such that

$$\mathbb{E}\left[\left(Z_{u,v} - Z_{\widetilde{u},\widetilde{v}}\right)^2\right] \leq \mathbb{E}\left[\left(Y_{u,v} - Y_{\widetilde{u},\widetilde{v}}\right)^2\right]$$

for all pairs (u,v) and $(\widetilde{u},\widetilde{v})$ in $\mathbb{T} := \mathbb{S}^{n-1} \times \mathbb{S}^{d-1}\left(\boldsymbol{\Sigma}^{-1}\right)$. Introducing the Gaussian process $Z_{u,v} := u^{\mathrm{T}}\mathbf{W}v = \langle\langle \mathbf{W}, uv^{\mathrm{T}}\rangle\rangle$, where we use $\langle\langle A, B\rangle\rangle := \sum_{j=1}^{n}\sum_{k=1}^{d} A_{jk}B_{jk}$ to denote the trace inner product.

Proof

We may assume without loss of generality that $\|v\|_2 \leq \|\tilde{v}\|_2$. Then

$$
\begin{aligned}
&\mathbb{E}\left[(Z_{u,v} - Z_{\tilde{u},\tilde{v}})^2\right] = \mathbb{E}\left[\left\langle\left\langle \mathbf{W}, uv^{\mathrm{T}} - \tilde{u}\tilde{v}^{\mathrm{T}} \right\rangle\right\rangle^2\right] = \left\| uv^{\mathrm{T}} - \tilde{u}\tilde{v}^{\mathrm{T}} \right\|_F^2 \\
&= \left\| u(v-\tilde{v})^{\mathrm{T}} + (u-\tilde{u})\tilde{v}^{\mathrm{T}} \right\|_F^2 \\
&= \left\| (u-\tilde{u})\tilde{v}^{\mathrm{T}} \right\|_F^2 + \left\| u(v-\tilde{v})^{\mathrm{T}} \right\|_F^2 + 2\left\langle\left\langle u(v-\tilde{v})^{\mathrm{T}}, (u-\tilde{u})\tilde{v}^{\mathrm{T}} \right\rangle\right\rangle \\
&\leq \underbrace{\|\tilde{v}\|_2^2}_{\bar{\sigma}^2_{\max}} \|u-\tilde{u}\|_2^2 + \underbrace{\|u\|_2^2}_{=1} \|v-\tilde{v}\|_2^2 + 2\underbrace{\left(\|u\|_2^2 - \langle u,\tilde{u}\rangle\right)}_{\geq 0}\underbrace{\left(\langle v,\tilde{v}\rangle - \|\tilde{v}\|_2^2\right)}_{\leq 0} \\
&\leq \bar{\sigma}^2_{\max}\|u-\tilde{u}\|_2^2 + \|v-\tilde{v}\|_2^2
\end{aligned}
$$

Proof

Motivated by this inequality, we define the Gaussian process $Y_{u,v} := \bar{\sigma}_{\max}\langle g, u\rangle + \langle h, v\rangle$, where $g \in \mathbb{R}^n$ and $h \in \mathbb{R}^d$ are both standard Gaussian random vectors and mutually independent. By construction, we have

$$\mathbb{E}\left[\left(Y_\theta - Y_{\widetilde{\theta}}\right)^2\right] = \bar{\sigma}^2_{\max}\|u - \widetilde{u}\|_2^2 + \|v - \widetilde{v}\|_2^2.$$

Thus, we may apply the Sudakov-Fernique bound to conclude that

$$\begin{aligned}
\mathbb{E}\left[\sigma_{\max}(\mathbf{X})\right] \le \mathbb{E}\left[\max_{(u,v)\in\mathbb{T}} Y_{u,v}\right] &= \bar{\sigma}_{\max}\mathbb{E}\left[\max_{u\in\mathbb{S}^{n-1}}\langle g, u\rangle\right] + \mathbb{E}\left[\max_{v\in\mathbb{S}^{d-1}(\mathbf{\Sigma}^{-1})}\langle h, v\rangle\right] \\
&= \bar{\sigma}_{\max}\mathbb{E}\left[\|g\|_2\right] + \mathbb{E}\left[\|\sqrt{\mathbf{\Sigma}}h\|_2\right] \\
&\le \sqrt{n}\bar{\sigma}_{\max} + \sqrt{\operatorname{tr}(\mathbf{\Sigma})}
\end{aligned}$$

which finishes the proof of the upper bound on $\sigma_{\max}(\mathbf{X})$.

Assume that the random vector $x_i \in \mathbb{R}^d$ is zero-mean, and sub-Gaussian with parameter at most σ, by which we mean that, for each fixed $v \in \mathbb{S}^{d-1}$,

$$\mathbb{E}\left[e^{\lambda\langle v, x_i\rangle}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}} \quad \text{for all } \lambda \in \mathbb{R}.$$

When the random matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ is formed by drawing each row $x_i \in \mathbb{R}^d$ in an i.i.d. manner from a σ-sub-Gaussian distribution, then we say that $\mathbf{X}$ is a sample from a row-wise σ-sub-Gaussian ensemble. For any such random matrix, we have the following result:

Theorem

There are universal constants $\{c_j\}_{j=0}^{3}$ *such that, for any row-wise* σ*-sub-Gaussian random matrix* $\mathbf{X} \in \mathbb{R}^{n\times d}$ *, the sample covariance* $\widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}$ *satisfies the bounds*

$$\mathbb{E}\left[e^{\lambda|||\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}|||_2}\right] \leq e^{c_0\frac{\lambda^2\sigma^4}{n}+4d} \qquad \text{for all } |\lambda| < \frac{n}{64e^2\sigma^2}.$$

and hence

$$\mathbb{P}\left[\frac{\lambda|||\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}|||_2}{\sigma^2} \geq c_1\left\{\sqrt{\frac{d}{n}}+\frac{d}{n}\right\}+\delta\right] \leq c_2 e^{-c_3 n\min\{\delta,\delta^2\}} \qquad \text{for all } \delta \geq 0.$$

Proof

We introduce the shorthand $\mathbf{Q} := \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$. Let $\left\{v^1, \ldots, v^N\right\}$ be a $\frac{1}{8}$-covering of the sphere $\mathbb{S}^{d-1}$ in the Euclidean norm; from Example 5.8, there exists such a covering with $N \leq 17^d$ vectors. Given any $v \in \mathbb{S}^{d-1}$, we can write $v = v^j + \Delta$ for some v^j in the cover, and an error vector Δ such that $\|\Delta\|_2 \leq \frac{1}{8}$, and hence

$$
\begin{aligned}
|\langle v, \mathbf{Q}v\rangle| &= |\left\langle v^j, \mathbf{Q}v^j\right\rangle + 2\left\langle \Delta, \mathbf{Q}v^j\right\rangle + \langle \Delta, \mathbf{Q}\Delta\rangle| \\
&\leq |\left\langle v^j, \mathbf{Q}v^j\right\rangle + 2\|\Delta\|_2 |||\mathbf{Q}|||_2 \left\|v^j\right\|_2 + |||\mathbf{Q}|||_2 \|\Delta\|_2^2 \\
&\leq |\left\langle v^j, \mathbf{Q}v^j\right\rangle| + \frac{1}{4}|||\mathbf{Q}|||_2 + \frac{1}{64}|||\mathbf{Q}|||_2 \\
&\leq |\left\langle v^j, \mathbf{Q}v^j\right\rangle| + \frac{1}{2}|||\mathbf{Q}|||_2
\end{aligned}
$$

Proof

$$|||\mathbf{Q}|||_2 = \max_{v \in S^{d-1}} |\langle v, \mathbf{Q}v\rangle| \le 2 \max_{j=1,\dots,N} \left|\left\langle v^j, \mathbf{Q}v^j\right\rangle\right|.$$

Consequently, we have

$$\begin{aligned}\mathbb{E}\left[e^{\lambda|||\mathbb{Q}|||_2}\right] &\le \mathbb{E}\left[\exp\left(2\lambda \max_{j=1,\dots,N}\left|\left\langle v^j, \mathbf{Q}v^j\right\rangle\right|\right)\right] \\ &\le \sum_{j=1}^{N}\left\{\mathbb{E}\left[e^{2\lambda\left\langle v^j,\mathbf{Q}v^j\right\rangle}\right] + \mathbb{E}\left[e^{-2\lambda\left\langle v^j,\mathbf{Q}v^j\right\rangle}\right]\right\}\end{aligned}$$

Next we claim that for any fixed unit vector $u \in \mathbb{S}^{d-1}$,

$$\mathbb{E}\left[e^{t\langle u,\mathbf{Q}u\rangle}\right] \le e^{512\frac{t^2}{n}e^4\sigma^4} \quad \text{for all } |t| \le \frac{n}{32e^2\sigma^2}.$$

Proof

Then we find that

$$\mathbb{E}\left[e^{\lambda|||\mathbf{Q}|||_2}\right] \leq 2Ne^{2048\frac{\lambda^2}{n}e^4\sigma^4} \leq e^{c_0\frac{\lambda^2\sigma^4}{n}+4d},$$

valid for all $|\lambda| < \frac{n}{64e^2\sigma^2}$, where the final step uses the fact that $2\left(17^d\right) \leq e^{4d}$. Now we proof the claim. We have

$$\mathbb{E}\left[e^{t\langle u,Qu\rangle}\right] = \prod_{i=1}^{n}\mathbb{E}\left[e^{\frac{t}{n}\{\langle x_i,u\rangle^2-\langle u,\mathbf{\Sigma}u\rangle\}}\right] = \left(\mathbb{E}\left[e^{\frac{t}{n}\{\langle x_1,u\rangle^2-\langle u,\mathbf{\Sigma}u\rangle\}}\right]\right)^n.$$

Letting $\varepsilon \in \{-1,+1\}$ denote a Rademacher variable, independent of x_1 , a standard symmetrization argument (see Proposition 4.11) implies that

Proof

$$
\begin{aligned}
\mathbb{E}_{x_1}\left[e^{\frac{t}{n}\{\langle x_1,u\rangle^2-\langle u,\mathbf{\Sigma}u\rangle\}}\right] &\le \mathbb{E}_{x_1,\varepsilon}\left[e^{\frac{2t}{n}\varepsilon\langle x_1,u\rangle^2}\right] \\
&\overset{\text{(i)}}{=} \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{2t}{n}\right)^k \mathbb{E}\left[\varepsilon^k \langle x_1,u\rangle^{2k}\right] \\
&\overset{\text{(ii)}}{=} 1+\sum_{\ell=1}^{\infty}\frac{1}{(2\ell)!}\left(\frac{2t}{n}\right)^{2\ell}\mathbb{E}\left[\langle x_1,u\rangle^{4\ell}\right]
\end{aligned}
$$

where step (i) follows by the power-series expansion of the exponential, and step (ii) follows since ε and x_1 are independent, and all odd moments of the Rademacher term vanish. By property (III) in Theorem 2.6 on equivalent characterizations of sub-Gaussian variables, we are guaranteed that

Proof

$$\mathbb{E}\left[\langle x_1, u\rangle^{4\ell}\right] \leq \frac{(4\ell)!}{2^{2\ell}(2\ell)!}(\sqrt{8}e\sigma)^{4\ell} \quad \text{for all } \ell = 1, 2, \ldots$$

and hence

$$\begin{aligned}
\mathbb{E}_{x_1}\left[e^{\frac{t}{n}\{\langle x_1,u\rangle^2-\langle u,\Sigma u\rangle\}}\right] &\leq 1 + \sum_{\ell=1}^{\infty}\frac{1}{(2\ell)!}\left(\frac{2t}{n}\right)^{2\ell}\frac{(4\ell)!}{2^{2\ell}(2\ell)!}(\sqrt{8}e\sigma)^{4\ell} \\
&\leq 1 + \sum_{\ell=1}^{\infty}(\underbrace{\frac{16t}{n}e^2\sigma^2}_{f(t)})^{2\ell}
\end{aligned}$$

Proof

where we have used the fact that $(4\ell)! \leq 2^{2\ell}[(2\ell)!]^2$. As long as $f(t) := \frac{16t}{n} e^2\sigma^2 < \frac{1}{2}$, we can write

$$1 + \sum_{\ell=1}^{\infty} \left[f^2(t)\right]^{\ell} \overset{\text{(i)}}{=} \frac{1}{1 - f^2(t)} \overset{\text{(ii)}}{\leq} \exp\left(2f^2(t)\right)$$

where step (i) follows by summing the geometric series, and step (ii) follows because $\frac{1}{1-a} \leq e^{2a}$ for all $a \in \left[0, \frac{1}{2}\right]$. Putting together the pieces and combining with our earlier bound (6.23), we have shown that $\mathbb{E}\left[e^{t\langle u, Qu\rangle}\right] \leq e^{2nf^2(t)}$, valid for all $|t| < \frac{n}{32e^2\sigma^2}$, which establishes the claim.

Proof

$$\mathbb{E}_X\left[\Phi\left(\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}}\right)\right] \leq \mathbb{E}_{X,\varepsilon}\left[\Phi\left(2\,\|\mathbb{S}_n\|_{\mathscr{F}}\right)\right]$$

$$\|\mathbb{S}_n\|_{\mathscr{F}} := \sup_{f \in \mathscr{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f\left(X_i\right) \right|$$

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} := \sup_{f \in \mathscr{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f\left(X_i\right) - \mathbb{E}[f(X)] \right|$$

6.4.1 Background on matrix analysis

Given a matrix $\mathbf{Q} \in \mathcal{S}^{d\times d}$, consider its eigendecomposition $\mathbf{Q} = \mathbf{U}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{U}$. Here the matrix $\mathbf{U} \in \mathbb{R}^{d\times d}$ is a unitary matrix, satisfying the relation $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_d$, whereas $\mathbf{\Gamma} := \mathrm{diag}(\gamma(\mathbf{Q}))$ is a diagonal matrix specified by the vector of eigenvalues $\gamma(\mathbf{Q}) \in \mathbb{R}^d$.

$$\mathbf{Q} \mapsto f(\mathbf{Q}) := \mathbf{U}^{\mathrm{T}} \mathrm{diag}\left(f\left(\gamma_1(\mathbf{Q})\right), \ldots, f\left(\gamma_d(\mathbf{Q})\right)\right) \mathbf{U}$$

By construction, this extension of f to $\mathcal{S}^{d\times d}$ is unitarily invariant, meaning that

$$f\left(\mathbf{V}^{\mathrm{T}}\mathbf{Q}\mathbf{V}\right) = \mathbf{V}^{\mathrm{T}} f(\mathbf{Q})\mathbf{V} \quad \text{for all unitary matrices } \mathbf{V} \in \mathbb{R}^{d\times d},$$

$$\gamma(f(\mathbf{Q})) = \left\{f\left(\gamma_j(\mathbf{Q})\right), j = 1, \ldots, d\right\}$$

The matrix exponential has the power-series expansion $e^{\mathbf{Q}} = \sum_{k=0}^{\infty} \frac{\mathbf{Q}^k}{k!}$. By the spectral mapping property, the eigenvalues of $e^{\mathbf{Q}}$ are positive.

6.4.2 Tail conditions for matrices

$\mathrm{var}(\mathbf{Q}) := \mathbb{E}\left[\mathbf{Q}^2\right] - (\mathbb{E}[\mathbf{Q}])^2$. The moment generating function of a random matrix $\mathbf{Q}$ is the matrix-valued mapping $\Psi_{\mathbf{Q}} : \mathbb{R} \to \mathcal{S}^{d\times d}$ given by

$$\Psi_{\mathbf{Q}}(\lambda) := \mathbb{E}\left[e^{\lambda \mathbf{Q}}\right] = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}\left[\mathbf{Q}^k\right].$$

Definition 6.6 A zero-mean symmetric random matrix $\mathbf{Q} \in \mathcal{S}^{d\times d}$ is sub-Gaussian with matrix parameter $\mathbf{V} \in \mathcal{S}_{+}^{d\times d}$ if

$$\Psi_{\mathbf{Q}}(\lambda) \preceq e^{\frac{\lambda^2 \mathbf{V}}{2}} \quad \text{for all } \lambda \in \mathbb{R}.$$

Sub-exponential random matrices

- **Definition 6.9** (Sub-exponential random matrices) A zero-mean random matrix is sub-exponential with parameters $(\mathbf{V}, \alpha)$ if

$$\Psi_{\mathbf{Q}}(\lambda) \leq e^{\frac{\lambda^2 \mathbf{V}}{2}} \quad \text{for all } |\lambda| < \frac{1}{\alpha}.$$

- **Definition 6.10** (Bernstein's condition for matrices) A zero-mean symmetric random matrix $\mathbf{Q}$ satisfies a Bernstein condition with parameter $b > 0$ if

$$\mathbb{E}\left[\mathbf{Q}^j\right] \leq \frac{1}{2} j! b^{j-2} \operatorname{var}(\mathbf{Q}) \quad \text{for } j = 3, 4, \ldots.$$

- We note that (a stronger form of) Bernstein's condition holds whenever the matrix $\mathbf{Q}$ has a bounded operator norm-say $\|\mathbf{Q}\|_2 \leq b$ almost surely.

$$\mathbb{E}\left[\mathbf{Q}^j\right] \leq b^{j-2} \operatorname{var}(\mathbf{Q}) \quad \text{for all } j = 3, 4, \ldots.$$

Bernstein condition

Lemma 6.11 For any symmetric zero-mean random matrix satisfying the Bernstein condition, we have

$$\Psi_{\mathbf{Q}}(\lambda) \leq \exp\left(\frac{\lambda^2 \operatorname{var}(\mathbf{Q})}{2(1-b|\lambda|)}\right) \quad \text{for all } |\lambda| < \frac{1}{b}.$$

- **Proof:** Since $\mathbb{E}[\mathbf{Q}] = 0$, applying the definition of the matrix exponential for a suitably small $\lambda \in \mathbb{R}$ yields($\mathbf{I}_d + \mathbf{A} \preceq e^{\mathbf{A}}$)

$$\mathbb{E}\left[e^{\lambda \mathbf{Q}}\right] = \mathbf{I}_d + \frac{\lambda^2 \operatorname{var}(\mathbf{Q})}{2} + \sum_{j=3}^{\infty} \frac{\lambda^j \mathbb{E}\left[\mathbf{Q}^j\right]}{j!}$$

$$\overset{\text{(i)}}{\preceq} \mathbf{I}_d + \frac{\lambda^2 \operatorname{var}(\mathbf{Q})}{2} \left\{\sum_{j=0}^{\infty} |\lambda|^j b^j\right\}$$

Matrix Chernoff approach and independent decompositions

Lemma 6.12 (Matrix Chernoff technique) Let $\mathbf{Q}$ be a zero-mean symmetric random matrix whose moment generating function Ψ_Q exists in an open interval $(-a, a)$. Then for any $\delta > 0$, we have

$$\mathbb{P}\left[\gamma_{\max}(\mathbf{Q}) \geq \delta\right] \leq \operatorname{tr}\left(\Psi_{\mathbf{Q}}(\lambda)\right) e^{-\lambda\delta} \quad \text{for all } \lambda \in [0, a),$$

where $\mathtt{tr}(\cdot)$ denotes the trace operator on matrices. Similarly, we have

$$\mathbb{P}\left[\|\mathbf{Q}\|_2 \geq \delta\right] \leq 2\operatorname{tr}\left(\Psi_{\mathbf{Q}}(\lambda)\right) e^{-\lambda\delta} \quad \text{for all } \lambda \in [0, a)$$

Proof:

- Applying Markov inequality.

Independent decompositions

Lemma 6.13 Let $\mathbf{Q}_1, \ldots, \mathbf{Q}_n$ be independent symmetric random matrices whose moment generating functions exist for all $\lambda \in I$, and define the sum $\mathbf{S}_n := \sum_{i=1}^n \mathbf{Q}_i$. Then

$$\operatorname{tr}\left(\Psi_{\mathbf{S}_n}(\lambda)\right) \leq \operatorname{tr}\left(e^{\sum_{i=1}^n \log \Psi_{Q_i}(\lambda)}\right) \quad \text{for all } \lambda \in I.$$

Independent decompositions

Proof: In order to prove this lemma, we require the following result due to Lieb (1973): for any fixed matrix $\mathbf{H} \in \mathcal{S}^{d\times d}$, the function $f : \mathcal{S}_+^{d\times d} \to \mathbb{R}$ given by

$$f(\mathbf{A}) := \operatorname{tr}\left(e^{\mathbf{H}+\log(\mathbf{A})}\right)$$

is concave. Introducing the shorthand notation $G(\lambda) := \operatorname{tr}\left(\Psi_{\mathbf{S}_n}(\lambda)\right)$, we note that, by linearity of trace and expectation, we have

$$G(\lambda) = \operatorname{tr}\left(\mathbb{E}\left[e^{\lambda\mathbf{S}_{n-1}+\log\exp(\lambda\mathbf{Q}_n)}\right]\right) = \mathbb{E}_{\mathbf{S}_{n-1}}\mathbb{E}_{\mathbf{Q}_n}\left[\operatorname{tr}\left(e^{\lambda\mathbf{S}_{n-1}+\log\exp(\lambda\mathbf{Q}_n)}\right)\right]$$

Using concavity of the function f with $\mathbf{H} = \lambda\mathbf{S}_{n-1}$ and $\mathbf{A} = e^{\lambda\mathbf{Q}_n}$, Jensen's inequality implies that

$$\mathbb{E}_{\mathbf{Q}_n}\left[\operatorname{tr}\left(e^{\lambda\mathbf{S}_{n-1}+\log\exp(\lambda\mathbf{Q}_n)}\right)\right] \le \operatorname{tr}\left(e^{\lambda\mathbf{S}_{n-1}+\log\mathbb{E}_{\mathbf{Q}_n}\exp(\lambda\mathbf{Q}_n)}\right),$$

so that we have shown that $G(\lambda) \le \mathbb{E}_{\mathbf{S}_{n-1}}\left[\operatorname{tr}\left(e^{\lambda\mathbf{S}_{n-1}+\log\Psi_{Q_n}(\lambda)}\right)\right]$.

Symmetrization

Example 6.14 (Rademacher symmetrization for random matrices) Let $\{\mathbf{A}_i\}_{i=1}^n$ be a sequence of independent symmetric random matrices, and suppose that our goal is to bound the maximum eigenvalue of the matrix sum $\sum_{i=1}^n (\mathbf{A}_i - \mathbb{E}[\mathbf{A}_i])$. $\widetilde{\mathbf{Q}}_i = \varepsilon_i \mathbf{A}_i$, where ε_i is an independent Rademacher variable. Let us now work through this reduction. By Markov's inequality, we have

$$\mathbb{P}\left[\gamma_{\max}\left(\sum_{i=1}^n \{\mathbf{A}_i - \mathbb{E}[\mathbf{A}_i]\}\right) \geq \delta\right] \leq \mathbb{E}\left[e^{\lambda \gamma_{\max}\left(\sum_{i=1}^n [\mathbf{A}_i - \mathbb{E}[\mathbf{A}_i])\right)}\right] e^{-\lambda\delta}.$$

The proof of symmetrization

By the variational representation of the maximum eigenvalue, we have

$$\begin{aligned}\mathbb{E}\left[e^{\lambda\gamma_{\max}\left(\sum_{i=1}^n(\mathbf{A}_i-\mathbb{E}[\mathbf{A}_i])\right)}\right] &= \mathbb{E}\left[\exp\left(\lambda\sup_{\|u\|_2=1}\left\langle u,\left(\sum_{i=1}^n\left(\mathbf{A}_i-\mathbb{E}\left[\mathbf{A}_i\right]\right)\right)u\right\rangle\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(2\lambda\sup_{\|u\|_2=1}\left\langle u,\left(\sum_{i=1}^n\varepsilon_i\mathbf{A}_i\right)u\right\rangle\right)\right] \\ &= \mathbb{E}\left[e^{2\lambda\gamma_{\max}\left(\sum_{i=1}^n\varepsilon_i\mathbf{A}_i\right)}\right] \\ &= \mathbb{E}\left[\gamma_{\max}\left(e^{2\lambda\sum_{i=1}^n\varepsilon_i\mathbf{A}_i}\right)\right],\end{aligned}$$

So that

$$\mathbb{E}\left[\gamma_{\max}\left(e^{2\lambda\sum_{i=1}^n\varepsilon_i\mathbf{A}_i}\right)\right]\leq \operatorname{tr}\left(\mathbb{E}\left[e^{2\lambda\sum_{i=1}^n\varepsilon_i\mathbf{A}_i}\right]\right)\leq \operatorname{tr}\left(e^{\sum_{i=1}^n\log\Psi_{\tilde{a}_i}(2\lambda)}\right)$$

6.4.4 Upper tail bounds for random matrices

(Theorem 6.15) (Hoeffding bound for random matrices) Let $\{\mathbf{Q}_i\}_{i=1}^n$ be a sequence of zero-mean independent symmetric random matrices that satisfy the sub-Gaussian condition with parameters $\{\mathbf{V}_i\}_{i=1}^n$. Then for all $\delta > 0$, we have the upper tail bound

$$\mathbb{P}\left[\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{Q}_i\right\|_2 \geq \delta\right] \leq 2\,\mathrm{rank}\left(\sum_{i=1}^n \mathbf{V}_i\right) e^{-\frac{n\sigma^2}{2\sigma^2}} \leq 2de^{-\frac{n\sigma^2}{2\sigma^2}},$$

where $\sigma^2 = \left\|\frac{1}{n}\sum_{i=1}^n \mathbf{V}_i\right\|_2$.

Proof: Lemma 6.13 and $\mathrm{tr}\left(e^{\mathbf{R}}\right) \leq de^{\|\mathbf{R}\|_2}$.

Bernstein bound for random matrices

Theorem 6.17 (Bernstein bound for random matrices) Let $\left\{\{\mathbf{Q}_i\}_{i=1}^n\right\}$ be a sequence of independent, zero-mean, symmetric random matrices that satisfy the Bernstein condition with parameter $b > 0$. Then for all $\delta \geq 0$, the operator norm satisfies the tail bound

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

where $\sigma^2 := \frac{1}{n}\left\|\sum_{j=1}^n \text{var}\left(\mathbf{Q}_j\right)\right\|_2$.

Proof: Lemma 6.13 and $\log \Psi_{\mathbf{Q}_i}(\lambda) \leq \frac{\lambda^2\,\text{var}(\mathbf{Q}_i)}{1-b|\lambda|}$

symmetrization

With d replaced by $(d_1 + d_2)$, a problem involving general zero-mean random matrices $\mathbf{A}_i \in \mathbb{R}^{d_1 \times d_2}$ can be transformed to a symmetric version by defining the $(d_1 + d_2)$-dimensional square matrices

$$\mathbf{Q}_i := \begin{bmatrix} \mathbf{0}_{d_1 \times d_1} & \mathbf{A}_i \\ \mathbf{A}_i^{\mathrm{T}} & \mathbf{0}_{d_2 \times d_2} \end{bmatrix},$$

and imposing some form of moment generating function bound-for instance, the sub-Gaussian condition on the symmetric matrices $\mathbf{Q}_i$.

The condition of **Theorem 6.17** is the quantity

$$\sigma^2 := \max \left\{ \left\| \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{A}_i \mathbf{A}_i^{\mathrm{T}}\right] \right\|_2 , \left\| \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{A}_i^{\mathrm{T}} \mathbf{A}_i\right] \right\| \|_2 \right\}$$

Bernstein bounds with sharpened dimension dependence

(Example 6.19) Consider a sequence of independent zero-mean random matrices $\mathbf{Q}_i$ bounded as $|||\mathbf{Q}_i|||_2 \leq 1$ almost surely, and suppose that our goal is to upper bound the maximum eigenvalue $\gamma_{\max}(\mathbf{S}_n)$ of the sum $\mathbf{S}_n := \sum_{i=1}^n \mathbf{Q}_i$. Defining the function $\phi(\lambda) := e^\lambda - \lambda - 1$.For any pair $\delta > 0$, we have

$$\mathbb{P}\left[\gamma_{\max}(\mathbf{S}_n) \geq \delta\right] \leq \frac{\operatorname{tr}(\overline{\mathbf{V}})}{\|\overline{\mathbf{V}}\|_2} \inf_{\lambda>0} \left\{ \frac{e^{\phi(\lambda)\|\overline{\mathbf{V}}\|_2}}{\phi(\lambda\delta)} \right\}.$$

where $\overline{\mathbf{V}} := \sum_{i=1}^n \operatorname{var}(\mathbf{Q}_i)$.

Proof:

- $\mathbb{P}\left[\gamma_{\max}(\mathbf{S}_n) \geq \delta\right] \leq \inf_{\lambda>0} \frac{\operatorname{tr}(\mathbb{E}[\phi(\lambda\mathbf{S}_n)])}{\phi(\lambda\delta)}$.
- $\log \Psi_{\mathbf{Q}_i}(\lambda) \leq \phi(\lambda) \operatorname{var}(\mathbf{Q}_i)$
- $\operatorname{tr}\left(\mathbb{E}\left[\phi\left(\lambda\mathbf{S}_n\right)\right]\right) \leq \frac{\operatorname{tr}(\overline{\mathbf{V}})}{\|\overline{\mathbf{V}}\|_2} e^{\phi(\lambda)\|\overline{\mathbf{V}}\|_2}$

6.4.5 Consequences for covariance matrices

(Corollary 6.20) Let $x_1, \ldots, x_n$ be i.i.d. zero-mean random vectors with covariance $\mathbf{\Sigma}$ such that $\left\|x_j\right\|_2 \leq \sqrt{b}$ almost surely. Then for all $\delta > 0$, the sample covariance matrix $\widehat{\mathbf{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^{\mathrm{T}}$ satisfies

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

Proof:$\mathbf{Q}_i := x_i x_i^{\mathrm{T}} - \mathbf{\Sigma}$

- $\|\mathbf{Q}_i\|_2 \leq \|x_i\|_2^2 + \|\mathbf{\Sigma}\|_2 \leq b + \|\mathbf{\Sigma}\|_2 \leq 2b$
- $\operatorname{var}(\mathrm{Q}_i) = \mathbb{E}\left[\left(x_i x_i^{\mathrm{T}}\right)^2\right] - \mathbf{\Sigma}^2 \leq \mathbb{E}\left[\|x_i\|_2^2\, x_i x_i^{\mathrm{T}}\right] \leq b\mathbf{\Sigma}$
- **Theorem 6.17** (Bernstein bound for random matrices)

Example 6.21 (Random vectors uniform on a sphere) Suppose that the random vectors x_i are chosen uniformly from the sphere $\mathbb{S}^{d-1}(\sqrt{d})$, so that $\|x_i\|_2 = \sqrt{d}$ for all $i = 1, \ldots, n$. By construction, we have $\mathbb{E}\left[x_i x_i^{\mathbf{T}}\right] = \mathbf{\Sigma} = \mathbf{I}_d$, and hence $\|\mathbf{\Sigma}\|_2 = 1$.

$$\mathbb{P}\left[\left\|\widehat{\mathbf{\Sigma}} - \mathbf{I}_d\right\|_2 \geq \delta\right] \leq 2de^{-\frac{n\delta^2}{2d+2d\delta}}$$

for all $\delta \geq 0$. This bound implies that

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{I}_d\right\|_2 \lesssim \sqrt{\frac{d \log d}{n}} + \frac{d \log d}{n}$$

Sparsity:6.5.1

- Suppose that the covariance matrix Σ is known to be relatively sparse, but that the positions of the non-zero entries are no longer known.Its zero pattern is captured by the adjacency matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ with entries $A_{j\ell} = \mathbb{I}\left[\Sigma_{j\ell} \neq 0\right]$. This adjacency matrix defines the edge structure of an undirected graph G on the vertices $\{1, 2, \ldots, d\}$, with edge (j, ℓ) included in the graph if and only if $\Sigma_{j\ell} \neq 0$, along with the self-edges (j, j) for each of the diagonal entries.

Sparsity

- The operator norm $\|\mathbf{A}\|_2$ of the adjacency matrix provides a natural measure of sparsity.In particular, it can be verified that $\|\mathbf{A}\|_2 \leq d$, with equality holding when G is fully connected, meaning that $\mathbf{\Sigma}$ has no zero entries. More generally, we have $\|\mathbf{A}\|_2 \leq s$ whenever $\mathbf{\Sigma}$ has at most s non-zero entries per row, or equivalently when the graph G has maximum degree at most $s - 1$.

6.5.1 Unknown sparsity and thresholding

Given a parameter $\lambda > 0$, the hard-thresholding operator is given by

$$T_\lambda(u) := u\mathbb{I}[|u| > \lambda] = \begin{cases} u & \text{if } |u| > \lambda \\ 0 & \text{otherwise} \end{cases}$$

- With a minor abuse of notation, for a matrix $\mathbf{M}$, we write $T_\lambda(\mathbf{M})$ for the matrix obtained by applying the thresholding operator to each element of $\mathbf{M}$.

Thresholding-based covariance estimation

Theorem 6.23 Let $\{x_i\}_{i=1}^n$ be an i.i.d. sequence of zero-mean random vectors with covariance matrix Σ, and suppose that each component x_{ij} is sub-Gaussian with parameter at most σ. If $n > \log d$, then for any $\delta > 0$, the thresholded sample covariance matrix $T_{\lambda_n}(\widehat{\mathbf{\Sigma}})$ with $\lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$ satisfies

$$\mathbb{P}\left[\left\| T_{\lambda_n}(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma} \right\|_2 \geq 2\|\mathbf{A}\|_2 \lambda_n\right] \leq 8e^{-\frac{n}{16}\min\{\delta,\delta^2\}}.$$

Proof:

- For any choice of λ_n such that $\|\mathbf{\Sigma} - \mathbf{\Sigma}\|_{\max} \leq \lambda_n$, we have $\left\| T_{\lambda_n}(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma} \right\|_2 \leq 2\|\mathbf{A}\|_2 \lambda_n$
- $\mathbf{B} := \left| T_{\lambda_n}(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma} \right| \Rightarrow \mathbf{B} \leq 2\lambda_n \mathbf{A} \Rightarrow \|\mathbf{B}\|_2 \leq 2\lambda_n \|\mathbf{A}\|_2$

The proof of the remainder of Theorem 6.23

The error matrix $\widehat{\boldsymbol{\Delta}} := \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$.

Lemma 6.26 Under the conditions of Theorem 6.23, we have

$$\mathbb{P}\left[\|\widehat{\Delta}\|_{\max}/\sigma^2 \geq t\right] \leq 8e^{-\frac{n}{16}\min\{t,t^2\}+2\log d} \quad \text{for all } t > 0.$$

Setting $t = \lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$

Proof:

- $\frac{1}{n}\sum_{i=1}^n x_{ij}^2$ is sub-exponential so that there are universal positive constants c_1, c_2 such that

$$\mathbb{P}\left[\left|\widehat{\Delta}_{jj}\right| \geq c_1\delta\right] \leq 2e^{-c_2 n\delta^2} \quad \text{for all } \delta \in (0,1)$$

- $2\widehat{\Delta}_{j\ell} = \frac{2}{n}\sum_{i=1}^n x_{ij}x_{i\ell} - 2\Sigma_{j\ell} = \frac{1}{n}\sum_{i=1}^n (x_{ij} + x_{i\ell})^2 - (\Sigma_{jj} + \Sigma_{\ell\ell} + 2\Sigma_{j\ell}) - \widehat{\Delta}_{jj} - \widehat{\Delta}_{\ell\ell}$
- $\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n (x_{ij} + x_{i\ell})^2 - (\Sigma_{jj} + \Sigma_{\ell\ell} + 2\Sigma_{j\ell})\right| \geq c_3\delta\right] \leq 2e^{-c_2 n\delta^2}$

- **Corollary 6.24** Suppose that, in addition to the conditions of Theorem 6.23, the covariance matrix $\mathbf{\Sigma}$ has at most s non-zero entries per row. Then with $\lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$ for some $\delta > 0$, we have

$$\mathbb{P}\left[\left\|T_{\lambda_n}(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\right\|_2 \geq 2s\lambda_n\right] \leq 8e^{-\frac{n}{16}\min\{\delta,\delta^2\}}.$$

Sparsity:6.5.2

More specifically, given a parameter $q \in [0, 1]$ and a radius R_q, we impose the constraint

$$\max_{j=1,\ldots,d} \sum_{\ell=1}^{d} \left|\Sigma_{j\ell}\right|^q \leq R_q$$

In the special case $q = 0$, this constraint is equivalent to requiring that each row of $\mathbf{\Sigma}$ have at most R_0 nonzero entries. For intermediate values $q \in (0, 1]$, it allows for many non-zero entries but requires that their absolute magnitudes (if ordered from largest to smallest) drop off relatively quickly.

Theorem 6.27 (Covariance estimation under ℓ_q-sparsity)

Suppose that the covariance matrix $\boldsymbol{\Sigma}$ satisfies the ℓ_q-sparsity constraint (6.58). Then for any λ_n such that $\|\widehat{\Sigma} - \Sigma\|_{\max} \leq \lambda_n/2$, we are guaranteed that

$$\left\| T_{\lambda_n}(\widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\Sigma} \right\|_2 \leq 4 R_q {\lambda_n}^{1-q}.$$

Consequently, if the sample covariance is formed using i.i.d. samples $\{x_i\}_{i=1}^n$ that are zero-mean with sub-Gaussian parameter at most σ, then with $\lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$, we have

$$\mathbb{P}\left[\left\| T_{\lambda_n}(\widehat{\boldsymbol{\Sigma}}) - \Sigma \right\|_2 \geq 4 R_q {\lambda_n}^{1-q} \right] \leq 8 e^{-\frac{n}{16}\min\{\delta, \delta^2\}} \quad \text{for all } \delta > 0$$

Proof: Fixing an index $j \in \{1, 2, \ldots, d\}$, define the set $S_j\left(\lambda_n/2\right) = \left\{\ell \in \{1, \ldots, d\} \quad \|\Sigma_{j\ell}| > \lambda_n/2\right\}$.

- $\left\|T_{\lambda_n}(\widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\Sigma}\right\|_2 \leq \max_{j=1,\ldots,d} \sum_{\ell=1}^{d} \left|T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \Sigma_{j\ell}\right|$
- $\sum_{\ell=1}^{d} \left|T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \Sigma_{j\ell}\right| = \sum_{\ell \in S_j(\lambda_n)} \left|T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \Sigma_{j\ell}\right| + \sum_{\ell \notin S_j(\lambda_n)} \left|T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \Sigma_{j\ell}\right| \leq \left|S_j\left(\lambda_n/2\right)\right| \frac{3}{2}\lambda_n + \sum_{\ell \notin S_j(\lambda_n)} \left|\Sigma_{j\ell}\right|$
- $\sum_{\ell \notin S_j(\lambda_n/2)} \left|\Sigma_{j\ell}\right| = \frac{\lambda_n}{2} \sum_{\ell \notin S_j(\lambda_n/2)} \frac{|\Sigma_{j\ell}|}{\lambda_n/2} \overset{\text{(i)}}{\leq} \frac{\lambda_n}{2} \sum_{\ell \notin S_j(\lambda_n/2)} \left(\frac{|\Sigma_{j\ell}|}{\lambda_n/2}\right)^q \overset{\text{(ii)}}{\leq} {\lambda_n}^{1-q} R_q$
- $R_q \geq \sum_{\ell=1}^{d} \left|\Sigma_{j\ell}\right|^q \geq \left|S_j\left(\lambda_n/2\right)\right| \left(\frac{\lambda_n}{2}\right)^q$