6. Random matrices and covariance estimation

担当：みーとみ

2021 年 7 月 1 日, 7 月 14 日

Table of Contents

- Notation とこの章で使う preliminary results の説明から.

- 行列 $A \in \mathbb{R}^{n \times m}$ with $n \geq m$ に対し, (順序付き) 特異値を

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

  と書く.
- 最小・最大特異値は次のように characterize される:

$$\sigma_{\max}(A) = \max_{v \in \mathbb{S}^{m-1}} \|Av\|_2 \quad \text{and} \quad \sigma_{\min}(A) = \min_{v \in \mathbb{S}^{m-1}} \|Av\|_2, \tag{6.1}$$

  ただし $\mathbb{S}^{d-1} := \left\{v \in \mathbb{R}^d \mid \|v\|_2 = 1\right\}$ は $\mathbb{R}^d$ 上の Euclidean unit sphere.
- また次の同値性が成り立つ: $|||A|||_2 = \sigma_{\max}(A)$.

- 対称行列の集合を $\mathcal{S}^{d\times d} := \left\{Q \in \mathbb{R}^{d\times d} \mid Q = Q^{\mathrm{T}}\right\}$ とし, その半正定値行列からなる部分集合を

$$\mathcal{S}_{+}^{d\times d} := \left\{Q \in \mathcal{S}^{d\times d} \mid Q \succeq 0\right\} \tag{6.2}$$

  と書く.

- 任意の対称行列 $Q \in \mathcal{S}^{d\times d}$ は対角化可能であり, その固有値を

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

  とする.

- このとき, $Q \succeq 0 \iff \gamma_{\min}(Q) \geq 0$.

- 最小・最大固有値の “Rayleigh – Ritz variational characterization”:

$$\gamma_{\max}(Q) = \max_{v \in \mathbb{S}^{d-1}} v^{\mathrm{T}} Q v \quad \text{and} \quad \gamma_{\min}(Q) = \min_{v \in \mathbb{S}^{d-1}} v^{\mathrm{T}} Q v. \tag{6.3}$$

- 任意の対称行列 Q に対し, その ℓ_2-operator norm は,

$$|||Q|||_2 = \max\{\gamma_{\max}(Q), |\gamma_{\min}(Q)|\} = \max_{v \in \mathbb{S}^{d-1}} \left|v^{\mathrm{T}} Q v\right|. \tag{6.4}$$

- 最後に, 行列 $A \in \mathbb{R}^{n \times m}$ with $n \geq m$ に対し, m-次元対称行列 $R := A^{\mathrm{T}} A$ を考えると,

$$\gamma_j(R) = (\sigma_j(A))^2 \quad \text{for } j = 1, \ldots, m.$$

- $\{x_1, \dots, x_m\}$ は, $\mathbb{R}^d$ 上の zero-mean・covariance $\Sigma = \mathrm{cov}(x_1) \in \mathbb{S}_+^{d\times d}$ なる分布からの n 個の i.i.d. サンプルとする.
- Σ の standard estimator は, 次の *sample covariance matrix* である:

$$\widehat{\Sigma} := \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}. \tag{6.5}$$

- 各 x_i は zero-mean なので $\mathbb{E}[x_i x_i^{\mathrm{T}}] = \Sigma$ であり, $\widehat{\Sigma}$ は Σ の unbiased estimator.
- したがって $\widehat{\Sigma} - \Sigma$ は期待値ゼロとなり, その ℓ_2-operator norm によって測った error の bound を求めることがこの章の goal となる.

- (6.4) の ℓ_2-operator norm の表現より, $|||\widehat{\Sigma} - \Sigma|||_2 \le \epsilon$ は以下と同値:

$$\max_{v \in \mathbb{S}^{d-1}} \left| \frac{1}{n} \sum_{i=1}^{n} \langle x_i, v_i \rangle^2 - v^{\mathrm{T}} \Sigma v \right| \le \epsilon. \tag{6.6}$$

- つまり, $|||\widehat{\Sigma} - \Sigma|||_2$ をコントロールすることは, v で indexed された関数クラス $x \mapsto \langle x, v \rangle^2$ の uniform law of large numbers を示すことと同値になる.

- その ℓ_2-operator norm をコントロールすることは, $\widehat{\Sigma}$ の固有値の一様収束も意味する: Weyl's theorem の corollary より,

$$\max_{j=1,\ldots,d} \left|\gamma_j(\widehat{\Sigma}) - \gamma_j\left(\Sigma\right)\right| \leq |||\widehat{\Sigma} - \Sigma|||_2. \tag{6.7}$$

- また最後に, ランダム行列 $X \in \mathbb{R}^{n\times d}$ が, 第 i 行に x_i^{T} を持つものとする

$$X = \begin{pmatrix} x_1^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{pmatrix} \in \mathbb{R}^{n\times d}$$

と,

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

なので, $\widehat{\Sigma}$ の固有値は $X/\sqrt{n}$ の特異値の 2 乗となる.

- サンプル x_i は, d-次元正規分布 $\mathcal{N}(0, \Sigma)$ から i.i.d. で引かれるとする.
- このとき,

$$X = \begin{pmatrix} x_1^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{pmatrix} \in \mathbb{R}^{n \times d}$$

  は, Σ-*Gaussian ensemble* から引かれると言う.
- Sample covariance $\widehat{\Sigma} = \frac{1}{n} X^{\mathrm{T}} X$ は, a multivariate Wishart distribution に従う.

Theorem 6.1

$X \in \mathbb{R}^{n\times d}$ は Σ-Gaussian ensemble から引かれるとする. このとき, 任意の $\delta > 0$ に対し, 最大特異値 $\sigma_{\max}(X)$ は以下の upper deviation inequality を満たす:

$$\mathbb{P}\left[\frac{\sigma_{\max}(X)}{\sqrt{n}} \geq \gamma_{\max}\left(\sqrt{\Sigma}\right)(1+\delta) + \sqrt{\frac{\mathrm{tr}(\Sigma)}{n}}\right] \leq \exp\left(-\frac{n\delta^2}{2}\right). \tag{6.8}$$

さらに $n \geq d$ なら, 最小特異値 $\sigma_{\min}(X)$ は以下の lower deviation inequality を満たす:

$$\mathbb{P}\left[\frac{\sigma_{\min}(X)}{\sqrt{n}} \leq \gamma_{\min}\left(\sqrt{\Sigma}\right)(1-\delta) - \sqrt{\frac{\mathrm{tr}(\Sigma)}{n}}\right] \leq \exp\left(-\frac{n\delta^2}{2}\right). \tag{6.9}$$

Example 6.2 (Operator norm bounds for the standard Gaussian ensemble)

- $W \in \mathbb{R}^{n \times d}$ は各成分が $\mathcal{N}(0,1)$ i.i.d. で引かれる random matrix とする（$\Sigma = I_d$）.
- Thm 6.1 より, $n \geq d$ なら, 確率 $1 - 2\exp\left(-\frac{n\delta^2}{2}\right)$ 以上で

$$\frac{\sigma_{\max}(W)}{\sqrt{n}} \leq 1 + \delta + \sqrt{\frac{d}{n}} \quad \text{and} \quad \frac{\sigma_{\min}(W)}{\sqrt{n}} \geq 1 - \delta - \sqrt{\frac{d}{n}} \tag{6.10}$$

  となる.
- よって, 同じ確率で

$$\left|\left|\left|\frac{1}{n}W^{\mathrm{T}}W - I_d\right|\right|\right|_2 \leq 2\epsilon + \epsilon^2, \quad \text{where} \epsilon = \sqrt{\frac{d}{n}} + \delta. \tag{6.11}$$

- したがって, $d/n \to 0$ なら, sample covariance $\widehat{\Sigma} = \frac{1}{n}W^{\mathrm{T}}W$ は identiry matrix I_d の一致推定量となる. ♣

Example 6.3 (Gaussian covariance estimation)

- $X \in \mathbb{R}^{n\times d}$ は Σ-Gaussian ensemble からの random matrix とする.
- このとき $X = W\sqrt{\Sigma}$ と書ける（$W \in \mathbb{R}^{n\times d}$ は standard Gaussian random matrix）ので,

$$\left|\!\left|\!\left| \frac{1}{n}X^{\mathrm{T}}X - \Sigma \right|\!\right|\!\right|_2 = \left|\!\left|\!\left| \sqrt{\Sigma}\left(\frac{1}{n}W^{\mathrm{T}}W - I_d\right) \right|\!\right|\!\right|_2 \le |||\Sigma|||_2 \left|\!\left|\!\left| \frac{1}{n}W^{\mathrm{T}}W - I_d \right|\!\right|\!\right|_2 .$$

- したがって (6.11) より, 任意の $\delta > 0$ に対して確率 $1 - 2\exp\left(-\frac{n\delta^2}{2}\right)$ で

$$\frac{|||\widehat{\Sigma} - \Sigma|||_2}{|||\Sigma|||_2} \le 2\sqrt{\frac{d}{n}} + 2\delta + \left(\sqrt{\frac{d}{n}} + \delta\right)^2 . \tag{6.12}$$

- よって, $|||\widehat{\Sigma} - \Sigma|||_2 / |||\Sigma|||_2$ は $d/n \to 0$ である限り 0 に収束する. ♣

Example 6.4 (Faster rates under trace constraints)

- $\{\gamma_j(\Sigma)\}_{j=1}^d$ は Σ の固有値列で, $\gamma_1(\Sigma)$ がそのうち最大のもの.
- Σ は, 次元に対して独立な定数 C に対し, 次の “trace constraint” を満たすとする:

$$\frac{\mathrm{tr}(\Sigma)}{|||\Sigma|||_2} = \frac{\sum_{j=1}^d \gamma_j(\Sigma)}{\gamma_1(\Sigma)} \le C. \tag{6.13}$$

- C は Σ の（実質的な）rank と見なせる（$\because$ (6.13) は $C = \mathrm{rank}(\Sigma)$ では常に成立. ）
- パラメータ $q \in [0, 1]$ と半径 $R_q > 0$ の the Schatten q-“balls” を, 以下で定義する:

$$\mathbb{B}_q(R_q) := \left\{ \Sigma \in S^{d\times d} \,\middle|\, \sum_{j=1}^d |\gamma_j(\Sigma)|^q \le R_q \right\}. \tag{6.14}$$

  - $q = 0$ なら, rank R_q 以下の対称行列の集合.
  - $q = 1$ なら, trace constraint になる.
- 任意の非零行列 $\Sigma \in \mathbb{B}_q(R_q)$ は, (6.13) を $C = R_q/(\gamma_1(\Sigma))^q$ で満たす.

- (6.13) を満たす任意の Σ に対し, Thm 6.1 は高確率で X の最大特異値が次のように抑えられることを保証する:

$$\frac{\sigma_{\max}(X)}{\sqrt{n}} \leq \gamma_{\max}(\sqrt{\Sigma}) \left(1 + \delta + \sqrt{\frac{C}{n}}\right). \tag{6.15}$$

- $\Sigma = I_d$ のときの bound (6.10) と比べると, C が d に置き換わって "実行的な rank" となっている. ♣

Proof of Theorem 6.1.

- Notation: $\overline{\sigma}_{\max} = \gamma_{\max}(\sqrt{\Sigma})$, $\overline{\sigma}_{\min} = \gamma_{\min}(\sqrt{\Sigma})$.
- 最大/最小特異値の upper/lower bound ともに以下の 2 段階で示す:
    1. 高確率で特異値が期待値に近いことを concentration inequality から示す（Ch.2）
    2. その期待値の bound の導出に Gaussian comparison inequality を用いる（Ch.5）

- ここでは最大特異値の upper bound のみを示す.（最小特異値の lower bound は大体似た方針で示せるがよりテクニカルなので Appendix (Section 6.6) にまわす.）

- $X = W\sqrt{\Sigma}$ と書ける, ただし $W \in \mathbb{R}^{n\times d}$ は i.i.d. $\mathcal{N}(0,1)$ entries をもつ.
- $W \mapsto \frac{\sigma_{\max}(W\sqrt{\Sigma})}{\sqrt{n}}$ を $\mathbb{R}^{nd}$ 上の実数値写像とみると, これは $L = \overline{\sigma}_{\max}/\sqrt{n}$ で Lipschitz w.r.t. Euclidean norm.（cf. Example 2.32）
- Gaussian r.v. に対する Lipschitz 関数の concentration inequality（Thm 2.26）より,

$$\mathbb{P}\left[\sigma_{\max}(X) \geq \mathbb{E}[\sigma_{\max}(X)] + \sqrt{n}\overline{\sigma}_{\max}\delta\right] \leq \exp\left(-\frac{n\delta^2}{2}\right).$$

- したがって, あとは以下を示せれば良い:

$$\mathbb{E}[\sigma_{\max}(X)] \leq \sqrt{n}\overline{\sigma}_{\max} + \sqrt{\mathrm{tr}(\Sigma)}. \tag{6.16}$$

- $\sigma_{\max}(X) = \max_{v' \in \mathbb{S}^{d-1}} \|Xv'\|_2$ で, $X = W\sqrt{\Sigma}$, $v = \sqrt{\Sigma}v'$ とすると次のように書ける:

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

ただし $\mathbb{S}^{d-1}(\Sigma^{-1}) := \{v \in \mathbb{R}^d \mid \|\Sigma^{-\frac{1}{2}} v\|_2 = 1\}$.

- $\{Z_{u,v},\ (u,v) \in \mathbb{T}\}$ where $\mathbb{T} := \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}(\Sigma^{-1})$ は zero-mean Gaussian process とみなせる.

- 別の Gaussian process $\{Y_{u,v},\ (u,v) \in \mathbb{T}\}$ で $\mathbb{E}[(Z_{u,v} - Z_{\tilde{u}\tilde{v}})^2] \leq \mathbb{E}[(Y_{u,v} - Y_{\tilde{u}\tilde{v}})^2]$ for all $(u,v), (u',v') \in \mathbb{T}$ となるようなものを construct することを考える.
- すると Sudakov-Fernique comparison（Thm. 5.27）から以下が言える:

$$\mathbb{E}[\sigma_{\max}(X)] = \mathbb{E}\left[\max_{(u,v) \in \mathbb{T}} Z_{u,v}\right] \leq \mathbb{E}\left[\max_{(u,v) \in \mathbb{T}} Y_{u,v}\right]. \tag{6.17}$$

- $(u, v), (\tilde{u}, \tilde{v}) \in \mathbb{T}$ を given とし, $\|v\|_2 \leq \|\tilde{v}\|_2$ とする.
- まず $Z_{u,v} = u^{\mathrm{T}} W v = \langle\langle W, uv^{\mathrm{T}} \rangle\rangle$ となる, where $\langle\langle A, B \rangle\rangle := \sum_{j=1}^{n} \sum_{k=1}^{d} A_{jk} B_{jk}$.
- W は i.i.d.$\mathcal{N}(0, 1)$ entries をもつので,

$$\mathbb{E}\left[(Z_{u,v} - Z_{\tilde{u}\tilde{v}})^2\right] = \mathbb{E}\left[\langle\langle W, uv^{\mathrm{T}} - \tilde{u}\tilde{v}^{\mathrm{T}} \rangle\rangle^2\right] = |||uv^{\mathrm{T}} - \tilde{u}\tilde{v}^{\mathrm{T}}|||_F^2.$$

- Frobenius norm を変形すると,

$$\begin{aligned}
&|||uv^{\mathrm{T}} - \tilde{u}\tilde{v}^{\mathrm{T}}|||_F^2 \\
=&|||u(v - \tilde{v})^{\mathrm{T}} - (u - \tilde{u})\tilde{v}^{\mathrm{T}}|||_F^2 \\
=&|||(u - \tilde{u})\tilde{v}^{\mathrm{T}}|||_F^2 + |||u(v - \tilde{v}) - \mathrm{T}|||_F^2 + 2\langle\langle u(v - \tilde{v})^{\mathrm{T}}, (u - \tilde{u})\tilde{v}^{\mathrm{T}} \rangle\rangle \\
\leq&\|\tilde{v}\|_2^2 \|u - \tilde{u}\|_2^2 + \|u\|_2^2 \|v - \tilde{v}\|_2^2 + 2(\|u\|_2^2 - \langle u, \tilde{u} \rangle)(\langle v, \tilde{v} \rangle - \|\tilde{v}\|_2^2).
\end{aligned}$$

- ここで, $\|u\|_2 = \|\tilde{u}\|_2 = 1$ より $\|u\|_2^2 - \langle u, \tilde{u} \rangle \geq 0$.
- 一方, Cauchy-Schwarz と仮定 $\|v\|_2 \leq \|\tilde{v}\|_2$ より, $|\langle v, \tilde{v} \rangle| \leq \|v\|_2 \|\tilde{v}\|_2 \leq \|\tilde{v}\|_2^2$.
- したがって,

$$|||uv^{\mathrm{T}} - \tilde{u}\tilde{v}^{\mathrm{T}}|||_F^2 \leq \|\tilde{v}\|_2^2 \|u - \tilde{u}\|_2^2 + \|v - \tilde{v}\|_2^2.$$

- $\mathbb{S}^{d-1}(\Sigma^{-1})$ の定義より, $\|\tilde{v}\|_2 \le \overline{\sigma} = \gamma_{\max}(\sqrt{\Sigma})$ なので,

$$\mathbb{E}[(Z_{u,v} - Z_{\tilde{u},\tilde{v}})^2] \le \overline{\sigma}_{\max}^2 \|u - \tilde{u}\|_2^2 + \|v - \tilde{v}\|_2^2.$$

- Gaussian process $Y_{u,v} := \overline{\sigma}_{\max}\langle g, u\rangle + \langle h, v\rangle$ を定義する（ただし $g \in \mathbb{R}^n, h \in \mathbb{R}^d$ は standard Gaussian rv's）と,

$$E[(Y_{u,v} - Y_{\tilde{u},\tilde{v}})^2] = \overline{\sigma}_{\max}^2 \|u - \tilde{u}\|_2^2 + \|v - \tilde{v}\|_2^2.$$

- よって Sudakov-Fernique bound (6.17) より,

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

- Jensen's inequality から, $\mathbb{E}[\|g\|_2] \le \sqrt{n}$ ans $\mathbb{E}[\|\sqrt{\Sigma}h\|_2] \le \sqrt{\mathbb{E}[h^{\mathrm{T}}\Sigma h]} = \sqrt{\mathrm{tr}(\Sigma)}$ となり, (6.16) が示された. $\square$

- Random vector $x_i \in \mathbb{R}^d$ は zero-mean で, sub-Gaussian with parameter at most σ, つまり各 $v \in \mathbb{S}^{d-1}$ に対し,

$$\mathbb{E}\left[\exp\left(\lambda\langle v, x_i\rangle\right)\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \quad \text{for all } \lambda \in \mathbb{R} \tag{6.18}$$

  が成り立つとする.
  - 例 1) $x_{ij} \in \mathbb{R}$ は zero-mean, sub-Gaussian with $\sigma = 1$ ($x_{ij} \sim \mathcal{N}(0,1)$ や Rademacher variable, サポート $[-1,1]$ の分布など)
  - 例 2) $x_i \sim \mathcal{N}(0,\Sigma)$ とすると, 任意の $v \in \mathbb{S}^{d-1}$ に対し $\langle v, x_i\rangle \sim \mathcal{N}(0, v^{\mathrm{T}}\Sigma v)$ で $v^{\mathrm{T}}\Sigma v \leq |||\Sigma|||_2$ より, x_i は sub-Gaussian with parameter at most $\sigma^2 = |||\Sigma|||_2$.
- このとき, $X \in \mathbb{R}^{n\times d}$ は *row-wise σ-sub-Gaussian ensemble* からのサンプルであるという.

Theorem 6.5

ある定数 c_0, c_1, c_2, c_3 が存在して, 任意の row-wise σ-sub-Gaussian ランダム行列 $X \in \mathbb{R}^{n\times d}$ について, 標本共分散 $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}}$ は次のの bound

$$\mathbb{E}\left[\exp\left(\lambda |||\widehat{\Sigma} - \Sigma|||_2\right)\right] \leq \exp\left(c_0 \frac{\lambda^2\sigma^2}{n} + 4d\right) \quad \text{for all } |\lambda| < \frac{n}{64e^2\sigma^2} \tag{6.19a}$$

を満たし, したがって,

$$\mathbb{P}\left[\frac{|||\widehat{\Sigma} - \Sigma|||_2}{\sigma^2} \geq c_1 \left\{\sqrt{\frac{d}{n}} + \frac{d}{n}\right\} + \delta\right] \leq c_2 \exp\left(-c_3 n \min\{\delta, \delta^2\}\right) \quad \text{for all } \delta \geq 0. \tag{6.19b}$$

Remarks:

- (6.19a) を given とすると, Chernoff technique (Ch.2) からただちに (6.19b) が示される.
- $\Sigma = I_d$ で x_i が sub-Gaussian w/ $\sigma = 1$ のとき, (6.19b) は高確率で

$$|||\widehat{\Sigma} - I_d|||_2 \precsim \sqrt{\frac{d}{n}} + \frac{d}{n}$$

となることを含意する.
- $n \geq d$ のとき, これは定数 $c' > 1$ について

$$1 - c'\sqrt{\frac{d}{n}} \leq \frac{\sigma_{\min}(X)}{\sqrt{n}} \leq \frac{\sigma_{\max}(X)}{\sqrt{n}} \leq 1 + c'\sqrt{\frac{d}{n}} \tag{6.20}$$

を意味し, standard Gaussian matrix についての result (6.10) の sub-Gaussian version とみなせる.

Proof

- $Q := \widehat{\Sigma} - \Sigma$ の ℓ_2-operator norm の moment 母関数の bound を求めたい.
- まず Section 6.1 より $|||Q|||_2 = \max_{v \in \mathbb{S}^{d-1}} |\langle v, Qv \rangle|$.
- Example 5.8 より, $\mathbb{S}^{d-1}$ には $N(\leq 17^d)$ 個のベクトルからなる $\frac{1}{8}$-covering が存在し, これを $\{v^1, \ldots, v^N\}$ とかく.
- 任意の $v \in \mathbb{S}^{d-1}$ は $v = v^j + \Delta$ where $\|\Delta\|_2 \leq \frac{1}{8}$ とかけ, よって

$$\langle v, Qv \rangle = \langle v^j, Qv^j \rangle + 2\langle \Delta, Qv^j \rangle + \langle \Delta, Q\Delta \rangle.$$

- 三角不等式と operator norm の定義から

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

- $v \in \mathbb{S}$ について sup をとると，

$$|||Q|||_2 = \max_{v \in \mathbb{S}^{d-1}} |\langle v, Qv \rangle| \le 2 \max_{j=1,\ldots,N} |\langle v^j, Qv^j \rangle|.$$

- よって，

$$\mathbb{E}\left[e^{\lambda |||Q|||_2}\right] \le \mathbb{E}\left[\exp\left(2\lambda \max_{j=1,\ldots,N} |\langle v^j, Qv^j \rangle|\right)\right] \le \sum_{j=1}^{N} \left\{\mathbb{E}\left[e^{2\lambda\langle v^j, Qv^j \rangle}\right] + \mathbb{E}\left[e^{-2\lambda\langle v^j, Qv^j \rangle}\right]\right\}. \tag{6.21}$$

- ここで，任意の $u \in \mathbb{S}^{d-1}$ に対して以下が成り立つ（証明は後で）：

$$\mathbb{E}\left[e^{t\langle u, Qu \rangle}\right] \le e^{512\frac{t^2}{n} e^4 \sigma^4} \quad \text{for all } |t| \le \frac{n}{32e^2\sigma^2}. \tag{6.22}$$

- (6.21) (6.22) より，

$$\mathbb{E}\left[e^{\lambda |||Q|||_2}\right] \le 2N e^{2048\frac{\lambda^2}{n} e^4 \sigma^4} \le \exp\left(c_0 \frac{\lambda^2 \sigma^4}{n} + 4d\right) \quad \text{for all } |\lambda| < \frac{n}{64e^2\sigma^2}$$

となり（2 つ目の不等号は $2 \cdot 17^d \le e^{4d}$ より），(6.19a) が示された．

Proof of the bound (6.22)

- $Q = \widehat{\Sigma} - \Sigma$ の定義と i.i.d. の仮定より,

$$\mathbb{E}\left[e^{t\langle u,Qu\rangle}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{\frac{t}{n}\{\langle x_i,u\rangle^2 - \langle u,\Sigma u\rangle\}}\right] = \left(\mathbb{E}\left[e^{\frac{t}{n}\{\langle x_1,u\rangle^2 - \langle u,\Sigma u\rangle\}}\right]\right)^n . \tag{6.23}$$

- $\epsilon \in \{-1, 1\}$ を Rademacher 変数とすると, symmetrization argument (Prop.4.11) より

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

$$\overset{\text{(ii)}}{=} 1 + \sum_{\ell=1}^{\infty} \frac{1}{(2\ell)!}\left(\frac{2t}{n}\right)^{2\ell} \mathbb{E}\left[\langle x_1, u\rangle^{4\ell}\right]$$

となる, ただし (i) は指数関数の冪乗展開, (ii) は奇数次項は Rademacher term が 0 になることより.

・ Thm.2.6 の sub-Gaussian の同値条件より,

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

が成り立つので,

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

となる, ただし最後の不等号は $(4\ell)! \le 2^{2\ell}[(2\ell)!]^2$ より.

- $f(t) = \frac{16t}{n} e^2 \sigma^2 < \frac{1}{2}$ なら

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

となる（(i) は $1/(1-a) \leq e^{2a}$ for all $a \in [0, 1/2]$ より）ので, (6.23) と合わせて $|t| < \frac{n}{32 e^2 \sigma^2}$ に対して $\mathbb{E}[e^{t\langle u, Qu \rangle}] \leq e^{2nf^2(t)}$, つまり (6.22) が示された. $\square$

- より一般的な条件下での bound を求める.
- そのために covariance matrices だけでなくより general な random matrices を考える.
- Main result の Theorem 6.15 と 6.17 は Hoeffding・Bernstein bounds の matrix-based analogs である.

- 対称行列 $Q \in \mathcal{S}^{d\times d}$ の対角化 $Q = U^{\mathrm{T}}\Gamma U$ を考える:
  - $U \in \mathbb{R}^{d\times d}$ はユニタリ行列 $U^{\mathrm{T}}U = I_d$.
  - $\Gamma := \mathrm{diag}(\gamma(Q))$ は固有値 $\gamma(Q) \in \mathbb{R}^d$ からなる対角行列.
- 関数 $f : \mathbb{R} \to \mathbb{R}$ を $\mathcal{S}^{d\times d}$ 上の関数に以下のように拡張する:

$$Q \mapsto f(Q) := U^{\mathrm{T}}\mathrm{diag}(f(\gamma_1(Q)), \dots, f(\gamma_d(Q)))U.$$

  - このとき, f はユニタリ不変, つまり

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

  - また, 固有値は次のように変換される (*spectral mapping property*):

$$\gamma(f(Q)) = \{f(\gamma_j(Q)),\ j = 1, \dots, d\}.$$

- 特に matrix exponential と matrix logarithm の 2 つの関数がこの章では重要.
- Matrix exponential:
  - Power-series expansion が成立: $e^Q = \sum_{k=0}^{\infty} \frac{Q^k}{k!}$.
  - Spectral mapping property より, e^Q の固有値は常に正, よって e^Q は常に正定値.
- Matrix logarithm は matrix exponential の inverse.

- 関数 f が単調であるとは, $Q \preceq R$ ならば $f(Q) \preceq f(R)$ が成り立つことをいう.
- Matrix logarithm は単調である（Lowner-Heinz theorem）
- Matrix exponential は単調でない.

- $f : \mathbb{R} \to \mathbb{R}$ が連続かつ非減少なら, 任意の対称行列 $Q \preceq R$ に対し,

$$\mathrm{tr}(f(Q)) \leq \mathrm{tr}(f(R)) \tag{6,25}$$

  が成り立つ（*trace inequality*）.

- 対称ランダム行列 $Q \in \mathcal{S}^{d\times d}$ に対し, polynomial moment $\mathbb{E}[Q^j]$ が存在すると仮定する.
- Q の variance は $\mathrm{var}(Q) := \mathbb{E}[Q^2] - (\mathbb{E}[Q])^2$ で, これは半正定値（Ex.6.6）
- Q の moment generating function $\Psi_Q : \mathbb{R} \to \mathcal{S}^{d\times d}$ は以下で与えられる:

$$\Psi_Q(\lambda) := \mathbb{E}[e^{\lambda Q}] = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathbb{E}[Q^k]. \tag{6.26}$$

- Ch.2 の議論と同様に, この moment generating function を用いて random matrix の sub-Gaussian と sub-exponential が定義される.

Definition 6.6

Zero-mean 対称ランダム行列 $Q \in \mathcal{S}^{d\times d}$ が sub-Gaussian with matrix parameter $V \in \mathcal{S}_+^{d\times d}$ であるとは, 以下が成り立つことをいう:

$$\Psi_Q(\lambda) \preceq e^{\frac{\lambda^2 V}{2}} \quad \text{for all } \lambda \in \mathbb{R}. \tag{6.27}$$

Example 6.7

- $Q = \epsilon B$ で, $\epsilon \in \{-1, 1\}$ は Rademacher 変数, $B \in \mathcal{S}^{d\times d}$ は fixed matrix とする.
- このとき, $\mathbb{E}[Q^{2k+1}] = 0$ かつ $\mathbb{E}[Q^{2k}] = B^{2k}$ なので,

$$\mathbb{E}[e^{\lambda Q}] = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} B^{2k} \preceq \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\lambda^2 B^2}{2}\right)^k = e^{\frac{\lambda^2 B^2}{2}}$$

  となり, Q は sub-Gaussian w/ $V = B^2 = \mathrm{var}(Q)$.
- より一般に, $Q = gB$ で $g \in \mathbb{R}$ が zero-mean σ-sub-Gaussian なら, Q は $V = \sigma^2 B^2$ で sub-Gaussian となる.

Example 6.8

- $Q = \epsilon C$ で, ϵ は Rademacher 変数, C は ϵ と独立で $|||C|||_2 \le b$ なるランダム行列とする.
- まず C を固定して ϵ について期待値をとると $\mathbb{E}_\epsilon[e^{\lambda \epsilon C}] \preceq e^{\frac{\lambda^2}{2} C^2}$.
- さらに $|||C|||_2 \le b$ より $e^{\frac{\lambda^2}{2} C^2} \preceq e^{\frac{\lambda^2}{2} b^2 I_d}$ となり, よって

$$\Psi_Q(\lambda) \preceq e^{\frac{\lambda^2}{2} b^2 I_d} \quad \text{for all } \lambda \in \mathbb{R}.$$

- したがって, Q は sub-Gaussian w/ matrix parameter $V = b^2 I_d$. ♣

Definition 6.9

Zero-mean ランダム行列 Q が sub-exponential with parameters (V, α) であるとは, 以下が成り立つことをいう:

$$\Psi_Q(\lambda) \preceq e^{\frac{\lambda^2 V}{2}} \quad \text{for all } |\lambda| < \frac{1}{\alpha}. \tag{6.28}$$

- 任意の sub-Gaussian 行列は sub-exponential w/ $(V, 0)$.
- Sub-exponential だが sub-Gaussian でないランダム行列はありうる:
    - 例）$M = \epsilon g^2 B$ where ϵ は Rademacher 変数, $g \sim \mathcal{N}(0, 1)$ でそれぞれ独立.

- Sub-exponential の 1 つの判定方法は, 次の Bernstein condition である.

Definition 6.10 (Bernstein's condition for matrices)

Zero-mean 対称ランダム行列 Q が Bernstein condition with parameter $b > 0$ を満たすとは, 以下が成り立つことをいう.

$$\mathbb{E}[Q^j] \preceq \frac{1}{2} j! b^{j-2} \mathrm{var}(Q) \quad \text{for } j = 3, 4, \ldots. \tag{6.29}$$

- Q が bounded operator norm を持つ, つまり $|||Q|||_2 \leq b$ almost surely の場合, 次が成り立つ.

$$\mathbb{E}[Q^j] \preceq b^{j-2} \mathrm{var}(Q) \quad \text{for all } j = 3, 4, \ldots. \tag{6.30}$$

- 次の Lemma は, Bernstein condition が sub-exponential condition を含意することを示す.

Lemma 6.11

Bernstein condition を満たす任意の zero-mean 対称行列に対し, 以下が成り立つ:

$$\Psi_Q(\lambda) \preceq \exp\left(\frac{\lambda^2 \mathrm{var}(Q)}{2(1-b|\lambda|)}\right) \quad \text{for all } |\lambda| < \frac{1}{b}. \tag{6.31}$$

Proof

- $\mathbb{E}[Q] = 0$ なので, matrix exponential の power-series expansion より

$$\begin{aligned}\mathbb{E}[e^{\lambda Q}] &= I_d + \frac{\lambda^2 \mathrm{var}(Q)}{2} + \sum_{j=3}^{\infty} \frac{\lambda^j \mathbb{E}[Q^j]}{j!} \\ &\overset{\text{(i)}}{\preceq} I_d + \frac{\lambda^2 \mathrm{var}(Q)}{2} \left\{ \sum_{j=0}^{\infty} |\lambda|^j b^j \right\} \\ &\overset{\text{(ii)}}{=} I_d + \frac{\lambda^2 \mathrm{var}(Q)}{2(1 - b|\lambda|)} \\ &\overset{\text{(iii)}}{\preceq} \exp\left(\frac{\lambda^2 \mathrm{var}(Q)}{2(1 - b|\lambda|)} \right),\end{aligned}$$

ただし (i) は Bernstein condition, (ii) は $|\lambda| < 1/b$ で成立, (iii) は matric inequality $I_d + A \preceq e^A$ (for any symmetric matrix A) より.

- まず Chernoff approach の matrix バージョンから.

Lemma 6.12 (Matrix Chernoff technique)

Q は zero-mean symmetric random matrix で, その moment generating function Ψ_Q は $\lambda \in (-a, a)$ の範囲で存在するものとする. このとき, 任意の $\delta > 0$ に対して以下が成り立つ:

$$\mathbb{P}\left[\gamma_{\max}(Q) \geq \delta\right] \leq \operatorname{tr}\left(\Psi_Q(\lambda)\right) e^{-\lambda\delta} \quad \text{for all } \lambda \in [0, a). \tag{6.32}$$

さらに同様に,

$$\mathbb{P}\left[\|Q\|_2 \geq \delta\right] \leq 2 \operatorname{tr}\left(\Psi_Q(\lambda)\right) e^{-\lambda\delta} \quad \text{for all } \lambda \in [0, a). \tag{6.33}$$

Proof

- 各 $\lambda \in [0, a)$ に対し, まず以下が成り立つ.

$$\mathbb{P}\left[\gamma_{\max}(Q) \geq \delta\right] = \mathbb{P}\left[e^{\gamma_{\max}(\lambda Q)} \geq e^{\lambda\delta}\right] \stackrel{\text{(i)}}{=} \mathbb{P}\left[\gamma_{\max}\left(e^{\lambda Q}\right) \geq e^{\lambda\delta}\right], \tag{6.34}$$

ただし (i) は行列関数の固有値の変換 (spectral mapping property) から.

- Marlkov's inequality より,

$$\mathbb{P}\left[\gamma_{\max}\left(e^{\lambda Q}\right) \geq e^{\lambda\delta}\right] \leq \mathbb{E}\left[\gamma_{\max}\left(e^{\lambda Q}\right)\right] e^{-\lambda\delta} \stackrel{\text{(i)}}{\leq} \mathbb{E}\left[\operatorname{tr}\left(e^{\lambda Q}\right)\right] e^{-\lambda\delta} \tag{6.35}$$

ただし (i) は $e^{\lambda Q}$ が positive definite から $\gamma_{\max}(e^{\lambda Q}) \leq \operatorname{tr}(e^{\lambda Q})$.

- Trace と $\mathbb{E}$ は交換可能なので,

$$\mathbb{E}\left[\operatorname{tr}\left(e^{\lambda Q}\right)\right] = \operatorname{tr}\left(\mathbb{E}\left[e^{\lambda Q}\right]\right) = \operatorname{tr}\left(\Psi_Q(\lambda)\right).$$

- 同じことが $\gamma(-Q) \geq \delta$, つまり $\gamma_{\min}(Q) \leq -\delta$ にも成り立ち, $|||Q|||_2 = \max\{\gamma_{\max}(Q)|\gamma_{\min}(Q)|\}$ なので, (6.33) も成り立つ. $\square$

Lemma 6.13

$Q_1, \ldots, Q_n$ は独立な対称ランダム行列で, moment generating function は $\lambda \in I$ に対し存在するものとし, $S_n := \sum_{i=1}^n Q_i$ とする. このとき以下が成り立つ.

$$\operatorname{tr}\left(\Psi_{\mathrm{S}_n}(\lambda)\right) \leq \operatorname{tr}\left(e^{\sum_{i=1}^n \log \Psi_{Q_i}(\lambda)}\right) \quad \text{for all } \lambda \in I. \tag{6.36}$$

Remark:

- Lemma 6.12 とあわせると, 独立なランダム行列の和の operator norm の tail bound が得られる, つまり,

$$\mathbb{P}\left[\left|\left|\left|\frac{1}{n}\sum_{i=1}^n \mathbf{Q}_i\right|\right|\right|_2 \geq \delta\right] \leq 2\operatorname{tr}\left(e^{\sum_{i=1}^n \log \Psi_{Q_i}(\lambda)}\right) e^{-\lambda n\delta} \quad \text{for all } \lambda \in [0, a).$$

Proof.

- Lieb(1973) より次の result を用いる: 任意の fixed matrix $H \in S^{d\times d}$ に対し, 次の関数 $f : S_+^{d\times d} \to \mathbb{R}$

$$f(A) := \mathrm{tr}(e^{H+\log(A)})$$

は concave である.
- $G(\lambda) := \mathrm{tr}(\Psi_{S_n}(\lambda))$ とかくと, trace と期待値の線形性から

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

- $H = \lambda S_{n-1},\ A = e^{\lambda Q_n}$ としたときの f の concavity と Jensen's inequality より,

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

- よって, $G(\lambda) \le \mathbb{E}_{S_{n-1}}[\mathrm{tr}(e^{\lambda S_{n-1}+\log\Psi_{Q_n}(\lambda)})]$.
- Q_{n-1} についても同様にすると, $G(\lambda) \le \mathbb{E}_{S_{n-2}}[\mathrm{tr}(e^{\lambda S_{n-2}+\log\Psi_{Q_{n-1}}(\lambda)+\log\Psi_{Q_n}(\lambda)})]$.
- これを繰り返していけばいい. $\square$

Example 6.14 (Rademacher symmetrization for random matrices)

- $\{A_i\}_{i=1}^n$ は独立な対称ランダム行列の列で, $\sum_i (A_i - \mathbb{E}[A_i])$ の最大固有値の bound を求めたいとする. まず, Markov inequality より,

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

- 最大固有値の variational representation から,

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

ただし (i) は Prop4.11(b) の symmetrization inequality w/ $\Psi(t) = e^{\lambda t}$, (ii) は spectral mapping property(6.24) より.

- さらに,

$$\mathbb{E}\left[\gamma_{\max}\left(e^{2\lambda\sum_{i=1}^{n}\varepsilon_i\mathbf{A}_i}\right)\right] \le \mathrm{tr}\left(\mathbb{E}\left[e^{2\lambda\sum_{i=1}^{n}\varepsilon_i\mathbf{A}_i}\right]\right) \le \mathrm{tr}\left(e^{\sum_{i=1}^{n}\log\Psi_{\widetilde{Q}_i}(2\lambda)}\right)$$

  ただし最後の不等号は $\widetilde{Q}_i = \epsilon_i A_i$ について Lemma 6.13 を適用.
- したがって, 2 の係数を除けば, A_i は zero-mean で distributionally symmetric around zero な行列に転換できる. ♣

Sub-Gaussian case

- Sub-Gaussian random matrix の Hoeffding-type tail bound から.

Theorem 6.15 (Hoeffding bound for random matrices)

$\{Q_i\}_{i=1}^n$ は zero-mean の独立対称ランダム行列の列で, それぞれ sub-Gaussian w/ parameters $\{V_i\}_{i=1}^n$ とする. このとき, 任意の $\delta > 0$ に対して次の upper tail bound が成り立つ:

$$\mathbb{P}\left[\left|\left|\left|\frac{1}{n}\sum_{i=1}^n \mathbf{Q}_i\right|\right|\right|_2 \geq \delta\right] \leq 2\,\mathrm{rank}\left(\sum_{i=1}^n \mathbf{V}_i\right) e^{-\frac{n\delta^2}{2\sigma^2}}, \tag{6.38}$$

ただし $\sigma^2 = |||\frac{1}{n}\sum_{i=1}^n V_i|||_2$.

Proof.

- まず $V := \sum_i V_i$ が full-rank のケースを考える.
- Sub-Gauusianity の定義と $\log$ の matrix monotonicity より,

$$\sum_{i=1}^{n} \log \Psi_{Q_i}(\lambda) \preceq \frac{\lambda^2}{2} \sum_{i=1}^{n} V_i$$

- $\exp$ 関数は increasing なので, (6.25) の trace inequality から,

$$\mathrm{tr}\left(e^{\sum_{i=1}^{n} \Psi_{Q_i}(\lambda)}\right) \leq \mathrm{tr}\left(e^{\frac{\lambda^2}{2} \sum_{i=1}^{n} V_i}\right).$$

- これと Chernoff bound (6.37) より,

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

- Fact: 任意の d-次元対称行列 R に対し, $\mathrm{tr}(e^R) \le d e^{|||R|||_2}$.
- $R = \frac{\lambda^2}{2}\sum_{i=1}^n V_i$ としてこれを使うと, $|||R|||_2 = \frac{\lambda}{2} n\sigma^2$ で,

$$\mathbb{P}\left[\left|\left|\left|\frac{1}{n}\sum_{i=1}^n Q_i\right|\right|\right|_2 \ge \delta\right] \le 2de^{\frac{\lambda^2}{2}n\sigma^2 - \lambda n\delta}.$$

- これは任意の $\lambda \ge 0$ について成り立つので, $\lambda = \delta/\sigma^2$ とすると claim を得る.

- 次に $V := \sum_i V_i$ が full-rank でなく, rank $r < d$ とする.
- V の固有値分解 $V = UDU^{\mathrm{T}}$ ($U \in \mathbb{R}^{d\times r}$ は正規直交列をもつ) を考え, $Q := \sum_{i=1}^n Q_i$ に対して r-次元行列 $\widetilde{Q} = U^{\mathrm{T}}QU$ をとると, $|||\widetilde{Q}|||_2 = |||Q|||_2$.
- $\widetilde{Q}$ に対して同様の議論を行えば, d のかわりに r として成り立つ. $\square$

- (6.38) は non-symmetric and/or non-square matrix での bound も導く.
- zero-mean random matrix $A_i \in\in \mathbb{R}^{d_1 \times d_2}$ に対し,

$$Q_i := \begin{bmatrix} 0_{d_1 \times d_2} & A_i \\ A_i^{\mathrm{T}} & 0_{d_2 \times d_1} \end{bmatrix}$$

を考えれば, 適切な条件下で対応する bound が得られる (see Exercise 6.10).

Example 6.16 (Looseness/sharpness of Theorem 6.15)

- 簡単のため, $n = d$ とする.
- 各 $i = 1, \dots, d$ について, $E_i \in S^{d\times d}$ を (i, i) だけ 1 で他は 0 である行列とする.
- $Q_i = y_i E_i$, ただし $\{y_i\}_{i=1}^n$ は i.i.d. なスカラーで 1-sub-Gaussian とする.
  (Rademacher 変数 $\epsilon_i \in \{-1, 1\}$ や standard Gaussian$N(0, 1)$ など)
- Q_i は $V_i = E_i$ で sub-Gaussian で, したがって $\sigma^2 = \|\frac{1}{d}\sum_{i=1}^d V_i\| = 1/d$.
- よって Thm6.15 より,

$$\mathbb{P}\left[\left|\left|\left|\frac{1}{d}\sum_{i=1}^{d} Q_i\right|\right|\right|_2 \geq \delta\right] \leq 2de^{-\frac{d^2\delta^2}{2}} \quad \text{for all } \delta > 0, \tag{6.40}$$

  したがって $|||\frac{1}{d}\sum_{i=1}^d Q_i|||_2 \precsim \frac{\sqrt{2\log(2d)}}{d}$ のオーダーとなる.

- 一方 y_i を Rademacher 変数 $y_i = \epsilon_i$ とすると,

$$\left|\!\left|\!\left|\sum_{i=1}^{d} Q_i\right|\!\right|\!\right|_2 = \max_{i=1,\ldots,d} \frac{|\epsilon_i|}{d} = \frac{1}{d}.$$

となり, (6.40) の bound は $\sqrt{\log d}$ のオーダーだけルーズである.

- y_i が standard Gaussian $y_i = g_i \sim N(0,1)$ なら,

$$\left|\!\left|\!\left|\sum_{i=1}^{d} Q_i\right|\!\right|\!\right|_2 = \max_{i=1,\ldots,d} \frac{|g_i|}{d} \simeq \frac{\sqrt{2\log d}}{d}$$

となり, Thm6.15 は d のオーダーに関してこれより improve できないことがわかる.

♣

Bernstein-type bounds for random matrices

- 次は Sub-exponential random matrices の Bernstein bound.

Theorem 6.17 (Bernstein bound for random matrices)

$\{Q_i\}_{i=1}^n$ は zero-mean な独立対称行列で Bernstein condition (6.29) を parameter $b > 0$ で満たすとする. このとき, 任意の $\delta \geq 0$ に対して以下が成り立つ:

$$\mathbb{P}\left[\frac{1}{n}\left|\left|\left|\sum_{i=1}^n Q_i\right|\right|\right|_2 \geq \delta\right] \leq 2\,\mathrm{rank}\left(\sum_{i=1}^n \mathrm{var}\left(Q_i\right)\right)\exp\left\{-\frac{n\delta^2}{2\left(\sigma^2 + b\delta\right)}\right\} \tag{6.42}$$

ただし, $\sigma^2 := \frac{1}{n}|||\sum_j \mathrm{var}(Q_j)|||_2$.

Proof.

- Lemma 6.13 より, $\mathrm{tr}(\Psi_{S_n}(\lambda)) \le \mathrm{tr}(e^{\sum \log \Psi_{Q_i}(\lambda)})$.
- Lemma 6.11 より, Bernstein condition から 任意の λ s.t. $|\lambda| < 1/b$ に対し $\log \Psi_{Q_i}(\lambda) \preceq \frac{\lambda^2 \mathrm{var}(Q_i)}{1-b|\lambda|}$.
- したがって,

$$\mathrm{tr}\left(\Psi_{S_n}(\lambda)\right) \le \mathrm{tr}\left(\exp\left(\frac{\lambda^2 \sum_{i=1}^n \mathrm{var}(Q_i)}{1-b|\lambda|}\right)\right) \le \mathrm{rank}\left(\sum_{i=1}^n \mathrm{var}(Q_i)\right) e^{\frac{n\lambda^2\sigma^2}{1-b|\lambda|}},$$

  ただし最後の不等号は Thm6.15 の証明と同様にして示せる.
- よって (6.37) と合わせると, 任意の $\lambda \in [0, 1/b)$ に対し,

$$\mathbb{P}\left[\left\|\frac{1}{n}\sum_{i=1}^n Q_i\right\|_2 \ge \delta\right] \le 2\mathrm{rank}\left(\sum_{i=1}^n \mathrm{var}(Q_i)\right) e^{\frac{n\sigma^2\lambda^2}{1-b|\lambda|} - \lambda n\delta}.$$

- $\lambda = \frac{\delta}{\sigma^2 + b\delta} \in (0, 1/b)$ とすると (6.42) を得る. □

Remarks

- Thm6.15 の場合と同様, non-squared matrix$\{A_i\}_{i=1}^n$ についての bound も出せる.
- たとえばもし $|||A_i|||_2 \le b$ almost surely なら, $\{Q_i\}$ defind by (6.39) は Bernstein condition w/ parameter b and quantity

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

  で Bernstein condition を満たす.
- Example 6.18 はこの transformation の具体例（省略）.

Example 6.19 (Bernstein bounds with sharpened dimension dependence)

- 独立な zero-mean random matrices Q_i で $|||Q_i|||_2 \leq 1$ almost surely とする.
- Sum $S_n := \sum_{i=1}^n Q_i$ の最大固有値 $\gamma_{\max}(S_n)$ の upper bound を求めたい.
- $\phi(\lambda) := e^\lambda - \lambda - 1$ は increasing on $\mathbb{R}_+$ で, Exercise 6.12 より

$$\mathbb{P}[\gamma_{\max}(S_n) \geq \delta] \leq \inf_{\lambda > 0} \frac{\mathrm{tr}(\mathbb{E}[\phi(\lambda S_n)])}{\phi(\lambda\delta)}. \tag{6.46}$$

- また $|||Q_i|||_2 \leq 1$ から同じ Exercise の result を使うと

$$\log \Psi_{Q_i}(\lambda) \preceq \phi(\lambda)\mathrm{var}(Q_i) \quad \text{and} \quad \mathrm{tr}(\mathbb{E}[\phi(\lambda S_n)]) \leq \frac{\mathrm{tr}(\bar{V})}{|||\bar{V}|||_2} e^{\phi(\lambda)|||\mathrm{var} V|||_2} \tag{6.47}$$

  where $\bar{V} = \sum_{i=1}^n \mathrm{var}(Q_i)$.
- (6.46) と合わせると

$$\mathbb{P}[\gamma_{\max}(S_n) \geq \delta] \leq \frac{\mathrm{tr}(\bar{V})}{|||\bar{V}|||_2} \inf_{\lambda > 0} \left\{ \frac{e^{\phi(\lambda)|||\bar{V}|||_2}}{\phi(\lambda\delta)} \right\} \tag{6.48}$$

  となり, 常に $\frac{\mathrm{tr}(\bar{V})}{|||\bar{V}|||_2} \leq \mathrm{rank}(\bar{V}) \leq d$ なので Thm6.17 の rank order より改善しうる.

♣

- Thm 6.17 から, covariance matrix の推定に有用な次の系が得られる.

Corollary 6.20

$x_1, \ldots, x_n$ は i.i.d. zero-mean random vectors で, covariance Σ, かつ $\|x_j\| \leq \sqrt{b}$ almost surely とする. このとき任意の $\delta > 0$ に対して, sample covariance $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}}$ は次を満たす:

$$\mathbb{P}\left[|||\widehat{\Sigma} - \Sigma|||_2 \geq \delta\right] \leq 2d \exp\left(-\frac{n\delta^2}{2b\left(|||\Sigma|||_2 + \delta\right)}\right). \tag{6.49}$$

Proof.

- Zero-mean random matrix $Q_i := x_i x_i^{\mathrm{T}} - \Sigma$ に対して Thm6.17 を適用する.
- 三角不等式より,

$$|||Q_i|||_2 \le \|x_i\|_2^2 + |||\Sigma|||_2 \le b + |||\Sigma|||_2.$$

- $\Sigma = \mathbb{E}[x_i x_i^{\mathrm{T}}]$ より, $|||\Sigma|||_2 = \max_{v \in \mathbb{S}^{d-1}} \mathbb{E}[\langle v, x_i \rangle^2] \le b$ で, よって $|||Q_i|||_2 \le 2b$.
- Q_i の分散については,

$$\mathrm{var}(Q_i) = \mathbb{E}[(x_i x_i^{\mathrm{T}})^2] - \Sigma^2 \preceq b\Sigma,$$

  で, よって $|||\mathrm{var}(Q_i)|||_2 \le b|||\Sigma|||_2$.
- これを (6.42) に入れると claim を得る. □

Example 6.21 (Random vectors uniform on a sphere)

- x_i は $S^{d-1}(\sqrt{d})$ から uniformly にひかれるとする（$\|x_i\| = \sqrt{d}$）.
- By construction, $\Sigma = \mathbb{E}[x_i^{\mathrm{T}} x_i] = I_d$, and hence $|||\Sigma|||_2 = 1$.
- Corollary 6.20 より

$$\mathbb{P}[|||\widehat{\Sigma} - I_d|||_2 \geq \delta] \leq 2de^{-\frac{n\delta^2}{2d+2d\delta}} \quad \text{for all } \delta \geq 0. \tag{6.50}$$

- したがって,

$$|||\widehat{\Sigma} - I_d|||_2 \preceq \sqrt{\frac{d \log d}{n}} + \frac{d \log d}{n} \tag{6.51}$$

with high probability となり, $\frac{d \log d}{n} \to 0$ であれば sample covariance は consistent.

- この結果はほぼ optimal だが, Thm6.5 を使うと $\log d$ 項は消去できる.

♣

Example 6.22 (“Spiked” random vectors)

- $x_i = \sqrt{d} e_{a(i)}$, where $a(i)$ は $\{1, \ldots, d\}$ から uniformly に決まる index.
- $\|x_i\|_2 = \sqrt{d}$ で $\mathbb{E}[x_i x_i^{\mathrm{T}}] = I_d$ なので, (6.50) の bound が同様になりたつ.
- このケースでは constant factor を除いて improve できない. ♣

- これまでは general・unstractured な設定で covariance matrix の推定を考えた.
- この章では sparse and/or graph-structured のもとではより早い収束が得られることを確認する.

- 最も簡単な設定では, covariance matrix は sparce で, その non-zero entry が分かってるとする:
- 例えば, covariance が diagonal なら, 各要素ごとの標本分散を求めて $\widehat{D} := \mathrm{diag}(\widehat{\Sigma}_{11}, \ldots, \widehat{\Sigma}_{dd})$ とするのが自然.
- このときは Exercise 6.15 より, sub-Gaussian variables なら estimation error のオーダーは $\sqrt{\frac{\log d}{n}}$ となり, unstructured setting の $\sqrt{\frac{d}{n}}$ よりよくなる.
- これに近い statement が違う形の sparsity のもとでも得られる.

6.5.1 Unknown sparsity and thresholding

- Σ は sparse であることは分かっているが, non-zero entries の position は分かっていないとする.
- パラメータ $\lambda > 0$ に対し, *hard-thresholding operator* は以下で定義される:

$$T_\lambda(u) := u\mathbb{I}[|u| > \lambda] = \begin{cases} u & \text{if } |u| > \lambda, \\ 0 & \text{otherwise.} \end{cases} \tag{6.52}$$

- 行列 M について, 各要素に $T_\lambda(\cdot)$ をかませた行列を $T_\lambda(M)$ と書くものとする.
- ここでは $T_{\lambda_n}(\widehat{\Sigma})$ の推定値を考えていく, ただし λ_n は n と d に依存して決まる.

- Σ の zero pattern は隣接行列 $A \in \mathbb{R}^{d\times d}$ with $A_{j\ell} = \mathbb{I}[\Sigma_{j\ell} \neq 0]$ で表せる.
- A は vertices が $\{1, 2, \ldots, d\}$ で edge が $\{(j, \ell) \mid \Sigma_{j\ell} \neq 0\}$ なる undirected graph G を表しているともよめる.
- $|||A|||_2$ は sparsity の measure とみることができ, $|||A|||_2 \leq d$(等号は fully conected のとき成立) である.
- より一般に, Σ が各行について最大で s の non-zero entry をもつなら $|||A|||_2 \leq s$ である.

Theorem 6.23 (Thresholding-based covariance estimation)

$\{x_i\}_{i=1}^n$ は zero-mean, covariance Σ の i.i.d. random vectors で, 各要素 x_{ij} はパラメータ最大 σ で sub-Gaussian とする. もし $n > \log d$ なら, 任意の $\delta > 0$ に対し, thresholded sample covariance matrix $T_{\lambda_n}(\widehat{\Sigma})$ w/ $\lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$ は以下を満たす:

$$\mathbb{P}\left[\left|\left|\left|T_{\lambda_n}(\widehat{\Sigma}) - \Sigma\right|\right|\right|_2 \geq 2|||\mathbf{A}|||_2\lambda_n\right] \leq 8e^{-\frac{n}{16}\min\{\delta,\delta^2\}}. \tag{6.53}$$

- 証明は次の (deterministic) result にもとづく:

$$\forall \lambda_n \geq \|\widehat{\Sigma} - \Sigma\|_{\max}, \quad |||T_{\lambda_n}(\widehat{\Sigma}) - \Sigma|||_2 \leq 2|||A|||_2 \lambda_n. \tag{6.54}$$

  - 任意の (j, ℓ) s.t. $\Sigma_{j\ell} = 0$ に対し, $\|\widehat{\Sigma} - \Sigma\|_{\max} \leq \lambda_n$ より $|\widehat{\Sigma}_{j\ell}| \leq \lambda_n$. $\therefore T_{\lambda_n}(\widehat{\Sigma}_{j\ell}) = 0$.
  - 一方任意の (j, ℓ) s.t. $\Sigma_{j\ell} \neq 0$ について,

$$\left| T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \Sigma_{j\ell} \right| \overset{\text{(i)}}{\leq} \left| T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \widehat{\Sigma}_{j\ell} \right| + \left| \widehat{\Sigma}_{j\ell} - \Sigma_{j\ell} \right| \overset{\text{(ii)}}{\leq} 2\lambda_n,$$

    ただし (i) は三角不等式, (ii) は $\left| T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \widehat{\Sigma}_{j\ell} \right| \leq \lambda_n$ と $\|\widehat{\Sigma} - \Sigma\|_{\max} \leq \lambda_n$ より.
  - よって $B := |T_{\lambda_n}(\widehat{\Sigma}) - \Sigma|$ は elementwise inequality $B \leq 2\lambda_n A$ を満たす.
  - B, A は non-negative より, $|||B|||_2 \leq 2\lambda_n |||A|||_2$ で (6.54) が示される.

Corollary 6.24

Thm6.23 の条件に加え, covariance Σ の各行は最大で s の non-zero entry を持つとする. このとき $\lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$ で,

$$\mathbb{P}[|||T_{\lambda_n}(\widehat{\Sigma}) - \Sigma \geq 2s\lambda_n] \leq 8e^{-\frac{n}{16}\min\{\delta,\delta^2\}}. \tag{6.55}$$

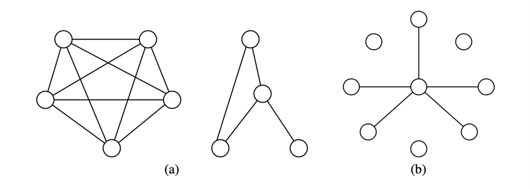
- $|||A|||_2 \leq s$ (see Exercise 6.2) より.

Example 6.25 (Sparsity and adjacency matrices)

- A が Figure 6.1(a) のような, 最大 $s-1$ degree で s-clique（s 個のノードの組でそれらすべてが違いにつながってる）をもつグラフの場合, $|||A|||_2 = s$ となり (6.53) と (6.55) は一致する.
- (b) のように 1 つのノードが s 個のノードとつながるハブのようになってる場合, $|||A|||_2 = 1 + \sqrt{s-1}$ で, Thm6.23 (6.53) より高確率で

$$|||T_{\lambda_n}(\widehat{\Sigma}) - \Sigma|||_2 \precsim \sqrt{\frac{s \log d}{n}}$$

と $\sqrt{s}$ のオーダーとなり, (6.55) より sharper になる. ♣



- Thm6.23 の証明のつづき.
- (6.54) から, $\widehat{\Delta} := \widehat{\Sigma} - \Sigma$ の infinity norm の bound を求めれば良い.

Lemma 6.26

Thm6.23 の条件のもとで, 以下が成り立つ:

$$\mathbb{P}[\|\widehat{\Delta}\|_{\max}/\sigma^2 \geq t] \leq 8e^{-\frac{n}{16}\min\{t,t^2\}+2\log d} \quad \text{for all } t > 0. \tag{6.56}$$

- (6.56) で $t = \lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$ とすると, $n > \log d$ より,

$$\mathbb{P}[\|\widehat{\Delta}\|_{\max} \geq \lambda_n] \leq 8e^{-\frac{n}{16}\min\{\delta,\delta^2\}},$$

となる.
- したがってあとは Lemma 6.26 を示せばよい.

- $\sigma = 1$ として一般性を失わない (x_i/σ は sub-Gaussian w/ at most 1 なので, あとで rescale すればよい).
- まず対角成分を考えると, Exercise 6.15(a) より定数 c_1, c_2 が存在して

$$\mathbb{P}[|\widehat{\Delta}_{jj}| \geq c_1\delta] \leq 2e^{-c_2 n\delta^2} \quad \text{for all} \delta \in (0,1). \tag{6.57}$$

- 非対角成分については次が成り立つ:

$$2\widehat{\Delta}_{j\ell} = \frac{2}{n}\sum_{i=1}^{n} x_{ij}x_{i\ell} - 2\Sigma_{j\ell} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij}+x_{i\ell})^2 - (\Sigma_{jj}+\Sigma_{\ell\ell}+2\Sigma_{j\ell}) - \widehat{\Delta}_{jj} - \widehat{\Delta}_{\ell\ell}.$$

- 各 x_{ij} は zero-mean sub-Gaussian with at most parameter σ なので, $x_{ij} + x_{i\ell}$ は zero-mean sub-Gaussian w/ parameter at most $2\sqrt{2}\sigma$ (see Exercise 2.13).
- よって, 定数 c_2, c_3 に対して, 任意の $\delta \in (0,1)$ について

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

  で, (6.57) と合わせると $\mathbb{P}[|\widehat{\Delta}_{j\ell}| \geq c_1'\delta] \leq 6e^{-c_2 n\delta^2}$ となる.
- (6.57) と合わせて d^2-entry について合わせると, claim (6.56) を得る.

- Thm 6.23 は, 厳密に 0 である entry が少ない場合は使い物にならない.
- 厳密に 0 ではなくとも多くの entry が “near zero” であるときを考える.

- Σ は, パラメータ $q \in [0,1]$ と半径 R_q に対して, 以下を満たすとする:

$$\max_{j=1,\dots,d} \sum_{\ell=1}^{d} |\Sigma_{j\ell}|^q \le R_q. \tag{6.58}$$

  - $q=0$ なら, 各行の non-zero entry が最大 R_q であることを示す.
  - Σ が (6.58) を満たすとき, ℓ_q-sparsity を満たすという.

Theorem 6.27 (Covariance estimation under ℓ_q-sparsity)

Covariance matrix Σ は ℓ_q-sparsity(6.58) を満たすとする．このとき任意の λ_n s.t. $\|\widehat{\Sigma} - \Sigma\|_{\max} \le \lambda_n/2$ に対して以下が成り立つ:

$$|||T_{\lambda_n}(\widehat{\Sigma}) - \Sigma|||_2 \le 4R_n\lambda_n^{1-q}. \tag{6.59a}$$

$\{x_i\}_{i=1}^n$ は zero-mean で sub-Gaussian w/ parameter at most σ からの i.i.d. サンプルなら, $\lambda_n/\sigma^2 = 8\sqrt{\frac{\log d}{n}} + \delta$ として以下が成り立つ:

$$\mathbb{P}\left[\left\|T_{\lambda_n}(\widehat{\Sigma}) - \Sigma\right\|_2 \ge 4R_q{\lambda_n}^{1-q}\right] \le 8e^{-\frac{n}{16}\min\{\delta,\delta^2\}} \quad \text{for all } \delta > 0 \tag{6.59b}$$

Proof.

- (6.59a) の deterministic claim を given とすると (6.59b) は sub-exponential 変数の tail bound から得られるので, (6.59a) を示す.
- Exercise 6.2 より, operator norm は次のように bound される:

$$|||T_{\lambda_n}(\widehat{\Sigma}) - \Sigma|||_2 \le \max_{j=1,\dots,d} \sum_{\ell=1}^{d} |T_{\lambda_n}(\widehat{\Sigma}_{j\ell}) - \Sigma_{j\ell}|.$$

- $j \in \{1,\dots,d\}$ を固定し, set $S_j(\lambda_n/2) := \{\ell \in \{1,\dots,d\} \mid |\Sigma_{j\ell} > \lambda_n/2|\}$ とする.
- 任意の $\ell \in S_j(\lambda_n/2)$ について,

$$\left|T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \Sigma_{j\ell}\right| \le \left|T_{\lambda_n}\left(\widehat{\Sigma}_{j\ell}\right) - \widehat{\Sigma}_{j\ell}\right| + \left|\widehat{\Sigma}_{j\ell} - \Sigma_{j\ell}\right| \le \frac{3}{2}\lambda_n.$$

- 一方 $\ell \notin S_j(\lambda_n/2)$ については, $T_{\lambda_n}(\Sigma_{j\ell}) = 0$ なので, $|T_{\lambda_n}(\widehat{\Sigma}_{j\ell}) - \Sigma_{j\ell}| = |\Sigma_{j\ell}|$.
- したがって,

$$\sum_{\ell=1}^{d} |T_{\lambda_n}(\widehat{\Sigma}_{j\ell} - \Sigma_{j\ell}| \le |S_j(\lambda_n/2)|\frac{3}{2}\lambda_n + \sum_{\ell \notin S_j(\lambda_n/2)} |\Sigma_{j\ell}|. \tag{6.60}$$

- ここで次が成り立つ:

$$\sum_{\ell \notin S_j(\lambda_n/2)} |\Sigma_{j\ell}| = \frac{\lambda_n}{2} \sum_{\ell \notin S_j(\lambda_n/2)} \frac{|\Sigma_{j\ell}|}{\lambda_n/2} \overset{\text{(i)}}{\leq} \frac{\lambda_n}{2} \sum_{\ell \notin S_j(\lambda_{n/2})} \left(\frac{|\Sigma_{j\ell}|}{\lambda_n/2}\right)^q \overset{\text{(ii)}}{\leq} \lambda_n^{1-q} R_q$$

  ただし (i) は $|\Sigma_{j\ell}| \leq \lambda_n/2$ と $q \in [0,1]$ より, (ii) は ℓ_q-sparsity から.
- 一方 $S_j(\lambda_n/2)$ と ℓ_q-sparsity から,

$$|S_j(\lambda_n/2)| \leq \left(\frac{\lambda_n}{2}\right)^{-q} R_n.$$

- よって, (6.60) は

$$\sum_{\ell=1}^{d} |T_{\lambda_n}(\widehat{\Sigma}_{j\ell} - \Sigma_{j\ell}| \leq 2^q R_q \lambda_n^{1-q} \frac{3}{2} + R_q \lambda_n^{1-q} \leq 4 R_q \lambda_n^{1-q}$$

  となる.
- これが全ての $j \in \{1, \ldots, d\}$ に対して成立するので, (6.59a) が示された. □