6. Random matrices and covariance estimation

担当：みーとみ

2021 年 6 月 30 日, 7 月 7 日

Table of Contents

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

6.1.1 Notation and basic facts

- 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,

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

- $\{x_1, \ldots, x_m\}$ 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, $|||\widehat{\Sigma} - \Sigma|||_2 \leq \epsilon$ は以下と同値:

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

- 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,

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

- 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, $\|v\|_2 \le \|\tilde{v}\|_2$ 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,

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

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

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

- $\mathbb{S}^{d-1}(\Sigma^{-1})$ 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,

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

- 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. □

- 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

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

を満たし, したがって,

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

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,

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

- よって,

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

- 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）:

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

- (6.21) (6.22) より,

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

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

- $Q = \widehat{\Sigma} - \Sigma$ 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

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

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,

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

が成り立つので,

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

となる, ただし最後の不等号は $(4\ell)! \leq 2^{2\ell}[(2\ell)!]^2$ 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 $\mathbb{E}[e^{t\langle u, Qu \rangle}] \leq e^{2n f^2(t)}$, つまり (6.22) が示された. □