

## 6. Random matrices and covariance estimation

---

担当：みーとみ

2021 年 7 月 1 日, 7 月 7 日

# Table of Contents

6.1 Some preliminaries

6.2 Wishart matrices and their behavior

6.3 Covariance matrices from sub-Gaussian ensembles

## 6.1 Some preliminaries

- 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:

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

ただし  $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d \mid \|v\|_2 = 1\}$  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  $\mathcal{S}^{d \times d} := \{Q \in \mathbb{R}^{d \times d} \mid Q = Q^T\}$  とし, その半正定値行列からなる部分集合を

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

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,  $Q \succeq 0 \Leftrightarrow \gamma_{\min}(Q) \geq 0$ .

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

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

- ・ 任意の対称行列  $Q$  に対し, その  $\ell_2$ -operator norm は,

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

- ・ 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  $R := A^T A$  を考えると,

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

## 6.1.2 Set-up of covariance estimation

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

$$\hat{\Sigma} := \frac{1}{n} \sum_{i=1}^n x_i x_i^T. \quad (6.5)$$

- 各  $x_i$  は zero-mean なので  $\mathbb{E}[x_i x_i^T] = \Sigma$  であり,  $\hat{\Sigma}$  は  $\Sigma$  の unbiased estimator.
- したがって  $\hat{\Sigma} - \Sigma$  は期待値ゼロとなり, その  $\ell_2$ -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  $\ell_2$ -operator norm の表現より,  $|||\hat{\Sigma} - \Sigma|||_2 \leq \epsilon$  は以下と同値:

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

- つまり,  $|||\hat{\Sigma} - \Sigma|||_2$  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  $\ell_2$ -operator norm をコントロールすることは,  $\hat{\Sigma}$  の固有値の一様収束も意味する: Weyl's theorem の corollary より,

$$\max_{j=1,\dots,d} \left| \gamma_j(\hat{\Sigma}) - \gamma_j(\Sigma) \right| \leq \|\hat{\Sigma} - \Sigma\|_2. \quad (6.7)$$

- ・ 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

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

と,

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

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

## 6.2 Wishart matrices and their behavior

- ・ 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,

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

は,  $\Sigma$ -Gaussian ensemble から引かれると言う.

- ・ Sample covariance  $\hat{\Sigma} = \frac{1}{n} X^T X$  は, a multivariate Wishart distribution に従う.

## Theorem 6.1

$X \in \mathbb{R}^{n \times d}$  は  $\Sigma$ -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:

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

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:

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

### 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

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

となる.

- よって, 同じ確率で

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

- したがって,  $d/n \rightarrow 0$  なら, sample covariance  $\hat{\Sigma} = \frac{1}{n} W^T W$  は identity 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  $\Sigma$ -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,

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

- 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

$$\frac{\|\hat{\Sigma} - \Sigma\|_2}{\|\Sigma\|_2} \leq 2\sqrt{\frac{d}{n}} + 2\delta + \left(\sqrt{\frac{d}{n}} + \delta\right)^2. \quad (6.12)$$

- よって,  $\|\hat{\Sigma} - \Sigma\|_2 / \|\Sigma\|_2$  は  $d/n \rightarrow 0$  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  $\Sigma$  の固有値列で,  $\gamma_1(\Sigma)$  がそのうち最大のもの.
- $\Sigma$  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:

$$\frac{\text{tr}(\Sigma)}{\|\Sigma\|_2} = \frac{\sum_{j=1}^d \gamma_j(\Sigma)}{\gamma_1(\Sigma)} \leq C. \quad (6.13)$$

- $C$  は  $\Sigma$  の (実質的な) rank と見なせる ( $\because$  (6.13) は  $C = \text{rank}(\Sigma)$  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:

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

- $q = 0$  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  $\Sigma$  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:

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

- $\Sigma = I_d$  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:  $\bar{\sigma}_{\max} = \gamma_{\max}(\sqrt{\Sigma})$ ,  $\bar{\sigma}_{\min} = \gamma_{\min}(\sqrt{\Sigma})$ .
- 最大/最小特異値の upper/lower bound とともに以下の 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  $L = \bar{\sigma}_{\max}/\sqrt{n}$  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,

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

- したがって, あとは以下を示せば良い:

$$\mathbb{E}[\sigma_{\max}(X)] \leq \sqrt{n}\bar{\sigma}_{\max} + \sqrt{\text{tr}(\Sigma)}. \quad (6.16)$$

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

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

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  $(u,v), (\tilde{u}, \tilde{v}) \in \mathbb{T}$  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:

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

- $(u, v), (\tilde{u}, \tilde{v}) \in \mathbb{T}$  を given とし,  $\|v\|_2 \leq \|\tilde{v}\|_2$  とする.
- まず  $Z_{u,v} = u^T W v = \langle \langle W, uv^T \rangle \rangle$  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,

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

- Frobenius norm を変形すると,

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

- 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,

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

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

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

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

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

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

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

- Jensen's inequality から,  $\mathbb{E}[\|g\|_2] \leq \sqrt{n}$  ans

$$\mathbb{E}[\|\sqrt{\Sigma}h\|_2] \leq \sqrt{\mathbb{E}[h^T \Sigma h]} = \sqrt{\text{tr}(\Sigma)} \text{ となり, (6.16) が示された.}$$

□

## 6.3 Covariance matrices from sub-Gaussian ensembles

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

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

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  $\langle v, x_i \rangle \sim \mathcal{N}(0, v^T \Sigma v)$  で  $v^T \Sigma v \leq |||\Sigma|||_2$  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  $\sigma$ -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  $\sigma$ -sub-Gaussian ランダム行列  $X \in \mathbb{R}^{n \times d}$  について, 標本共分散  $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$  は次の bound

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

を満たし, したがって,

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

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

$$|||\hat{\Sigma} - I_d|||_2 \lesssim \sqrt{\frac{d}{n}} + \frac{d}{n}$$

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

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

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

- $Q := \hat{\Sigma} - \Sigma$  の  $\ell_2$ -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  $\{v^1, \dots, v^N\}$  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,

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

- よって,

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

- 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) :

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

- (6.21) (6.22) より,

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

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.

## Proof of the bound (6.22)

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

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

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

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

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,

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

が成り立つので,

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

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

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

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

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  $|t| < \frac{n}{32e^2\sigma^2}$  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.  $\square$

## 6.4 Bounds for general matrices

- 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.

## 6.4.1 Background on matrix analysis

- ・ 対称行列  $Q \in \mathcal{S}^{d \times d}$  の対角化  $Q = U^T \Gamma U$  を考える:
  - ・  $U \in \mathbb{R}^{d \times d}$  はユニタリ行列  $U^T U = I_d$ .
  - ・  $\Gamma := \text{diag}(\gamma(Q))$  は固有値  $\gamma(Q) \in \mathbb{R}^d$  からなる対角行列.

- ・ 関数  $f : \mathbb{R} \rightarrow \mathbb{R}$  を  $\mathcal{S}^{d \times d}$  上の関数に以下のように拡張する:

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

- ・ このとき,  $f$  はユニタリ不変, つまり

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

- ・ 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 (Löwner-Heinz theorem)
- Matrix exponential は単調でない.
- $f: \mathbb{R} \rightarrow \mathbb{R}$  が連続かつ非減少なら, 任意の対称行列  $Q \preceq R$  に対し,

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

が成り立つ (*trace inequality*) .

## 6.4.2 Tail conditions for matrices

- ・ 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  $\text{var}(Q) := \mathbb{E}[Q^2] - (\mathbb{E}[Q])^2$  で, これは半正定値 (Ex.6.6)
- ・  $Q$  の moment generating function  $\Psi_Q : \mathbb{R} \rightarrow \mathcal{S}^{d \times d}$  は以下で与えられる:

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

- ・ 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:

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

## 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/  $V = B^2 = \text{var}(Q)$ .

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

### Example 6.8

- $Q = \epsilon C$  で,  $\epsilon$  は Rademacher 変数,  $C$  は  $\epsilon$  と独立で  $\|C\|_2 \leq b$  なるランダム行列とする.
- まず  $C$  を固定して  $\epsilon$  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  $\|C\|_2 \leq b$  より  $e^{\frac{\lambda^2}{2}C^2} \preceq e^{\frac{\lambda^2}{2}b^2I_d}$  となり, よって

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

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



### Definition 6.9

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

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

- 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  $\epsilon$  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.

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

- $Q$  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.

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

- ・ 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:

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

## Proof

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

$$\mathbb{E}[e^{\lambda Q}] = I_d + \frac{\lambda^2 \text{var}(Q)}{2} + \sum_{j=3}^{\infty} \frac{\lambda^j \mathbb{E}[Q^j]}{j!}$$

$$\stackrel{(i)}{\preceq} I_d + \frac{\lambda^2 \text{var}(Q)}{2} \left\{ \sum_{j=0}^{\infty} |\lambda|^j b^j \right\}$$

$$\stackrel{(ii)}{=} I_d + \frac{\lambda^2 \text{var}(Q)}{2(1 - b|\lambda|)}$$

$$\stackrel{(iii)}{\preceq} \exp \left( \frac{\lambda^2 \text{var}(Q)}{2(1 - b|\lambda|)} \right),$$

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

## 6.4.3 Matrix Chernoff approach and independent decompositions

- ・ 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  $\Psi_Q$  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:

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

さらに同様に,

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

## Proof

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

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

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

- Markov's inequality より,

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

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

- Trace と  $\mathbb{E}$  は交換可能なので,

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

- 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.



### Lemma 6.13

$Q_1, \dots, Q_n$  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.

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

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,

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

## 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  $f : S_+^{d \times d} \rightarrow \mathbb{R}$

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

は concave である.

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

$$G(\lambda) = \text{tr} \left( \mathbb{E} \left[ e^{\lambda S_{n-1} + \log \exp(\lambda Q_n)} \right] \right) = \mathbb{E}_{S_{n-1}} \mathbb{E}_{Q_n} \left[ \text{tr} \left( e^{\lambda S_{n-1} + \log \exp(\lambda Q_n)} \right) \right].$$

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

$$\mathbb{E}_{Q_n} \left[ \text{tr} \left( e^{\lambda S_{n-1} + \log \exp(\lambda Q_n)} \right) \right] \leq \text{tr} \left( e^{\lambda S_{n-1} + \log \mathbb{E}_{Q_n} \exp(\lambda Q_n)} \right).$$

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

### Example 6.14 (Rademacher symmetrization for random matrices)

-