Supp. 1 Random matrices and covariance estimation

28. september 2025

1 Preliminaries

Given any rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ with $n \geq m$, its ordered singular values are written as

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

and we have

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

where $\mathbb{S}^{m-1} \triangleq \{\boldsymbol{v} \in \mathbb{R}^m | \|\boldsymbol{v}\|_2 = 1\}$ is the Euclidean unit sphere in $\mathbb{R}^m$. Note that $|||\boldsymbol{A}|||_2 = \sigma_{\max}(\boldsymbol{A})$.

Denote the set of symmetric matrices in $\mathbb{R}^d$ as $\mathscr{S}^{d\times d} = \left\{\boldsymbol{Q} \in \mathbb{R}^{d\times d} | \boldsymbol{Q}^T = \boldsymbol{Q}\right\}$ and the set of positive semi-definite matrices as $\mathscr{S}_+^{d\times d} = \left\{\boldsymbol{Q} \in \mathscr{S}^{d\times d} | \boldsymbol{Q} \succeq 0\right\}$.

For any symmetric matrix $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$, its ordered eigenvalues are written as

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

The Rayleigh-Ritz variational characterization of the min and max eigenvalues is defined as

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

For any symmetric matrix $\boldsymbol{Q}$, there are two equivalent forms of ℓ_2 operator norm of $\boldsymbol{Q}$, i.e.,

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

(ℓ_2-operator norm in maximization)

For any rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ with $n \geq m$, we define $\boldsymbol{R} = \boldsymbol{A}^T\boldsymbol{A}$. The following inequalities hold

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

Theorem 1.1 (Courant-Fisher min-max theorem): *For any* $\boldsymbol{Q} \in \mathcal{S}^{d\times d}$,

$$\gamma_k = \max_{\mathtt{dim}(\mathbb{E})=k} \min_{\substack{\boldsymbol{v}\in\mathbb{E} \\ \|\boldsymbol{v}\|=1}} \boldsymbol{v}^T \boldsymbol{Q} \boldsymbol{v},$$

where $\mathbb{E}$ *is a subspace of* $\mathbb{R}^d$.

2 Covariance Estimation

Let $\{X_1, X_2, \ldots, X_n\}$ be a collection of n independent and identically distributed samples from a distribution with zero mean and covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. A unbiased estimator of $\boldsymbol{\Sigma}$ is the sample covariance matrix

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T.$$

Since $\mathbb{E}[X_i] = \boldsymbol{0}$, $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}] = \boldsymbol{\Sigma}$.

Our goal is to obtain bounds on the error $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ measured in ℓ_2 operator norm. From (ℓ_2-operator norm in maximization), we have

$$|||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}|||_2 \le \epsilon \Longleftrightarrow \max_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \left| \frac{1}{n} \sum_{i=1}^{n} \langle X_i, \boldsymbol{v}\rangle^2 - \boldsymbol{v}^T \boldsymbol{\Sigma} \boldsymbol{v} \right| \le \epsilon$$

By the Weyl's theorem, we have the following corollary.

Corollary 2.1: *For any symmetric matrices* $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{S}^{d\times d}$,

$$\max_{k=1,2,\ldots,d} |\gamma_k(\boldsymbol{A}) - \gamma_k(\boldsymbol{B})| \le |||\boldsymbol{A} - \boldsymbol{B}|||_2$$

holds.

Similar, for general rectangular matrices

$$|\sigma_k(\boldsymbol{A}) - \sigma_k(\boldsymbol{B})| \le |||\boldsymbol{A} - \boldsymbol{B}|||_2$$

holds, where $\sigma_k(\cdot)$ *is the* k*-largest singular value.*

Hence, control in the operator norm guarantees the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ are uniformly close to those of $\boldsymbol{\Sigma}$.

Let $\boldsymbol{X} \triangleq [X_1; X_2; \ldots, X_n] \in \mathbb{R}^{n\times d}$ be a random matrix and $\{\sigma_k(\boldsymbol{X})\}_{k=1}^{\min\{n,d\}}$ be the collection of singular values of $\boldsymbol{X}$, we have

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T = \frac{1}{n} \boldsymbol{X}^T \boldsymbol{X}.$$

and hence it follows that

$$\gamma_k(\widehat{\boldsymbol{\Sigma}}) = \sigma_k^2\left(\boldsymbol{X}/\sqrt{n}\right) \quad k = 1, 2, \ldots, \min\{n, d\}.$$

Definition 2.1 ($\boldsymbol{\Sigma}$-Gaussian ensemble): *Let $X_i \overset{i.i.d}{\sim} \mathcal{N}(0,\boldsymbol{\Sigma})$. We say that the random matrix $\boldsymbol{X} \triangleq [X_1; X_2; \dots, X_n] \in \mathbb{R}^{n\times d}$ is drawn from the $\boldsymbol{\Sigma}$-Gaussian ensemble and the sample covariance matrix $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$ is said to follow a multivariate Wishart distribution.*

Theorem 2.1: *Let $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ be drawn according to the $\boldsymbol{\Sigma}$-Gaussian ensemble. Then for all $\delta > 0$, the maximum singular value $\sigma_{\max}(\boldsymbol{X})$ satisfies the upper deviation inequality*

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

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

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

Example 2.1 (Operator norm bounds for the standard Gaussian ensemble): *Consider a random matrix $\boldsymbol{W} \in \mathbb{R}^{n\times d}$ drawn from $\boldsymbol{I}_d$-Gaussian ensemble, i.e., $\boldsymbol{W}$ has i.i.d $\mathcal{N}(0,1)$ entries. By specializing Theorem 2.1, we conclude that for $n \geq d$, we have*

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

held with probability greater than $1-2e^{-n\delta^2/2}$, which implies

$$\left\|\left\|\frac{1}{n}\boldsymbol{W}^T\boldsymbol{W} - \boldsymbol{I}_d\right\|\right\|_2 \leq 2\epsilon+\epsilon^2 \quad \text{where } \epsilon = \sqrt{\frac{d}{n}}+\delta,$$

with the same probability. Consequently, the sample covariance $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{W}^T\boldsymbol{W}$ is a consistent estimate of $\boldsymbol{I}_d$ as $d/n \to 0$.

Proof. Let $\boldsymbol{R} \triangleq \frac{1}{n}\boldsymbol{W}^d\boldsymbol{W}$. Since

$$\gamma_k(\boldsymbol{R}) = \sigma_k^2\left(\boldsymbol{W}/\sqrt{n}\right) \quad \text{for } k = 1,2,\dots,\min\{n,d\},$$

we have

$$\gamma_{\max}(\boldsymbol{R}) \leq \left(1+\delta+\sqrt{\frac{d}{n}}\right)^2 \quad \text{and} \quad \gamma_{\min}(\boldsymbol{R}) \geq \left(1-\delta-\sqrt{\frac{d}{n}}\right)^2$$

held with probability greater than $1-2e^{-n\delta^2/2}$.

By (ℓ_2-operator norm in maximization)

$$\|\|\boldsymbol{R}-\boldsymbol{I}_d\|\|_2 = \max\{\gamma_{\max}(\boldsymbol{R})-1, |\gamma_{\min}(\boldsymbol{R})-1|\}.$$

Moreover,

$$\gamma_{\min} - 1 \le \gamma_{\max} - 1 \le 2\left(\delta + \sqrt{\frac{d}{n}}\right) + \left(\delta + \sqrt{\frac{d}{n}}\right)^2,$$

and

$$\begin{aligned}\gamma_{\min} - 1 &\ge -2\left(\delta + \sqrt{\frac{d}{n}}\right) + \left(\delta + \sqrt{\frac{d}{n}}\right)^2 \\ &\ge -2\left(\delta + \sqrt{\frac{d}{n}}\right) - \left(\delta + \sqrt{\frac{d}{n}}\right)^2\end{aligned}$$

Putting all these inequalities together, we conclude that

$$|||\boldsymbol{R} - \boldsymbol{I}_d|||_2 = \max\{\gamma_{\max}(\boldsymbol{R}) - 1, |\gamma_{\min}(\boldsymbol{R}) - 1|\} \le 2\left(\delta + \sqrt{\frac{d}{n}}\right) + \left(\delta + \sqrt{\frac{d}{n}}\right)^2$$

holds with the probability greater than $1 - 2e^{-n\delta^2/2}$. □

Example 2.2 (Gaussian covariance estimation): *Let $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ be a random matrix from the $\boldsymbol{\Sigma}$-Gaussian ensemble. By standard properties of the multivariate Gaussian, we can write $\boldsymbol{X} = \boldsymbol{W}\sqrt{\boldsymbol{\Sigma}}$, where $\boldsymbol{W}$ is a standard Gaussian random matrix, and hence*

$$\left|\left|\left|\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{\Sigma}\right|\right|\right|_2 = \left|\left|\left|\sqrt{\boldsymbol{\Sigma}}\left(\frac{1}{n}\boldsymbol{W}^T\boldsymbol{W} - \boldsymbol{I}_d\right)\sqrt{\boldsymbol{\Sigma}}\right|\right|\right|_2 \le |||\boldsymbol{\Sigma}|||_2 \left|\left|\left|\frac{1}{n}\boldsymbol{W}^T\boldsymbol{W} - \boldsymbol{I}_d\right|\right|\right|_2.$$

Consequently, from Example 2.1 we conclude that, for $n \ge d$

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

with probability at least $1 - e^{-n\delta^2/2}$ for all $\delta > 0$, which implies the relative error $\frac{|||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}|||_2}{|||\boldsymbol{\Sigma}|||_2}$ converges to zero as long the ratio d/n converges to zero.

Example 2.3 (Faster rates under trace constraints): *Let $\{\gamma_j(\boldsymbol{\Sigma})\}_{j=1}^d$ be the ordered eigenvalues of $\boldsymbol{\Sigma}$, with $\gamma_1(\Sigma)$ being the maximum eigenvalue. Now consider a non-zero covariance matrix $\boldsymbol{\Sigma}$ that satisfies a "trace constraint" of the form*

$$\frac{\mathtt{tr}(\boldsymbol{\Sigma})}{\|\boldsymbol{\Sigma}\|_2} = \frac{\sum_{j=1}^d \gamma_j(\boldsymbol{\Sigma})}{\gamma_1(\boldsymbol{\Sigma})} \leq C,$$

where C is some constant independent of dimension. The above inequality always holds with $C = \mathtt{rank}(\boldsymbol{\Sigma})$, providing a lower bound of the matrix rank.
For any matrix be drawn according to the $\boldsymbol{\Sigma}$-Gaussian ensemble with its covariance matrix satisfied the "trace constraint", Theorem 2.1 guarantees that, with high probability, the maximum singular value is bounded above as

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

By comparison to the earlier bound for $\boldsymbol{\Sigma} = \boldsymbol{I}_d$ in Example 2.1, we know that the parameter C acts as the effective dimension.

A more general of the "trace constraint" is the Schatten q-"balls" of matrics, which take the form

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

with $q \in [0,1]$ is a given parameter, and $R_q > 0$ is the radius. These matrices exhibit a relatively fast eigendecay.

Before justifing Theorem 2.1, we present the Sudakov-Fernique inequality

Theorem 2.2 (Sudakov-Fernique Inequality): *Given a pair of zero-mean N-dimensional Gaussian vectors $(X_1,\ldots,X_N)$ and $(Y_1,\ldots,Y_N)$, suppose that*

$$\mathbb{E}[(X_i - X_i)^2] \leq \mathbb{E}[(Y_j - Y_j)^2] \quad \textit{for all } (i,j) \in [N]\times[N].$$

Then

$$\mathbb{E}[\max_{j\in[N]} X_j] \leq \mathbb{E}[\max_{j\in[N]} Y_j].$$

Q: Does Sudakov-Fernique inequality hold in infinite dimension? A:Yes!

Proof of Theorem 2.1. For brevity, Let $\bar{\sigma}$ denote $\gamma_{\max}(\sqrt{\boldsymbol{\Sigma}})$. We rewrite $\boldsymbol{X}$ as $\boldsymbol{W}\sqrt{\boldsymbol{\Sigma}}$, where the random matrix $\boldsymbol{W} \in \mathbb{R}^{n\times d}$ has $\mathcal{N}(0,1)$ entries.

Viewing the mapping $\boldsymbol{W} \mapsto \frac{\sigma_{\max}(\boldsymbol{W}\sqrt{\boldsymbol{\Sigma}})}{\sqrt{n}}$ as a function of matrices. By Corollary 2.1, we have

$$\left|\frac{\sigma_{\max}(\boldsymbol{P}\sqrt{\boldsymbol{\Sigma}})}{\sqrt{n}} - \frac{\sigma_{\max}(\boldsymbol{Q}\sqrt{\boldsymbol{\Sigma}})}{\sqrt{n}}\right| \leq \frac{1}{\sqrt{n}}\|(\boldsymbol{P}-\boldsymbol{Q})\sqrt{\boldsymbol{\Sigma}}\|_2 \leq \frac{\bar{\sigma}}{\sqrt{n}}\|\boldsymbol{P}-\boldsymbol{Q}\|_2 \leq \frac{\bar{\sigma}}{\sqrt{n}}\|\boldsymbol{P}-\boldsymbol{Q}\|_F.$$

Hence, such function is Lipschitz continuous with constant $\frac{\bar{\sigma}}{\sqrt{n}}$. By concentration measure for Lipschitz functions Gaussian random vectors, we conclude that

$$\mathbb{P}[\sigma_{\max}(\boldsymbol{X}) \geq \mathbb{E}[\sigma_{\max}(\boldsymbol{X})] + \sqrt{n}\bar{\sigma}\delta] \leq e^{-n\delta^2/2}(e^{-n^2\delta^2/2}?),$$

Therefore, it suffices to show that

$$\mathbb{E}[\sigma_{\max}(\boldsymbol{X})] \leq \sqrt{n}\bar{\sigma} + \sqrt{\mathtt{tr}(\boldsymbol{\Sigma})}. \qquad \text{(Upper Bound of Max Singular Value)}$$

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

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

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

Let $\mathbb{T} = \mathbb{S}^{n-1} \times \mathbb{S}^{d-1}(\boldsymbol{\Sigma}^{-1})$. To proof (Upper Bound of Max Singular Value), we construct another Gaussian process $\{Y_{\boldsymbol{u},\boldsymbol{v}} : (\boldsymbol{u},\boldsymbol{v}) \in \mathbb{T}\}$ such that

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

and subsequently applying the Sudakov-Fernique inequality, we have

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

By the symmetry of $\left(Z_{(\boldsymbol{u},\boldsymbol{v})} - Z_{(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})}\right)^2$ w.r.t the argument pairs $(\boldsymbol{u},\boldsymbol{v})$ and $(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})$, we assume that $\|\boldsymbol{v}\|_2 \leq \|\tilde{\boldsymbol{v}}\|_2$.

$$\begin{aligned}
\mathbb{E}\left[\left(Z_{(\boldsymbol{u},\boldsymbol{v})} - Z_{(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})}\right)^2\right] &= \mathbb{E}\left[\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{v} - \tilde{\boldsymbol{u}}^T\boldsymbol{W}\tilde{\boldsymbol{v}}\right]^2 \\
&= \mathbb{E}\left[\left(\langle\langle \boldsymbol{W}, \boldsymbol{u}\boldsymbol{v}^T\rangle\rangle - \langle\langle \boldsymbol{W}, \tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}^T\rangle\rangle\right)^2\right] \\
&= \mathbb{E}\left[\langle\langle \boldsymbol{W}, \boldsymbol{u}\boldsymbol{v}^T - \tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}^T\rangle\rangle^2\right]
\end{aligned}$$

Since the matrix $\boldsymbol{W}$ has i.i.d. $\mathcal{N}(0,1)$ entries, and by the expansion of inner produce $\langle\langle \boldsymbol{A}, \boldsymbol{B}\rangle\rangle = \sum_{i,j=1} \boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$, we show that

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

Moreover,

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

and

$$\begin{aligned}
\langle\langle (\boldsymbol{u} - \tilde{\boldsymbol{u}})\tilde{\boldsymbol{v}}^T, \boldsymbol{u}(\boldsymbol{v} - \tilde{\boldsymbol{v}})^T\rangle\rangle &= \boldsymbol{u}^T(\boldsymbol{u} - \tilde{\boldsymbol{u}})\tilde{\boldsymbol{v}}^T(\boldsymbol{v} - \tilde{\boldsymbol{v}}) \\
&= (\|\boldsymbol{u}\|_2^2 - \boldsymbol{u}^T\tilde{\boldsymbol{u}}) \cdot (\tilde{\boldsymbol{v}}^T\boldsymbol{v} - \|\tilde{\boldsymbol{v}}\|_2^2)
\end{aligned}$$

Note that $\|\boldsymbol{u}\|_2 = \|\tilde{\boldsymbol{u}}\|_2 = 1$, $\|\boldsymbol{u}\|_2^2 - \boldsymbol{u}^T\tilde{\boldsymbol{u}}^T \geq 0$. By the Cauchy-Schwartz inequality and $\|\boldsymbol{v}\|_2 \leq \|\tilde{\boldsymbol{v}}\|_2$, we obtain

$$\langle \tilde{\boldsymbol{v}}, \boldsymbol{v} \rangle \leq \|\tilde{\boldsymbol{v}}\|_2 \cdot \|\boldsymbol{v}\| \leq \|\tilde{\boldsymbol{v}}\|_2^2$$

Since $\boldsymbol{v} \in \mathbb{S}^{d-1}(\boldsymbol{\Sigma})$ and $\bar{\sigma} = \|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_2 = \max_{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{v}\|_2=1} \|\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{v}\|$, we have $\|\boldsymbol{v}\|_2 \leq \bar{\sigma}$.

Putting together, we have

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

Define the Gaussian process $Y_{\boldsymbol{u},\boldsymbol{v}} := \bar{\sigma}\langle \boldsymbol{g}, \boldsymbol{u}\rangle + \langle \boldsymbol{h}, \boldsymbol{v}\rangle$, where $\boldsymbol{g} \in \mathbb{R}^n$ and $\boldsymbol{h} \in \mathbb{R}^d$ are both standard Gaussian random vectors, and mutually independent.

Computing the expectation directly, we have

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

Recalling the Sudakov-Fernique inequality, we obtain

$$\begin{aligned}
\mathbb{E}[\sigma_{\max}(\boldsymbol{X})] = \mathbb{E}\left[\max_{(\boldsymbol{u},\boldsymbol{v})\in\mathbb{T}} Z_{\boldsymbol{u},\boldsymbol{v}}\right] &\leq \mathbb{E}\left[\max_{(\boldsymbol{u},\boldsymbol{v})\in\mathbb{T}} Y_{\boldsymbol{u},\boldsymbol{v}}\right] \\
&= \bar{\sigma}\mathbb{E}[\|\boldsymbol{g}\|_2] + \mathbb{E}[\|\sqrt{\boldsymbol{\Sigma}}\boldsymbol{h}\|_2] \\
&\overset{(i)}{\leq} \bar{\sigma}\sqrt{n} + \mathbb{E}[\|\sqrt{\boldsymbol{\Sigma}}\boldsymbol{h}\|_2] \\
&\overset{(ii)}{\leq} \bar{\sigma}\sqrt{n} + \sqrt{\mathtt{tr}(\boldsymbol{\Sigma})}
\end{aligned}$$

where both (i) and (ii) are results of the Jensen's inequality.

□

3 Covariance matrices from sub-Gaussian ensembles

Definition 3.1: *A zero-mean random vector $\boldsymbol{x} \in \mathbb{R}^d$ is said to be sub-Gaussian with parameter at most σ, if for each fixed $\boldsymbol{v} \in \mathbb{S}^{d-1}$, we have*

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

- Suppose that the matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ has i.i.d. $\mathcal{N}(0,1)$ entries. Hence each entry x_{ij} is zero-mean and sub-Gaussian with parameter $\sigma = 1$. In all of rows x_i of $\boldsymbol{X}$, for any vector $\boldsymbol{v} \in \mathbb{S}^{d-1}$, $\langle \boldsymbol{v}, \boldsymbol{x}_i\rangle \sim \mathcal{N}(0,1)$ and therefore is sub-Gaussian with parameter at most σ.
- Suppose that $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. For any $\boldsymbol{v} \in \mathbb{S}^{d-1}$, we have $\langle \boldsymbol{v}, \boldsymbol{x}\rangle \sim \mathcal{N}(0, \boldsymbol{v}^T\boldsymbol{\Sigma}\boldsymbol{v})$. By definition of operator norm for positive semi-definite matrices, we have $\boldsymbol{v}^T\boldsymbol{\Sigma}\boldsymbol{v} \leq |||\boldsymbol{\Sigma}|||$. Hence, $\boldsymbol{x}$ is sub-Gaussian with parameter at most $\sigma^2 = |||\boldsymbol{\Sigma}|||$.

Definition 3.2 (Row-wise σ-sub-Gaussian ensemble): *If a random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ is formed by drawing each row $\boldsymbol{x}_i \in \mathbb{R}^d$ in an i.i.d. manner from a σ-sub-Gaussian distribution, then we say that $\boldsymbol{X}$ is a sample from a row-wise σ-sub-Gaussian ensemble.*

Theorem 3.1: *There are universal constants $\{c_j\}_{j=0}^3$ such that, for any row-wise σ-sub-Gaussian random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$, the sample covariance $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$ satisfies the bouds*

$$\mathbb{E}\left[e^{\lambda |||\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|||_2}\right] \le e^{c_0 \frac{\lambda^2\sigma^4}{n}+4d} \quad \textit{for all } |\lambda| < \frac{n}{64e^2\sigma^2},$$

and hence

$$\mathbb{P}\left[\frac{|||\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|||_2}{\sigma^2} \ge c_1\left\{\sqrt{\frac{d}{n}}+\frac{d}{n}\right\}+\delta\right] \le c_2 e^{c_3 n \min\{\delta,\delta^2\}} \quad \textit{for all } \delta \ge 0.$$

- **Suppose $\boldsymbol{\Sigma} = \boldsymbol{I}_d$ and each $\boldsymbol{x}_i$ is sub-Gaussian with parameter $\sigma = 1$, the second tai bound in Theorem 3.1 implies that**

  $$|||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I}_d|||_2 \lesssim \sqrt{\frac{d}{n}} + \frac{d}{n}$$

  with high probability.

  Since

  $$|||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I}_d|||_2 = \max\{\gamma_{\max}(\widehat{\boldsymbol{\Sigma}})-1, |\gamma_{\min}(\widehat{\boldsymbol{\Sigma}})-1|\}$$

  and note that $\gamma_k(\widehat{\boldsymbol{\Sigma}}) = \sigma_k^2(\boldsymbol{X}/\sqrt{n})$ for $k = 1,2,\ldots,\min\{n,d\}$, we have

  $$1-\sqrt{\frac{d}{n}}-\frac{d}{n} \le \sigma_{\min}^2(\boldsymbol{X}/\sqrt{n}) \le \sigma_{\max}^2(\boldsymbol{X}/\sqrt{n}) \le 1+\sqrt{\frac{d}{n}}+\frac{d}{n}.$$

  If $n \ge c'^2 d$ for some $c' > 1$, we can conclude that

  $$1-c'\sqrt{\frac{d}{n}} \lesssim \frac{\sigma_{\min}(\boldsymbol{X})}{\sqrt{n}} \lesssim \frac{\sigma_{\max}(\boldsymbol{X})}{\sqrt{n}} \lesssim 1+c'\sqrt{\frac{d}{n}}.$$

  This result is similar to Example 2.1, except that the constant c' is larger than one.

Proof of Theorem 3.1. For brevity, let $\boldsymbol{Q} := \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ and note that $|||\boldsymbol{Q}|||_2 = \max_{\boldsymbol{v}\in\mathbb{S}^{d-1}} |\langle \boldsymbol{v}, \boldsymbol{Q}\boldsymbol{v}\rangle|$. There exists a collection of vectors $\mathscr{C} \triangleq \{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_N\}$ as an $\frac{1}{8}$-cover of $\mathbb{S}^{d-1}$ in the Euclidean norm, where the size of the collection $N \le 17^d$. For any $\boldsymbol{v} \in \mathbb{S}^{d-1}$, there exists $\boldsymbol{v}_i \in \mathscr{C}$ such that $\|\boldsymbol{v}-\boldsymbol{v}_i\|_2 \le \frac{1}{8}$, which implies $\boldsymbol{v}$ can be represented as $\boldsymbol{v} = \boldsymbol{v}_i + \Delta$ with $\|\Delta\|_2 \le \frac{1}{8}$.

Hence

$$\langle \boldsymbol{v}, \boldsymbol{Q}\boldsymbol{v}\rangle = \langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle + 2\langle \boldsymbol{v}_i, \boldsymbol{Q}\Delta\rangle + \langle \Delta, \boldsymbol{Q}\Delta\rangle.$$

From the triangle inequality, we get

$$\begin{aligned}
|\langle \boldsymbol{v}, \boldsymbol{Q}\boldsymbol{v}\rangle| &= |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle + 2\langle \boldsymbol{v}_i, \boldsymbol{Q}\Delta\rangle + \langle \Delta, \boldsymbol{Q}\Delta\rangle| \\
&\le |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle| + 2|\langle \boldsymbol{v}_i, \boldsymbol{Q}\Delta\rangle| + |\langle \Delta, \boldsymbol{Q}\Delta\rangle| \\
&\le |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle| + 2|||\boldsymbol{Q}|||_2\|\Delta\|_2 + \|\Delta\|_2^2|||\boldsymbol{Q}|||_2 \\
&\le |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle| + 2\cdot\frac{1}{8}|||\boldsymbol{Q}|||_2 + \left(\frac{1}{8}\right)^2|||\boldsymbol{Q}|||_2 \\
&\le |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle| + \frac{1}{2}|||\boldsymbol{Q}|||_2
\end{aligned}$$

We conclude that, for any $\boldsymbol{v} \in \mathbb{S}^{d-1}$,

$$|||\boldsymbol{Q}|||_2 \triangleq \max_{\boldsymbol{v}\in\mathbb{S}^{d-1}} |\langle \boldsymbol{v}, \boldsymbol{Q}\boldsymbol{v}\rangle| \leq 2\max_{\boldsymbol{v}_i\in\mathscr{C}} |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle|.$$

And therefore, for $\lambda \geq 0$,

$$\mathbb{E}\left[e^{\lambda|||\boldsymbol{Q}|||_2}\right] \leq \mathbb{E}\left[e^{2\lambda \max_{\boldsymbol{v}_i\in\mathscr{C}} |\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle|}\right] \leq \sum_{i=1}^{N}\left(\mathbb{E}\left[e^{2\lambda\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle}\right] + \mathbb{E}\left[e^{-2\lambda\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle}\right]\right).$$

By the symmetry of λ in the last inequality, we have

$$\mathbb{E}\left[e^{\lambda|||\boldsymbol{Q}|||_2}\right] \leq \sum_{i=1}^{N}\left(\mathbb{E}\left[e^{2\lambda\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle}\right] + \mathbb{E}\left[e^{-2\lambda\langle \boldsymbol{v}_i, \boldsymbol{Q}\boldsymbol{v}_i\rangle}\right]\right) \quad \text{for all } \lambda \in \mathbb{R}.$$

We claim that for any fixed $\boldsymbol{u} \in \mathbb{S}^{d-1}$,

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

Applying the inequality (♠) twice with $t = 2\lambda$ and $t = -2\lambda$, respectively, we obtain

$$\mathbb{E}\left[e^{\lambda|||\boldsymbol{Q}|||_2}\right] \leq 2Ne^{2048\frac{\lambda^2}{n}e^4\sigma^4} \quad \text{for all } |\lambda| \leq \frac{n}{64e^2\sigma^2}.$$

Since $2N \leq 2(17)^d \leq e^{4d}$,

$$\mathbb{E}\left[e^{\lambda|||\boldsymbol{Q}|||_2}\right] \leq e^{2048\frac{\lambda^2}{n}e^4\sigma^4+4d} \quad \text{for all } |\lambda| \leq \frac{n}{64e^2\sigma^2}.$$

Let $c_0 = 2048e^4$, we prove the first inequality in the theorem. The second inequality is a result of the Chernoff trick.

The remaining detail is to prove the bound (♠).

By the definition of $\boldsymbol{Q}$ and the i.i.d

$$\begin{aligned}
\mathbb{E}\left[e^{t\langle \boldsymbol{u}, \boldsymbol{Q}\boldsymbol{u}\rangle}\right] &= \mathbb{E}\left[e^{\frac{t}{n}\sum_{i=1}^{n}(\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle^2 - \langle \boldsymbol{u}, \boldsymbol{\Sigma}\boldsymbol{u}\rangle)}\right] \\
&= \prod_{i=1}^{n}\mathbb{E}\left[e^{\frac{t}{n}\{\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle^2 - \langle \boldsymbol{u}, \boldsymbol{\Sigma}\boldsymbol{u}\rangle\}}\right] \\
&= \left(\mathbb{E}\left[e^{\frac{t}{n}\{\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2 - \langle \boldsymbol{u}, \boldsymbol{\Sigma}\boldsymbol{u}\rangle\}}\right]\right)^n
\end{aligned}$$

Letting $\epsilon \in \{-1, +1\}$ denote a Rademacher variable, independent of x_1. By the symmetrization technique, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{t}{n}\{\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2 - \langle \boldsymbol{u}, \boldsymbol{\Sigma}\boldsymbol{u}\rangle\}}\right] &= \mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{t}{n}\mathbb{E}_{\boldsymbol{x}'}\{\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2 - \langle \boldsymbol{x}', \boldsymbol{u}\rangle^2\}}\right] \\
&\leq \mathbb{E}_{\boldsymbol{x}_1, \boldsymbol{x}'}\left[e^{\frac{t}{n}(\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2 - \langle \boldsymbol{x}', \boldsymbol{u}\rangle^2)}\right] \\
&= \mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{t}{n}\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2}\right] \cdot \mathbb{E}_{\boldsymbol{x}_1}\left[e^{-\frac{t}{n}\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2}\right] \\
&\leq \sqrt{\mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{2t}{n}\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2}\right]} \cdot \sqrt{\mathbb{E}_{\boldsymbol{x}_1}\left[e^{-\frac{2t}{n}\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2}\right]} \\
&\leq \frac{\mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{2t}{n}\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2}\right] + \mathbb{E}_{\boldsymbol{x}_1}\left[e^{-\frac{2t}{n}\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle^2}\right]}{2}.
\end{aligned}$$

By the Taylor series of e^x, we show that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{t}{n}\{\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2-\langle \boldsymbol{u},\boldsymbol{\Sigma}\boldsymbol{u}\rangle\}}\right] &\leq \frac{\mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{2t}{n}\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2}\right]+\mathbb{E}_{\boldsymbol{x}_1}\left[e^{-\frac{2t}{n}\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2}\right]}{2} \\
&\leq \frac{1}{2}e^{\frac{2t}{n}\mathbb{E}\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2}+\frac{1}{2}e^{-\frac{2t}{n}\mathbb{E}\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2} \\
&= \frac{1}{2}\left(\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{2t}{n}\right)^k\left[\mathbb{E}\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2\right]^k+\sum_{k=0}^{\infty}\frac{1}{k!}\left(-\frac{2t}{n}\right)^k\left[\mathbb{E}\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2\right]^k\right) \\
&= 1+\sum_{\ell=1}^{\infty}\frac{1}{(2\ell)!}\left(\frac{2t}{n}\right)^{2\ell}\mathbb{E}[\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2]^{2\ell} \\
&\leq 1+\sum_{\ell=1}^{\infty}\frac{1}{(2\ell)!}\left(\frac{2t}{n}\right)^{2\ell}\mathbb{E}[\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^{4\ell}]
\end{aligned}$$

Using the equivalent characterization of sub-Gaussian variables, we have

$$\mathbb{E}[\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^{4\ell}] \leq \frac{(4\ell)!}{2^{2\ell}(2\ell)!}\left(\sqrt{8}e\sigma\right)^{4\ell} \qquad \ell=1,2,\cdots.$$

Hence

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_1}\left[e^{\frac{t}{n}\{\langle \boldsymbol{x}_1,\boldsymbol{u}\rangle^2-\langle \boldsymbol{u},\boldsymbol{\Sigma}\boldsymbol{u}\rangle\}}\right] &\leq 1+\sum_{\ell=1}^{\infty}\frac{1}{(2\ell)!}\left(\frac{2t}{n}\right)^{2\ell}\frac{(4\ell)!}{2^{2\ell}(2\ell)!}\left(\sqrt{8}e\sigma\right)^{4\ell} \\
&\leq 1+\sum_{\ell=1}^{\infty}\left(\underbrace{\frac{16t}{n}e^2\sigma^2}_{f(t)}\right)^{2\ell}
\end{aligned}$$

by the fact that $(4\ell)! \leq 2^{2\ell}[(2\ell)!]^2 (\ell=1,2,\cdots)$.

By the Taylor series of $\frac{1}{1-x}$ and the inequality $\frac{1}{1-a} \leq e^{2a}$ for $a \in [0,\frac{1}{2}]$, we have

$$1+\sum_{\ell=1}^{\infty}\frac{16t}{n}e^2\sigma^2 = \frac{1}{1-f^2(t)} \leq e^{2f^2(t)} \quad \text{if } \left|\frac{16t}{n}e^2\sigma^2\right| < \frac{\sqrt{2}}{2}.$$

Putting together, we conclude that

$$\mathbb{E}\left[e^{t\langle \boldsymbol{u},\boldsymbol{Q}\boldsymbol{u}\rangle}\right] \leq 2e^{nf^2(t)} \quad \text{if } |t| < \frac{\sqrt{2}n}{32e^2\sigma^2}.$$

which of course holds when $|t| \leq \frac{n}{32e^2\sigma^2}$. □