Wishart matrices and their behaviour

Kaveh S. Nobari

Lectures in High-Dimensional Statistics

Department of Mathematics and Statistics
Lancaster University

Contents

Motivation

The issue of covariance estimation is intertwined with random matrix theory, since sample covariance is a particular type of random matrix. These slides follow the structure of chapter 6 of Wainwright (2019) to shed light on random matrices in a non-asymptotic setting, with the aim of obtaining explicit deviation inequalities that hold for all sample sizes and matrix dimensions.

In the classical framework of covariance matrix estimation, the sample size n tends to infinity while the matrix dimension d is fixed; in this setting the behavior of sample covariance matrix is characterized by the usual limit theory. In contrast, in high-dimensional settings the data dimension is either comparable to the sample size ($d \asymp n$) or possibly much larger than the sample size $d \gg n$.

We begin with the simplest case, namely ensembles of Gaussian random matrices, and we then discuss more general sub-Gaussian ensembles, before moving to milder tail conditions.

First, let us consider rectangular matrices, for instance matrix $A \in \mathbb{R}^{n\times m}$ with $n \geq m$, the ordered singular values are written as follows

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

The maximum and minimum singular values are obtained by maximizing the "blow-up factor"

$$\sigma_{\max}(A) = \max_{\forall x} \frac{\|Ax\|_2}{\|x\|_2}, \quad \sigma_{\min}(A) = \min_{\forall x} \frac{\|Ax\|_2}{\|x\|_2}$$

which are obtained when x is the largest and smallest singular vectors respectively - i.e.

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

noting that $\|v\|_2 = 1$, since $\mathbb{S}^{m-1} := \{v \in \mathbb{R}^m \mid \|v\|_2 = 1\}$ is the Euclidean unit sphere in $\mathbb{R}^m$. We may denote

$$|||A|||_2 = \sigma_{\max}(A)$$

However, covariance matrices are square symmetric matrices, thus we must also focus on symmetric matrices in $\mathbb{R}^d$, denoted $\mathbb{S}^{d\times d} := \{Q \in \mathbb{R}^{d\times d} \mid Q = Q'\}$, as well as subset of semi-definite matrices given by

$$\mathbb{S}_+^{d\times d} := \{Q \in \mathbb{S}^{d\times d} \mid Q \geq 0\}.$$

Any matrix $Q \in \mathbb{S}^{d\times d}$ is diagonalizable via unitary transformation, and let us denote the vector of eigenvalues of Q by $\gamma(Q) \in \mathbb{R}^d$ ordered as

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

Note the matrix Q is positive semi-definite, which may be expressed as $Q \geq 0$, iff $\gamma_{\min}(Q) \geq 0$.
The Rayleigh-Ritz variational characterization of the minimum and maximum eigenvalues

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

For symmetric matrix Q, the l_2 norm can be expressed as

$$|||Q|||_2 = \max\{\gamma_{\max}(Q), |\gamma_{\min}(Q)|\} := \max_{v\in\mathbb{S}^{d-1}} |v'Qv|$$

Finally, suppose we have a rectangular matrix $A \in \mathbb{R}^{n\times m}$, with $n \geq m$. We know that any rectangular matrix can be expressed using singular value decomposition (SVD hereafter), as follows

$$A = U\Sigma V'$$

where U is an $n \times n$ unitary matrix, Σ is an $n \times m$ rectangular diagonal matrix with non-negative real numbers on the diagonal up and V is an $n \times n$ unitary matrix. Using SVD, we can express $A'A$ where

$$A'A = V\Sigma'U'U\Sigma V'$$

and since U is an orthogonal matrix, we know that $U'U = I$ where I is the identity matrix.

$$A'A = V(\Sigma'\Sigma)V'$$

Therefore, as the diagonal matrix Σ contains the eigenvalues of matrix A, hence, $\Sigma'\Sigma$ contains the eigenvalues of $A'A$ and it can be thus concluded

$$\gamma_j(A'A) = (\sigma_j(A))^2, \quad j = 1, \cdots, m$$

Let $\{x_1, \cdots, x_n\}$ be a collection of n i.i.d samples from a distribution in $\mathbb{R}^d$ with zero mean and the covariance matrix Σ. A standard estimator of sample covariance matrix is

$$\hat{\Sigma} := \frac{1}{n}\sum_{i=1}^{n} x_i x_i'.$$

Since, each x_i for $i = 1, \cdots, n$ has zero mean, it is guaranteed that

$$\mathbb{E}[x_i x_i'] = \Sigma$$

and the random matrix $\hat{\Sigma}$ is an unbiased estimator of the population covariance Σ. Consequently the error matrix $\hat{\Sigma} - \Sigma$ has mean zero, and goal is to obtain bounds on the error measures in l_2-norm. We are essentially seeking a band of the form

$$\left|\left|\left|\hat{\Sigma} - \Sigma\right|\right|\right|_2 \leq \varepsilon,$$

where,

$$\begin{aligned}
\left|\!\left|\!\left|\hat{\Sigma} - \Sigma\right|\!\right|\!\right|_2 &= \max_{v \in \mathbb{S}^{d-1}} \left| v' \left\{ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' - \Sigma \right\} v \right| \\
&= \max_{v \in \mathbb{S}^{d-1}} \left| \frac{1}{n} \sum_{i=1}^{n} v' x_i x_i' v - v' \Sigma v \right| \\
&= \max_{v \in \mathbb{S}^{d-1}} \left| \frac{1}{n} \sum_{i=1}^{n} \langle x_i, v_i \rangle^2 - v' \Sigma v \right| \leq \varepsilon
\end{aligned}$$

which suggests that controlling the deviation $\left|\!\left|\!\left|\hat{\Sigma} - \Sigma\right|\!\right|\!\right|_2$ is equivalent to establishing a ULLN for the class of functions $x \to \langle x, v \rangle^2$, indexed by vectors $v \in \mathbb{S}^{d-1}$.

Definition (Weyl's Inequality)

(I) Given any real symmetric matrices A, B,

$$\gamma_1(A+B) \geq \gamma_1(A) + \gamma_1(B)$$
$$\gamma_n(A+B) \leq \gamma_n(A) + \gamma_n(B)$$

(II) Given any real symmetric matrices A, B,

$$|\gamma_k(A) - \gamma_k(B)| \leq |||(A-B)|||_2.$$

[see DasGupta (2008)]

,

Control in the operator norm further guarantees that the eigenvalues of $\hat{\Sigma}$ are uniformly close to those of Σ. Furthermore, given Weyl's inequality II above, we have

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

Note that the random matrix $X \in \mathbb{R}^{n \times d}$ has the vectors x_i' on its i^{th} row and singular values denotes by $\{\sigma_j(X)\}_{j=1}^{\min n,d}$. Thus,

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

and hence, the eigenvalues of $\hat{\Sigma}$ are the squares of the singular values of $X/\sqrt{n}$.

Definition (Gaussian ensembles and Wishart distribution)

Suppose that each sample x_i of a matrix $X \in \mathbb{R}^{n \times d}$ is drawn from an i.i.d multivariate $N(0, \Sigma)$ distribution. In this case we say that the associated matrix $X \in n \times d$, with x_i' and its i^{th} row, is drawn from the Σ-Gaussian ensemble. The associated sample covariance $\hat{\Sigma} = \frac{1}{n} X'X$ is said to follow a multivariate Wishart distribution.

Following Wainwright (2019), we present deviation inequalities for Σ-Gaussian ensembles and present a few examples before proving said inequalities.

Theorem

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

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

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

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

Example (Norm bounds for standard Gaussian ensemble):
Consider $W \in \mathbb{R}^{n\times d}$ generated with i.i.d $N(0,1)$ entries, which leads to the I_d-Gaussian ensemble. Given the above Theorem, it can be concluded that for $n \geq d$

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

Now it is evident that

$$1-P\left[\frac{\sigma_{\max}(W)}{\sqrt{n}} \geq 1+\delta+\sqrt{\frac{d}{n}}\right] = P\left[\frac{\sigma_{\max}(W)}{\sqrt{n}} \leq 1+\delta+\sqrt{\frac{d}{n}}\right]$$

thus according to the earlier Theorem,

$$P\left[\frac{\sigma_{\max}(W)}{\sqrt{n}} \leq 1+\delta+\sqrt{\frac{d}{n}}\right] \geq 1-\exp\left(-\frac{n\delta^2}{2}\right)$$

and similarly

$$P\left[\frac{\sigma_{\min}(W)}{\sqrt{n}} \geq 1-\delta-\sqrt{\frac{d}{n}}\right] \geq 1-\exp\left(-\frac{n\delta^2}{2}\right)$$

Thus, it can easily be seen that both bounds hold with probability greater than $1-2\exp\left(-\frac{n\delta^2}{2}\right)$. As we recall, the eigenvalues of the symmetric covariance matrix $\hat{\Sigma}$ is the square of the singular values $W/\sqrt{n}$. Furthermore,

$$\begin{aligned}\left|\left|\left|\hat{\Sigma}-\Sigma\right|\right|\right|_2 &= \max_{v\in\mathbb{S}^{d-1}}\left|v'\left\{\frac{1}{n}W'W-I_d\right\}v\right| \\ &= \max_{v\in\mathbb{S}^{d-1}}\left|\frac{1}{n}v'(W'W)v-v'I_dv\right|\end{aligned}$$

Note that $v' I_d v = \|v\|_2^2 = 1$. Thus,

$$\left|\!\left|\!\left| \hat{\Sigma} - \Sigma \right|\!\right|\!\right|_2 = \left|\!\left|\!\left| \frac{1}{n} W' W - I_d \right|\!\right|\!\right|_2 = \max_{v \in \mathbb{S}^{d-1}} \left| \frac{1}{n} v'(W'W)v - 1 \right|$$

Moreover, we have

$$\frac{\sigma_{\max}(W)}{\sqrt{n}} \leq 1 + \delta + \sqrt{\frac{d}{n}}$$

or

$$\frac{(\sigma_{\max}(W))^2}{n} \leq 1 + 2\left(\underbrace{\delta + \sqrt{\frac{d}{n}}}_{\varepsilon}\right) + \left(\underbrace{\delta + \sqrt{\frac{d}{n}}}_{\varepsilon}\right)^2$$

$$\left\{ \frac{(\sigma_{\max}(W))^2}{n} - 1 \right\} \leq 2\varepsilon + \varepsilon^2$$

thus,

$$\left|\left|\left|\frac{1}{n}W'W - I_d\right|\right|\right|_2 \leq 2\varepsilon + \varepsilon^2$$

Note that $\frac{d}{n} \to 0$, thus, the sample covariance matrix $\hat{\Sigma}$ is a consistent estimate of the identity matrix I_d.

Example (Gaussian covariance estimation):
Let $X \in \mathbb{R}^{n\times d}$ be a random matrix from the Σ-Gaussian ensemble. Noting that a if $X \sim N(0,\Sigma)$ it can equivalently be written as $X \sim \sqrt{\Sigma}N(0,I_d)$. So assuming that $W \sim N(0,I_d)$, we may express X as $X = W\sqrt{\Sigma}$. Moreover,

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

Thus, given the earlier example we know that

$$\left|\left|\left|\frac{1}{n}W'W - I_d\right|\right|\right|_2 \leq 2\varepsilon + \varepsilon^2,$$

where $\varepsilon = \delta + \sqrt{\frac{d}{n}}$. Therefore,

$$\frac{\left\|\left|\hat{\Sigma} - \Sigma\right|\right\|_2}{\||\Sigma|\|_2} \leq 2\varepsilon + \varepsilon^2$$

with probability at least $1 - 2\exp\left(\frac{-n\delta^2}{2}\right)$. Therefore, the relative error above converges to zero, so long as $d/n \to 0$.

To show the proof for the earlier Theorem first we recap a concept from the concentration inequalities chapter:

Recap (Theorem 2.26 of Wainwright):

Let $(X_1, \cdots, X_n)$ be a vector of i.i.d standard Gaussian variables, and let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz with respect to the Euclidean norm. Then the variable $f(X) - \mathbb{E}[f(X)]$ is sub-Gaussian with parameter at most L, and hence

$$P[|f(X) - E[f(X)]| \geq t] \leq 2 \exp\left(-\frac{t^2}{2L^2}\right), \quad \forall t \geq 0$$

Example (Singular values of Gaussian random matrices):
For $n > d$, let $X \in \mathbb{R}^{n \times d}$ be a random matrix with i.i.d. $N(0, 1)$ entries, and let

$$\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_d(X) \geq 0$$

are the ordered singular values of the matrix X. Referring to Weyl's inequality II, and given another matrix $Y \in \mathbb{R}^{n \times d}$, we have

$$\max_{k=1,\cdots,d} |\sigma_k(X) - \sigma_k(Y)| \leq |||X - Y|||_2 \leq |||X - Y|||_F$$

where $|||.|||_F$ denotes the Frobenius norm. Recalling that an L-Lipschitz function is one for which

$$|f(X) - f(Y)| \leq L|||X - Y|||_2$$

it can be suggested that $\sigma_k(X)$ for each k is a 1-Lipschitz function of random matrix. Furthermore, from Theorem 2.26 of Wainwright it can be shown that

$$P[|\sigma_k(X) - \mathbb{E}[\sigma_k(X)]| \geq \delta] \leq 2\exp\left(-\frac{\delta^2}{2}\right), \quad \forall \delta \geq 0$$

Now we wish to show that for $X \in \mathbb{R}^{n\times d}$ that is drawn according to the Σ-Gaussian ensemble, the maximum singular value $\sigma_{\max}(X)$ satisfies the upper deviation inequality

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

Let us denote $\bar{\sigma}_{\max} = \gamma_{\max}(\sqrt{\Sigma})$ and recall that we can write $X = W\sqrt{\Sigma}$, where $W \in \mathbb{R}^{n\times d}$ has i.i.d. $N(0,1)$ entries.
Let us view the mapping $W \to \frac{\sigma_{\max}(W\sqrt{\Sigma})}{\sqrt{n}}$ as a real-valued function on $\mathbb{R}^{nd}$. Noting that

$$\begin{aligned}\frac{\sigma_{\max}(W\sqrt{\Sigma})}{\sqrt{n}} &:= \frac{\left\|\left|W\sqrt{\Sigma}\right|\right\|_2}{\sqrt{n}} \\ &\leq \frac{\||W|\|_2 \left\|\left|\sqrt{\Sigma}\right|\right\|_2}{\sqrt{n}}\end{aligned}$$

Thus, it is evident that this function is Lipschitz function with respect to the Euclidean norm with constant at most $L = \bar{\sigma}_{\max}/\sqrt{n}$. Hence, by concentration of measure for Lipschitz functions of Gaussian random vectors, we conclude that

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

Substituting $\bar{\sigma}_{\max}(X)/\sqrt{n}$ for L and multiplying both sides of the inequality in the probability by $\sqrt{n}$, we obtain

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

Therefore, it is sufficient to show that

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

Recall that the maximum singular value has the variational representation

$$\sigma_{\max}(X) = \max_{v' \in \mathbb{S}^{d-1}} \|Xv'\|_2,$$

where $\mathbb{S}^{d-1}$ denotes the Euclidean unit sphere in $\mathbb{R}^d$. Since $X = W\sqrt{\Sigma}$, we may write the above expression as follows

$$\begin{aligned}
\sigma_{\max}(X) &= \max_{v' \in \mathbb{S}^{d-1}} \|W \underbrace{\sqrt{\Sigma}v'}_{v}\|_2 \\
&= \max_{v \in \mathbb{S}^{d-1}(\Sigma^{-1})} \|Wv\|_2 \\
&= \max_{u \in \mathbb{S}^{n-1}} \max_{v \in \mathbb{S}^{d-1}(\Sigma^{-1})} \underbrace{u'Wv}_{Z_{u,v}}
\end{aligned}$$

where $\mathbb{S}^{d-1}(\Sigma^{-1}) := \{v \in \mathbb{R}^d \mid \|\Sigma^{-\frac{1}{2}}v\|\}_2 = 1\}$ is an ellipse. Hence, obtaining bounds on the maximum singular value corresponds to controlling the supremum of the zero-mean Gaussian process $\{Z_{u,v}, (u,v) \in \mathbb{T}\}$ indexed by the set $\mathbb{T} := \mathbb{S}^{n-1} \times \mathbb{S}^{d-1}(\Sigma^{-1})$.

Let us now construct another Gaussian process, say $\{Y_{u,v},(u,v) \in \mathbb{T}\}$ such that

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

Theorem (Sudakov-Fernique)

Given a pair of zero-mean n-dimensional Gaussian vectors $(X_1,\cdots,X_n)$ *and* $Y_1,\cdots,Y_n$*, suppose that*

$$\mathbb{E}[(X_i - X_j)^2] \leq \mathbb{E}[(Y_i - Y_j)^2], \quad \forall(i,j) \in [n] \times [n].$$

Then $\mathbb{E}[\max_{j=1,\cdots,n} X_j] \leq \mathbb{E}[\max_{j=1,\cdots,n} Y_j]$.

Thus, from the results of the above Theorem, we can conclude that

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

Introducing the Gaussian process $Z_{u,v} := u'Wv$, let us first compute the induced pseudo-metric ρ_Z. For the two pairs (u, v) and $(\tilde{u}, \tilde{v})$, we may assume $\|v\|_2 \le \|\tilde{v}\|_2$. Furthermore, let $\langle\langle ., . \rangle\rangle$ be the trace inner product, which is defined as follows

Definition (Trace inner product)

For any $n \times n$ matrix A, the trace is the sum of the diagonal entries, -i.e. $Tr(A) = \sum_i a_{ii}$. On the other hand, for two $m \times n$ matrices A and B, the Frobenius or Trace inner product is

$$\langle\langle A, B \rangle\rangle = \sum_{ij} a_{ij} b_{ij}.$$

Furthermore,

$$\langle\langle A, B \rangle\rangle = Tr(A'B) = Tr(BA').$$

It can be observed that $Z_{u,v} := \langle\langle W, uv' \rangle\rangle$. Since the matrix W has i.i.d $N(0,1)$ entries, we have

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

We may rearrange and expand the above Frobenius norm

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

By the definition of set $\mathbb{T}$, we know that $\|u\|_2 = \|\tilde{u}\|_2 = 1$, and we further have $\|u\|_2^2 - \langle u, \tilde{u}\rangle \geq 0$. Recall the Cauchy-Schwarz inequality

Definition (Cauchy-Schwarz inequality)

If $p = q = 2$ and $1/p + 1/q = 1$, and if $\mathbb{E}[|Y|^p] < \infty$ and $\mathbb{E}[|Z|^q] < \infty$, then

$$\mathbb{E}[|YZ|] \leq [\mathbb{E}[|Y|]^2]^{1/2}[\mathbb{E}[|Z|]^2]^{1/2}$$

[see White (2014)]

Thus, using the Cauchy-Schwarz inequality, we have

$$|\langle v, \tilde{v} \rangle| \leq \|v\|_2 \|\tilde{v}\|_2$$

Since from our earlier assumption $\|v\|_2 \leq \|\tilde{v}\|_2$, it can further be concluded that

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

Combining the above and the earlier results, it can be concluded that

$$2(\|u\|_2^2 - \langle u, \tilde{u} \rangle)(\langle v, \tilde{v} \rangle - \|\tilde{v}\|_2^2) \leq 0.$$

Putting these findings together with the expansion of the Frobenius norm, we obtain

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

Furthermore, by the definition of the set $\mathbb{S}^{d-1}(\Sigma^{-1})$, we have $\|\tilde{v}\|_2 \leq \bar{\sigma}_{\max} = \gamma_{\max}(\sqrt{\Sigma})$, and as a result

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

Now let us define the Gaussian process $Y_{u,z} := \bar{\sigma}_{\max}\langle g, u\rangle + \langle h, v\rangle$, where $g \in \mathbb{R}^n$ and $h \in \mathbb{R}^d$ are standard Gaussian random vectors. By construction,

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

Once again using the Sudakov-Fernique bound we conclude that

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

By Jensen's inequality, we have

$$\mathbb{E}[\|g\|_2] \leq \sqrt{n}$$

and similarly

$$\mathbb{E}[\|\sqrt{\Sigma}h\|_2] \leq \sqrt{\mathbb{E}[h'\Sigma h]} = \sqrt{Tr(\Sigma)}$$

which establishes the result of the Theorem.

References

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Science & Business Media.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.