RMT - Covariance Matrix Estimation

Yessin Moakher

February 2025

Contents

1 Introduction and motivation

Notations. - *Singular Values :* $\sigma_{max}(A) = \sigma_1(A) \geq \ldots \geq \sigma_n(A) \geq 0$.
A real symetric - Eigenvalues : $\lambda_{max}(A) = \lambda_1(A) \geq \ldots \geq \lambda_n(A)$.

2 Classical Approach

For the demonstration of the theorem, it is important to recall some elements regarding the operator norm of symmetric matrices and introduce the concept of the ε-Net, which will be very useful for proving the theorem.

Lemma 1. *If $A \in \mathbb{R}^{d\times d}$ is a symmetric matrix, then*

$$\|A\| = \max_{\|x\|_2=1} |\langle Ax, x\rangle|.$$

Definition 1 (ϵ-net)**.** *Let (X, d) be a metric space. Let $K \subset X$ and $\epsilon > 0$. A set $\mathcal{N} \subset K$ is an ϵ-net on K if and only if*

$$K \subset \bigcup_{x\in\mathcal{N}} B(x, \epsilon).$$

Proposition 1. *Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix, and let $\epsilon \in \left[0, \frac{1}{2}\right)$. For any ϵ-net $\mathcal{N}$ of the unit sphere S^{d-1}, we have*

$$\|A\| \leq \frac{1}{1-2\epsilon} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle| .$$

Proof. According to Lemma 1, there exists $x \in S^{d-1}$ such that $|\langle Ax, x\rangle| = \|Ax\|$, and by the definition, there exists $x_0 \in \mathcal{N}$ such that $\|x - x_0\|_2 \leq \epsilon$.

By the triangle inequality, we get:

$$\begin{aligned} \|A\| - |\langle Ax_0, x_0\rangle| &= |\langle Ax, x\rangle| - |\langle Ax_0, x_0\rangle| \\ &\leq |\langle Ax, x\rangle - \langle Ax_0, x_0\rangle| \\ &= |\langle Ax, x - x_0\rangle + \langle A(x - x_0), x_0\rangle| \\ &\leq |\langle Ax, x - x_0\rangle| + |\langle A(x - x_0), x_0\rangle| . \end{aligned}$$

Using the Cauchy-Schwarz inequality, we get:

$$\begin{aligned} \|A\| - |\langle Ax_0, x_0\rangle| &\leq \|A\|_2\|x - x_0\|_2 + \|A(x - x_0)\|_2\|x_0\|_2 \\ &\leq 2\epsilon\|A\|. \end{aligned}$$

Thus:

$$\begin{aligned} \|A\| &\leq \frac{1}{1-2\epsilon} |\langle Ax_0, x_0\rangle| \\ &\leq \frac{1}{1-2\epsilon} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle| . \end{aligned}$$

□

Proposition 2. *For every $\epsilon > 0$, we can find an ϵ-net $\mathcal{N}$ on S^{d-1} such that its cardinality satisfies*

$$|\mathcal{N}| \leq 9^d.$$

Proposition 3 (Hanson-Wright Inequality [?])**.** *Let $Z = (Z_1, \ldots, Z_m) \in \mathbb{R}^m$ be a random vector with independent sub-Gaussian components Z_i that satisfy $\mathbb{E}[Z_i] = 0$ and $\|Z_i\|_{\psi_2}^2 \leq K$. Let A be an $m \times m$ matrix. Then, for all $t \geq 0$,*

$$\mathbb{P}\left(|Z^\top AZ - \mathbb{E}[Z^\top AZ]| > t\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|}\right)\right).$$

This inequality can be rewritten as,

$$\mathbb{P}\left(|Z^\top AZ - \mathbb{E}[Z^\top AZ]| > K^2\max(\frac{\|A\|t}{c}, \|A\|_F\sqrt{\frac{t}{c}})\right) \leq 2\exp(-t)$$

or alternatively by introducing $C = \frac{1}{\min(c,\sqrt{c})}$

$$\mathbb{P}\left(|Z^\top AZ - \mathbb{E}[Z^\top AZ]| > CK^2\max(\|A\|t, \|A\|_F\sqrt{t})\right) \leq 2\exp(-t)$$

2.0.1 Proof of the Theorem

Proof. [?] Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be realizations of a centered Gaussian vector X with covariance matrix $\Sigma = \mathbb{E}(XX^T)$. Define

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} X_iX_i^T.$$

1) **Approximation using an ϵ-net:** We take $\epsilon = \frac{1}{4}$. By Proposition 2, consider an ϵ-net $\mathcal{N}$ of the unit sphere S^{d-1} with cardinality at most 9^d.

Since $\Sigma_n - \Sigma$ is a symmetric matrix, by Proposition 1, we have

$$\|\Sigma_n - \Sigma\| \leq 2 \sup_{u \in \mathcal{N}} |\langle(\Sigma_n - \Sigma)u, u\rangle|$$

which expands to

$$2 \sup_{u \in \mathcal{N}} \left| \frac{1}{n} \sum_{i=1}^{n} \langle X_i X_i^T u, u\rangle - \mathbb{E}\left[\langle X X^T u, u\rangle\right] \right|.$$

Let

$$g_n(u) = \frac{1}{n} \sum_{i=1}^{n} \langle X_i X_i^T u, u\rangle - \mathbb{E}\left[\langle X X^T u, u\rangle\right].$$

By rearranging inner products, we obtain

$$g_n(u) = \frac{1}{n} \sum_{i=1}^{n} \left(\langle X_i, u\rangle^2 - \mathbb{E}\left(\langle X, u\rangle^2\right)\right).$$

Now, setting $X_i = \Sigma^{\frac{1}{2}} Z_i$, where Z_i is a standardized Gaussian vector, we substitute X_i to get

$$g_n(u) = \frac{1}{n} \sum_{i=1}^{n} \left(\langle Z_i, \Sigma^{\frac{1}{2}} u\rangle^2 - \mathbb{E}\left(\langle Z_i, \Sigma^{\frac{1}{2}} u\rangle^2\right)\right).$$

Writing this in matrix form:

$$g_n(u) = \frac{1}{n} \sum_{i=1}^{n} \left(Z_i^T \Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T Z_i - \mathbb{E}\left(Z_i^T \Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T Z_i\right)\right).$$

Defining $Z = (Z_1^T, \ldots, Z_n^T)^T$ and $A(u)$ as the block diagonal matrix of size $nd \times nd$ with n blocks $\Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T$, we obtain

$$g_n(u) = \frac{1}{n}\left(Z^T A(u) Z - \mathbb{E}(Z^T A(u) Z)\right).$$

2) **Concentration inequality:** Let $t > 0$ and $u \in S^{d-1}$. Using the Hanson-Wright inequality (Proposition 3) for $nt > 0$,

$$\mathbb{P}\left(|Z^\top A(u) Z - \mathbb{E}[Z^\top A(u) Z]| > CK^2 \max(\|A(u)\| nt, \|A(u)\|_F \sqrt{nt})\right) \leq 2\exp(-nt).$$

Since $A(u)$ is symmetric,

$$\|A(u)\| = \max\{|\lambda_1(A(u))|, |\lambda_{nd}(A(u))|\} = \max\{|\lambda_1(\Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T)|, |\lambda_d(\Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T)|\}.$$

Thus,

$$\|A(u)\| = \|\Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T\| \leq \|\Sigma\| \|u\|^2,$$

and

$$\|A(u)\|_F^2 = n\|\Sigma^{\frac{1}{2}} u (\Sigma^{\frac{1}{2}} u)^T\|_F^2 \leq n(\|\Sigma\| \|u\|^2)^2.$$

Bounding the left-hand side, the inequality simplifies to

$$\mathbb{P}\left(|g_n(u)| > \|\Sigma\| CK^2 \max(t, \sqrt{t})\right) \leq 2\exp(-nt).$$

3) **Union bound:** Finally, we conclude by bounding the operator norm. For any $t > 0$,

$$\mathbb{P}\left(\|\Sigma_n - \Sigma\| > 2\|\Sigma\|CK^2 \max(t, \sqrt{t})\right) \leq 2 \times 9^d \exp(-nt).$$

$\square$

Proposition 4 (Weyl's Inequality)**.** *For two symmetric matrices A, B in $\mathbb{R}^{d\times d}$, we have the following result regarding eigenvalues:*

$$\forall i \in \{1, \dots, d\}, \quad |\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|.$$

3 Non-Asymtotic Viewpoint

To do : Motivate why a non asymtotic analysis is important in high dimensions.

3.1 Eigenvalues of Wishart Matricies (Sample Covariance matrix of random matrix with Gaussian rows)

3.1.1 Preliminaries

We have the following results for Gaussian processes

Lemma 2 (Lipschitz functions Gaussian concentration inequality)**.** *Let $(x_i) \sim \mathcal{N}(0,1)$ i.i.d. and $f : \mathbb{R}^n \to \mathbb{R}$, we suppose that f is L-Lipschitz in $\|.\|_2$. Then*

$$\forall t \geq 0, \mathbb{P}(|f(X_1, \dots, X_n) - \mathbb{E}(f(X_1, \dots, X_n))| \geq t) \leq 2exp(-\frac{t^2}{2L^2})$$

Lemma 3 (Sudakov-Fernique)**.** *Given $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ two zero-mean n-dimensional Gaussian vectors, suppose that*

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

Then

$$\mathbb{E}(max_{i=1,\dots,n} X_i) \leq \mathbb{E}(max_{i=1,\dots,n} Y_i)$$

3.2 Result

We had $(x_i) \sim \mathcal{N}(0, \Sigma)$ i.i.d. , where $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix. We have

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \in \mathbb{R}^{n\times d}, \quad \hat{\Sigma} = \frac{1}{n} X^T X.$$

Theorem 1. *for all $\delta > 0$, we have :*

$$1)\ \mathbb{P}(\frac{\sigma_{max}(X)}{\sqrt{n}} \geq \lambda_{max}(\sqrt{\Sigma})(1+\delta) + \sqrt{\frac{tr(\Sigma)}{n}}) \leq e^{-n\delta^2/2}$$

$$2)\ if\ n \geq d\ , \mathbb{P}(\frac{\sigma_{min}(X)}{\sqrt{n}} \leq \lambda_{min}(\sqrt{\Sigma})(1+\delta) - \sqrt{\frac{tr(\Sigma)}{n}}) \leq e^{-n\delta^2/2}$$

Proof. Notation: $\bar{\sigma}_{\max} = \gamma_{\max}(\sqrt{\Sigma})$

- Step 1 : Concentration

By standard properties of the multivariate Gaussian distribution, we can write

$$\mathbf{X} = \mathbf{W}\sqrt{\Sigma},$$

where the random matrix $\mathbf{W} \in \mathbb{R}^{n\times d}$ has i.i.d. $\mathcal{N}(0,1)$ entries.

Using Weyl's theorem, given another matrix $\mathbf{W}' \in \mathbb{R}^{n\times d}$, we have:

$$\left|\sigma_{\max}(\mathbf{W}\sqrt{\Sigma}) - \sigma_{\max}(\mathbf{W}'\sqrt{\Sigma})\right| \leq \|(\mathbf{W}-\mathbf{W}')\sqrt{\Sigma}\|_2 \leq \|\mathbf{W}-\mathbf{W}'\|_2 \lambda_{\max}(\sqrt{\Sigma}).$$

So, the mapping

$$\mathbf{W} \mapsto \frac{\sigma_{\max}(\mathbf{W}\sqrt{\Sigma})}{\sqrt{n}}$$

viewed as a real-valued function on $\mathbb{R}^{nd}$, is Lipschitz with respect to the Euclidean norm with constant at most $L = \bar{\sigma}_{\max}/\sqrt{n}$.

Using Lemma 2, the function satisfies a Gaussian concentration inequality

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

Consequently, it suffices to show that

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

- Step 2 : Bounding the expected value

We first use a variational formulation of σ_{max}

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

where

$$S^{d-1}(\Sigma^{-1}) := \left\{v \in \mathbb{R}^d \mid \|\Sigma^{-1/2}v\|_2 = 1\right\}.$$

Thus, it suffices to bound the expectation of the zero-mean Gaussian process $\{Z_{u,v}, (u,v) \in T\}$ indexed by the set $T := S^{n-1} \times S^{d-1}(\Sigma^{-1})$.

We want to apply the Sudakov-Fernique Lemma, but first, we need to find another Gaussian process $\{Y_{u,v}, (u,v) \in T\}$ such that

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

Applying the Sudakov-Fernique Lemma, we'll obtain:

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

Given two pairs (u,v) and $(\tilde{u},\tilde{v})$, assume $\|v\|_2 \leq \|\tilde{v}\|_2$ (otherwise reverse their roles). Then,

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

the last equality is because W has i.i.d. N(0, 1) entries

Expanding the Frobenius norm:

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

Using $\|v\|_2 \leq \|\tilde{v}\|_2$, we obtain:

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

So since $\|u\|_2^2 = \|\tilde{u}\|_2^2 (= 1)$,

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

which simplifies to:

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

By definition of $S^{d-1}(\Sigma^{-1})$, we have $\|\tilde{v}\|_2 \leq \bar{\sigma}_{\max} = \gamma_{\max}(\sqrt{\Sigma})$, leading to:

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

Motivated by this inequality, we define the Gaussian process

$$Y_{u,v} := \bar{\sigma}_{\max}\langle g, u\rangle + \langle h, v\rangle,$$

where $g \in \mathbb{R}^n$ and $h \in \mathbb{R}^d$ are standard Gaussian random vectors with i.i.d. $\mathcal{N}(0,1)$ entries, and mutually independent. By construction,

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

Applying the Sudakov–Fernique bound (Lemma 3), we obtain:

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

By Jensen's inequality,

$$\mathbb{E}[\|g\|_2] \leq \sqrt{n}, \quad \mathbb{E}[\|\sqrt{\Sigma}h\|_2] \leq \sqrt{\mathbb{E}[h^T\Sigma h]} = \sqrt{\mathrm{tr}(\Sigma)}.$$

This proves inequality (1) of the theorem.

It remains to prove the lower bound (2) on the minimal singular value. It is based on a similar argument, but requires somewhat more technical work, so we'll make the simplifying assumption that $\Sigma = I_d$.

Again with Lemma 2,

$$\mathbb{P}\left[\sigma_{\min}(\mathbf{X}) \leq \mathbb{E}[\sigma_{\min}(\mathbf{X})] + \sqrt{n}\delta\right] \leq e^{-n\delta^2/2}.$$

So it suffices to show that

$$\mathbb{E}[\sigma_{\min}(X)] \geq \sqrt{n} - \sqrt{d}.$$

For $n \geq d$, we use the variational representation:

$$\sigma_{\min}(X) = \min_{v\in S^{d-1}}\max_{u\in S^{n-1}}\langle u, Xv\rangle.$$

We'll need the following Gaussian process inequality which is a sort of generalization of Sudakov-Fernique

Theorem 14.2 (Gordon's inequality). Let $(Z_{s,t})_{s\in S, t\in T}$ and $(Y_{s,t})_{s\in S, t\in T}$ be two Gaussian processes with $\mathbb{E}[Z_{s,t}] = \mathbb{E}[Y_{s,t}]$, satisfying:

$$\mathbb{E}[(Z_{s,t_1} - Z_{s,t_2})^2] \geq \mathbb{E}[(Y_{s,t_1} - Y_{s,t_2})^2], \quad \forall t_1, t_2 \in T, s \in S,$$

$$\mathbb{E}[(Z_{s_1,t} - Z_{s_2,t})^2] \leq \mathbb{E}[(Y_{s_1,t} - Y_{s_2,t})^2], \quad \forall s_1 \neq s_2, t \in T.$$

Then,

$$\mathbb{E}\left[\max_{s\in S, t\in T} Z_{s,t}\right] \leq \mathbb{E}\left[\max_{s\in S, t\in T} Y_{s,t}\right].$$

Taking $Y_{u,v} = \langle g, u\rangle + \langle h, v\rangle$ where g and h are iid Gaussian random vectors, we check that $Z_{u,v}$ and $Y_{u,v}$ satisfy the conditions of the theorem. Then,

$$\begin{aligned}
-\mathbb{E}[\sigma_{\min}(X)] &= \mathbb{E}\left[\max_{v\in S^{d-1}} -\|Xv\|_2\right] \\
&= \mathbb{E}\left[\max_{v\in S^{d-1}} \min_{u\in S^{n-1}} \langle u, -Xv\rangle\right] \\
&\leq \mathbb{E}\left[\max_{v\in S^{d-1}} \min_{u\in S^{n-1}} (\langle g, u\rangle + \langle h, v\rangle)\right] \\
&= \mathbb{E}\left[\max_{v\in S^{d-1}} \langle h, v\rangle\right] + \mathbb{E}\left[\min_{u\in S^{n-1}} \langle g, u\rangle\right] \\
&= \mathbb{E}[\|h\|_2] - \mathbb{E}[\|g\|_2] \approx \sqrt{d} - \sqrt{n}.
\end{aligned}$$

(by properties of chi-squared distributions)

Thus,

$$\mathbb{E}[\sigma_{\min}(X)] \geq \sqrt{n} - \sqrt{d}.$$

□

4 Large n,d regime

Goal: To study the asymptotic distribution of eigenvalues of the sample covariance matrix.

To rigorously define the convergence of eigenvalues, we introduce the following concept:

Definition 2 (Empirical Spectral Distribution (ESD)). *For a matrix $A \in \mathcal{M}_d(\mathbb{C})$, the empirical spectral distribution (ESD) is defined as the probability measure:*

$$\hat{\mu}(A) := \frac{1}{d} \sum_{\lambda \in Sp(A)} \delta_\lambda,$$

where $Sp(A)$ denotes the spectrum of A, and δ_λ is the Dirac delta function at λ. This corresponds to selecting an eigenvalue uniformly at random from the spectrum of A.

To characterize the convergence of a sequence of probability measures, we employ a common technique in random matrix theory: studying the convergence of their Stieltjes transforms.

Definition 3 (Stieltjes Transform). *For a probability measure μ, the Stieltjes transform is defined as:*

$$m_\mu(z) = \int \frac{1}{t - z}\, d\mu(t), \quad \textit{for } z \textit{ such that } \operatorname{Im}(z) > 0.$$

The utility of the Stieltjes transform is demonstrated by the following result:

Lemma 4. *Let μ_n be a sequence of probability measures and μ a probability measure. If*

$$\forall z,\ \mathrm{Im}(z) > 0, \quad m_{\mu_n}(z) \xrightarrow[n\to+\infty]{} m_{\mu}(z),$$

then the sequence μ_n converges weakly to μ:

$$\mu_n \xrightarrow[n\to+\infty]{W} \mu.$$

The Stieltjes transform of the empirical spectral distribution (ESD) of a matrix A, having real eigenvalues, has a closed-form expression. Let

$$\mu_n = \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(A)},$$

where $\lambda_i(A)$ are the eigenvalues of A. For all z with $\mathrm{Im}(z) > 0$ (in particular not real), the Stieltjes transform of μ_n is given by:

$$m_{\mu_n}(z) = \frac{1}{d}\sum_{i=1}^{d} \frac{1}{\lambda_i(A) - z} = \frac{1}{d}\mathrm{Tr}\left((A - zI_d)^{-1}\right).$$

Definition 4 (Resolvent)**.** *Let A be a square matrix of size $n \times n$. The resolvent R_A of the matrix A is defined on $\mathbb{C} \setminus Sp(A)$ by:*

$$Q_A(z) = (A - zI)^{-1}.$$

It shall be denoted Q or (Q_n) when it is non-ambiguous.

Returning to our problem, consider $(x_i) \sim \mathcal{N}(0, \Sigma)$ i.i.d., where $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix. Define

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \in \mathbb{R}^{n\times d}, \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n}X^T X.$$

We aim to study the convergence of the eigenvalue distribution of the sample covariance matrix $\hat{\Sigma}$. To achieve this, we analyze its Stieltjes transform:

$$m_{\mu_n}(z) = \frac{1}{d}\mathrm{Tr}(Q(z)),$$

where $Q(z) = (\hat{\Sigma} - zI_d)^{-1}$ is the resolvent of the sample covariance matrix $\hat{\Sigma}$.

Assumption 1 (Large n,d regime)**.** *As $n \to \infty$, we have $\frac{d}{n} \to r \in (0, \infty)$.*

In the classical regime, where d **is fixed**, the law of large numbers implies:

$$Q(z) = \left(\frac{1}{n}XX^T - zI_d\right)^{-1} \xrightarrow[n\to+\infty]{\text{a.s.}} (\Sigma - zI_d)^{-1}.$$

This is unfortunately not true anymore when d also tends to infinity. As an example for that when $\Sigma = I_d$, we have the Marchenko-Pastur Law ...

Let's find the ...

Suppose, it exists $Q(z) \in \mathbb{C}^{d\times d}$, such that $\frac{1}{d}(Q_n(z) - Q(z)) \to 0$.

Step 1: concentration of $Q_n(z)$

We can write $X = W\sqrt{\Sigma}$, where $W \in \mathbb{R}^{n\times d}$ has i.i.d. $\mathcal{N}(0, 1)$ entries. Now let's consider the mapping $S : W \to \mathrm{Tr}((\frac{1}{n}\sqrt{\Sigma}W^TW\sqrt{\Sigma} - zI_d)^{-1})$ as a real-valued function on $\mathbb{R}^{nd}$ and let's show that it is Lipchitz. Let $B(W) = (\frac{1}{n}\sqrt{\Sigma}W^TW\sqrt{\Sigma} - zI_d)$

$$\begin{aligned}
d_S(W)[H] &= d_{\mathrm{Tr}}(B(W)^{-1}) \circ d_{.^{-1}}(B(W))[H] \\
&= \mathrm{Tr}(d_{.^{-1}}(B(W)) \circ d_{B(.)}(W)[H]) \\
&= \mathrm{Tr}(-B(W)^{-1}d_{B(.)}(W)[H]B(W)^{-1}) \\
&= \mathrm{Tr}(-d_{B(.)}(W)[H]B(W)^{-2}) \\
&= \mathrm{Tr}\left(-\frac{1}{n}(H^T\sqrt{\Sigma}W + W^T\sqrt{\Sigma}H)B(W)^{-2}\right) \\
&= \left\langle \frac{-2}{n}\sqrt{\Sigma}WB(W)^{-2}, H \right\rangle_F
\end{aligned}$$

Thus :

$$\nabla_W S(W) = \frac{-2}{n}\sqrt{\Sigma}WB(W)^{-2}.$$

$$\begin{aligned}
\|\nabla_W S(W)\|_F^2 &= \|\frac{-2}{n}\sqrt{\Sigma}WB(W)^{-2}\|_F^2 \\
&= \|\frac{-2}{n}U\Lambda V^T(\frac{1}{n}V\Lambda U^TU\Lambda V^T - zI_d)^{-2}\|_F^2 \\
&= \|\frac{-2}{n}\Lambda(\frac{1}{n}\Lambda^2 - zI_d)^{-2}\|_F^2, \text{ } U \text{ and } V \text{ are orthogonal matrices} \\
&= \frac{4}{n^2}\sum_{i=1}^{d}\left|\frac{\sigma_i(X)}{\left(\frac{1}{n}\sigma_i(X)^2 - \mathrm{Re}(z) + i\,\mathrm{Im}(z)\right)^2}\right|^2 \\
&= \frac{4}{n^2}\sum_{i=1}^{d}\frac{\sigma_i(X)^2}{\left(\frac{1}{n}\sigma_i(X)^2 - \mathrm{Re}(z)\right)^2 + \mathrm{Im}(z)^2}
\end{aligned}$$

let $g(x) = \frac{x}{(\frac{x}{n}-a)^2+b^2}$, $g(0) = 0$ and $\lim_{x\to+\infty} g(x) = 0$.

$$g'(x) = 0 \iff -\frac{x^2}{n^2} + a^2 + b^2 = 0$$

.

Thus :

$$\|\nabla_W S(W)\|_F^2 \leq \frac{4d}{n^2}\frac{n|z|}{(|z| - \mathrm{Re})^2 + \mathrm{Im}(z)^2} = \frac{2d}{n(|z| - \mathrm{Re}(z))}$$

According to the lemma 2 (Lipschitz functions Gaussian concentration inequality), we have

$$\forall t \geq 0, \mathbb{P}(|Q_n - \mathbb{E}Q_n| \geq t) \leq 2exp(-\frac{t^2n(|z| - \mathrm{Re}(z))}{4d})$$

Step 2: Controlling $\bar{Q} - \mathbb{E}(Q_n)$:

The difference between two resolvent matrices has a nice form :

$$\bar{Q} - Q_n = \bar{Q} - Q_n(\Sigma' - zI_d)\bar{Q} = Q_n(\hat{\Sigma} - zI_d - \Sigma + zI_d)\bar{Q} = Q_n(\hat{\Sigma} - \Sigma')\bar{Q}$$

. Thus

$$\bar{Q} - \mathbb{E}Q_n = \mathbb{E}(Q_n(\hat{\Sigma} - \Sigma')\bar{Q}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Q_n(x_ix_i^T - \Sigma')\bar{Q}). \tag{1}$$

Let us denote $X_{-i} \in \mathcal{M}_{n-1,d}$, the matrix X deprived of it's i-th row, and let us define $\hat{\Sigma}_{-i} = \frac{1}{n} X_{-i}^T X_{-i}$, and $Q_i = (\hat{\Sigma}_{-i} - zI_d)^{-1}$.

Applying the Sherman-Morrison identities, which state:

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u},$$

and

$$(A + uv^T)^{-1} u = \frac{A^{-1}u}{1 + v^T A^{-1} u}$$

Here we have:

$$A = \hat{\Sigma}_{-i} - zI_d, \quad u = \frac{1}{\sqrt{n}} x_i, \quad v^T = \frac{1}{\sqrt{n}} x_i^T.$$

Thus:

$$Q_n = Q_i - \frac{Q_i x_i x_i^T Q_i}{n + x_i^T Q_i x_i}, \text{ and } Q_n x_i = \frac{Q_{-i} x_i}{1 + \frac{1}{n} x_i^T Q_{-i} x_i}.$$

Substituting in equation1 gives as :

$$\bar{Q} - \mathbb{E}Q_n = \frac{1}{n} \sum_{i=1}^{n} (\mathbb{E}[Q_{-i}(\frac{x_i x_i^T}{1 + \frac{1}{n} x_i^T Q_{-i} x_i} - \Sigma')\bar{Q}] + \frac{1}{n^2}\mathbb{E}[Q_{-i} x_i x_i^T Q_n \Sigma' \bar{Q}]).$$

- Concentration of $\delta_i = x_i^T Q_{-i} x_i$:

$$\begin{aligned}
\mathbb{E}(x_i^T Q_{-i} x_i) &= \mathbb{E}(\mathbb{E}(x_i^T Q_{-i} x_i | Q_{-i})) \\
&= \mathbb{E}(\mathbb{E}(\mathrm{Tr}(x_i^T Q_{-i} x_i) | Q_{-i})), \text{because it is a } 1 \times 1 \text{ matrix} \\
&= \mathbb{E}(\mathrm{Tr}(\Sigma Q_{-i})), \text{ because } x_i \text{ and } Q_{-i} \text{ are independent} \\
&= \mathrm{Tr}(\Sigma \mathbb{E}(Q_{-i}))
\end{aligned}$$

Conditional on Q_{-i}, δ_i is a quadratic form on Gaussian random vector, by applying Hanson-Wright inequality (on the standardized Gaussian vector for independence), we get :

$$\mathbb{P}(|x_i^T Q_{-i} - \mathrm{Tr}(\Sigma Q_{-i})| \geq nt | Q_{-i}) \leq 2\exp(-c\min(\frac{n^2 t^2}{\|\Sigma\|_F^2 \|Q_{-i}\|_F^2}, \frac{nt}{\|\Sigma\| \|Q_{-i}\|})$$

Since $\hat{\Sigma}_{-i}$ is positive semi-definite, its eigenvalues are real.

Thus,

$$\|Q_{-i}\| = \frac{1}{\sigma_{\min}(\hat{\Sigma}_{-i} - zI_d)} \leq \frac{1}{\mathrm{Im}(z)}.$$

And,

$$\|Q_{-i}\|_F^2 = \sum_{i=1}^{d} \sigma_i (\hat{\Sigma}_{-i} - zI_d)^{-1})^2 = \sum_{i=1}^{d} \frac{1}{|\lambda_i(\hat{\Sigma}_i) - z|^2} \leq \sum_{i=1}^{d} \frac{1}{\mathrm{Im}(z)^2} = \frac{d}{\mathrm{Im}(z)^2}$$

Rearranging the terms,

$$\mathbb{P}\left(\left|x_i^T Q_{-i} x_i - \mathrm{Tr}(\Sigma Q_{-i})\right| \geq nt \mid Q_{-i}\right) \leq 2\exp\left(-c\min\left(\frac{n^2 t^2 \,\mathrm{Im}(z)^2}{d\|\Sigma\|_F^2}, \frac{nt\,\mathrm{Im}(z)}{\|\Sigma\|}\right)\right).$$

5 Eigenvectors estimation

5.1 Proof of Theorem ??

5.1.1 Preliminaries

Lemma 5. *If $A \in \mathbb{R}^{d\times d}$ is a symmetric matrix, then the resolvent of A is given by*

$$Q_A(z) = \sum_{j=1}^{d} \frac{u_j u_j^\top}{\lambda_j(A) - z},$$

where $\lambda_j(A)$ are the eigenvalues of A and u_j are the corresponding eigenvectors.

Proposition 5. *If $A \in \mathbb{R}^{d\times d}$ is a symmetric matrix, whose eigenvalues are denoted as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$, then the projector onto the eigenspace associated with λ_i is given by:*

$$P_{\lambda_i} = \frac{1}{2\pi i} \oint_{\Gamma_i} Q_A(\eta)\, d\eta,$$

where Γ_i is a closed contour surrounding λ_i, but not other eigenvalues λ_j such that $\lambda_i \neq \lambda_j$.

Proof. The result follows by applying Cauchy's theorem to the resolvent formula obtained in Lemma 5. □

Lemma 6. *Let $\Sigma, \Sigma_n \in \mathbb{R}^{d\times d}$ be two real symmetric matrices, whose eigenvalues are denoted respectively as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_d$. Let $\{\lambda_i\}_{i=s}^{r}$ be the eigenvalues of Σ whose corresponding eigenvectors span the eigenspace associated with the eigenvalue λ_j. Define $\lambda_0 = \infty$ and $\lambda_{d+1} = -\infty$. If*

$$\|\Sigma - \Sigma_n\| < \frac{\min(\lambda_r - \lambda_{r+1}, \lambda_{s-1} - \lambda_s)}{2},$$

then, defining $g_{\lambda_j} = \min(\lambda_r - \lambda_{r+1}, \lambda_{s-1} - \lambda_s)$, we have:

$$\begin{aligned}
\lambda_i &\notin D(\lambda_j, \frac{g_{\lambda_j}}{2}) \text{ for all } i \notin \{s, \ldots, r\},\\
\hat{\lambda}_i &\in D(\lambda_j, \frac{g_{\lambda_j}}{2}) \text{ for all } i \in \{s, \ldots, r\},\\
\hat{\lambda}_i &\notin D(\lambda_j, \frac{g_{\lambda_j}}{2}) \text{ for all } i \notin \{s, \ldots, r\},
\end{aligned}$$

where $D(\lambda_j, \frac{g_{\lambda_j}}{2})$ denotes the open disk of radius $\frac{g_{\lambda_j}}{2}$ centered at λ_j.

Proof. Let $g_{\lambda_j} = \min(\lambda_r - \lambda_{r+1}, \lambda_{s-1} - \lambda_s)$ and assume the lemma's condition holds.

Since the eigenvalues are arranged in descending order, it is immediate that $\lambda_i \notin D(\lambda_j, \frac{g_{\lambda_j}}{2})$ for all $i \notin \{s, \ldots, r\}$. Using Weyl's inequality (Proposition 4), we also get $\hat{\lambda}_i \in D(\lambda_j, \frac{g_{\lambda_j}}{2})$ for all $i \in \{s, \ldots, r\}$.

Furthermore, by similar arguments, for any $i \notin \{s, \ldots, r\}$:

$$|\lambda_j - \hat{\lambda}_i| \geq |\lambda_j - \lambda_i| - |\lambda_i - \hat{\lambda}_i| > r - \frac{r}{2} = \frac{r}{2}$$

which concludes the proof. □

5.1.2 Proof of the Theorem

Proof. In the following, we denote

$$g_{\lambda_j} = \min(\lambda_r - \lambda_{r+1}, \lambda_{s-1} - \lambda_s),$$

and assume that $\|\Sigma - \Sigma_n\| < \frac{g_{\lambda_j}}{4}$.

Fix $z \in \partial D(\lambda_j, \frac{g_{\lambda_j}}{2})$, the circle of radius $\frac{g_{\lambda_j}}{2}$ centered at λ_j, and consider:

$$Q_\Sigma(z) - Q_{\Sigma_n}(z) = (zI - \Sigma)^{-1} - (zI - \Sigma - (\Sigma_n - \Sigma))^{-1}.$$

Note that

$$|z - \lambda_i| \geq ||\lambda_i - \lambda_j| - |\lambda_j - z|| = \left||\lambda_i - \lambda_j| - \frac{r}{2}\right| \geq \frac{g_{\lambda_j}}{2}.$$

Thus,

$$\|R_\Sigma(z)\| = \max_i \frac{1}{|z - \lambda_i|} = \frac{1}{\min_i |z - \lambda_i|} \leq \frac{2}{g_{\lambda_j}}.$$

Finally,

$$\|(\Sigma_n - \Sigma) \cdot Q_\Sigma(z)\| \leq \|\Sigma_n - \Sigma\| \cdot \|Q_\Sigma(z)\| \leq \frac{g_{\lambda_j}}{4} \cdot \frac{2}{g_{\lambda_j}} = \frac{1}{2} < 1.$$

Since $(M_d(\mathbb{R}), \|.\|)$ is a unital Banach algebra, we can write:

$$Q_{\Sigma_n}(z) = (zI - \Sigma - (\Sigma_n - \Sigma))^{-1} = Q_\Sigma(z) \cdot \sum_{k=0}^{\infty} \left((\Sigma_n - \Sigma) \cdot Q_\Sigma(z)\right)^k.$$

Thus, we obtain:

$$\begin{aligned}
\|Q_\Sigma(z) - Q_{\Sigma_n}(z)\| &= \left\|(zI - \Sigma)^{-1} \cdot \sum_{k=1}^{\infty} \left((\Sigma_n - \Sigma)(zI - \Sigma)^{-1}\right)^k\right\| \\
&\leq \left\|(zI - \Sigma)^{-1}\right\| \cdot \sum_{k=1}^{\infty} \left(\|\Sigma_n - \Sigma\| \cdot \left\|(zI - \Sigma)^{-1}\right\|\right)^k \\
&\leq \frac{\|\Sigma_n - \Sigma\| \cdot \left\|(zI - \Sigma)^{-1}\right\|^2}{1 - (\|\Sigma_n - \Sigma\| \cdot \|(zI - \Sigma)^{-1}\|)} \\
&\leq \frac{8 \cdot \|\Sigma_n - \Sigma\|}{g_{\lambda_j}^2}.
\end{aligned}$$

Using Proposition **??** and Lemma 6, it follows that:

$$\begin{aligned}
\left\|P_{\lambda_j}(\Sigma) - P_{\widehat{\lambda}_j}(\Sigma_n)\right\| &= \frac{1}{2\pi} \left\| \oint_{\partial D(\lambda_j, \frac{g_{\lambda_j}}{2})} Q_\Sigma(z) - Q_{\Sigma_n}(z) dz \right\| \\
&\leq \frac{1}{2\pi} \oint_{\partial D(\lambda_j, \frac{g_{\lambda_j}}{2})} \|Q_\Sigma(z) - Q_{\Sigma_n}(z)\| \, |dz| \\
&\leq \frac{1}{2\pi} \oint_{\partial D(\lambda_j, \frac{g_{\lambda_j}}{2})} \frac{8 \cdot \|\Sigma_n - \Sigma\|}{g_{\lambda_j}^2} |dz| \\
&= \frac{4 \cdot \|\Sigma_n - \Sigma\|}{\min(\lambda_r - \lambda_{r+1}, \lambda_{s-1} - \lambda_s)}.
\end{aligned}$$

□