Introduction to random matrices and tensors

Mohamed El Amine Seddik
melaseddik.github.io

March 2023

Abstract

This series of lectures aims to provide a brief introduction to the theory of random matrices and tensors, with a focus on fundamental concepts and techniques for their analysis. In the first part of the course, we will introduce the notion of the Stieltjes transform, a classical tool for characterizing the spectral behavior of large random matrices. We will then present the proof of the classical semi-circle law and the Marčenko-Pastur law in random matrices, using the resolvent method. Moving on, we will delve into the analysis of spiked random matrices, which arise in various fields such as statistical physics, signal processing, and machine learning. We will present techniques for computing their spectral properties and characterize the related phase transitions. Finally, we will present an extension to the analysis of random tensors, representing higher-order matrices generalizations. We will introduce the notions of singular values and vectors for tensors to define a set of associated random matrix ensemble to random tensors, based on which we will provide an analysis of spiked random tensors using the resolvent method. The lectures will be self-contained and require only a basic knowledge of linear algebra, probability theory, and functional analysis.

1 Properties of the Stieltjes transform and proof of the semi-circle law with the resolvent method

This lecture introduces the notion of the Stieltjes transform and the related properties and provides proof of the semicircle law for the GOE relying on the resolvent method.

1.1 Tools

1.1.1 Weak convergence

We start with the following definition of the notion of weak convergence which shall be used subsequently. Let $\mathcal{C}_b(\mathbb{R})$ be the class of bounded functions on $\mathbb{R}$.

Definition 1.1 (Weak convergence)**.** We say that a sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ if

$$\int_{\mathbb{R}} f(x) d\mu_n(x) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} f(x) d\mu(x)$$

almost surely (a.s.) for any $f \in \mathcal{C}_b(\mathbb{R})$ and we write $\mu_n \xrightarrow[n\to\infty]{w} \mu$.

Property 1.2. If μ_n is a probability measure and $\mu_n \xrightarrow[n\to\infty]{w} \mu$ then μ is also a probability measure. Indeed,

$$1 = \int_{\mathbb{R}} 1 d\mu_n(x) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} 1 d\mu(x) = \mu(\mathbb{R})$$

so we get $\mu(\mathbb{R}) = 1$ by uniqueness of the limit.

1.1.2 The Stieltjes Transform

We will now introduce the main ingredient to study the spectral behavior of large random matrices.

Definition 1.3. Let μ be a probability measure on $\mathbb{R}$. The Stieltjes transform denoted g_μ of μ is defined by

$$\begin{aligned} g_\mu : \quad & \mathbb{C}_+ \to \mathbb{C} \\ & z \mapsto g_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z} \end{aligned}$$

where $\mathbb{C}_+ = \{z \in \mathbb{C} \mid \Im[z] > 0\}$.

Let $z = x + iy$ with $x, y \in \mathbb{R}$. We have that the mapping

$$\lambda \mapsto \frac{1}{\lambda - z} = \frac{\lambda - x}{(\lambda - x)^2 + y^2} + i \frac{y}{(\lambda - x)^2 + y^2} = \Re\left[\frac{1}{\lambda - z}\right] + i\Im\left[\frac{1}{\lambda - z}\right]$$

is continuous and bounded. Since μ is finite, then $g_\mu(z)$ is well-defined over $\mathbb{C}_+$. Furthermore, a Stieltjes transform is identified by the fact that $\Im[g_\mu(z)] > 0$ for all $z \in \mathbb{C}_+$.

Example 1.4. We have the following examples:

1. Let $\mu = \delta_x$ the Dirac measure at point x, then $g_\mu(z) = \frac{1}{x-z}$.
2. Let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$ where λ_i are the eigenvalues of a symmetric matrix M. Then

$$g_{\mu_n}(z) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i - z} = \frac{1}{n}\operatorname{Tr}\left[(M - zI)^{-1}\right]$$

   where $G(z) = (M - zI)^{-1}$ is the so-called *resolvent* matrix of M.

We have the following theorem which shows the equivalence between the weak convergence of a sequence of probability measures and the almost sure convergence of the underlying Stieltjes transforms.

Theorem 1.5. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures. Then*

$$\mu_n \xrightarrow[n\to\infty]{w} \mu \quad \Leftrightarrow \quad g_{\mu_n}(z) \xrightarrow[n\to\infty]{a.s.} g_\mu(z) \quad \text{for all} \quad z \in \mathbb{C}_+$$

An essential aspect of the Stieltjes transform is that there exists an inverse formula that allows the characterization of the underlying probability measure. Indeed, we have the following theorem.

Theorem 1.6 (Inversion Formula)**.** *Let μ be a probability measure. Then, for all $a, b \in \mathbb{R}$, we have*

$$\frac{1}{2}\left[\mu(\{a\}) + \mu(\{b\})\right] + \mu(]a, b[) = \lim_{y\to 0^+} \frac{1}{\pi}\int_a^b \Im\left[g_\mu(x + iy)\right] dx$$

Specifically, the density function of μ is given by $\mu(dx) = \lim_{y\to0^+} \frac{1}{\pi}\Im\left[g_\mu(x + iy)\right] dx$.

Proof. We already know that

$$\Im\left[g_\mu(x + iy)\right] = \int_{\mathbb{R}} \Im\left[\frac{1}{\lambda - x - iy}\right] d\mu(\lambda) = \int_{\mathbb{R}} \frac{y}{(\lambda - x)^2 + y^2} d\mu(\lambda)$$

Thus,

$$\int_a^b \Im\left[g_\mu(x + iy)\right] dx = \int_{\mathbb{R}}\int_a^b \frac{y}{(\lambda - x)^2 + y^2} dx d\mu(\lambda) = \int_{\mathbb{R}} f(y, \lambda) d\mu(\lambda).$$

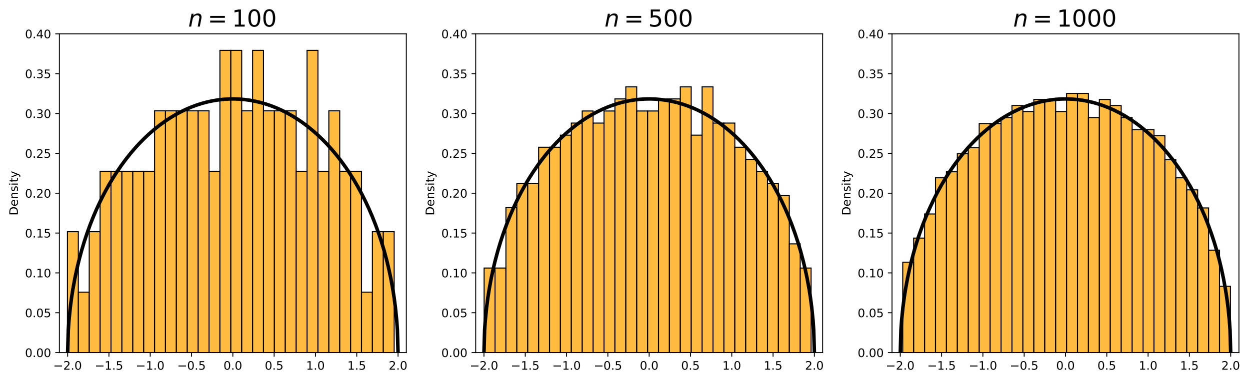

Figure 1: Empirical histogram of the eigenvalues of M from GOE for different values of the dimension n and the limiting semicircle law.

where $f(y,\lambda) = \arctan\left(\frac{b-\lambda}{y}\right) - \arctan\left(\frac{\lambda-a}{y}\right)$. Note that $|f(y,\lambda)| \leq \pi$ for all $y > 0$ and $\lambda \in \mathbb{R}$, and we have $f(y,\lambda) \to f(\lambda)$ as $y \to 0^+$, where

$$f(\lambda) = \begin{cases} 0 & \text{if} \quad \lambda \notin [a,b] \\ \pi/2 & \text{if} \quad \lambda \in \{a,b\} \\ \pi & \text{if} \quad \lambda \in]a,b[\end{cases}$$

Therefore, by the dominated convergence theorem, we have

$$\lim_{y\to 0^+} \frac{1}{\pi} \int_a^b \Im\left[g_\mu(x+iy)\right] dx = \lim_{y\to 0^+} \frac{1}{\pi} \int_{\mathbb{R}} f(y,\lambda) d\mu(\lambda) = \frac{1}{\pi} \int_{\mathbb{R}} f(\lambda) d\mu(\lambda) = \frac{1}{2}\left[\mu(\{a\}) + \mu(\{b\})\right] + \mu(]a,b[).$$

$\square$

1.2 Semicircle law for GOE

We consider the Gaussian Orthogonal Ensemble (GOE) which stands for the ensemble of symmetric random matrices with standard Gaussian i.i.d. entries. Specifically, let $M = \frac{1}{\sqrt{n}}(m_{ij})_{i,j=1}^n$ such that $(m_{ij})_{j\leq i}$ are independent standard Gaussian random variable $\mathcal{N}(0,1)$.

Therefore, we are interested in describing the limiting measure of the *empirical spectral measure* μ_n of M when $n \to \infty$, where

$$\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$$

with $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of M. Indeed, our aim is to prove that

$$\mu_n \xrightarrow[n\to\infty]{w} \mu \quad \text{(a.s.)}$$

where μ is the semicircle law with density function

$$\mu(dx) = \frac{1}{2\pi}\sqrt{4-x^2}dx$$

defined in the compact support $[-2,2]$. Figure 1 illustrates the convergence of μ_n to μ when n grows large.

As we saw from Theorem 1.5, proving $\mu_n \xrightarrow[n\to\infty]{w} \mu$ is equivalent to show that for all $z \in \mathbb{C}_+$

$$g_{\mu_n}(z) \xrightarrow[n\to\infty]{\text{a.s.}} g(z) \tag{1}$$

where g is the Stieltjes transform of μ given by

$$g(z) = \frac{-z+\sqrt{z^2-4}}{2} \quad z \notin [-2,2]$$

1.2.1 Strategy of the proof

The proof of equation 2 requires two steps:

1. First, we show that $g_{\mu_n}(z)$ behaves asymptotically as $\mathbb{E}\left[g_{\mu_n}(z)\right]$ for all $z \in \mathbb{C}_+$, i.e.

$$|g_{\mu_n}(z) - \mathbb{E}\left[g_{\mu_n}(z)\right]| \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

   which can be obtained by concentration inequalities (we will show this in Lecture 2).

2. Prove that, for all $z \in \mathbb{C}_+$, $\mathbb{E}\left[g_{\mu_n}(z)\right] \xrightarrow[n\to\infty]{} g(z)$.

1.2.2 Derivation of step 2

We now consider the proof of step 2. Indeed, by definition, we have

$$\mathbb{E}\left[g_{\mu_n}(z)\right] = \frac{1}{n}\mathbb{E}\left[\operatorname{Tr} G(z)\right]$$

where $G(z) = (M - zI)^{-1}$. Moreover, we have

$$(M - zI)G(z) = I \quad \Rightarrow \quad MG(z) - zG(z) = I$$

Applying the normalized trace operator and tacking the expectation, we get[1]

$$\frac{1}{n}\mathbb{E}\left[\operatorname{Tr}(MG)\right] - \frac{z}{n}\mathbb{E}\left[\operatorname{Tr} G\right] = 1 \quad \Leftrightarrow \quad \frac{1}{n}\mathbb{E}\left[\operatorname{Tr}(MG)\right] - z\mathbb{E}\left[g_{\mu_n}(z)\right] = 1$$

Therefore, we need to develop the term $\mathbb{E}\left[\operatorname{Tr}(MG)\right]$. Indeed, we have

$$\mathbb{E}\left[\operatorname{Tr}(MG)\right] = \frac{1}{\sqrt{n}}\sum_{ij}\mathbb{E}[m_{ij}G_{ji}] = \frac{1}{\sqrt{n}}\sum_{ij}\mathbb{E}\left[\frac{\partial G_{ji}}{\partial m_{ij}}\right]$$

where the last equality is obtained by Stein's lemma, i.e., $\mathbb{E}[Xf(X)] = \mathbb{E}[f'(X)]$ if $X \sim \mathcal{N}(0,1)$. Moreover, using the fact that

$$\frac{\partial G}{\partial m_{ij}} = -G\frac{\partial M}{\partial m_{ij}}G$$

we have $\frac{\partial G_{kl}}{\partial m_{ij}} = -\frac{1}{\sqrt{n}}\left(G_{ki}G_{jl} + G_{li}G_{jk}\right)$ for $i \neq j$. Therefore, we find

$$\mathbb{E}\left[\operatorname{Tr}(MG)\right] = -\frac{1}{n}\sum_{ij}\mathbb{E}\left[G_{ii}G_{jj} + G_{ij}G_{ji}\right] = -\frac{1}{n}\mathbb{E}\left[\operatorname{Tr}(G)^2\right] - \frac{1}{n}\mathbb{E}\left[\operatorname{Tr}(G^2)\right]$$

Besides, since $\frac{1}{(\lambda_i - z)^2} \leq \frac{1}{\Im[z]^2}$, we have

$$\left|\frac{1}{n^2}\mathbb{E}\left[\operatorname{Tr}(G^2)\right]\right| = \left|\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{(\lambda_i - z)^2}\right| \leq \frac{1}{n\Im[z]^2} \xrightarrow[n\to\infty]{} 0$$

and as we will see in Lecture 2 by concentration arguments, the contribution of the term $\mathbb{E}\left[\left(\frac{1}{n}\operatorname{Tr} G\right)^2\right]$ is the same as $\mathbb{E}\left[\left(\frac{1}{n}\operatorname{Tr} G\right)\right]^2$. Plugging all together, we, therefore, find that

$$\mathbb{E}\left[g_{\mu_n}(z)\right]^2 + z\mathbb{E}\left[g_{\mu_n}(z)\right] + 1 \xrightarrow[n\to\infty]{} 0 \quad \text{for all} \quad z \in \mathbb{C}_+$$

[1] We will omit the dependence of G on z in the notations for convenience.

Moreover, be the definition of the Stieltjes transform, $g_{\mu_n}(z)$ is constrained in the closed ball of radius $\frac{1}{|\Im[z]|}$. Then there exists a sub-sequence $g_{\mu_{n_k}}(z)$ that converges to a limit $g(z)$ satisfying

$$g(z)^2 + zg(z) + 1 = 0 \quad \Rightarrow \quad g(z) = \frac{-z \pm \sqrt{z^2-4}}{2}$$

and the Stieltjes transform is $g(z) = \frac{-z+\sqrt{z^2-4}}{2}$ since $\Im[g(z)] > 0$ for $\Im[z] > 0$. Hence, we find

$$\mathbb{E}\left[g_{\mu_n}(z)\right] \xrightarrow[n\to\infty]{} g(z).$$

2 Proof of the Marchenko-Pastur law using the resolvent method in the Gaussian case and universality

2.1 Tools

We recall Stein's lemma in the following Lemma.

Lemma 2.1 (Gaussian integration by part). *Let $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$ be some differentiable function with at most polynomial growth. For $X_1, \ldots, X_n$ being i.i.d. $\mathcal{N}(0,1)$ random variables, we have*

$$\mathbb{E}\left[X_i \mathcal{F}(X_1, \ldots, X_n)\right] = \mathbb{E}\left[\frac{\partial \mathcal{F}}{\partial X_i}\right]$$

The above lemma is a simple generalization of the one-dimensional case. In fact, if $X \sim \mathcal{N}(0,1)$, then

$$\mathbb{E}\left[X\mathcal{F}(X)\right] = \int_{\mathbb{R}} x\mathcal{F}(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \left[-\mathcal{F}(x)e^{-x^2/2}\right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} \mathcal{F}'(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \mathbb{E}\left[\mathcal{F}'(X)\right]$$

We will also be interested in controlling the variance of some $\mathcal{F}(X_1, \ldots, X_n)$ which can be obtained thanks to Poincare's inequality.

Lemma 2.2 (Poincare's inequality). *Let $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$ be some differentiable function with at most polynomial growth. For $X_1, \ldots, X_n$ being i.i.d. $\mathcal{N}(0,1)$ random variables, we have*

$$\mathrm{Var}\left[\mathcal{F}(X_1, \ldots, X_n)\right] \leq \sum_{i=1}^n \mathbb{E}\left|\frac{\partial \mathcal{F}}{\partial X_i}\right|^2$$

Lemma 2.3 (Borel-Cantelli lemma). *Let $(E_n)_n$ be a sequence of events in Ω. If $\sum_{n=1}^\infty \mathbb{P}\left[E_n\right] < \infty$ then $\mathbb{P}\left[\limsup_n E_n\right] = 0$.*

Lemma 2.4 (Block matrix inverse formula). *Let A, D invertible matrices and B, C rectangular matrices, we have*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(A - BD^{-1}C)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}$$

Lemma 2.5 (Trace Lemma). *Let $x \in \mathbb{R}^p$ a random vector with i.i.d. entries with zero mean, unit variance and finite $2k$ moments. For $A \in \mathbb{R}^{p\times p}$ deterministic (or independent from x), we have*

$$\mathbb{E}\left[\left|\frac{1}{p}x^\top Ax - \frac{1}{p}\operatorname{Tr} A\right|^k\right] \leq O\left(\frac{\|A\|^p}{p^{k/2}}\right)$$

In particular, if the spectral norm of A is bounded, then $\frac{1}{p}x^\top Ax - \frac{1}{p}\operatorname{Tr} A \to 0$ as $p \to \infty$ almost surely.

Lemma 2.6 (Rank-one perturbation). *Let $A, B \in \mathbb{R}^p$ symmetric matrices and let μ be the empirical spectral measure of A. For $x \in \mathbb{R}^p$, $\alpha > 0$ and $z \in \mathbb{C} \setminus \mathcal{S}(\mu)$*

$$\left|\frac{1}{p}\operatorname{Tr}\left[B\left(A + \alpha xx^\top - zI_p\right)^{-1}\right] - \frac{1}{p}\operatorname{Tr}\left[B\left(A - zI_p\right)^{-1}\right]\right| \leq \frac{1}{p}\frac{\|B\|}{\mathtt{dist}(z, \mathcal{S}(\mu))}$$

In particular, if the spectral norm of B is bounded, then as $p \to \infty$

$$\frac{1}{p}\operatorname{Tr}\left[B\left(A + \alpha xx^\top - zI_p\right)^{-1}\right] - \frac{1}{p}\operatorname{Tr}\left[B\left(A - zI_p\right)^{-1}\right] \to 0$$

2.2 Concentration of the empirical Stieltjes transform: the case of semicircle law

In this section, we aim to prove the following parts to complete the proof of the semicircle law.

1. Concentration of the empirical Stieltjes transform around its expectation, i.e. for all $z \in \mathbb{C}_+$

$$|g_{\mu_n}(z) - \mathbb{E}\left[g_{\mu_n}(z)\right]| \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

2. Show that $\mathbb{E}\left[g_{\mu_n}(z)^2\right]$ has the same contribution as $\mathbb{E}\left[g_{\mu_n}(z)\right]^2$.

We would like to prove that for all $z \in \mathbb{C}_+$ and any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left[|g_{\mu_n}(z) - \mathbb{E}\left[g_{\mu_n}(z)\right]| \geq \varepsilon\right] < \infty$$

Indeed, through Poincare's inequality, we have

$$\begin{aligned}
\operatorname{Var}\left[g_{\mu_n}(z)\right] &\leq \frac{1}{n^2}\sum_{ij}\mathbb{E}\left|\frac{\partial \operatorname{Tr} G}{\partial m_{ij}}\right|^2 = \frac{1}{n^2}\sum_{ij}\mathbb{E}\left|\sum_{k}\frac{\partial G_{kk}}{\partial m_{ij}}\right|^2 \leq \frac{4}{n^3}\sum_{ij}\mathbb{E}\left|\sum_{k} G_{ki}G_{kj}\right|^2 \\
&= \frac{4}{n^3}\sum_{ij}\mathbb{E}\left|[G^2]_{ij}\right|^2 = \frac{4}{n^3}\mathbb{E}\left[\sum_{i}[G^4]_{ii}\right] = \frac{4}{n^3}\mathbb{E}\left[\operatorname{Tr}(G(z)^4)\right] \leq \frac{4}{n^2\Im[z]^4}
\end{aligned}$$

By Markov's inequality and the Borel-Cantelli lemma, we obtain

$$|g_{\mu_n}(z) - \mathbb{E}\left[g_{\mu_n}(z)\right]| \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

Moreover, since $\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, we obtain

$$\mathbb{E}\left[g_{\mu_n}(z)^2\right] = \mathbb{E}\left[g_{\mu_n}(z)\right]^2 + O_z\left(n^{-2}\right)$$

2.3 Sample covariance in low versus high dimension

In this section, we illustrate the behavior of the sample covariance matrix in a low versus a high dimensional regime. Specifically, let $x_1, \ldots, x_n \in \mathbb{R}^p$ be a set of i.i.d. random vectors such that $x_i \sim \mathcal{N}(0, I_p)$. The maximum likelihood estimator of the population covariance matrix (here I_p) is given by the sample covariance matrix

$$M = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top = \frac{1}{n}XX^\top$$

where $X = [x_1, \ldots, x_n] \in \mathbb{R}^{p\times n}$. With the strong law of large numbers, $M \xrightarrow[n\to\infty]{} I_p$ almost surely or equivalently, the spectral norm $\|M - I_p\| \xrightarrow[n\to\infty]{} 0$. Besides, if $\frac{p}{n} \xrightarrow[n\to\infty]{} \gamma \in (0, \infty)$, then

$$\|M - I_p\| \not\to 0$$

In particular, suppose that $\gamma > 1$, then we have the joint wise convergence

$$\max_{i,j}|(M - I_p)_{ij}| = \max_{i,j}\left|\frac{1}{n}X_{j,:}X_{i,:}^\top - \delta_{ij}\right| \xrightarrow[n\to\infty]{} 0 \quad \text{(a.s.)}$$

However, there is an eigenvalue mismatch since

$$\begin{aligned}
&0 = \lambda_1(M) = \cdots = \lambda_{p-n}(M) \leq \lambda_{p-n+1}(M) \leq \cdots \leq \lambda_p(M) \\
&1 = \lambda_1(I_p) = \cdots = \lambda_p(I_p)
\end{aligned}$$

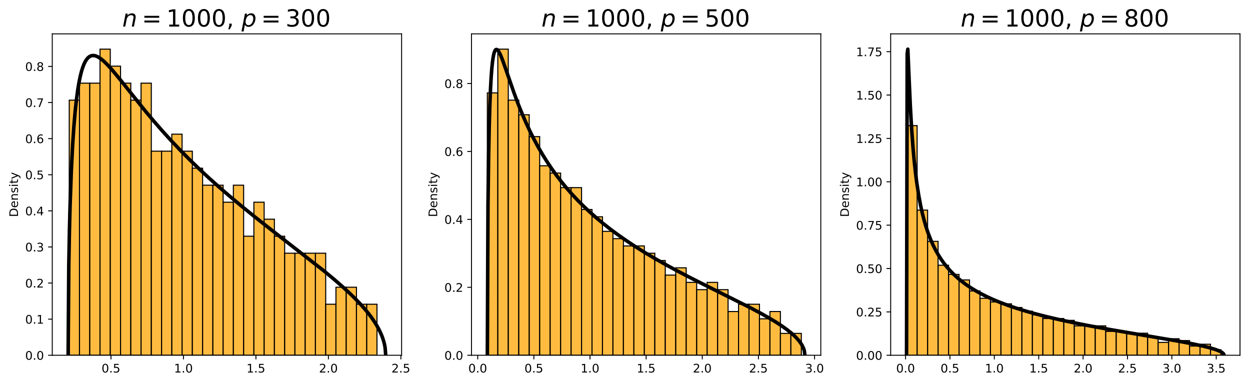

Figure 2: Empirical histogram of the eigenvalues of M from Wishart distribution for different values of the dimensions n, p and the limiting Marchenko-Pastur law.

2.4 Marchenko-Pastur Law

We consider now the following random matrix model

$$M = \frac{1}{n} X X^\top$$

where $X = (X_{ij})_{i,j=1}^{p,n} \in \mathbb{R}^{p \times n}$ is a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Therefore, we are interested in describing the limiting measure of the *empirical spectral measure* μ_n of M when $n \to \infty$ with $\frac{p}{n} \xrightarrow[n\to\infty]{} \gamma \in (0, \infty)$, where

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}$$

with $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of M. Indeed, our aim is to prove that

$$\mu_n \xrightarrow[n\to\infty]{w} \mu_\gamma \quad \text{(a.s.)}$$

where μ_γ is the Marchenko-Pastur law with density function

$$\mu_\gamma(dx) = \left(1 - \frac{1}{\gamma}\right)_+ \delta_0(dx) + \frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2\pi\gamma x} 1_{[\lambda_-, \lambda_+]}(x) dx$$

where $\lambda^\pm = (1 \pm \sqrt{\gamma})^2$. Figure 2 illustrates the convergence of μ_n to μ_γ when n grows large with different ratios p/n.

As we saw from Theorem 1.5, proving $\mu_n \xrightarrow[n\to\infty]{w} \mu_\gamma$ is equivalent to show that for all $z \in \mathbb{C}_+$

$$g_{\mu_n}(z) \xrightarrow[n\to\infty]{} g(z) \quad \text{(a.s.)} \tag{2}$$

where g is the Stieltjes transform of μ_γ given by

$$g(z) = \frac{-[z + (\gamma - 1)] + \sqrt{(z - \lambda^+)(z - \lambda^-)}}{2z\gamma} \qquad z \notin [\lambda_-, \lambda_+]$$

Again, we denote the resolvent of M by

$$G(z) = (M - zI_p)^{-1} \quad \text{for} \quad z \in \mathbb{C}_+$$

We make the assumption that $\frac{p}{n} \xrightarrow[n\to\infty]{} \gamma \in (0,\infty)$. We have the identity

$$MG(z) - zG(z) = I_p \quad \Rightarrow \quad \frac{1}{p}\mathbb{E}\left[\mathrm{Tr}\,(MG)\right] - z\mathbb{E}\left[g_{\mu_n}(z)\right] = 1$$

where $g_{\mu_n}(z) = \frac{1}{p}\operatorname{Tr} G(z)$. Developing the expectation $\frac{1}{p}\mathbb{E}\left[\mathrm{Tr}\,(MG)\right]$, we get

$$\begin{aligned}
\frac{1}{p}\mathbb{E}\left[\mathrm{Tr}\,(MG)\right] &= \frac{1}{np}\sum_{ijk}\mathbb{E}\left[X_{ik}X_{jk}G_{ji}\right] = \frac{1}{np}\sum_{ijk}\mathbb{E}\left[\frac{\partial X_{jk}}{\partial X_{ik}}G_{ji}\right] + \frac{1}{np}\sum_{ijk}\mathbb{E}\left[X_{jk}\frac{\partial G_{ji}}{\partial X_{ik}}\right] \\
&= \frac{1}{np}\sum_{ijk}\mathbb{E}\left[\delta_{ij}G_{ji}\right] + \frac{1}{np}\sum_{ijk}\mathbb{E}\left[X_{jk}\frac{\partial G_{ji}}{\partial X_{ik}}\right] \\
&= \mathbb{E}\left[g_{\mu_n}(z)\right] + \frac{1}{np}\sum_{ijk}\mathbb{E}\left[X_{jk}\frac{\partial G_{ji}}{\partial X_{ik}}\right]
\end{aligned}$$

And from the identity

$$\frac{\partial G}{\partial X_{ij}} = -G\frac{\partial M}{\partial X_{ij}}G$$

We first have

$$\frac{\partial M_{kl}}{\partial X_{ij}} = \frac{1}{n}\sum_m \frac{\partial (X_{km}X_{lm})}{\partial X_{ij}} = \frac{1}{n}\sum_m \frac{\partial X_{km}}{\partial X_{ij}}X_{lm} + \frac{1}{n}\sum_m X_{km}\frac{\partial X_{lm}}{\partial X_{ij}} = \frac{1}{n}\delta_{ik}X_{lj} + \frac{1}{n}X_{kj}\delta_{il}$$

Hence

$$\begin{aligned}
\frac{\partial G_{ab}}{\partial X_{ij}} &= -\sum_{kl} G_{ak}\frac{\partial M_{kl}}{\partial X_{ij}}G_{lb} = -\frac{1}{n}\sum_{kl} G_{ak}\delta_{ik}X_{lj}G_{lb} - \frac{1}{n}\sum_{kl} G_{ak}X_{kj}\delta_{il}G_{lb} \\
&= -\frac{1}{n}G_{ai}(GX)_{bj} - \frac{1}{n}G_{bi}(GX)_{aj}
\end{aligned}$$

Therefore

$$\frac{\partial G_{ij}}{\partial X_{ik}} = -\frac{1}{n}G_{ii}(GX)_{jk} - \frac{1}{n}G_{ij}(GX)_{ik}$$

So, we find

$$\begin{aligned}
&\frac{1}{np}\sum_{ijk}\mathbb{E}\left[X_{jk}\frac{\partial G_{ji}}{\partial X_{ik}}\right] = -\frac{1}{n^2p}\sum_{ijk}\mathbb{E}\left[X_{jk}G_{ii}(GX)_{jk}\right] - \frac{1}{n^2p}\sum_{ijk}\mathbb{E}\left[X_{jk}G_{ij}(GX)_{ik}\right] \\
&= -\frac{1}{n^2p}\mathbb{E}\left[\mathrm{Tr}(G)\,\mathrm{Tr}(GXX^\top)\right] - \frac{1}{n^2p}\mathbb{E}\left[\mathrm{Tr}(GXX^\top G))\right] = -\mathbb{E}\left[\frac{1}{p}\,\mathrm{Tr}(G)\frac{1}{n}\,\mathrm{Tr}\left(G\frac{XX^\top}{n}\right)\right] - \underbrace{\frac{1}{n}\mathbb{E}\left[\frac{1}{p}\,\mathrm{Tr}\left(G\frac{XX^\top}{n}G\right)\right]}_{\xrightarrow[n\to\infty]{}0} \\
&= -\mathbb{E}\left[\frac{1}{p}\,\mathrm{Tr}(G)\frac{1}{n}\,\mathrm{Tr}\left(G\left(M - zI_p + zI_p\right)\right)\right] + O_z(n^{-1}) = -\mathbb{E}\left[\frac{1}{p}\,\mathrm{Tr}(G)\left(\frac{1}{n}\,\mathrm{Tr}(I_p) + \frac{z}{n}\,\mathrm{Tr}(G)\right)\right] + O_z(n^{-1}) \\
&= -\mathbb{E}\left[g_{\mu_n}(z)\left(\frac{p}{n} + \frac{zp}{n}g_{\mu_n}(z)\right)\right] + O_z(n^{-1}) = -\frac{p}{n}\mathbb{E}\left[g_{\mu_n}(z)\right] - z\frac{p}{n}\mathbb{E}\left[g_{\mu_n}(z)^2\right] + O_z(n^{-1})
\end{aligned}$$

Therefore, we find that the limiting Stieltjes transform satisfies the equation

$$\boxed{z\gamma g(z)^2 + (z - 1 + \gamma)g(z) + 1 = 0 \quad \Rightarrow \quad g(z) = \frac{1 - z - \gamma + \sqrt{(z-1+\gamma)^2 - 4z\gamma}}{2z\gamma}}$$

where

$$
\begin{aligned}
(z-1+\gamma)^2-4z\gamma &= z^2+2z(\gamma-1)+(\gamma-1)^2-4z\gamma \\
&= z^2-2z(\gamma+1)+(\gamma-1)^2 \\
&= (z-(\gamma+1))^2-4\gamma \\
&= [z-(\gamma+1+2\sqrt{\gamma})]\,[z-(\gamma+1-2\sqrt{\gamma})] \\
&= (z-\lambda^+)(z-\lambda^-)
\end{aligned}
$$

where $\lambda^{\pm}=(1\pm\sqrt{\gamma})^2$. Hence, we find

$$
g(z)=\frac{-[z+(\gamma-1)]+\sqrt{(z-\lambda^+)(z-\lambda^-)}}{2z\gamma}
$$

Now, we show that

$$
\mu_\gamma(dx)=\left(1-\frac{1}{\gamma}\right)_+\delta_0(dx)+\frac{\sqrt{(\lambda^+-x)(x-\lambda^-)}}{2\pi\gamma x}1_{[\lambda_-,\lambda_+]}(x)dx
$$

- We first show that $\mu_\gamma(\{0\})=\left(1-\frac{1}{\gamma}\right)_+$. Indeed, we have[2] $\mu_\gamma(\{0\})=\lim_{y\to 0}y\Im[g(iy)]$, and since

  $$
  \lim_{y\to 0}\sqrt{(iy-\lambda^+)(iy-\lambda^-)}=-\sqrt{\lambda^+\lambda^-}
  $$

  We have

  $$
  g(iy)=\frac{-[iy+(\gamma-1)]-\sqrt{\lambda^+\lambda^-}+o(y)}{2iy\gamma}
  $$

  Thus

  $$
  \lim_{y\to 0}y\Im[g(iy)]=\frac{\gamma-1}{2\gamma}+\frac{\sqrt{\lambda^+\lambda^-}}{2\gamma}=\frac{1}{2}\left[\left(1-\frac{1}{\gamma}\right)+\left|1-\frac{1}{\gamma}\right|\right]=\left(1-\frac{1}{\gamma}\right)_+
  $$

- Now we compute $\mu_\gamma(]\lambda_-,\lambda_+])$. Let $x\in]\lambda_-,\lambda_+]$ and $z\in\mathbb{C}_+$ with $z\to x$. Thus,

  $$
  |(z-\lambda^+)(z-\lambda^-)|\to(\lambda^+-x)(x-\lambda^-)\quad\text{and}\quad\arg[(z-\lambda^+)(z-\lambda^-)]\to\pi
  $$

  Thus $\lim_{z\to x}\sqrt{(z-\lambda^+)(z-\lambda^-)}=i\sqrt{(\lambda^+-x)(x-\lambda^-)}$, and we find

  $$
  \lim_{z\to x}\frac{1}{\pi}\Im[g(z)]=\frac{\sqrt{(\lambda^+-x)(x-\lambda^-)}}{2\gamma\pi x}
  $$

2.5 Universality with concentration results

This section considers a different approach to proving the convergence of μ_n to the Marchenko-Pastur Law while relaxing the Gaussianity assumption to any distribution with a bounded fourth-order moment. Indeed,

$$
g_{\mu_n}(z)=\frac{1}{p}\operatorname{Tr}(M-zI_p)^{-1}=\frac{1}{p}\sum_{i=1}^{p}\left[\left(\frac{1}{n}XX^\top-zI_p\right)^{-1}\right]_{ii}
$$

Denoting

$$
X=\begin{bmatrix}y^\top\\ Y_{-1}\end{bmatrix}\in\mathbb{R}^{p\times n}
$$

[2]For $z=\rho e^{i\theta}$, $\sqrt{z}=\sqrt{\rho}e^{i\frac{\theta}{2}}$ for $\theta\in(0,2\pi)$. Moreover, $\lim_{\theta\to 0}\sqrt{z}=\sqrt{\rho}$ and $\lim_{\theta\to 2\pi}\sqrt{z}=-\sqrt{\rho}$.

We have

$$\left(\frac{1}{n}XX^\top - zI_p\right)^{-1} = \left(\begin{bmatrix} \frac{1}{n}y^\top y - z & \frac{1}{n}y^\top Y_{-1} \\ \frac{1}{n}Y_{-1}y & \frac{1}{n}Y_{-1}Y_{-1}^\top - zI_{p-1}\end{bmatrix}\right)^{-1}$$

Therefore, by block matrix inverse formula, we have

$$\left[\left(\frac{1}{n}XX^\top - zI_p\right)^{-1}\right]_{11} = \frac{1}{-z - z\frac{1}{n}y^\top\left(\frac{1}{n}Y_{-1}^\top Y_{-1} - zI_n\right)^{-1}y}$$

And, by trace Lemma, as $n, p \to \infty$ we have

$$\left[\left(\frac{1}{n}XX^\top - zI_p\right)^{-1}\right]_{11} - \frac{1}{-z - z\frac{1}{n}\operatorname{Tr}\left(\frac{1}{n}Y_{-1}^\top Y_{-1} - zI_n\right)^{-1}} \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

Since, $X^\top X = Y_{-1}^\top Y_{-1} + yy^\top$, by rank-one perturbation Lemma, we have

$$\left[\left(\frac{1}{n}XX^\top - zI_p\right)^{-1}\right]_{11} - \frac{1}{-z - z\frac{1}{n}\operatorname{Tr}\left(\frac{1}{n}X^\top X - zI_n\right)^{-1}} \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

And from[3] $\frac{1}{n}\operatorname{Tr}\left(\frac{1}{n}X^\top X - zI_n\right)^{-1} = \frac{1}{n}\operatorname{Tr}\left(\frac{1}{n}XX^\top - zI_p\right)^{-1} - \frac{n-p}{n}\frac{1}{z}$, we find

$$\left[\left(\frac{1}{n}XX^\top - zI_p\right)^{-1}\right]_{11} - \frac{1}{1 - \frac{p}{n} - z - z\frac{1}{n}\operatorname{Tr}\left(\frac{1}{n}XX^\top - zI_p\right)^{-1}} \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

Therefore, by repeating for the remaining entries and averaging, we obtain

$$g_{\mu_n}(z) - \frac{1}{1 - \frac{p}{n} - z - z\frac{p}{n}g_{\mu_n}(z)} \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

Hence, $g_{\mu_n}(z) \xrightarrow[n\to\infty]{\text{a.s.}} g(z)$, with $g(z)$ solution of

$$\boxed{g(z) = \frac{1}{1 - \gamma - z - z\gamma g(z)} \quad \Leftrightarrow \quad z\gamma g(z)^2 + (z - 1 + \gamma)g(z) + 1 = 0}$$

3 Analysis of random spiked matrices and characterization of the related phase transitions

3.1 Notations

We denote by $\mathcal{D}(\mu, \sigma^2)$ any distribution with mean μ, variance σ^2 and bounded $2k$ moments for any $k \in \mathbb{N}$. We denote by $\mathcal{B}_n = \{u \in \mathbb{R}^n \mid \|u\| < \infty\}$ the set of vectors in $\mathbb{R}^n$ with bounded ℓ_2 norms. We denote by $\mathbb{S}^{n-1} = \{u \in \mathbb{R}^n \mid \|u\| = 1\}$ the unit sphere in $\mathbb{R}^n$.

3.2 Tools

Definition 3.1 (Deterministic equivalent)**.** Let $G(z) \in \mathbb{R}^{n\times n}$ and $\bar{G}(z) \in \mathbb{R}^{n\times n}$ be random and deterministic resolvents respectively. We say that $\bar{G}(z)$ is a deterministic equivalent of $G(z)$ and we denote $G(z) \sim \bar{G}(z)$ if for any $u, v \in \mathcal{B}_n$ independent of $G(z)$

$$u^\top G(z)v - u^\top \bar{G}(z)v \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

[3]The matrices $XX^\top$ and $X^\top X$ share the same non-zero eigenvalues and $\sum_{i=1}^p \frac{1}{\lambda_i(XX^\top)-z} = \sum_{i=1}^n \frac{1}{\lambda_i(X^\top X)-z} + (p-n)\frac{1}{0-z}$.

Lemma 3.2 (Woodbury matrix identity)**.** *Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{k\times k}$ invertible matrices and $U \in \mathbb{R}^{n\times k}$ and $V \in \mathbb{R}^{k\times n}$. Then*

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

Lemma 3.3 (Sherman-Morison)**.** *Let $A \in \mathbb{R}^{n\times n}$ invertible, $\alpha > 0$ and $u, v \in \mathbb{R}^n$. Then*

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

3.3 Spiked random matrices

This section presents the behavior of spiked random matrices of the form $M = L + W$ where L is a low-rank (signal) matrix and W is random (noise) matrix. These models find many applications in signal processing and modern machine learning. Typically, one is interested in estimating the low-rank component L from a realization of the random matrix M. RMT allows the characterization of the necessary conditions (phase transition) under which signal recovery is possible and further characterizes the performance of such estimation of L.

For instance, for $L = \beta xx^\top$ and W a Wigner random matrix. Computing the principal component of $M = \beta xx^\top + W$, i.e., the eigenvector of M corresponding to its largest eigenvalue provides an estimator u of x and we are typically interested in studying the following:

- Phase transition: Is there a minimum value β_c for β below which signal recovery is impossible?
- Performance recovery: Evaluate the alignment $|u^\top x|^2$ between the true signal x and its estimator u in terms of β.

We will see that RMT allows addressing the above questions. In particular, given λ the largest eigenvalue of M, the alignment $|u^\top x|^2$ can be expressed through a Cauchy integral involving the resolvent of M. Precisely,

$$|u^\top x|^2 = -\frac{1}{2i\pi}\oint_{\mathcal{C}_\lambda} x^\top G(z)x dz$$

where $G(z) = (M - zI_n)^{-1}$ and $\mathcal{C}_\lambda$ is a positively oriented contour surrounding the eigenvalue λ. Specifically, RMT will be useful for providing a consistent estimation of the quadratic form $x^\top G(z)x$, through the notion of *deterministic equivalents* as per Definition 3.1.

3.3.1 Wigner model

We start by analyzing the rank-one spiked Wigner model which is given by

$$M_\beta = \beta xx^\top + W \quad \text{with} \quad W = \frac{1}{\sqrt{n}}X \in \mathbb{R}^{n\times n}$$

where $X_{ij} = X_{ji} \sim \mathcal{D}(0, 1)$, $x \in \mathbb{S}^{n-1}$ and $\beta \geq 0$. We further denote the resolvent of M_β by

$$G_\beta(z) = (M_\beta - zI_n)^{-1}$$

Deterministic equivalent: Our aim now is to find a deterministic equivalent for $G_\beta(z)$. In particular, as we saw in Lecture 1, we first have for any $z \in \mathbb{C}\setminus]-2, 2[$

$$G_0(z) \sim g(z)I_n \quad \text{with} \quad g(z) = \frac{-z + \sqrt{z^2 - 4}}{2}$$

And by Sherman-Morison Lemma, we have

$$G_\beta(z) = G_0(z) - \frac{\beta G_0(z)xx^\top G_0(z)}{1 + \beta x^\top G_0(z)x}$$

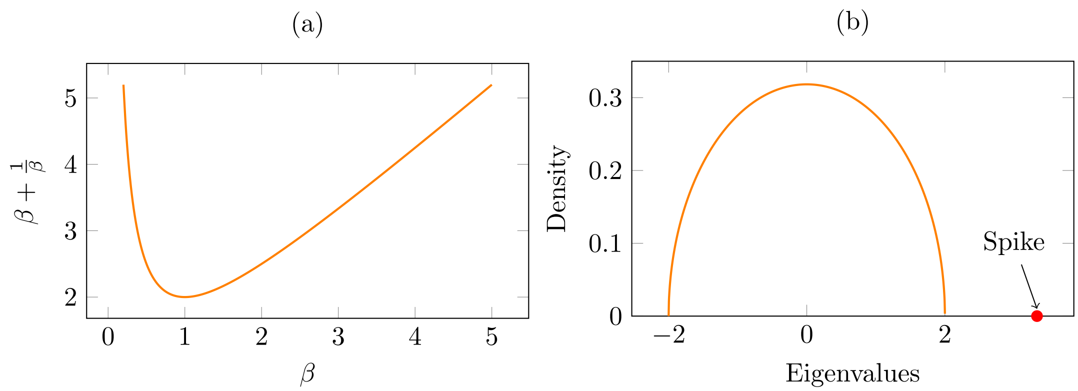

Figure 3: Wigner model - (a) The function $\beta \mapsto \beta + \frac{1}{\beta}$ varying β. (b) Limiting spectral measure and isolated spike for $\beta = 3$.

Let $u, v \in \mathcal{B}_n$, we therefore have

$$u^\top G_\beta(z) v = u^\top G_0(z) v - \frac{\beta u^\top G_0(z) x x^\top G_0(z) v}{1 + \beta x^\top G_0(z) x} \xrightarrow[n \to \infty]{\text{a.s.}} g(z) u^\top v - \frac{\beta g^2(z) u^\top x x^\top v}{1 + \beta g(z) \|x\|^2}$$

Hence, we obtain (since $\|x\| = 1$)

$$\boxed{G_\beta(z) \sim g(z) I_n - \frac{\beta g^2(z)}{1 + \beta g(z)} x x^\top}$$

Phase transition: An isolated eigenvalue (or spike) appears in the spectrum of M_β when the above deterministic equivalent gets singular. Indeed, to find the position of the spike $\lambda_{\max}$, it suffices to solve the equation $1 + \beta g(z) = 0$ for $z \notin (-2, 2)$. In fact,

$$1 + \beta g(z) = 0 \quad \Rightarrow \quad \sqrt{z^2 - 4} = z - \frac{2}{\beta} \quad \Rightarrow \quad \frac{z}{\beta} = \frac{1}{\beta^2} + 1 \quad \Rightarrow \quad \boxed{\lambda_{\max} = \beta + \frac{1}{\beta}}$$

Moreover, to obtain the critical value β_c of β above which a spike appears in the spectrum of M_β, we remark that $\lambda_{\max}$ is a convex function of β (see Fig. 3), therefore β_c corresponds to $\lambda'_{\max}(\beta) = 0$ implying $\boxed{\beta_c = 1}$.

Asymptotic alignment:

$$|u^\top x|^2 = \frac{-1}{2i\pi} \oint_{\mathcal{C}_{\lambda_{\max}}} x^\top G_\beta(z) x dz \xrightarrow[n \to \infty]{\text{a.s.}} \frac{-1}{2i\pi} \oint_{\mathcal{C}_{\lambda_{\max}}} \frac{g(z)}{1 + \beta g(z)} dz = -\texttt{Res}_{\lambda_{\max}} \left\{ \frac{g(z)}{1 + \beta g(z)} \right\}$$

where

$$\texttt{Res}_{\lambda_{\max}} \left\{ \frac{g(z)}{1 + \beta g(z)} \right\} = \lim_{z \to \lambda_{\max}} \frac{(z - \lambda_{\max}) g(z)}{1 + \beta g(z)} = \frac{g(\lambda_{\max})}{\beta g'(\lambda_{\max})}$$

From $1 + \beta g(\lambda_{\max}) = 0$, we have $g(\lambda_{\max}) = -\frac{1}{\beta}$. Since $g(z)$ satisfies $g^2(z) + z g(z) + 1 = 0$, taking the derivative, we have $2g'(z) g(z) + g(z) + z g'(z) = 0$, hence $g'(z) = \frac{-g(z)}{z + 2g(z)}$. Therefore,

$$g'(\lambda_{\max}) = \frac{1/\beta}{\beta + 1/\beta - 2/\beta} = \frac{1}{\beta^2 - 1}$$

From which we obtain

$$\texttt{Res}_{\lambda_{\max}} \left\{ \frac{g(z)}{1 + \beta g(z)} \right\} = \frac{-1/\beta}{\beta / (\beta^2 - 1)} = \frac{1 - \beta^2}{\beta^2} = \frac{1}{\beta^2} - 1$$

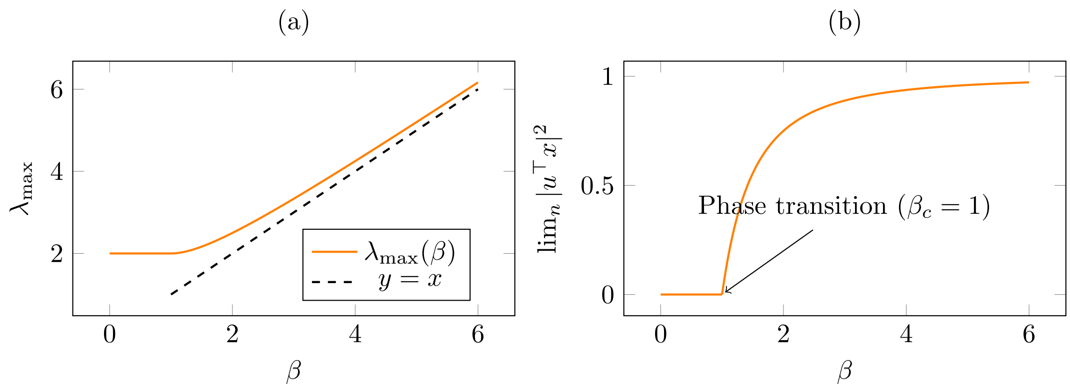

Figure 4: Wigner model - (a) Asymptotic spike $\lambda_{\max}(\beta)$ in terms of β. (b) Asymptotic alignment in terms of β.

Therefore, for $\beta \geq 1$

$$\boxed{|u^\top x|^2 \xrightarrow[n\to\infty]{\text{a.s.}} \left(1 - \frac{1}{\beta^2}\right) 1_{\beta \geq 1}}$$

3.3.2 Wishart model

We now consider the rank-one spiked Wishart model which is given by

$$\begin{aligned} Y_\beta &= \beta x y^\top + \frac{1}{\sqrt{n}} X \in \mathbb{R}^{p\times n} \\ M_\beta &= Y_\beta Y_\beta^\top \end{aligned}$$

where $X_{ij} \sim \mathcal{D}(0,1)$, $x \in \mathbb{S}^{p-1}$, $y \in \mathbb{S}^{n-1}$ and $\beta \geq 0$. Again, we denote the resolvent of M_β by, for $z \in \mathbb{C}_+$

$$G_\beta(z) = (M_\beta - zI_p)^{-1}$$

We further assume that $\frac{p}{n} \xrightarrow[n\to\infty]{} \gamma \in (0, \infty)$ and denote $\lambda^\pm = (1 \pm \sqrt{\gamma})^2$.

Deterministic equivalent: As we saw in Lecture 2, we have for any $z \in \mathbb{C} \setminus (\lambda^-, \lambda^+)$

$$G_0(z) \sim g(z) I_p \quad \text{with} \quad g(z) = \frac{-(z + (\gamma - 1)) + \sqrt{(z-\lambda^+)(z-\lambda^-)}}{2\gamma z}$$

Denote

$$\phi = \frac{1}{\sqrt{n}} X y, \quad U = [\beta x, \phi] \in \mathbb{R}^{p\times 2}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad B^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$$

The matrix M_β expresses as

$$M_\beta = \frac{1}{n} X X^\top + U B U^\top$$

Therefore, by Woodbury matrix identity, we have

$$\begin{aligned} G_\beta(z) &= G_0(z) - G_0(z) U \left[B^{-1} + U^\top G_0(z) U\right]^{-1} U^\top G_0(z) \\ &= G_0(z) - G_0(z) U \begin{pmatrix} \beta^2 x^\top G_0(z) x & 1 + \beta x^\top G_0(z)\phi \\ 1 + \beta x^\top G_0(z)\phi & -1 + \phi^\top G_0(z)\phi \end{pmatrix}^{-1} U^\top G_0(z) \end{aligned}$$

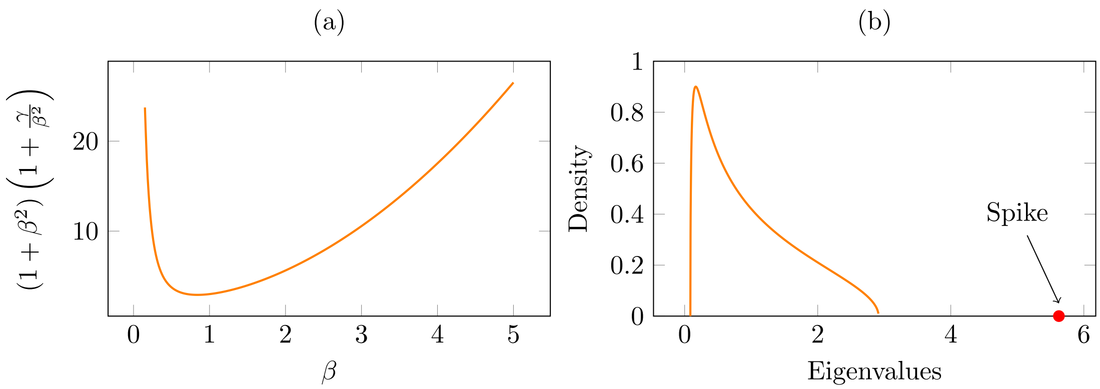

Figure 5: Wishart model - (a) The function $\beta \mapsto (1+\beta^2)\left(1+\frac{\gamma}{\beta^2}\right)$ varying β. (b) Limiting spectral measure and isolated spike for $\beta = 2$ and $\gamma = \frac{1}{2}$.

Denote $\tilde{G}_0(z) = \left(\frac{1}{n}X^\top X - zI_n\right)^{-1}$, we have

$$\begin{aligned} x^\top G_0(z)x &\xrightarrow[n\to\infty]{\text{a.s.}} g(z)x^\top x = g(z) \\ x^\top G_0(z)\phi &= o(1) \\ \phi^\top G_0(z)\phi &= \frac{1}{n}y^\top X^\top G_0(z)Xy = \frac{1}{n}y^\top \tilde{G}_0(z)X^\top Xy = y^\top \tilde{G}_0(z)\left(\frac{1}{n}X^\top X - zI_n + zI_n\right)y \\ &= \|y\|^2 + zy^\top \tilde{G}_0(z)y = 1 + zy^\top \tilde{G}_0(z)y \end{aligned}$$

We can show as in Lecture 2, that $\tilde{G}_0(z) \sim \tilde{g}(z)I_n$ with $\tilde{g}(z) = \gamma g(z) + \frac{\gamma-1}{z}$. Hence, we further have

$$y^\top \tilde{G}_0(z)y \xrightarrow[n\to\infty]{\text{a.s.}} \tilde{g}(z)y^\top y = \tilde{g}(z)$$

From this, we get

$$\begin{pmatrix} \beta^2 x^\top G_0(z)x & 1+\beta x^\top G_0(z)\phi \\ 1+\beta x^\top G_0(z)\phi & -1+\phi^\top G_0(z)\phi \end{pmatrix}^{-1} \xrightarrow[n\to\infty]{\text{a.s.}} \begin{pmatrix} \beta^2 g(z) & 1 \\ 1 & z\tilde{g}(z) \end{pmatrix}^{-1} = \frac{1}{z\beta^2 g(z)\tilde{g}(z) - 1}\begin{pmatrix} z\tilde{g}(z) & -1 \\ -1 & \beta^2 g(z) \end{pmatrix}$$

Therefore, we find the following deterministic equivalent for $G_\beta(z)$

$$G_\beta(z) \sim g(z)I_p - \frac{\beta^2 g^2(z) z\tilde{g}(z)}{\beta^2 z g(z)\tilde{g}(z) - 1}xx^\top$$

Moreover, multiplying $\tilde{g}(z) = \gamma g(z) + \frac{\gamma-1}{z}$ by $zg(z)$, we have

$$zg(z)\tilde{g}(z) = z\gamma g^2(z) + (\gamma-1)g(z)$$

and by the definition of $g(z)$, it satisfies $z\gamma g^2(z) = -1 + (1-z-\gamma)g(z)$. Hence,

$$zg(z)\tilde{g}(z) = -1 + (1-z-\gamma)g(z) + (\gamma-1)g(z) = -(1+zg(z))$$

Thus,

$$\frac{\beta^2 g^2(z) z\tilde{g}(z)}{\beta^2 z g(z)\tilde{g}(z) - 1} = \frac{\beta^2 g(z)(1+zg(z))}{\beta^2(1+zg(z)) + 1} = \frac{\beta^2 g^2(z)}{\beta^2 g(z) + \frac{g(z)}{1+zg(z)}} \quad \text{multiplying } \cdot/\cdot \text{ by } \frac{g(z)}{1+zg(z)}$$

Besides $\frac{g(z)}{1+zg(z)} = 1 + \gamma g(z)$ and we finally get

$$\boxed{G_\beta(z) \sim g(z)I_p - \frac{\beta^2 g^2(z)}{(\beta^2+\gamma)g(z)+1}xx^\top}$$

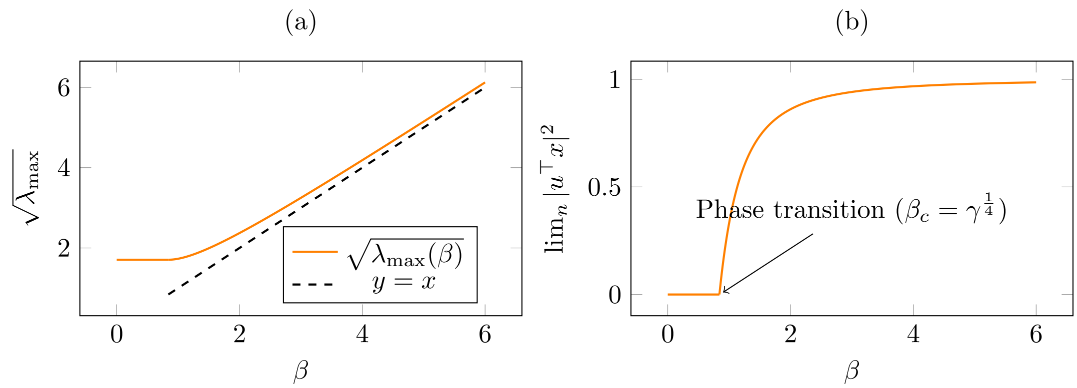

Figure 6: Wishart model - (a) Asymptotic spike $\lambda_{\max}(\beta)$ in terms of β. (b) Asymptotic alignment in terms of β. We choose $\gamma = \frac{1}{2}$.

Phase transition: To find the position of the spike $\lambda_{\max}$, we solve the equation $(\beta^2 + \gamma)g(z) + 1 = 0$ for $z \notin (\lambda^-, \lambda^+)$. We therefore get

$$1 + (\beta^2 + \gamma)g(z) = 0 \quad \Rightarrow \quad z = \beta^2 + \gamma + 1 + \frac{\gamma}{\beta^2} \quad \Rightarrow \quad \boxed{\lambda_{\max} = \left(1 + \beta^2\right)\left(1 + \frac{\gamma}{\beta^2}\right)}$$

And the phase transition can be obtained by solving $\frac{\partial \lambda_{\max}}{\partial \beta} = 0$ (see Fig. 5) which implies $\boxed{\beta_c = \gamma^{\frac{1}{4}}}$.

Asymptotic alignment: Let u be the eigenvector of M_β corresponding to its largest eigenvalue, then

$$|u^\top x|^2 = \frac{-1}{2i\pi} \oint_{\mathcal{C}_{\lambda_{\max}}} x^\top G_\beta(z) x dz \xrightarrow[n\to\infty]{\text{a.s.}} \frac{-1}{2i\pi} \oint_{\mathcal{C}_{\lambda_{\max}}} \frac{(1 + \gamma g(z))g(z)}{1 + (\beta^2 + \gamma)g(z)} dz = -\texttt{Res}_{\lambda_{\max}} \left\{ \frac{(1 + \gamma g(z))g(z)}{1 + (\beta^2 + \gamma)g(z)} \right\}$$

where

$$\texttt{Res}_{\lambda_{\max}} \left\{ \frac{(1 + \gamma g(z))g(z)}{1 + (\beta^2 + \gamma)g(z)} \right\} = \lim_{z \to \lambda_{\max}} \frac{(z - \lambda_{\max})(1 + \gamma g(z))g(z)}{1 + (\beta^2 + \gamma)g(z)} = \frac{(1 + \gamma g(\lambda_{\max}))g(\lambda_{\max})}{(\beta^2 + \gamma)g'(\lambda_{\max})}$$

From $1 + (\beta^2 + \gamma)g(\lambda_{\max}) = 0$, we have $g(\lambda_{\max}) = \frac{-1}{\beta^2 + \gamma}$. Since $g(z)$ satisfies $z\gamma g^2(z) + (\gamma + z - 1)g(z) + 1 = 0$, taking the derivative w.r.t. z, we get $g'(z) = \frac{-g(z)(1 + \gamma g(z))}{2z\gamma g(z) + \gamma + z - 1}$. Putting all together, we obtain

$$\texttt{Res}_{\lambda_{\max}} \left\{ \frac{(1 + \gamma g(z))g(z)}{1 + (\beta^2 + \gamma)g(z)} \right\} = \frac{\gamma - \beta^4}{\beta^2(\beta^2 + \gamma)}$$

Finally,

$$\boxed{|u^\top x|^2 \xrightarrow[n\to\infty]{\text{a.s.}} \frac{1 - \gamma\beta^{-4}}{1 + \gamma\beta^{-2}} 1_{\beta \geq \gamma^{\frac{1}{4}}}}$$

4 Extension to spiked random tensors

Refer to the following material for the derivations:

- Paper: `https://arxiv.org/abs/2112.12348`
- Slides: `https://melaseddik.github.io/files/slides/slides_random_tensors.pdf`