Extremal eigenvalues of sample covariance matrices with general population

JINWOONG KWAK\textsuperscript{1,*}, JI OON LEE\textsuperscript{1,**} and JAEWHI PARK\textsuperscript{1,†}

\textsuperscript{1}Department of Mathematical Sciences, KAIST, Republic of Korea.
E-mail: *jw-kwak@kaist.ac.kr; **joon.lee@kaist.edu; †jw-park@kaist.ac.kr

We consider the eigenvalues of sample covariance matrices of the form $Q = (\Sigma^{1/2}X)(\Sigma^{1/2}X)^*$. The sample $X$ is an $M \times N$ rectangular random matrix with real independent entries and the population covariance matrix $\Sigma$ is a positive definite diagonal matrix independent of $X$. Assuming that the limiting spectral density of $\Sigma$ exhibits convex decay at the right edge of the spectrum, in the limit $M, N \to \infty$ with $N/M \to d \in (0, \infty)$, we find a certain threshold $d_+$ such that for $d > d_+$ the limiting spectral distribution of $Q$ also exhibits convex decay at the right edge of the spectrum. In this case, the largest eigenvalues of $Q$ are determined by the order statistics of the eigenvalues of $\Sigma$, and in particular, the limiting distribution of the largest eigenvalue of $Q$ is given by a Weibull distribution. In case $d < d_+$, we also prove that the limiting distribution of the largest eigenvalue of $Q$ is Gaussian if the entries of $\Sigma$ are i.i.d. random variables. While $\Sigma$ is considered to be random mostly, the results also hold for deterministic $\Sigma$ with some additional assumptions.

\textit{MSC 2010 subject classifications:} 60B20, 62H10, 15B52.
\textit{Keywords:} Sample covariance matrix, deformed Marchenko–Pastur distribution, largest eigenvalue.

1. Introduction

For a vector-valued, centered random variable $y \in \mathbb{R}^M$, its population covariance matrix is given by $\Sigma := \mathbb{E}[yy^T]$. For $N$ independent samples $(y_1, \ldots, y_N)$ of $y$, the sample covariance matrix $\frac{1}{N} \sum_{i=1}^N y_i y_i^T$ can be a simple and unbiased estimator of $\Sigma$ when $N$ is much larger than $M$. On the other hand, if the sample number $N$ is comparable to the population size $M$, the sample covariance matrix is no more a reasonable estimator for the population covariance matrix. Nevertheless, even in such a case, the characteristic of the population covariance matrix may appear in the sample covariance matrix, as we consider in this paper.

We are interested in a matrix of the form

$$Q = (\Sigma^{1/2}X)(\Sigma^{1/2}X)^*, \quad (1.1)$$

where the sample $X$ is an $M \times N$ matrix whose entries are independent real random variables with variance $1/N$, and the general population covariance $\Sigma$ is an $M \times M$ real
diagonal positive definite matrix. We focus on the case that \( M \) and \( N \) tend to infinity simultaneously with \( d := N/M \rightarrow d \in (0, \infty) \), as \( M, N \rightarrow \infty \).

The asymptotic behavior of the empirical spectral distribution (ESD) of sample covariance matrices was first considered by Marchenko and Pastur [22]; they derived a core structure of the limiting spectral distribution (LSD) for a class of sample covariance matrices and the LSD is called the Marchenko–Pastur (MP) law. In the null case, \( \Sigma = I \), the distribution of the rescaled largest eigenvalue converges to the Tracy–Widom law [13, 15, 16, 25]. For the non-null case, i.e. \( \Sigma \neq I \), the location and the distribution of the outlier eigenvalues, including the celebrated BBP transition, have been studied extensively when \( \Sigma \) is a finite rank perturbation of the identity. For more detail, we refer to [2, 4, 5, 23, 24, 27] and references therein.

When \( \Sigma \) has more complicated structure, e.g., the LSD of \( \Sigma \) has no atoms, the limiting distribution of the largest eigenvalue is given by the Tracy–Widom distribution under certain conditions. It was first proved by El Karoui [6] for complex sample covariance matrices and extended to the real case [3, 17, 20]. In these works, one of the key assumptions is that the LSD exhibits the “square-root” type behavior at the right edge of the spectrum, which also appears in the Wigner semicircle distribution or the Marchenko–Pastur distribution. It is then natural to consider the local behavior of the eigenvalues when square-root type behavior is absent. Note that if the LSD of \( \Sigma \) decays concavely at the right edge the LSD of \( Q \) exhibits the square-root behavior at the right edge [14].

**Main contribution**

The sample covariance matrix, as \( N \) gets relatively larger than \( M \), approximates the population covariance matrix more accurately. Thus, it is natural to conjecture that the behavior of the largest eigenvalues of the sample covariance matrix must be similar to that of \( \Sigma \) if \( d \) is above a certain threshold. Our main result of this paper establishes the conjecture rigorously. We find that there exists \( d_{+} \) such that for \( d \) \( d_{+} \)

- The LSD of \( Q \) is convex near the right edge of its support (Theorem 2.7), and
- the distribution of the largest eigenvalues of \( Q \) are determined by the order statistics of the eigenvalues of \( \Sigma \) (Theorem 2.8).

We also prove that the largest eigenvalue of \( Q \) converges to a Gaussian for \( d < d_{+} \), when the entries of \( \Sigma \) are i.i.d. (Theorem 2.9)

**Main idea of the proof**

In the first step, we prove general properties of the LSD of \( Q \). The proof is based on the fact that the LSD of \( Q \) can be defined by a functional equation whose unique solution is the Stieltjes transform of LSD of \( Q \); see also [22].

In the second step, we prove a local law for the resolvents of \( Q \) and \( Q \). The main technical difficulty of the proof stems from that it is not applicable the usual approach based on the self-consistent equation as in [17]. Technically, this is due to the lack of the stability bounds as in equation A.8 of [17], which is not guaranteed when the LSD of \( \Sigma \) decays convexly at the edge. Thus, we adapt the strategy of [19] for deformed Wigner matrices in the analysis of the self-consistent equation. For the analysis of the resolvents, we use the linearization of \( Q \) whose inverse is conveniently related to the resolvents of
Extremal eigenvalues of sample covariance matrices

\( Q \) and \( Q \). Together with Schur’s complement formula and other useful formulas for the resolvents of \( Q \) or \( Q \), we prove a priori estimates for the local law. In the last step, we apply the “fluctuation averaging” argument to control the imaginary part of the resolvent of \( Q \) on much smaller scale than \( N^{-1/2} \). The analysis is different from other works involving the same idea such as [10, 25], due to the unboundedness of the diagonal entries of the resolvent of \( Q \). Finally, by precisely controlling the imaginary part of the argument in the resolvent, we track the location of the eigenvalues at the edge.

**Related works**

In the context of Wigner matrices, the edge behavior of the LSD of a Wigner matrix can be altered by deforming it. The deformed Wigner matrix is of the form \( H = W + \lambda V \) where \( W \) is a Wigner matrix and \( V \) is a real diagonal matrix independent of \( W \). If \( \lambda \) is chosen so that the spectral norm of \( W \) is of comparable order with that of \( V \), and the LSD of \( V \) has convex decay at the edge of its spectrum, then the LSD of \( H \) also exhibits the same decay at the edge if the strength of the deformation \( \lambda \) is above a certain threshold. In that case, the limiting fluctuation of the largest eigenvalues is given by a Weibull distribution instead of the Tracy–Widom distribution. See [18, 19] for more precise statements.

The largest eigenvalues of sample covariance matrices are frequently used in the analysis of signal-plus-noise models. In systems biology, the largest eigenvalues derived from single-cell data sets can be used for identification of biological information [1]. In the context of machine learning, the behavior of the largest eigenvalues indicate different phases of training in deep neural networks [21].

**Organization of the paper**

The rest of the paper is organized as follows: In section 2, we define the model and state the main results. In section 3, we introduce basic notations and tools that will be used in the analysis. In section 4, we prove main theorems. Section 5 is devoted to the proof of Proposition 4.6, one of the key results used in the proof of main theorems. Proofs of some technical lemmas are collected in Supplementary Material.

2. **Definition and Results**

In this section, we define our model and state the main result.

2.1. **Definition of the model**

**Definition 2.1** (Sample covariance matrix with general population). A sample covariance matrix with general population \( \Sigma \) is a matrix of the form

\[
Q := (\Sigma^{1/2}X)(\Sigma^{1/2}X)^*,
\]

(2.1)

where \( X \) and \( \Sigma \) are given as follows:
Let $X$ be an $M \times N$ real random matrix whose entries $(x_{ij})$ are independent, zero-mean random variables with variance $1/N$ and satisfying

$$\mathbb{E}[|x_{ij}|^p] \leq \frac{c_p}{N^{p/2}} \quad (2.2)$$

for some positive constants $c_p > 0$ depending only on $p \in \mathbb{N}$.

Let $\Sigma$ be an $M \times M$ real diagonal matrix whose LSD is $\nu$, entries $(\sigma_\alpha)$ are nonnegative and independent from $X$. Also, the measure $\nu$ has density

$$\rho_\nu(t) = Z^{-1}(1-t)^b f(t) I_{[l,1]}(t), \quad 0 < l < 1 \quad (2.3)$$

where $b > -1$, $f \in C^1[l,1]$ with $f(t) > 0$ for $t \in [l,1]$, and $Z$ is a normalizing constant.

The dimensions $N = N(M)$ and

$$\hat{d} = \frac{N}{M} \to d \in (0, \infty), \quad (2.4)$$

as $n \to \infty$. (For simplicity, we assume that $\hat{d}$ is constant, so we use $d$ instead of $\hat{d}$.)

We denote the eigenvalues of $Q$ by $(\lambda_i)$ with the ordering $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$.

The measure $\nu$ is called a Jacobi measure. We remark that the measure $\nu$ has support $[l,1]$ for some $l > 0$. In this paper, we only consider the case $b > 1$ in (2.3).

Note that in Definition 2.1, we only assume the independence of the entries $(x_{ij})$ and do not assume that $(x_{ij})$ are identically distributed. We mainly assume that $\Sigma$ is random.

**Remark 2.2.** In the sequel, we often interchange $N$ to $CM$ in the middle of several inequalities for some absolute constant $C$ reasoning that $M$ and $N$ have the same order.

**Remark 2.3.** With the assumption on the Jacobi measure, we have that $\liminf \sigma_M \geq l$ and $\limsup \sigma_1 \leq 1$, which were also assumed in [6].

**Remark 2.4.** Let $Q := X^* \Sigma X$, then $Q$ is an $M \times M$ matrix and $Q$ is an $N \times N$. The eigenvalues of $Q$ can be described as the following; $Q$ shares the nonzero eigenvalues with $Q$ and has 0 eigenvalue with multiplicity $N - M$ when $N \geq M$. Thus, we denote the eigenvalues of $Q$ by $(\lambda_i)_{i=1}^N$ where $\lambda_i = 0$ for $M + 1 \leq i \leq N$.

### 2.2. Assumptions on $\Sigma$

For our main result, Theorem 2.8, to hold, it requires that the gaps between the largest eigenvalues $(\sigma_\alpha)$, $\alpha \in [1,n_0]$, of $\Sigma$ must not be too small. Heuristically, when $b > 1$ in (2.3), the Jacobi measure has convex decay at the edge so that we can regard that the distance of immediate eigenvalues is typically large near the edge. Due to the distance, a few largest eigenvalues of $\Sigma$ significantly affect the edge of LSD of $Q$ more than any other small eigenvalues. In order to describe the condition mathematically, we introduce the following event $\Omega$, which is a “good configuration” of the largest eigenvalues of $\Sigma$. 
Denote by $b$ the constant
\begin{equation}
    b := \frac{1}{2} - \frac{1}{b + 1} = \frac{b - 1}{2(b + 1)} = \frac{b}{b + 1} - \frac{1}{2},
\end{equation}
which depends only on $b$ in (2.3). Fix some (small) $\phi > 0$ satisfying
\begin{equation}
    \phi < \left( \frac{10 + b + 1}{b} \right) b,
\end{equation}
and define the domain $D_\phi$ of the spectral parameter $z$ by
\begin{equation}
    D_\phi := \{ z = E + i\eta \in \mathbb{C}^+ : l \leq E \leq 2 + \tau, \ M^{-1/2-\phi} \leq \eta \leq M^{-1/(b+1)+\phi} \}.
\end{equation}
Further, we define $N$-dependent constants $\kappa_0$ and $\eta_0$ by
\begin{equation}
    \kappa_0 := M^{-1/(b+1)}, \quad \eta_0 := \frac{M^{-\phi}}{\sqrt{M}},
\end{equation}
where typical choices for $z \equiv L_+ - \kappa + i\eta$ will be $\kappa$ and $\eta$ with $\kappa \leq M^\phi \kappa_0$ and $\eta \geq \eta_0$.

We are now ready to give the definition of a “good configuration” $\Omega$. Let $\mu_{fc}$ be the limiting spectral measure of $Q$ and $m_{fc}$ the Stieltjes transform of $\mu_{fc}$. (See section 3.2 for the precise definition.) Without loss of generality, we assume that the entries of $\Sigma$ satisfy the following inequality,
\begin{equation}
    \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_M \geq 0.
\end{equation}

**Definition 2.5.** Let $n_0 > 10$ be a fixed positive integer independent of $M$. We define $\Omega$ to be the event on $\Sigma$ which for any $\gamma \in [1, n_0 - 1]$, the following conditions:

1. The $\gamma$-th largest eigenvalue $\sigma_\gamma$ satisfies, for all $\beta \in [1, n_0]$ with $\beta \neq \gamma$,
\begin{equation}
    M^{-\phi} \kappa_0 < |\sigma_\beta - \sigma_\gamma| < (\log M) \kappa_0.
\end{equation}

   In addition, for $\gamma = 1$, we have
\begin{equation}
    M^{-\phi} \kappa_0 < |1 - \sigma_1| < (\log M) \kappa_0,
\end{equation}

   hence for $\alpha \in [n_0 + 1, M]$,
\begin{equation}
    M^{-\phi} \kappa_0 < |\sigma_\alpha - \sigma_\gamma|.
\end{equation}

2. There exists a constant $\epsilon < 1$ independent of $M$ such that for any $z \in D_\phi$ satisfying
\begin{equation}
    \min_{\alpha \in [1, M]} \left| \operatorname{Re} \left( 1 + \frac{1}{\sigma_\alpha m_{fc}} \right) \right| = \left| \operatorname{Re} \left( 1 + \frac{1}{\sigma_\gamma m_{fc}} \right) \right|,
\end{equation}

\[
\frac{1}{N} \sum_{\alpha=1}^{M} \frac{\sigma_{\alpha}^2 |m_{fc}|^2}{|1 + \sigma_{\alpha} m_{fc}|^2} < \epsilon < 1.
\]

(2.14)

We remark that together with (2.10) and (2.11), (2.13) implies that for all \(\alpha \neq \gamma\),

\[
\left| \text{Re} \left(1 + \frac{1}{\sigma_{\alpha} m_{fc}}\right) \right| \geq \frac{M^{-\phi} \kappa_0}{2},
\]

(2.15)

3. For any \(\epsilon > 0\), there exists \(C_\epsilon > 0\) and \(M_\epsilon\) (large) such that for any \(z \in D_\phi\) and \(M \geq M_\epsilon\),

\[
\left| \frac{1}{N} \sum_{\alpha=1}^{M} \sigma_{\alpha} \sigma_{\alpha} m_{fc} + 1 - d^{-1} \int t d\nu(t) \right| \leq \frac{C_\epsilon M^{\phi + \epsilon}}{\sqrt{M}}.
\]

(2.16)

Throughout the paper, we assume that \(\Sigma\) satisfy Definition 2.5, and ESD of \(\Sigma\) converges weakly to a Jacobi measure with \(b > 1\).

**Assumption 2.6.** Let \(\Sigma\) be \(M \times M\) real diagonal random matrix satisfying the conditions in Definition 2.1 with \(b > 1\). We assume that:

i. When \(\Sigma\) is deterministic, \(\Sigma\) satisfies \(\Omega\) in Definition 2.5.

ii. When \(\Sigma\) is random, \(\mathbb{P}(\Omega) \to 1\) as \(M \to \infty\).

We remark that if \(\Sigma\) is a diagonal random matrix whose entries are i.i.d Jacobi measure \(\nu\) with \(b > 1\), the Glivenko–Cantelli theorem asserts that the LSD of \(\Sigma\) converges to \(\nu\) itself. Moreover, the “bad configuration” \(\Omega^c\) occurs rarely, since we show that

\[
\mathbb{P}(\Omega) \geq 1 - C(|\log M|)^{1+2b} M^{-\phi}
\]

(2.17)

in Appendix A. In other words, when \(\Sigma\) has i.i.d. entries with law \(\nu\), it automatically satisfies the properties in Definition 2.5 with high probability. For the non i.i.d random or deterministic \(\Sigma\), we assume Assumption 2.6.

**2.3. Main results**

Our first main result is about the behavior of the limiting spectral measure of \(\mathcal{Q}, \mu_{fc}\), near its right edge. The following theorem establishes not only the explicit location of the right edge of \(\mu_{fc}\) but also the local behavior of \(\mu_{fc}\) near the right edge. We denote by \(L_+\) the right endpoint of the support of \(\mu_{fc}\) and \(\kappa \equiv \kappa(E) := |E - L_+|\) where \(z = E + i\eta\).

**Theorem 2.7.** Suppose that \(\mathcal{Q}\) is a sample covariance matrix with general population \(\Sigma\) defined in Definition 2.1. Let \(\nu\) be a Jacobi measure defined in (2.3) with \(b > 1\). Define

\[
d_+ := \int t^2 d\nu(t) \quad \text{and} \quad \tau_+ := d^{-1} \int t d\nu(t).
\]

(2.18)
If \( d > d_+ \), then \( L_+ = 1 + \tau_+ \). Moreover, for \( 0 \leq \kappa \leq L_+ \), there exists a constant \( C > 1 \) such that

\[
C^{-1} \kappa^b \leq \mu_{fc}(L_+ - \kappa) \leq C\kappa^b.
\] (2.19)

We prove Theorem 2.7 in Section 4.1.

Our second result concerns the locations of the largest eigenvalues of \( Q \) in the super-critical case, which are determined by the order statistics of the eigenvalues of \( \Sigma \). In the following, we fix some \( n_0 \in \mathbb{N} \) independent of \( M \) and consider the largest eigenvalues \( (\lambda_\alpha)_{\alpha=1}^{n_0} \) of \( Q \).

**Theorem 2.8.** Suppose that Assumption 2.6, assumptions in Theorem 2.7 and \( d > d_+ \) hold. Let \( n_0 > 10 \) be a fixed constant independent of \( M \) and let \( 1 \leq \gamma < n_0 \). Then the joint distribution function of the \( \gamma \) largest rescaled eigenvalues,

\[
\mathbb{P}\left(M^1/(b+1)(L_+ - \lambda_1) \leq s_1, M^1/(b+1)(L_+ - \lambda_2) \leq s_2, \ldots, M^1/(b+1)(L_+ - \lambda_\gamma) \leq s_\gamma\right),
\] (2.20)

converges to the joint distribution function of the \( \gamma \) largest rescaled order statistics of \( (\sigma_\alpha) \),

\[
\mathbb{P}\left(C_d M^1/(b+1)(1 - \sigma_1) \leq s_1, C_d M^1/(b+1)(1 - \sigma_2) \leq s_2, \ldots, C_d M^1/(b+1)(1 - \sigma_\gamma) \leq s_\gamma\right),
\] (2.21)

as \( N \to \infty \), where \( C_d = \frac{d-d_+}{d} \). In particular, when \( \Sigma \) has i.i.d. entries with law \( \nu \), the cumulative distribution function of the rescaled largest eigenvalue \( M^1/(b+1)(L_+ - \lambda_1) \) converges to the cumulative distribution function of the Weibull distribution,

\[
G_{b+1}(s) := 1 - \exp\left(-\frac{C_\nu s^{b+1}}{(b+1)}\right),
\] (2.22)

where

\[
C_\nu := \left(\frac{d}{d-d_+}\right)^{b+1} \lim_{t \to 1} \frac{\rho_\nu(t)}{(1-t)^b}.
\] (2.23)

Our third result states that the largest eigenvalue of \( Q \) exhibits Gaussian fluctuation when \( d < d_+ \) and the eigenvalues of \( \Sigma \) are i.i.d. random variables.

**Theorem 2.9** (Gaussian fluctuation for the regime \( d < d_+ \)). Suppose that assumptions in Theorem 2.7 hold except that \( d < d_+ \). Further, assume that the eigenvalues of \( \Sigma \) are i.i.d. random variables. Then, the rescaled fluctuation \( M^{1/2}(\lambda_1 - L_+) \) converges in distribution as \( N \to \infty \) to a centered Gaussian random variable with variance

\[
(d^2 M)^{-1} \left( \int \frac{t\tau}{t+\tau} \, d\nu(t) - \left( \int \frac{t\tau}{t+\tau} \, d\nu(t) \right)^2 \right)
\] (2.24)

where \( \tau \) and \( L_+ \) are defined in the proof.
We prove Theorems 2.8 and 2.9 in Section 4. If \( X \) is Gaussian, our main results still hold for general, non-diagonal \( \Sigma \) satisfying Definition 2.5.

**Corollary 2.10.** Suppose that assumptions in Theorem 2.7 hold with the following changes: the entries of \( X \) are Gaussian, and \( \Sigma \) is not necessarily diagonal, with eigenvalues \( (\sigma_\alpha) \). Then, the results in Theorems 2.7, 2.8, and 2.9 hold without any change.

### 2.4. Numerical experiment

We conduct numerical simulations to observe the local behavior of the empirical spectral distribution of deformed sample covariance matrices. In each simulation done with MATLAB, we generate 10 sample covariance matrices of the form

\[
Q = \frac{1}{N} X^* \Sigma X
\]  

under fixed \( \Sigma \) and plot the histograms of non-zero eigenvalues of \( Q \) to find the behavior of the ESD of \( Q \) at the right edge.

#### 2.4.1. Convex, super-critical case, \( b > 1 \)

We first generate 4000 \( \times \) 6000 matrices \( X \) with i.i.d standard Gaussian entries and a 4000 \( \times \) 4000 diagonal matrix \( \Sigma \) with i.i.d. entries sampled from the density function

\[
f_1(t) = Z_1^{-1} e^t(1 - t)^3 \mathbb{1}_{[1/10,1]}(t) \tag{2.26}
\]

with a normalization constant \( Z_1 \). In this setting, \( d = N/M = 1.5 \) and \( d_+ \approx 0.703908 \). The histogram of nonzero eigenvalues of \( Q \) can be seen from Figure 1a, which shows that the ESD exhibits convex decay at the right edge.

#### 2.4.2. Concave, sub-critical case, \( b > 1 \)

We next generate 4000 \( \times \) 2000 matrices \( X \) with i.i.d. standard Gaussian entries. The diagonal matrix \( \Sigma \) is the same as in Section 2.4.1. In this setting, \( d < d_+ \), and the ESD exhibits concave decay at the right edge, as can be seen from Figure 1b.

#### 2.4.3. Concave case, \( b < 1 \)

In this setting, we generate 4000 \( \times \) 6000 matrices \( X \) with i.i.d standard Gaussian entries again as in Section 2.4.1, but we use a 4000 \( \times \) 4000 diagonal matrix \( \Sigma \) with i.i.d. entries sampled from the density function \( f_2 \) given by

\[
f_2(t) = Z_2^{-1} e^t(1 - t)^{1/2} \mathbb{1}_{[1/10,1]}(t) \tag{2.27}
\]

with a normalization constant \( Z_2 \). Formally, \( d_+ = \infty \) in this case, and the ESD exhibits concave decay at the right edge as in Figure 1c.
3. Preliminaries

In this section, we collect some basic notations and identities.

3.1. Notations

We adopt the following notation introduced in [8] for high-probability estimates:

**Definition 3.1 (Stochastic dominance).** Consider two families

\[ X = (X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}), \quad Y = (Y^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}) \]  

(3.1)

of nonnegative random variables, where \( U^{(N)} \) is a (possibly \( N \)-dependent) parameter set. We say \( X \) is stochastically dominated by \( Y \), uniformly in \( u \), if for all (small) \( \epsilon > 0 \) and (large) \( D > 0 \),

\[ \sup_{u \in U^{(N)}} \mathbb{P}[X^{(N)}(u) > N Y^{(N)}(u)] \leq N^{-D}, \]  

(3.2)

for sufficiently large \( N \geq N_0(\epsilon, D) \). If \( X \) is stochastically dominated by \( Y \), uniformly in \( u \), we write \( X \prec Y \). We also write \( X = O_{\prec}(Y) \) if \(|X| \prec Y \) for some complex family \( X \).

We remark that the relation \( \prec \) is a partial ordering with the arithmetic rules of an order relation; e.g., if \( X_1 \prec Y_1 \) and \( X_2 \prec Y_2 \) then \( X_1 + X_2 \prec Y_1 + Y_2 \) and \( X_1 X_2 \prec Y_1 Y_2 \).

**Definition 3.2 (high probability event).** We say an event \( \Omega \) occurs with high probability if for given \( D > 0 \), \( \mathbb{P}(\Omega) \geq 1 - N^{-D} \) whenever \( N \geq N_0(D) \). Also, we say an event \( \Omega_2 \) occurs with high probability on \( \Omega_1 \) if for given \( D > 0 \), \( \mathbb{P}(\Omega_2 | \Omega_1) \geq 1 - N^{-D} \) whenever \( N \geq N_0(D) \).

Equivalently, \( \Omega \) holds with high probability if \( 1 - \mathbb{1}(\Omega) \prec 0 \). For convenience, we use double brackets to denote the index set, i.e., for \( n_1, n_2 \in \mathbb{R} \),

\[ \llbracket n_1, n_2 \rrbracket := [n_1, n_2] \cap \mathbb{Z}. \]  

(3.3)
Throughout the paper, we use lowercase Latin letters $a, b, \cdots$ for indices in $[1, N]$, uppercase letters $A, B, \cdots$ for indices in $[1, N+M]$, and Greek letters $\alpha, \beta, \cdots$ for indices in $[1, M]$. We also use Greek letters with tilde for indices in $[N+1, N+M]$, e.g., $\tilde{\alpha} = N + \alpha$.

We use the symbols $O(\cdot)$ and $o(\cdot)$ for the standard big-O and little-o notation. The notations $O$, $o$, $\ll$, $\gg$, refer to the limit $N \to \infty$ unless stated otherwise, where the notation $a \ll b$ means $a = o(b)$. We use $c$ and $C$ to denote positive constants that are independent on $N$. Their values may change line by line but in general we do not track the change. We write $a \sim b$, if there is $C \geq 1$ such that $C^{-1} |b| \leq |a| \leq C |b|$.

### 3.2. Deformed Marchenko–Pastur law

As shown in [22], if the empirical spectral distribution (ESD) of $\Sigma$, $\nu_N$, converges in distribution to some probability measure $\nu$, then the ESD of $Q$ converges weakly in probability to a certain deterministic distribution $\mu_{fc}$ which is called the **deformed Marchenko–Pastur law**. It was also proved in [22] that $\mu_{fc}$ can be expressed in terms of its Stieltjes transform as follows:

For a (probability) measure $\omega$ on $\mathbb{R}$, its Stieltjes transform is defined by

$$m_\omega(z) := \int_{\mathbb{R}} \frac{d\omega(x)}{x-z}, \quad (z \in \mathbb{C}^+). \quad (3.4)$$

Note that $m_\omega(z)$ is an analytic function in the upper half plane and $\text{Im} \ m_\omega(z) \geq 0$ for $z \in \mathbb{C}^+$. Let $m_{fc}$ be the Stieltjes transform of $\mu_{fc}$. It was proved in [22] that $m_{fc}$ satisfies the self-consistent equation

$$m_{fc}(z) = \left\{-z + d^{-1} \int_{\mathbb{R}} \frac{t d\nu(t)}{1 + tm_{fc}(z)}\right\}^{-1}, \quad \text{Im} \ m_{fc}(z) \geq 0, \quad (z \in \mathbb{C}^+), \quad (3.5)$$

where $\nu$ is the limiting spectral distribution (LSD) of $\Sigma$. It was also shown that (3.5) has a unique solution. Moreover, $\lim \sup_{\eta \searrow 0} \text{Im} \ m_{fc}(E+i\eta) < \infty$, and $m_{fc}(z)$ determines an absolutely continuous probability measure $\mu_{fc}$ whose density is given by

$$\rho_{fc}(E) = \frac{1}{\pi} \lim_{\eta \searrow 0} \text{Im} \ m_{fc}(E+i\eta), \quad (E \in \mathbb{R}). \quad (3.6)$$

For the properties of $\mu_{fc}$, we refer to [26]. We remark that the density $\rho_{fc}$ is analytic inside its support.

**Remark 3.3.** The measure $\mu_{fc}$ is identified with the multiplicative free convolution of the Marchenko–Pastur measure $\mu_{MP}$ and the measure $\nu$, and denoted by $\mu_{fc} := \nu \boxtimes \mu_{MP}$.

### 3.3. Resolvent and Linearization of $Q$

We define the resolvent, or Green function $G_Q(z)$ of $Q$ and its normalized trace $m_Q(z)$ by

$$G_Q(z) = ((G_Q)_{AB}(z)) := (Q-z)^{-1}, \quad m_Q(z) := \frac{1}{N} \text{Tr} \ G_Q(z), \quad (z \in \mathbb{C}^+). \quad (3.7)$$
We refer to $z$ as the spectral parameter and set $z = E + i\eta$, $E \in \mathbb{R}$, $\eta > 0$.

For the analysis of the resolvent $G_Q(z)$, we use the following linearization trick as in [20]. Define a partitioned $(N + M) \times (N + M)$ matrix

$$H(z) := \begin{bmatrix} -zI_N & X^* \\ X & -\Sigma^{-1} \end{bmatrix}, \quad z \in \mathbb{C}^+$$

(3.8)

where $I_N$ is the $N \times N$ identity matrix. Note that $H$ is invertible, as proved in [20]. Set $G(z) := H(z)^{-1}$ and define the normalized (partial) traces, $m$ and $\tilde{m}$, of $G$ by

$$m(z) := \frac{1}{N} \sum_{a=1}^{N} G_{aa}(z), \quad \tilde{m}(z) := \frac{1}{M} \sum_{\tilde{\alpha} = N+1}^{N+M} G_{\tilde{\alpha}\tilde{\alpha}}. \quad (3.9)$$

With abuse of notation, when we use Greek indices with tilde such as $G_{\tilde{\alpha}\tilde{\alpha}} = G_{N+a,N+a}$, we omit the tilde and set $G_{\alpha\alpha} \equiv G_{\tilde{\alpha}\tilde{\alpha}}$ if it does not causes any confusion.

Frequently, we abbreviate $G \equiv G(z)$, $m \equiv m(z)$, etc. Also, $m(z) = m_Q(z)$ holds as a consequence of the Schur complement formula, see [20]. Furthermore, from (4.1) of [17] and Remark 2.4, we have

$$m(z) = \frac{1}{Nz} \sum_{\alpha} \sigma_{\alpha}^{-1} G_{\alpha\alpha} - \frac{N-M}{Nz}. \quad (3.10)$$

### 3.4. Minors

For $T \subset [1, N + M]$, the matrix minor $H^{(T)}$ of $H$ is defined as

$$(H^{(T)})_{AB} = \mathbb{1}(A \notin T)\mathbb{1}(B \notin T)H_{AB}, \quad (3.11)$$

i.e., the entries in the $T$-indexed columns/rows are replaced by zeros. We define the resolvent $G^{(T)}(z)$ of $H^{(T)}$ by

$$G^{(T)}_{AB}(z) := \left( \frac{1}{H^{(T)} - z} \right)_{AB}. \quad (3.12)$$

For simplicity, we use the notations

$$\sum_{a}^{(T)} := \sum_{a=1}^{N} \mathbb{1}(a \notin T), \quad \sum_{a \neq b}^{(T)} := \sum_{a=1, b=1}^{N} \mathbb{1}(a \neq b, a, b \notin T), \quad \sum_{a}^{(T)} := \sum_{a=1}^{M} \mathbb{1}(a \notin T), \quad \sum_{a \neq \beta}^{(T)} := \sum_{a=1, \beta=1}^{M} \mathbb{1}(a \neq \beta, a, \beta \notin T) \quad (3.13)$$

and abbreviate $(A) = \{A\}$, $(TA) = (T \cup \{A\})$. In Green function entries $(G^{(T)}_{AB})$ we refer to $\{A, B\}$ as lower indices and to $T$ as upper indices.

Finally, we set

$$m^{(T)} := \frac{1}{N} \sum_{a}^{(T)} G^{(T)}_{aa}, \quad \tilde{m}^{(T)} := \frac{1}{M} \sum_{\alpha}^{(T)} G^{(T)}_{\alpha\alpha}. \quad (3.14)$$

Note that we use the normalization $N^{-1}$ instead of $(N - |T|)^{-1}$. 

---

*Extremal eigenvalues of sample covariance matrices*
3.5. Resolvent identities

The next lemma collects the main identities between the matrix elements of $G$ and its minor $G^{(\Sigma)}$.

**Lemma 3.4.** Let $G(z) = H^{-1}(z)$, $z \in \mathbb{C}^+$ be a Green function defined by (3.8) and $\Sigma$ is diagonal. For $a, b \in [1, N]$, $\alpha, \beta \in [1, M]$, $A, B, C \in [1, N + M]$, the following identities hold:

- **Schur complement/Feshbach formula:** For any $a$ and $\alpha$,
  \[ G_{aa} = \frac{1}{-z - \sum_{\alpha, \beta} x_{aa} G_{\alpha \beta}^{(a)} x_{\alpha a}} \quad \text{and} \quad G_{\alpha \alpha} = \frac{1}{-\sigma_{\alpha}^{-1} - \sum_{a, b} x_{aa} G_{ab}^{(a)} x_{\alpha b}}. \]  
  \( (3.15) \)

- For $a \neq b$,
  \[ G_{ab} = -G_{aa} \sum_{\alpha} x_{aa} G_{ab}^{(a)} = -G_{bb} \sum_{\beta} G_{ab}^{(b)} x_{\beta b}. \]  
  \( (3.16) \)

- For $\alpha \neq \beta$,
  \[ G_{\alpha \beta} = -G_{\alpha \alpha} \sum_{a} x_{aa} G_{\alpha \beta}^{(a)} = -G_{\beta \beta} \sum_{b} G_{ab}^{(\beta)} x_{\beta b}. \]  
  \( (3.17) \)

- For any $a$ and $\alpha$,
  \[ G_{aa} = -G_{a \alpha} \sum_{\beta} x_{\beta a} G_{\alpha \beta}^{(a)} = -G_{\alpha \alpha} \sum_{b} G_{ab}^{(\alpha)} x_{ab}. \]  
  \( (3.18) \)

- For $A, B \neq C$,
  \[ G_{AB} = G_{AB}^{(C)} + \frac{G_{AC} G_{CB}}{G_{CC}}. \]  
  \( (3.19) \)

- **Ward identity:** For any $a$,
  \[ \sum_{b} |G_{ab}|^2 = \frac{\text{Im} G_{aa}}{\eta}, \]  
  \( (3.20) \)

where $\eta = \text{Im} z$.

For the proof of Lemma 3.4, we refer to Lemma 4.2 in [11], Lemma 6.10 in [9], and equation (3.31) in [12].

Denote by $E_A$ the partial expectation with respect to the $A$-th column/row of $H$ and set

\[ Z_a := (1 - E_a) (X^* G^{(a)} X)_{aa}, \quad Z_{\alpha} := (1 - E_{\alpha}) (X G^{(\alpha)} X^*)_{\alpha \alpha}. \]  
\( (3.21) \)

Using $Z_A$, we can rewrite $G_{AA}$ as $G_{aa}^{-1} = -z - d^{-1} m^{(a)} - Z_a$ and $G_{\alpha \alpha}^{-1} = -\sigma_{\alpha}^{-1} - m^{(\alpha)} - Z_{\alpha}$.
Lemma 3.5. There is a constant $C$ such that for any $z \in \mathbb{C}^+$ and $A \in [1, N + M]$, 
\[
|m(z) - m^{(A)}(z)| \leq \frac{C}{N\eta}.
\] (3.22)

Furthermore, since $C^{-1}N \leq M \leq CN$, for some constant $C > 0$, we also have 
\[
|m(z) - m^{(A)}(z)| \leq \frac{C}{M\eta}.
\] (3.23)

The lemma follows from Cauchy’s interlacing property of eigenvalues of $H$ and its minor $H^{(A)}$. For a detailed proof we refer to [7]. For $T \subset [1, N + M]$ with, say, $|T| \leq 10$, we obtain 
\[
|m - m^{(T)}| \leq \frac{C}{N\eta}.
\]

3.6. Concentration estimates

For $i \in [1, N]$, let $(X_i)$ and $(Y_i)$, be two families of random variables that 
\[
E R_i = 0, \quad E|R_i|^2 = 1, \quad E|R_i|^p \leq c_p \quad (p \geq 3), \quad (3.24)
\]

$R_i = X_i Y_i$, for all $p \in \mathbb{N}$ and some constants $c_p$, uniformly in $i \in [1, N]$. We collect here some useful concentration estimates.

Lemma 3.6. Let $(X_i)$ and $(Y_i)$ be independent families of random variables and let $(a_{ij})$ and $(b_i)$, $i, j \in [1, N]$, be families of complex numbers. Suppose that all entries $(X_i)$ and $(Y_i)$ are independent and satisfy (3.24). Then we have the bounds 
\[
\left| \sum_i b_i X_i \right| \prec \left( \sum_i |b_i|^2 \right)^{1/2}, \quad (3.25)
\]
\[
\left| \sum_i \sum_j a_{ij} X_i Y_j \right| \prec \left( \sum_{i,j} |a_{ij}|^2 \right)^{1/2}, \quad (3.26)
\]
\[
\left| \sum_i \sum_j a_{ij} X_i X_j - \sum_i a_{ii} X_i^2 \right| \prec \left( \sum_{i,j} |a_{ij}|^2 \right)^{1/2}. \quad (3.27)
\]

If the coefficients $a_{ij}$ and $b_i$ are depend on an additional parameter $u$, then all of the above estimates are uniform in $u$; the threshold $N_0 = N_0(\varepsilon, D)$ in the definition of $\prec$ depends only on the family $(c_p)$ from (3.24). In particular, $N_0$ does not depend on $u$.

We also remark that the following are easily obtained from (2.2) and Remark 2.2: 
\[
|x_{ij}| \prec \frac{1}{\sqrt{N}}, \quad |x_{ij}| \prec \frac{1}{\sqrt{M}}. \quad (3.28)
\]
4. Proof of Main Results

We begin this section by briefly outlining the idea of the proof.

- To prove Theorem 2.7, we follow the strategy in [18]. Instead of directly analyzing the self-consistent equation (3.5), we convert it into an equation of $z$. Then, the location of the right edge of $\mu_{fc}$ and its local behavior can be proved by analyzing the behavior of $z$, which is considered as a function of $m_{fc}$, the Stieltjes transform of $\mu_{fc}$.

- To prove Theorem 2.8, we approximate $m$, the normalized trace of the resolvent, by $m_{fc}$ (Lemma 4.4 and Proposition 5.1). In the approximation, we introduce an intermediate random object $\hat{m}_{fc}$, which can be used to locate the extremal eigenvalues (Proposition 4.6). Combining it with the approximate linearity of $m_{fc}$ (Lemma 4.1), we can prove Theorem 2.8.

- To prove Theorem 2.9, we first show that the location of the right edge of the spectrum exhibits a Gaussian fluctuation of order $M^{-1/2}$ by applying the central limit theorem for a function of the eigenvalues of $\Sigma$. We conclude the proof by showing that the distance between the largest eigenvalue and the right edge is of order $N^{-2/3}$ and hence negligible.

4.1. Proof of Theorem 2.7

**Proof of Theorem 2.7.** Recall (3.5), which we rewrite as follows:

$$z = -\frac{1}{m_{fc}(z)} + d^{-1} \int_{\mathbb{R}} \frac{td\nu(t)}{1 + tm_{fc}(z)},$$  \hspace{1cm} (4.1)

Let $\tau := 1/m_{fc}$, and consider $z$ as a function of $\tau$, which we call $F(\tau)$. We then have

$$F(\tau) := -\tau + d^{-1} \int_{\mathbb{R}} \frac{t\tau d\nu(t)}{\tau + t}. \hspace{1cm} (4.2)$$

Taking imaginary part on the both sides, then

$$\text{Im} F(\tau) = -\text{Im} \tau \left\{ 1 - d^{-1} \int_{\mathbb{R}} \frac{t^2 d\nu(t)}{(\text{Re} \tau + t)^2 + (\text{Im} \tau)^2} \right\}. \hspace{1cm} (4.3)$$

Let

$$H(\tau) := d^{-1} \int_{\mathbb{R}} \frac{t^2 d\nu(t)}{(\text{Re} \tau + t)^2 + (\text{Im} \tau)^2}. \hspace{1cm} (4.4)$$

For any fixed $\text{Re} \tau \in (-1, 0)$, $H(\tau) \to 0$ as $|\text{Im} \tau| \to \infty$, and $H(\tau) \to \infty$ as $|\text{Im} \tau| \to 0$. By monotonicity, there is a unique $\text{Im} \tau > 0$ such that $H(\tau) = 1$ so that $\text{Im} F(\tau) = 0$. 
which corresponds to the bulk of the spectrum. On the other hand, for any fixed $\text{Re} \tau \in (-\infty, -1)$, $H(\tau)$ is monotone decreasing function of $|\text{Im} \tau|$, which implies

$$
\sup_{\text{Re} \tau \in (-\infty, -1)} H(\tau) = H(-1) = d^{-1} \int_{l}^{1} \frac{t^2 \text{d}\nu(t)}{(-1 + t)^2} = \frac{d}{d} < 1, \quad (4.5)
$$

where $l = \inf\{x \in \mathbb{R} : x \in \text{supp } \nu\}$. We thus find that there is no solution of $\text{Im } F(\tau) = 0$ when $\text{Re } \tau \in (-\infty, -1)$, which corresponds to the outside of the spectrum. This shows that $\tau = -1$ at the right edge of the spectrum. It is immediate from (4.1) that $F(-1) = 1 + \tau$, which is the end point we denoted by $L_+$. This proves the first part of Theorem 2.7.

4.2. Definition of $\hat{m}_{fc}$

In this subsection, we introduce $\hat{m}_{fc}$, which will be used as an intermediate random object in the comparison between $m$ and $m_{fc}$. The key property of $\hat{m}_{fc}$ is that it directly depends on $\Sigma$ unlike $m_{fc}$, but it does not depend on $X$.

Let $\hat{\nu}$ be the ESD of $\Sigma$, i.e.,

$$
\hat{\nu} := \frac{1}{M} \sum_{\alpha=1}^{M} \delta_{\sigma_{\alpha}}. \quad (4.6)
$$

We define $\hat{m}_{fc} = \hat{m}_{fc}(z)$ as a solution to the self-consistent equation

$$
\hat{m}_{fc}(z) = \left\{ -z + \frac{1}{N} \sum_{\alpha=1}^{M} \frac{\sigma_{\alpha}}{\sigma_{\alpha} \hat{m}_{fc}(z) + 1} \right\}^{-1}, \quad \text{Im } \hat{m}_{fc}(z) \geq 0, \quad (z \in \mathbb{C}^+). \quad (4.7)
$$

Similarly to (3.5), equation (4.7) also has the unique solution, which is the Stieltjes transform of a probability measure, $\hat{\mu}_{fc}$, which is absolutely continuous. The random measure $\hat{\nu} \boxtimes \mu_{MP}$, which is the multiplicative free convolution between $\hat{\nu}$ and the Marchenko–Pastur law $\mu_{MP}$, can be recovered from $\hat{m}_{fc}$ by using the Stieltjes inversion formula (3.6).

4.3. Properties of $m_{fc}$ and $\hat{m}_{fc}$

Recall the definitions of $m_{fc}$ and $\hat{m}_{fc}$. Let

$$
R_2(z) := d^{-1} \int \frac{t^2|m_{fc}|^2 \text{d}\nu(t)}{|tm_{fc}(z) + 1|^2}, \quad \hat{R}_2(z) := \frac{1}{N} \sum_{\alpha=1}^{M} \frac{\sigma_{\alpha}^2 |\hat{m}_{fc}|^2}{|\sigma_{\alpha} \hat{m}_{fc}(z) + 1|^2}, \quad (z \in \mathbb{C}^+). \quad (4.8)
$$

Recall from (3.5) that

$$
\frac{1}{m_{fc}} = -z + d^{-1} \int \frac{t \text{d}\nu(t)}{tm_{fc} + 1}. \quad (4.9)
$$
Taking imaginary part and rearranging, we have that
\[ 1 = \text{Im} \, z \cdot \left| m_{fc} \right|^2 + d^{-1} \int t^2 \left| m_{fc} \right|^2 d\nu(t) \left/ \|m_{fc}(z) + 1\|^2 \right. \]  \quad (4.10)
This in particular shows that \( 0 \leq R_2(z) < 1 \), and by similar manner we also find that \( 0 \leq \hat{R}_2(z) < 1 \). We also note that the self-consistent equation (3.5) implies \( |m_{fc}| \sim 1 \).

In the following lemma, we show that \( 1/m_{fc} \) is approximately a linear function of \( z \) near the right edge.

**Lemma 4.1.** Let \( z = L_+ - \kappa + i\eta \in \mathcal{D}_\phi \). Then,
\[
\frac{1}{m_{fc}(z)} = -1 + \frac{d}{d - d_+}(L_+ - z) + O \left( (\log M)(\kappa + \eta)^{\min\{b, 2\}} \right). \quad (4.11)
\]
Similarly, if \( z, z' \in \mathcal{D}_\phi \), then
\[
\frac{1}{m_{fc}(z)} - \frac{1}{m_{fc}(z')} = -\frac{d}{d - d_+}(z - z') + O \left( (\log M)^2 (N^{-1/(b+1)})^{\min\{b-1, 1\}}|z - z'| \right). \quad (4.12)
\]

The proof of Lemma 4.1 is presented in Appendix B.

**Remark 4.2.** Lemma 4.1 reveals the local behavior of \( 1/m_{fc} \) at the right edge. For \( z_\alpha := L_+ - \frac{d - d_+}{d}(1 - \sigma_\alpha) + i\eta \), we obtain
\[
\frac{1}{m_{fc}(z_\alpha)} = -\sigma_\alpha + i\frac{d}{d - d_+}\eta + O \left( (\log M) M^{-\min\{b, 2\} / (b+1) + 2\phi} \right). \quad (4.13)
\]

We consider the following subset of \( \mathcal{D}_\phi \) to estimate the difference \( |\hat{m}_{fc} - m_{fc}| \).

**Definition 4.3.** Let \( A := [n_0, M] \). We define the domain \( \mathcal{D}'_\phi \) of the spectral parameter \( z \) as
\[
\mathcal{D}'_\phi = \left\{ z \in \mathcal{D}_\phi : \left| 1 + \frac{1}{\sigma_\alpha m_{fc}} \right| > \frac{1}{2} M^{-1/(b+1) - \phi}, \forall \alpha \in A \right\}. \quad (4.14)
\]

In the sequel, we show that \( \mathcal{D}'_\phi \) contains \( z = \lambda_\alpha + i\eta_0 \in \mathbb{C}^+ \) for \( \alpha \in [1, n_0 - 1] \) with high probability. See Remark 4.7.

Recall that \( \sigma_1 > \sigma_2 > \ldots > \sigma_M \). We now show that \( \hat{m}_{fc}(z) \) approximates \( m_{fc}(z) \) well for \( z \in \mathcal{D}'_\phi \). For technical reason, we compare the reciprocals of \( m_{fc} \) and \( \hat{m}_{fc} \), which makes the estimate more convenient when compared to estimating \( |m - \hat{m}_{fc}| \) directly.

**Lemma 4.4.** For any \( z \in \mathcal{D}'_\phi \),
\[
\left| \frac{1}{\hat{m}_{fc}(z)} - \frac{1}{m_{fc}(z)} \right| < \frac{1}{M\eta_0} = M^{-1/2 + \phi}. \quad (4.15)
\]
The detail of the proof is given in Appendix C.

**Remark 4.5.** The estimate on $|m_{fc} - \hat{m}_{fc}|$ easily follows from Lemma 4.4. To see this, we first observe that $m_{fc} \sim 1$ implies $m_{fc}^{-1} \sim 1$. Combining with Lemma 4.4 above, we also find that $\hat{m}_{fc} \sim 1$. Since $|m_{fc} - \hat{m}_{fc}| < M^{-1/2+\phi}$, we get the following estimate with high probability:

$$|m_{fc} - \hat{m}_{fc}| \leq C M^{-1/2+2\phi}. \quad (4.16)$$

### 4.4. Proof of Theorem 2.8

In this subsection, we prove Proposition 4.8, which would directly imply Theorem 2.8. The key idea is that we can approximate $(\lambda_\gamma)$ in terms of $(\sigma_\gamma)$ by applying the properties of $\hat{m}_{fc}$ in Section 4.3 and hence we can estimate the locations of the largest eigenvalues $(\lambda_\gamma)$ of $Q$ by $(\sigma_\gamma)$. The precise statement for the idea is the following proposition.

**Proposition 4.6.** Let $n_0 > 10$ be a fixed integer independent of $M$ and $\gamma \in [1, n_0 - 1]$. Suppose that the assumptions in Theorem 2.8 hold. Then, with $\eta_0$ defined in (2.8), the following holds with high probability:

$$\text{Re} \frac{1}{\hat{m}_{fc}(\lambda_\gamma + i\eta_0)} = -\sigma_\gamma + O(M^{-1/2+3\phi}), \quad (4.17)$$

We postpone the proof of Proposition 4.6 to Section 5.

**Remark 4.7.** Since $|\sigma_\alpha - \sigma_\gamma| \geq M^{-\phi} \kappa_0 \gg M^{-1/2+3\phi}$ for all $\alpha \neq \gamma$ by (2.10), Proposition 4.6 implies that

$$\left| 1 + \text{Re} \frac{1}{\sigma_\alpha \hat{m}_{fc}(\lambda_\gamma + i\eta_0)} \right| \geq \left| \text{Re} \frac{1}{\sigma_\alpha \hat{m}_{fc}(\lambda_\gamma + i\eta_0)} - \text{Re} \frac{1}{\sigma_\gamma \hat{m}_{fc}(\lambda_\gamma + i\eta_0)} \right| \geq \frac{\kappa_0}{2}. \quad (4.18)$$

Hence, we find that $\lambda_\gamma + i\eta_0 \in \mathcal{D}_{\phi}', \gamma \in [1, n_0 - 1]$ with high probability.

We now prove Theorem 2.8 by proving the following proposition.

**Proposition 4.8.** Suppose that the assumptions in Proposition 4.6 hold. Then there exists a constant $C$ such that with high probability

$$\left| \lambda_\gamma - \left( L_+ - \frac{d - d_+}{d} (1 - \sigma_\gamma) \right) \right| \leq \frac{C}{M^{1/(b+1)}} \left( \frac{M^{3\phi}}{M^b} + \frac{(\log M)^2}{M^{1/(b+1)}} \right). \quad (4.19)$$
Proof of Theorem 2.8 and Proposition 4.8. From Lemma 4.4 and Proposition 4.6, with high probability
\[
\text{Re} \left( \frac{1}{m_{fc}(\lambda_\gamma + i\eta_0)} \right) = -\sigma_\gamma + O(M^{-\frac{1}{2} + 3\phi}). \tag{4.20}
\]
Recall we have proved in Lemma 4.1 that
\[
\frac{1}{m_{fc}(\lambda_\gamma + i\eta_0)} = -1 + \frac{d}{d-d_+}(L_+ - \lambda_\gamma - i\eta_0) + O(\kappa_0^{\min\{b,2\}}(\log M)^2). \tag{4.21}
\]
Thus,
\[
\text{Re} \frac{1}{m_{fc}(\lambda_\gamma + i\eta_0)} = -1 + \frac{d}{d-d_+}(L_+ - \lambda_\gamma) + O(\kappa_0^{\min\{b,2\}}(\log M)^2). \tag{4.22}
\]
We now have with high probability that
\[
\lambda_\gamma = -(1-\sigma_\gamma)\frac{d-d_+}{d} + L_+ + O(\kappa_0^{\min\{b,2\}}(\log M)^2) + O(M^{-1/2 + 3\phi}), \tag{4.23}
\]
which completes the proof of Proposition 4.8.

To prove Theorem 2.8, we notice that the distribution of the largest eigenvalue of \(\Sigma\) is given by the order statistics of \((\sigma_\alpha)\). The Fisher–Tippett–Gnedenko theorem asserts that the limiting distribution of the largest eigenvalue of \(\Sigma\) is a member of either Gumbel, Fréchet or Weibull family, and in our case it is the Weibull distribution. This completes the proof of Theorem 2.8. \(\square\)

The following corollary provides an estimate on the speed of the convergence

Corollary 4.9. Suppose that the assumptions in Proposition 4.6 hold. Then, there exists a constant \(C_1\) such that for \(s \in \mathbb{R}^+\) and any sufficiently large \(N\),
\[
\mathbb{P} \left( M^{1/(b+1)} \frac{d-d_+}{d}(1-\sigma_\gamma) \leq s - C_1 \left( \frac{M^{3\phi}}{M^{b + (1/b+1)}} \right) \right) - C_1 \frac{(\log M)^{1+2b}}{M^\phi} \leq \mathbb{P} \left( M^{1/(b+1)} (L_+ - \lambda_\gamma) \leq s \right) \leq \mathbb{P} \left( M^{1/(b+1)} \frac{d-d_+}{d}(1-\sigma_\gamma) \leq s + C_1 \left( \frac{M^{3\phi}}{M^{b + (1/b+1)}} \right) \right) + C_1 \frac{(\log M)^{1+2b}}{M^\phi}, \tag{4.24}
\]

Remark 4.10. The constants in Proposition 4.8 and Corollary 4.9 depend only on \(d\), the measure \(\nu\), and the constant \(c_p\) in (2.2); in particular, they do not depend on the detailed structure of the sample \(X\).
4.5. Proof of Theorem 2.9

In this subsection, we prove Theorem 2.9 that holds in the case $d < d_+$ and the entries of $\Sigma$ are i.i.d. random variables. Recall that $\tilde{\mu}_{fc} := \tilde{\nu} \boxtimes \mu_{MP}$ and $L_+$ is the right edge of the support of $\mu_{fc}$.

**Proof.** Following the proof in [6, 18], we find that $L_+$ is the solution of the equations

\[
\frac{1}{m_{fc}(L_+)} = -L_+ + d^{-1} \int \frac{td\nu(t)}{1 + tm_{fc}(L_+)} , \quad d^{-1} \int \left| \frac{tm_{fc}(L_+)}{1 + tm_{fc}(L_+)} \right|^2 d\nu(t) = 1 \quad (4.25)
\]

and similarly $\tilde{L}_+$ is the solution of the equations

\[
\frac{1}{\tilde{m}_{fc}(\tilde{L}_+)} = -\tilde{L}_+ + \frac{1}{N} \sum_{\alpha=1}^M \frac{\sigma_\alpha}{1 + \sigma_\alpha \tilde{m}_{fc}(\tilde{L}_+)} , \quad \frac{1}{N} \sum_{\alpha=1}^M \left| \frac{\sigma_\alpha \tilde{m}_{fc}(\tilde{L}_+)}{1 + \sigma_\alpha \tilde{m}_{fc}(\tilde{L}_+)} \right|^2 = 1 . \quad (4.26)
\]

Let $\tau = 1/m_{fc}(L_+)$ and $\tilde{\tau} := 1/\tilde{m}_{fc}(\tilde{L}_+)$. Since $d < d_+$, we assume that

\[
d^{-1} \int \frac{t^2d\nu(t)}{(1 - t)^2} > 1 + \delta , \quad \frac{1}{N} \sum_{\alpha=1}^M \frac{\sigma_\alpha^2}{(1 - \sigma_\alpha)^2} > 1 + \delta \quad (4.27)
\]

for some $\delta > 0$, where the second inequality holds with high probability. From the assumption, we find that $\tau, \tilde{\tau} < -1$. Thus,

\[
0 = \frac{1}{N} \sum_{\alpha=1}^M \frac{\sigma_\alpha^2}{(\bar{\tau} + \sigma_\alpha)^2} - 1 = \frac{1}{N} \sum_{\alpha=1}^M \frac{\sigma_\alpha^2}{(\bar{\tau} + \sigma_\alpha)^2} - \frac{1}{N} \sum_{\alpha=1}^M \frac{\sigma_\alpha^2}{(\bar{\tau} + \sigma_\alpha)^2} + O(M^{-1/2})
\]

\[
= \frac{1}{N} \sum_{\alpha=1}^M \frac{(2\sigma_\alpha + \tau + \tilde{\tau})(\tau - \tilde{\tau})}{(\bar{\tau} + \sigma_\alpha)^2(\sigma_\alpha)^2} + O(M^{-1/2}) . \quad (4.28)
\]

We also notice that $2\sigma_\alpha + \tau + \tilde{\tau} < 0$. Further, with high probability, $|\{\sigma_\alpha : \sigma_\alpha < 1/2\}| > cN$ for some constant $c > 0$ independent of $N$. Hence,

\[
- \frac{1}{N} \sum_{\alpha=1}^M \frac{2\sigma_\alpha + \tau + \tilde{\tau}}{(\bar{\tau} + \sigma_\alpha)^2(\sigma_\alpha)^2} > c' > 0 , \quad (4.29)
\]

for some constant $c'$ independent of $N$. Together with (4.28), we thus find that

\[
\tau - \tilde{\tau} = O(M^{-1/2}) . \quad (4.30)
\]

We now have that

\[
\tilde{\tau} + \tilde{L}_+ = \frac{1}{N} \sum_{\alpha=1}^M \frac{\tilde{\tau}_\alpha}{\tilde{\tau}_\alpha} = \frac{1}{N} \sum_{\alpha=1}^M \frac{\tau_\alpha}{\tau + \sigma_\alpha} + \frac{1}{N} \sum_{\alpha=1}^M \frac{\sigma_\alpha^2}{(\tau + \sigma_\alpha)^2} (\tilde{\tau} - \tau) + O(M^{-1})
\]

\[
= L_+ + \tau + Y + (\tilde{\tau} - \tau) + O(M^{-1}) , \quad (4.31)
\]
where the random variable $Y$ is defined by
\[
Y := \frac{1}{N} \sum_{\alpha=1}^{M} \frac{\tau \sigma_{\alpha}}{\tau + \sigma_{\alpha}} - d^{-1} \int \frac{t \tau}{t + \tau} d\nu(t) = \frac{1}{N} \sum_{\alpha=1}^{M} \left( \frac{\tau \sigma_{\alpha}}{\tau + \sigma_{\alpha}} - \mathbb{E} \left[ \frac{\tau \sigma_{\alpha}}{\tau + \sigma_{\alpha}} \right] \right)
\]
\[
= \frac{d^{-1}}{M} \sum_{\alpha=1}^{M} \left( \frac{\tau \sigma_{\alpha}}{\tau + \sigma_{\alpha}} - \mathbb{E} \left[ \frac{\tau \sigma_{\alpha}}{\tau + \sigma_{\alpha}} \right] \right)
\] (4.32)

By the central limit theorem, $Y$ converges to a centered Gaussian random variable with variance
\[
(d^2 M)^{-1} \left\{ \int \left| \frac{t \tau}{t + \tau} \right|^2 d\nu(t) - \left( \int \frac{t \tau}{t + \tau} d\nu(t) \right)^2 \right\}.
\] (4.33)

Since $\hat{L}_+ - L_+ = Y + O(M^{-1})$, this completes the proof of the desired lemma.

With Lemma 4.4, adapting the idea of the proof of Lemma A.4 in [18], we find that $1 + tm_{fc}(z) \sim 1$ and hence $1 + \sigma_{\gamma} \hat{m}_{fc}(z) \sim 1$. Thus, our model satisfies Condition 1.1 in [3] so that Theorem 4.1 therein holds and we get
\[
|L_+ - \lambda_1| \asymp M^{-2/3}.
\] (4.34)

From $|\hat{L}_+ - L_+| \sim M^{-1/2}$, we find that the fluctuation of the largest eigenvalue is dominated by the Gaussian distribution in Theorem 2.9. Moreover, we also have proved the sharp transition between the Gaussian limit and Weibull limit as $d$ crosses $d_+$.

5. Estimates on the Location of the Eigenvalues

In this section, our main object is the proof of Proposition 4.6. Let $\hat{E}_\gamma \in \mathbb{R}$ be a solution $E = \hat{E}_\gamma$ to the equation
\[
1 + \text{Re} \frac{1}{\sigma_{\gamma} \hat{m}_{fc}(E + i\eta_0)} = 0
\] (5.1)
where $\gamma \in [1, n_0 - 1]$ and $\eta_0$ is defined in (2.8). Considering Lemma 4.1 and Lemma 4.4, it is easy to check that there is at least one such $\hat{E}_\gamma$. If there are multiple solutions to (5.1), we choose the largest one as $\hat{E}_\gamma$ and set $\hat{z}_\gamma := \hat{E}_\gamma + i\eta_0$.

The key argument in the proof of Proposition 4.6 is similar to that of section 5 of [19]. The main idea is that when $\mu_{fc}$ has a convex decay (see Theorem 2.7.), the imaginary part of $m(z)$ has a peak if and only if
\[
\text{Im} \left( \frac{\sigma_{\gamma} \hat{m}_{fc}(z)}{1 + \sigma_{\gamma} \hat{m}_{fc}(z)} \right), \quad (z \in \mathbb{C}^+),
\] (5.2)
becomes large enough for some $\gamma \in [1, n_0 - 1]$. Furthermore, since the locations of the eigenvalues $\lambda_\gamma$ are correspond to the positions of the peaks of $\text{Im} m$, we are able to estimate the location of the $\gamma$-th largest eigenvalue in terms of $\sigma_\gamma$. 

\[ \hat{z}_\gamma \]
5.1. Properties of $\hat{m}_{fc}$ and $m$

In order to prove Proposition 4.6, we need an a priori estimate on the difference between $m(z)$ and $\hat{m}_{fc}(z)$ so-called “local law” where $z$ is close to the edge. However, it is more convenient to consider the difference between their reciprocal rather than dealing with $|m(z) - \hat{m}_{fc}(z)|$ directly. After that, we can use the fact that $|\hat{m}_{fc}|$ is bounded away from zero to recover the order of $|m(z) - \hat{m}_{fc}(z)|$. Recall the constant $\phi > 0$ in (2.6) and the definition of the domain $D'_{\phi}$ in (4.14). In the proof of Proposition 4.6, we will use the following local law as an a priori estimate.

Proposition 5.1. [Local law near the edge] We have on $\Omega$ that

$$\left| \frac{1}{m(z)} - \frac{1}{\hat{m}_{fc}(z)} \right| < \frac{1}{M\eta_0},$$

(5.3)

for all $z \in D'_{\phi}$.

Remark 5.2. Since we have $\hat{m}_{fc} \sim 1$, the Proposition 5.1 implies

$$|m(z) - \hat{m}_{fc}(z)| \lesssim \frac{1}{M\eta_0}. \quad (5.4)$$

The proposition is proved in Appendix C. In the rest of this section, we gather some properties of $\hat{m}_{fc}(z)$ and estimate $\Im m(z)$ when $z = E + i\eta_0 \in D'_{\phi}$.

Recall the definitions of $(\hat{z}_\gamma)$ in (5.1). We begin by deriving a basic property of $\hat{m}_{fc}(z)$ near $(\hat{z}_\gamma)$. Recall the definition of $\eta_0$ in (2.8).

Lemma 5.3. For $z = E + i\eta_0 \in D'_{\phi}$, the following hold on $\Omega$:

1. if $|z - \hat{z}_\gamma| \geq M^{-1/2+3\phi}$ for all $\gamma \in \{1, n_0 - 1\}$, then there exists a constant $C > 1$ such that

$$C^{-1}\eta_0 \leq -\Im \frac{1}{\hat{m}_{fc}(z)} \leq C\eta_0. \quad (5.5)$$

2. if $z = \hat{z}_\gamma$ for some $\gamma \in \{1, n_0 - 1\}$, then there exists a constant $C > 1$ such that

$$C^{-1}M^{-1/2} \leq -\Im \frac{1}{\hat{m}_{fc}(z)} \leq CM^{-1/2}. \quad (5.6)$$

The proof of Lemma 5.3 is given in Appendix C. Now we estimate the imaginary part of $m(z)$ for the smallest $\eta = \eta_0$.

Lemma 5.4. We have on $\Omega$ that, for all $z = E + i\eta_0 \in D'_{\phi}$,

$$\Im m(z) \prec \frac{1}{M\eta_0}. \quad (5.7)$$
Since the proof of Lemma 5.4 is closely related to that of Proposition 5.1, we present it in Appendix C also. As a corollary of above lemma, we have a bound for \(Z_a\) and \(Z_{\alpha}\) in (3.21). The concentration estimate implies that

\[
|Z_{\alpha}| \ll \sqrt{\frac{\text{Im} m^{(\alpha)}}{N\eta}}, \quad |Z_a| \ll \sqrt{\frac{\text{Im} \tilde{m}^{(\alpha)}}{M\eta}}.
\]  (5.8)

The relation (3.10) and Lemma 3.5 (the Cauchy interlacing property) implies that

\[
|Z_{\alpha}| \ll \sqrt{\frac{\text{Im} m}{N\eta} + \frac{1}{N\eta}}, \quad |Z_a| \ll \sqrt{\frac{\text{Im} m}{M\eta} + \frac{1}{M\eta}}.
\]  (5.9)

Hence, as a corollary of Lemma 5.4, we obtain:

**Corollary 5.5.** We have on \(\Omega\) that for all \(z = E + i\eta_0 \in D'_\phi\),

\[
\max_A |Z_A(z)| \ll \frac{1}{M\eta_0}, \quad \max_A |Z_A^{(B)}(z)| \ll \frac{1}{M\eta_0}, \quad (B \in \llbracket 1, N + M \rrbracket). \]  (5.10)

### 5.2. Estimates on \(|\tilde{m} - \tilde{m}^{(\alpha)}|\)

Since we need a more precise estimate on the difference \(|\text{Im} m(z) - \text{Im} \tilde{m}_{fc}(z)|\), we construct tighter estimates on \(|\tilde{m} - \tilde{m}^{(\alpha)}|\) and \(N^{-1} \sum Z_A\). In this section, we provide enhanced bound on the difference \(|\tilde{m} - \tilde{m}^{(\alpha)}|\).

**Lemma 5.6.** The following bounds hold on \(\Omega\) for all \(z = E + i\eta_0 \in D'_\phi\): For given \(z\), choose \(\gamma \in \llbracket 1, n_0 - 1 \rrbracket\) such that (2.13) is satisfied. Then, for any \(\alpha \neq \gamma, \alpha \in \llbracket 1, M \rrbracket\),

\[
|\tilde{m} - \tilde{m}^{(\gamma)}| \ll (M\eta_0)^{-1} = M^{-1/2 + \phi},
\]  (5.11)

\[
|\tilde{m}(z) - \tilde{m}^{(\alpha)}(z)| \ll M^{-1 + 1/(b+1) + 4\phi},
\]  (5.12)

and \( |\tilde{m}^{(\gamma)}(z) - \tilde{m}^{(\gamma\alpha)}(z)| \ll M^{-1 + 1/(b+1) + 4\phi} \).  (5.13)

The proof of Lemma 5.6 is given in Appendix E.

### 5.3. Estimates on \(N^{-1} \sum Z_a\) and \(N^{-1} \sum Z_{\alpha}\)

Recall that \(n_0 > 10\) is an integer independent of \(M\). In the following lemmas, we control the fluctuation averages \(\frac{1}{N} \sum_{a=1}^N Z_a\), \(\frac{1}{N} \sum_{\alpha=n_0}^M Z_{\alpha}\) and other weighted average sums.

**Lemma 5.7.** For all \(z = E + i\eta \in D'_\phi\), the following bound holds on \(\Omega\):

\[
\left| \frac{1}{N} \sum_a Z_a \right| \ll \left( \frac{1}{M\eta_0} \right)^2.
\]  (5.14)
Lemma 5.8. For all $z \in \mathcal{D}_\phi'$, the following bounds hold on $\Omega$:

$$\left| \frac{1}{N} \sum_{\alpha=n_0}^{M} Z_\alpha(z) \right| \prec M^{-1/2-b/2+2\phi}, \quad (5.15)$$

and, for $\gamma \in \{1, n_0 - 1\}$,

$$\left| \frac{1}{N} \sum_{\alpha=n_0, \alpha \neq \gamma}^{M} Z_\alpha^{(\gamma)}(z) \right| \prec M^{-1/2-b/2+2\phi}. \quad (5.16)$$

Corollary 5.9. For all $z \in \mathcal{D}_\phi'$, the following bounds hold on $\Omega$:

$$\left| \frac{1}{N} \sum_{\alpha=n_0}^{M} \frac{\tilde{m}_{fc}(z)^2 Z_\alpha(z)}{(\sigma_\alpha^{-1} + \tilde{m}_{fc}(z))^2} \right| \prec M^{-1/2-b/2+2\phi}, \quad (5.17)$$

and, for $\gamma \in \{1, n_0 - 1\}$,

$$\left| \frac{1}{N} \sum_{\alpha=n_0, \alpha \neq \gamma}^{M} \frac{\tilde{m}_{fc}(z)^2 Z_\alpha^{(\gamma)}(z)}{(\sigma_\alpha^{-1} + \tilde{m}_{fc}(z))^2} \right| \prec M^{-1/2-b/2+2\phi}. \quad (5.18)$$

Lemma 5.7, Lemma 5.8 and Corollary 5.9 are proved in Appendix D.

Remark 5.10. The bounds we obtained in Lemma 5.6, Lemma 5.7, Lemma 5.8, and Corollary 5.9 are $o(\eta)$. This will be used on several occasions in the next section.

5.4. Proof of Proposition 4.6

Recall the definition of $(\hat{z}_\gamma)$ in (5.1). We first estimate $\text{Im} \ m(z)$ for $z = E + i\eta_0$ satisfying $|z - \hat{z}_\gamma| \geq M^{-1/2+3\phi}$, for all $\gamma \in \{1, n_0 - 1\}$.

Lemma 5.11. There exists a constant $C > 1$ such that the following bound holds with high probability on $\Omega$: For any $z = E + i\eta_0 \in \mathcal{D}_\phi'$, satisfying $|z - \hat{z}_\gamma| \geq M^{-1/2+3\phi}$ for all $\gamma \in \{1, n_0 - 1\}$, we have

$$C^{-1}\eta_0 \leq \text{Im} \ m(z) \leq C\eta_0. \quad (5.19)$$

This implies that the order of the imaginary part of $m(z)$ is $\eta$ when $z$ is sufficiently far from $\hat{z}_\gamma$. The proof is postponed to Appendix E.

As a next step, we show that $\text{Im} \ m^{(\gamma)}(z) \sim \eta$ even though when $z$ is close to $\hat{z}_\gamma$. Furthermore, we find a point $\tilde{z}_\gamma$ close to $\hat{z}_\gamma$ such that the imaginary part of $m(\tilde{z}_\gamma)$ is much larger than $\eta$. The proof is similar to that of Lemma 5.11 and Lemma 5.12 can be found in Appendix E.
Lemma 5.12. There exists a constant $C > 1$ such that the following bound holds with high probability on $\Omega$, for all $z = E + i\eta_0 \in \mathcal{D}_\phi'$: For given $z$, choose $\gamma \in [1, n_0 - 1]$ such that (2.13) is satisfied. Then, we have
\[
C^{-1}\eta_0 \leq \text{Im} m^{(\gamma)}(z) \leq C\eta_0.
\] (5.20)

Corollary 5.13. The following bound holds on $\Omega$, for all $z = E + i\eta_0 \in \mathcal{D}_\phi'$: For given $z$, choose $\gamma \in [1, n_0 - 1]$ such that (2.13) is satisfied. Then, we have
\[
|Z_\gamma| \ll \frac{1}{\sqrt{M}}.
\] (5.21)

Now we are able to locate the points for which $\text{Im} m(z) \gg \eta_0$ near the edge.

Lemma 5.14. For any $\gamma \in [1, n_0 - 1]$, there exists $\tilde{E}_\gamma \in \mathbb{R}$ such that the following holds with high probability on $\Omega$: If we let $\tilde{z}_\gamma := \tilde{E}_\gamma + i\eta_0$, then $|\tilde{z}_\gamma - \tilde{z}_\gamma| \leq M^{-1/2 + 3\phi}$ and $\text{Im} m(\tilde{z}_\gamma) \gg \eta_0$.

Proof. From the assumption in (2.13), Corollary 5.5, and Proposition 5.1, we attain that
\[
d^{-1}\tilde{m} = \frac{G_{\gamma\gamma}}{N} + \frac{1}{N} \sum_{\alpha}^{(\gamma)} \frac{-1}{\sigma_\alpha^{-1} + m^{(\alpha)}} + Z_\alpha = \frac{G_{\gamma\gamma}}{N} + \frac{1}{N} \sum_{\alpha}^{(\gamma)} \frac{-1}{\sigma_\alpha^{-1} + m} + o(\eta_0)
\] (5.22)

with high probability on $\Omega$. (We refer Appendix E.2 for details.) Consider
\[
-\frac{1}{G_{\gamma\gamma}} = \sigma_\gamma^{-1} + m^{(\gamma)} + Z_\gamma.
\] (5.23)

Setting $z^+ := \tilde{z}_\gamma + N^{-1/2 + 3\phi}$, Lemma 4.1 shows that
\[
\text{Re} \frac{1}{m_{fc}(z^+_\gamma)} - \text{Re} \frac{1}{m_{fc}(\tilde{z}_\gamma)} \leq -CM^{-1/2 + 3\phi},
\] (5.24)
on $\Omega$. Thus, from Lemma 4.4 and the definition of $\tilde{z}_k$, we find that
\[
\text{Re} \frac{1}{m_{fc}(z^+_\gamma)} + \sigma_\gamma \leq -CM^{-1/2 + 3\phi},
\] (5.25)
on $\Omega$. Similarly, if we let $z^- := \tilde{z}_\gamma - M^{-1/2 + 3\phi}$, we have that
\[
\text{Re} \frac{1}{m_{fc}(z^-_\gamma)} + \sigma_\gamma \geq CM^{-1/2 + 3\phi},
\] (5.26)
on $\Omega$. Since
\[
-\frac{1}{G_{\gamma\gamma}} = \frac{\tilde{m}_{fc}}{\sigma_\gamma} \left( \sigma_\gamma + \frac{1}{\tilde{m}_{fc}} + o(M^{-1/2 + 3\phi}) \right),
\] (5.27)
Extremal eigenvalues of sample covariance matrices

with high probability on \( \Omega \), we find that there exists \( \tilde{z}_\gamma = \tilde{E}_\gamma + i\eta_0 \) with \( \tilde{E}_\gamma \in (\tilde{E}_\gamma - M^{-1/2+3\delta}, \tilde{E}_\gamma + M^{-1/2+3\delta}) \) such that \( \text{Re} G_{\gamma\gamma}(\tilde{z}_\gamma) = 0 \). When \( z = \tilde{z}_\gamma \), we have from Lemma 5.12 and Corollary 5.13 that on \( \Omega \),

\[
\text{Im} G_{\gamma\gamma}(\tilde{z}_\gamma) = \frac{1}{|\text{Im} m^{(\gamma)}(\tilde{z}_\gamma) + \text{Im} Z(\tilde{z}_\gamma)|} \geq M^{1/2-\delta}, \quad \text{Re} G_{\gamma\gamma}(\tilde{z}_\gamma) = 0. \tag{5.28}
\]

From (5.22), we obtain that

\[
d^{-1}\text{Im} \tilde{m}(\tilde{z}_\gamma) = \frac{\text{Im} G_{\gamma\gamma}(\tilde{z}_\gamma)}{N} + 1 \sum_{\alpha}^{(\gamma)} \frac{\text{Im} m(\tilde{z}_\gamma)}{\sigma_\alpha^{-1} + m(\tilde{z}_\gamma)} + o(\eta_0). \tag{5.29}
\]

Considering the definition of \( G_{aa} \), Lemma 5.6, Lemma 5.7 and Lemma 5.8, we have that

\[
-\text{Im} \frac{1}{m(\tilde{z}_\gamma)} = \eta_0 + \text{Im} d^{-1} \tilde{m}(\tilde{z}_\gamma) + o(\eta_0). \tag{5.30}
\]

(We refer Appendix E.2 for details.) Hence we have

\[
(1 - K^{(\gamma)}_{m})\text{Im} \left\{ -\frac{1}{m(\tilde{z}_\gamma)} \right\} = \eta + \frac{\text{Im} G_{\gamma\gamma}(\tilde{z}_\gamma)}{N} + o(\eta_0). \tag{5.31}
\]

Since \( K^{(\gamma)}_{m} < c < 1 \) for some constant \( c \), with high probability on \( \Omega \), we get

\[
-\text{Im} \frac{1}{m(\tilde{z}_\gamma)} \geq M^{-\delta/2}M^{-1/2} \gg \eta_0, \tag{5.32}
\]

from (5.29), with high probability on \( \Omega \), which was to be proved.

We now turn to the proof of Proposition 4.6. Recall that we denote by \( \lambda_\gamma \) the \( \gamma \)-th largest eigenvalue of \( Q \), \( \gamma \in \{1, n_0 - 1\} \). Also recall that \( \kappa_0 = M^{-1/(b+1)} \); see (2.8).

**Proof of Proposition 4.6.** First, we consider the case \( \gamma = 1 \). From the spectral decomposition of \( Q \), we have

\[
\text{Im} m(E + i\eta_0) = \frac{1}{N} \sum_{i=1}^{N} \frac{\eta_0}{(\lambda_i - E)^2 + \eta_0^2}, \tag{5.33}
\]

and \( \text{Im} m(\lambda_1 + i\eta_0) \geq (M\eta_0)^{-1} \gg \eta_0 \). Recall the definition of \( \tilde{z}_1 = \tilde{E}_1 + i\eta_0 \) in (5.1). Since, with high probability on \( \Omega \), \( \text{Im} m(z) \sim \eta_0 \) for \( z \in D_0 \) satisfying \( |z - \tilde{z}_1| \geq M^{-1/2+3\delta} \), as we proved in Lemma 5.11, we obtain that \( \lambda_1 < \tilde{E}_1 + M^{-1/2+3\delta} \).

Recall the definitions for \( \tilde{z}_1 \) and \( z_1^- \) in the proof of Lemma 5.14. Assume \( \lambda_1 < \tilde{E}_1 - M^{-1/2+3\delta} \), then \( \text{Im} m(E + i\eta_0) \) is a decreasing function of \( E \) on the interval \( (\tilde{E}_1 - M^{-1/2+3\delta}, \tilde{E}_1 + M^{-1/2+3\delta}) \). However, we already have shown in Lemma 5.11 and Lemma 5.14 that with high probability, \( \text{Im} m(\tilde{z}_1) \gg \eta_0 \), \( \text{Im} m(z_1^-) \sim \eta_0 \), and...
Re $\tilde{z}_1 > \text{Re} \tilde{z}_1^-$. It contradicts to previous assumption, so $\lambda_1 \geq \hat{E}_1 - M^{-1/2+3\phi}$. Now Lemma 4.1 and Lemma 4.4, together with Lemma 5.3 conclude that

$$\frac{1}{\hat{m}_{fc}(\lambda_1 + i\eta_0)} = \frac{1}{\hat{m}_{fc}(\tilde{z}_1)} + O(M^{-1/2+3\phi}) = -\sigma_1 + O(M^{-1/2+3\phi}),$$

which proves the proposition for the special choice $\gamma = 1$.

Next, we consider the case $\gamma = 2$; with induction, the other cases can be shown by similar manner. Consider $H^{(1)}$, the minor of $H$ obtained by removing the first row and column and denote the largest eigenvalue of $H^{(1)}$ by $\lambda_1^{(1)}$. The Cauchy’s interlacing property implies $\lambda_2 \leq \lambda_1^{(1)}$. Following the first part of the proof, we obtain

$$\hat{E}_2 - M^{-1/2+3\phi} \leq \lambda_1^{(1)} \leq \hat{E}_2 + M^{-1/2+3\phi},$$

where we let $\hat{z}_2 = \hat{E}_2 + i\eta_0$ be a solution to the equation

$$\sigma_2 + \text{Re} \frac{1}{\hat{m}_{fc}(\hat{z}_2)} = 0.$$ (5.36)

This shows that

$$\lambda_2 \leq \hat{E}_2 + M^{-1/2+3\phi}.$$ (5.37)

To prove the lower bound, we may argue as in the first part of the proof. Recall that we have proved in Lemma 5.11 and Lemma 5.14 that with high probability on $\Omega$,

1. For $z = \tilde{z}_2 - M^{-1/2+3\phi}$, we have $\text{Im} m(z) \leq C\eta_0$.
2. There exists $\tilde{z}_2 = \tilde{E}_2 + i\eta_0$ such that $|\tilde{z}_2 - \tilde{z}_2| \leq M^{-1/2+3\phi}$ with $\text{Im} m(\tilde{z}_2) \gg \eta_0$.

If $\lambda_2 < \hat{E}_2 - M^{-1/2+3\phi}$, then

$$\text{Im} m(E + i\eta_0) - \frac{1}{N} \frac{\eta_0}{(\lambda_i - E)^2 + \eta_0^2} = \frac{1}{N} \sum_{i=2}^{N} \frac{\eta_0}{(\lambda_i - E)^2 + \eta_0^2}$$ (5.38)

is a decreasing function of $E$. Since we know that with high probability on $\Omega$,

$$\frac{1}{N} \frac{\eta_0}{(\lambda_1 - E)^2 + \eta_0^2} \leq \frac{1}{N} \frac{C\eta_0}{M^{-2\phi} \kappa_0^2} \ll \eta_0,$$ (5.39)

we have $\text{Im} m(\tilde{z}_2) \leq C\eta_0$, which contradicts to the definition of $\tilde{z}_2$. Thus, we find that $\lambda_2 \geq \hat{E}_2 - M^{-1/2+3\phi}$ with high probability on $\Omega$.

We now proceed as above to conclude that, with high probability on $\Omega$,

$$\frac{1}{\hat{m}_{fc}(\lambda_2 + i\eta_0)} = \frac{1}{\hat{m}_{fc}(\tilde{z}_2)} + O(M^{-1/2+3\phi}) = -\sigma_2 + O(M^{-1/2+3\phi}),$$ (5.40)

which proves the proposition for $\gamma = 2$. The general case is proven in the same way. □
Acknowledgements

We thank Paul Jung for helpful discussions. The work of J. Kwak was partially supported by National Research Foundation of Korea under grant number NRF-2017R1A2B2001952. The work of J. O. Lee and J. Park was partially supported by National Research Foundation of Korea under grant number NRF-2019R1A5A1028324.

References

[1] L. Aparicio, M. Bordyuh, A. J. Blumberg, and R. Rabadan. A random matrix theory approach to denoise single-cell data. *Patterns*, 1(3):100035, 2020.
[2] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.
[3] Z. Bao, G. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.*, 43(1):382–421, 2015.
[4] A. Bloemendal and B. Virág. Limits of spiked random matrices I. *Probab. Theory Related Fields*, 156(3-4):795–825, 2013.
[5] A. Bloemendal and B. Virág. Limits of spiked random matrices II. *Ann. Probab.*, 44(4):2726–2769, 2016.
[6] N. El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.
[7] L. Erdös. Universality of Wigner random matrices. In *XVIIth International Congress on Mathematical Physics*, pages 86–105. World Sci. Publ., Hackensack, NJ, 2010.
[8] L. Erdös, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14(8):1837–1926, 2013.
[9] L. Erdös, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*, 314(3):587–640, 2012.
[10] L. Erdös, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab.*, 18:no. 59, 58, 2013.
[11] L. Erdös, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.
[12] L. Erdös, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.
[13] P. J. Forrester. The spectrum edge of random matrix ensembles. *Nuclear Phys. B*, 402(3):709–728, 1993.
[14] H. C. Ji. Regularity properties of free multiplicative convolution on the positive line. *Int. Math. Res. Not. IMRN*, 2020.
[15] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.
[16] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.
[17] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probab. Theory Related Fields*, 169(1-2):257–352, 2017.

[18] J. O. Lee and K. Schnelli. Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. *J. Math. Phys.*, 54(10):103504, 62, 2013.

[19] J. O. Lee and K. Schnelli. Extremal eigenvalues and eigenvectors of deformed Wigner matrices. *Probab. Theory Related Fields*, 164(1-2):165–241, 2016.

[20] J. O. Lee and K. Schnelli. Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.*, 26(6):3786–3839, 2016.

[21] M. Mahoney and C. Martin. Traditional and heavy tailed self regularization in neural network models. In *International Conference on Machine Learning*, 4284–4293, 2019.

[22] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math USSR Sb*, 1(4):457–483, apr 1967.

[23] M. Y. Mo. Rank 1 real Wishart spiked model. *Comm. Pure Appl. Math.*, 65(11):1528–1638, 2012.

[24] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642, 2007.

[25] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001, 2014.

[26] J. W. Silverstein and S.-I. Choi. Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.

[27] D. Wang. The largest eigenvalue of real symmetric, Hermitian and Hermitian self-dual random matrix models with rank one external source, Part I. *J. Stat. Phys.*, 146(4):719–761, 2012.