The Annals of Applied Probability
2016, Vol. 26, No. 6, 3786–3839
DOI: 10.1214/16-AAP1193

TRACY–WIDOM DISTRIBUTION FOR THE LARGEST EIGENVALUE OF REAL SAMPLE COVARIANCE MATRICES WITH GENERAL POPULATION

BY JI OON LEE[1] AND KEVIN SCHNELLI[2]

KAIST and IST Austria

We consider sample covariance matrices of the form $\mathcal{Q} = (\Sigma^{1/2}X)(\Sigma^{1/2}X)^*$, where the sample X is an $M \times N$ random matrix whose entries are real independent random variables with variance $1/N$ and where Σ is an $M \times M$ positive-definite deterministic matrix. We analyze the asymptotic fluctuations of the largest rescaled eigenvalue of $\mathcal{Q}$ when both M and N tend to infinity with $N/M \to d \in (0, \infty)$. For a large class of populations Σ in the sub-critical regime, we show that the distribution of the largest rescaled eigenvalue of $\mathcal{Q}$ is given by the type-1 Tracy–Widom distribution under the additional assumptions that (1) either the entries of X are i.i.d. Gaussians or (2) that Σ is diagonal and that the entries of X have a sub-exponential decay.

CONTENTS

Received June 2015; revised January 2016.
[1]Supported in part by TJ Park Junior Faculty Fellowship.
[2]Supported by ERC Advanced Grant RANMAT No. 338804.
MSC2010 subject classifications. 60B20, 15B52, 62H10.
Key words and phrases. Sample covariance matrix, Tracy–Widom distribution, edge universality.

1. Introduction. Covariance matrices are fundamental objects in multivariate statistics whose study is an integral part of various fields such as signal processing, genomics, financial mathematics, etc. Sample covariance matrices are the simplest estimators for population covariance matrices: The population covariance matrix of a mean-zero random variable $\mathbf{y} \in \mathbb{R}^M$ is $\Sigma := \mathbb{E}\mathbf{y}\mathbf{y}'$. Given N independent samples $(\mathbf{y}_1, \dots, \mathbf{y}_N)$ of $\mathbf{y}$, Σ may be estimated through the sample covariance matrix $\mathcal{Q} := \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i{}'$. Indeed, since $\mathbb{E}\mathcal{Q} = \Sigma$, $\mathcal{Q}$ converges, for fixed M, almost surely to Σ as N tends to infinity. However, in many modern applications the population size M may be as large or even larger than N, and hence, one may take M and N simultaneously to infinity in an asymptotic analysis. In this setting, Σ cannot be estimated through $\mathcal{Q}$ due to the high dimensionality. Yet, some properties of Σ may be inferred from spectral statistics of $\mathcal{Q}$, for example, the limiting behavior of the largest eigenvalues of $\mathcal{Q}$ is frequently used in hypothesis testing for the structure of Σ.

In this paper, we investigate the limiting behavior of the largest eigenvalues of the form

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

where the sample or data matrix, X, is an $M \times N$ matrix whose entries are a collection of independent real or complex random variables of variance $1/N$ and where the general population covariance, Σ, is an $M \times M$ real positive-definite deterministic matrix. We are interested in the high-dimensional case, where $\widehat{d} := N/M \to d \in (0, \infty)$, as $N \to \infty$. We further mainly focus on the real setting, where X is a real data matrix, since, mathematically, the complex case is easier to deal with. Also, the real case is of primary interest in statistics, although complex data matrices arise in some applications. For detailed discussions of this model, we refer to, for example, [3, 6, 14, 28, 29]. In Section 2.4 we outline an application of this model. We denote the eigenvalues of $\mathcal{Q}$ and Σ in decreasing order by $(\mu_i)_{i=1}^M$ and $(\sigma_m)_{m=1}^M$, respectively.

The main results of this paper show that the limiting distribution of the largest rescaled eigenvalue of $\mathcal{Q}$ is given by the Tracy–Widom distribution, that is,

$$\lim_{N\to\infty} \mathbb{P}\big(\gamma_0 N^{2/3}(\mu_1 - E_+) \le s\big) = F_1(s) \qquad (s \in \mathbb{R}), \tag{1.2}$$

where $\gamma_0 \equiv \gamma_0(N)$ and $E_+ \equiv E_+(N)$ depend only on the sequence $(\sigma_m)_{m=1}^M$ and the ratio $\widehat{d}$. Here, F_1 denotes the cumulative distribution function (CDF) of the type-1 Tracy–Widom distribution [46, 47] which arises as the limiting CDF of the largest rescaled eigenvalue of the Gaussian orthogonal ensemble (GOE). More precisely, we show that (1.2) holds in the "sub-critical regime" where the largest eigenvalues of Σ are close to the bulk of the spectrum of Σ (for a precise statement see Assumption 2.2 below) and if either of the following holds:

(1) the entries of X are i.i.d. real Gaussians (Corollary 2.7), or

(2) the general population Σ is diagonal and the entries of X have a sub-exponential decay (Theorem 2.4).

Our results are also valid in the setting of complex data matrices X. In that setup, one replaces F_1 in (1.2) by F_2, the CDF of the type-2 Tracy–Widom distribution which arises as the limiting CDF of the largest rescaled eigenvalue of the Gaussian unitary ensemble (GUE).

To situate our result in the literature, we first recall that the limiting spectral distribution of the model (1.1) was derived for general Σ by Marchenko and Pastur [34]. When X has i.i.d. Gaussian entries, $\mathcal{Q}$ is called a Wishart matrix. For Wishart matrices with identity population covariance, often referred to as the null case, it is well known that the limiting distribution of the largest rescaled eigenvalue coincides with the corresponding distribution of the GOE and GUE, respectively: in the null case, (1.2) was obtained in [28] for real Wishart matrices and in [27] for complex Wishart matrices.

In the nonnull case, where Σ is not a multiple of the identity matrix, first results were obtained for spiked population models introduced in [28], where Σ is a finite rank perturbation of the identity matrix. Complex spiked Wishart matrices were studied in [4], where an interesting phase transition in the asymptotic behavior of the largest rescaled eigenvalue as a function of the spikes was observed. In particular, it was shown that the largest rescaled eigenvalue follows the Tracy–Widom distribution F_2 in the sub-critical regime, that is, for small finite rank perturbations. These results rely on an explicit formula—the Baik–Ben Arous–Johansson–Péché (BBJP)-formula—for the joint eigenvalue distribution of complex Wishart matrices. For real Wishart matrices, the counterpart of the BBJP-formula is not available, due to the lack of an analogue of the Harish–Chandra–Itzykson–Zuber integral for the orthogonal group. Relying on quite different methods, almost sure convergence of the largest eigenvalues was derived in [5] and Tracy–Widom fluctuations of the largest eigenvalue of spiked population models were obtained in [23]. The equivalent results of the aforementioned phase transition for finite rank perturbations were obtained in the real setting in [10, 11, 25, 36].

In the general nonnull case, sufficient conditions for the validity of (1.2) in the sub-critical regime were given in [14] for the nonsingular case $d \in (0, \infty)$, $d \neq 1$ and in [39] for the singular case $d = 1$. Yet, these results rely on the BBJP-formula and are thus limited to complex Wishart matrices. Corollary 2.7 below establishes under similar assumptions the limiting behavior (1.2) for real Wishart matrices with general population.

The aforementioned results are believed to be universal in the sense that they are independent of the details of the distributions of the entries of X (provided they decay sufficiently fast). This phenomenon is referred to as edge universality. It was established in the null case in [22, 41, 44] for symmetric distributions and subsequently in [50] for distributions with vanishing third moment. This third moment condition was removed in [42]. For spiked sample covariance matrices, universality results were obtained in [23] under the assumption that the entries' distribution of X are symmetric. This condition was removed in [9]. For full rank deformed populations matrices Σ, universality results were obtained in [6] under the assumption that Σ is either diagonal or that the first four moments of the entries' distribution of X match those of the standard Gaussian distribution in case Σ is nondiagonal. Recently, the edge universality was established in [31] for general Σ. Once the edge universality for general sample covariance matrices has been established, the limiting CDF of the rescaled largest eigenvalue may then be identified in the complex setting with F_2 via the results of [14, 39]. In the real setting, this identification was only possible in the null case and finite rank deformations thereof. Our main new results allow this identification in the real setting with general population covariances, that is, it allows to identify F_1 as the limiting CDF of the rescaled largest eigenvalue.

Our proof of (1.2) is based on a comparison of Green functions. Discrete Green function comparison via Lindeberg's replacement strategy [13, 45] was used to prove the edge universalities of Wigner matrices [21, 45] and of null sample covariance matrices [42]. Continuous Green function comparison was used to establish CLT results for linear statistics of null sample covariance matrices [33], and more recently, to derive estimates on the Green function itself, that is, local laws, for nonnull sample covariance matrices [31]. However, as for the deformed Wigner matrices considered in [32], a direct application of discrete or continuous Green function comparison does not work for nonnull sample covariance matrices. We thus adopt the new approach developed in [32]: we consider a continuous interpolation between the given sample covariance matrix and a null sample covariance matrix. We follow the associated Green function flow and estimate its change over time. This change is then offset by rescaling the matrix.

Our analysis requires as an a priori ingredient a local law for the Green function, that is, an optimal estimate on the entries of the Green function on scales slightly below $N^{-2/3}$ at the upper edge (see Lemma 3.3 below for a precise statement). Optimal local laws in the bulk and at the edges of the spectrum were obtained for Wigner matrices in [17, 19, 20]. Using a similar approach, optimal local laws for sample covariance matrices with $\Sigma = \mathbb{1}$ were obtained in [42]; see also [8, 18]. These results were extended to sample covariance matrices with general population under a four moment matching condition in [6]. The four moment matching condition was very recently removed in [31].

This paper is organized as follows: In Section 2 we define the model, present the main results of the paper and outline some applications. In Section 3 we collect the tools and known results we need in our proofs. In Section 4 we prove the main theorems using our essential new technical result, Proposition 4.1, the Green function comparison theorem at the edge. In Sections 5 and 6 we outline the ideas of the proof of the Green function comparison theorem. Its technical details can be found in the Appendices A, B and C. Some results required in these Appendices are adaptations from [32].

2. Definitions and main result.

2.1. *Sample covariance matrix with general population.*

DEFINITION 2.1. Let $X = (x_{ij})$ be an $M \times N$ matrix whose entries $\{x_{ij} : 1 \le i \le M, 1 \le j \le N\}$ are a collection of independent real random variables such that

$$\mathbb{E}x_{ij} = 0, \qquad \mathbb{E}|x_{ij}|^2 = \frac{1}{N}. \tag{2.1}$$

Moreover, we assume that $(\sqrt{N}x_{ij})$ have a sub-exponential tail, that is, there are C and $\vartheta > 0$ such that

$$\mathbb{P}\big(|\sqrt{N}x_{ij}| > t\big) \le C\mathrm{e}^{-t^{\vartheta}}, \tag{2.2}$$

for all i, j.

Further, $M \equiv M(N)$ with

$$\widehat{d} = \frac{N}{M} \to d \in (0, \infty), \tag{2.3}$$

as $N \to \infty$. For simplicity, we assume that N/M is constant, hence we use d instead of $\widehat{d}$.

Note that we do not require in Definition 2.1 that the entries or columns of X are identically distributed.

Let Σ be an $M \times M$ real positive-definite deterministic matrix. We denote by $\widehat{\rho}$ the empirical eigenvalue distribution of Σ, that is, if we let $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_M \ge 0$ be the eigenvalues of Σ, then

$$\widehat{\rho} := \frac{1}{M}\sum_{j=1}^{M} \delta_{\sigma_j}. \tag{2.4}$$

We then form the sample covariance matrix

$$\mathcal{Q} := (\Sigma^{1/2}X)(\Sigma^{1/2}X)^*, \tag{2.5}$$

and denote its eigenvalues in decreasing order by $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_M$. Note that the $M \times M$ matrix $\mathcal{Q}$ and the $N \times N$ matrix

$$Q := X^* \Sigma X \tag{2.6}$$

share the same nonzero eigenvalues. Since we are interested in behavior of the largest eigenvalues, we focus on Q since it is for technical reasons more amenable than $\mathcal{Q}$. With some abuse of terminology, we also call Q a sample covariance matrix and we denote its M largest eigenvalues by $(\mu_i)_{i=1}^M$, also.

2.2. *Deformed Marchenko Pastur law.* Assuming that the empirical spectral distribution $\widehat{\rho}$ of Σ converges weakly to some distribution ρ, it was shown in [34] that the empirical eigenvalue distribution of Q converges weakly in probability to a deterministic distribution, ρ_{fc}, referred to as the "deformed Marchenko–Pastur law" below, which depends on ρ and the ratio d. It can be described in terms of its Stieltjes transform: For a (probability) measure ω on the real line, we define its Stieltjes transform by

$$m_\omega(z) := \int_{\mathbb{R}} \frac{\mathrm{d}\omega(v)}{v - z} \qquad \bigl(z = E + \mathrm{i}\eta \in \mathbb{C}^+\bigr). \tag{2.7}$$

Here and below, we write $z = E + \mathrm{i}\eta$, with $E \in \mathbb{R}$, $\eta \geq 0$. Note that m_ω is an analytic function in the upper half-plane and that $\operatorname{Im} m_\omega(z) \geq 0$, $\operatorname{Im} z > 0$. Assuming that ω is absolutely continuous with respect to Lebesgue measure, we can recover the density of ω from m_ω by the inversion formula

$$\omega(E) = \lim_{\eta \searrow 0} \frac{1}{\pi} \operatorname{Im} m_\omega(E + \mathrm{i}\eta) \qquad (E \in \mathbb{R}). \tag{2.8}$$

We use the same symbols to denote measures and their densities.

Choosing ω to be the standard Marchenko–Pastur law ρ_{MP}, the Stieltjes transform $m_{\rho_{\mathrm{MP}}} \equiv m_{\mathrm{MP}}$ can be computed explicitly and one checks that m_{MP} satisfies the relation

$$m_{\mathrm{MP}}(z) = \frac{1}{-z + d^{-1} \frac{1}{m_{\mathrm{MP}}(z)+1}}, \qquad \operatorname{Im} m_{\mathrm{MP}}(z) \geq 0 \qquad \bigl(z \in \mathbb{C}^+\bigr). \tag{2.9}$$

The deformed Marchenko–Pastur law ρ_{fc} is defined as follows. Assume that $\widehat{\rho}$ converges weakly to ρ as N goes to infinity. Then the Stieltjes transform of the deformed Marchenko–Pastur law, m_{fc}, is obtained as the unique solution to the self-consistent equation

$$\begin{aligned} m_{\mathrm{fc}}(z) &= \frac{1}{-z + d^{-1} \int_{\mathbb{R}} \frac{t}{t m_{\mathrm{fc}}(z)+1} \,\mathrm{d}\rho(t)}, \\ \operatorname{Im} m_{\mathrm{fc}}(z) &\geq 0 \qquad \bigl(z \in \mathbb{C}^+\bigr). \end{aligned} \tag{2.10}$$

It is well known [34] that the functional equation (2.10) has a unique solution that is uniformly bounded on the upper half-plane. The density of the deformed Marchenko–Pastur law ρ_{fc} is obtained from m_{fc} by the Stieltjes inversion formula (2.8). The measure ρ_{fc} has been studied in [43], for example, it was shown that the density of ρ_{fc} is an analytic function inside its support. The measure ρ_{fc} is also called the multiplicative free convolution of the Marchenko–Pastur law and the measure ρ; we refer to, for example, [2, 49].

For finite N, we let $\widehat{m}_{\mathrm{fc}}$ denote the unique solution to

$$\widehat{m}_{\mathrm{fc}}(z) = \frac{1}{-z + d^{-1}\int_{\mathbb{R}} \frac{t}{t\widehat{m}_{\mathrm{fc}}(z)+1}\,\mathrm{d}\widehat{\rho}(t)}, \qquad \operatorname{Im}\widehat{m}_{\mathrm{fc}}(z) \geq 0 \qquad (z \in \mathbb{C}^+), \tag{2.11}$$

and let $\widehat{\rho}_{\mathrm{fc}}$ denote the measure obtained from $\widehat{m}_{\mathrm{fc}}(z)$ through (2.8). It is easy to check that $\widehat{\rho}_{\mathrm{fc}}$ is a well-defined probability measure with a continuous density.

The rightmost endpoint of the support of $\widehat{\rho}_{\mathrm{fc}}$ is determined as follows. Define ξ_+ as the largest solution to

$$\int_{\mathbb{R}} \left(\frac{t\xi_+}{1-t\xi_+}\right)^2 \mathrm{d}\widehat{\rho}(t) = d, \tag{2.12}$$

with $d = \frac{N}{M}$. Note that ξ_+ is unique and that $\xi_+ \in [0, \sigma_1^{-1}]$. We also introduce E_+ by setting

$$E_+ := \frac{1}{\xi_+}\left(1 + d^{-1}\int_{\mathbb{R}} \frac{t\xi_+}{1-t\xi_+}\,\mathrm{d}\widehat{\rho}(t)\right). \tag{2.13}$$

Considering the imaginary part of (2.11) in the limit $\eta \searrow 0$, one infers from [43] that the rightmost edge of $\widehat{\rho}_{\mathrm{fc}}$, that is, the rightmost endpoint of the support of $\widehat{\rho}_{\mathrm{fc}}$, is given by E_+ and that

$$\xi_+ = -\lim_{\eta\to 0} \widehat{m}_{\mathrm{fc}}(E_+ + \mathrm{i}\eta) = -\widehat{m}_{\mathrm{fc}}(E_+). \tag{2.14}$$

The following assumption is required to establish our main results. It appeared previously in [6, 14].

ASSUMPTION 2.2. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M$ denote the eigenvalues of Σ. Then, we assume that $\liminf_N \sigma_M > 0$, $\limsup_N \sigma_1 < \infty$ and

$$\limsup_N \sigma_1 \xi_+ < 1. \tag{2.15}$$

REMARK 2.3. We remark that Assumption 2.2 was used in [6, 31] to derive the local deformed Marchenko–Pastur law for Q. The inequality (2.15) guarantees that the distribution $\widehat{\rho}_{\mathrm{fc}}(E)$ exhibits a square-root type behavior at the rightmost endpoint of its support; see Lemma 3.2 below.

2.3. *Main result*. The main result of this paper is as follows.

THEOREM 2.4. *Let $Q = X^*\Sigma X$ be an $N \times N$ sample covariance matrix with sample X and population Σ, where X is a real random matrix satisfying the assumptions in Definition 2.1 and Σ is a real diagonal deterministic matrix satisfying Assumption 2.2. Recall that F_1 denotes the cumulative distribution function of the type-1 Tracy–Widom distribution.*

Let μ_1 be the largest eigenvalue of Q. Then, there exist $\gamma_0 \equiv \gamma_0(N)$ depending only on the empirical eigenvalue distribution $\widehat{\rho}$ of Σ and the ratio d such that the distribution of the largest rescaled eigenvalue of Q converges to the Tracy–Widom distribution, that is,

$$\lim_{N\to\infty} \mathbb{P}\big(\gamma_0 N^{2/3}(\mu_1 - E_+) \le s\big) = F_1(s), \tag{2.16}$$

for all $s \in \mathbb{R}$, where $E_+ \equiv E_+(N)$ is given in (2.13).

REMARK 2.5. The scaling factor $\gamma_0 \equiv \gamma_0(N)$ is given by [14]

$$\frac{1}{\gamma_0^3} = \frac{1}{d}\int_{\mathbb{R}} \left(\frac{t}{1 - t\xi_+}\right)^3 \mathrm{d}\widehat{\rho}(t) + \frac{1}{\xi_+^3}. \tag{2.17}$$

The factor γ_0 has the following meaning: It was shown in [43] that the deformed Marchenko–Pastur law ρ_{fc} exhibits a square-root behavior, that is,

$$\rho_{\mathrm{fc}}(E) = C_0\sqrt{E_+ - E}\big(1 + O(\sqrt{E_+ - E})\big) \qquad (E \le E_+),$$

at the upper edge E_+, with

$$C_0 = \frac{1}{\pi}\left(\frac{2}{g''(\xi_+)}\right)^{1/2}, \qquad g(x) = \frac{1}{x} + \frac{1}{d}\int_{\mathbb{R}} \frac{t}{1 - tx}\,\mathrm{d}\rho(t).$$

Thus, the choice of γ_0 in (2.17) makes the "curvature" at the upper edge coincide with that of Wigner semicircle law, $1/\pi$. Moreover, it follows from Assumption 2.2 that $\gamma_0 \sim 1$.

REMARK 2.6. Theorem 2.4 can be extended to correlation functions of the extremal eigenvalues as follows: Let W^{GOE} be an $N \times N$ random matrix belonging to the Gaussian Orthogonal Ensemble (GOE); see [2, 35]. The joint distributions of $\mu_1^{\mathrm{GOE}} \ge \mu_2^{\mathrm{GOE}} \ge \cdots \ge \mu_N^{\mathrm{GOE}}$, the eigenvalues of W^{GOE}, are explicit and the joint distribution of the k largest eigenvalues can be written in terms of the Airy kernel [24] for any fixed k. The generalization of (2.16) to the k largest eigenvalues of Q then reads

$$\begin{aligned}&\lim_{N\to\infty} \mathbb{P}\big(\big(\gamma_0 N^{2/3}(\mu_i - E_+) \le s_i\big)_{1\le i\le k}\big)\\ &\quad= \lim_{N\to\infty} \mathbb{P}\big(\big(N^{2/3}(\mu_i^{\mathrm{GOE}} - 2) \le s_i\big)_{1\le i\le k}\big),\end{aligned} \tag{2.18}$$

for all $s_1, s_2, \ldots, s_k \in \mathbb{R}$.

If the entries of X are Gaussian, the result in Theorem 2.4 holds for general, nondiagonal Σ.

COROLLARY 2.7. *Let $Q = X^*\Sigma X$ be an $N \times N$ sample covariance matrix with sample X and general population Σ, where X is a real random matrix with independent Gaussian entries satisfying the assumptions in Definition 2.1 and Σ is a real positive-definite deterministic matrix satisfying Assumption 2.2. Let μ_1 be the largest eigenvalue of Q.*

Then the distribution of the largest rescaled eigenvalue of Q converges to the type-1 Tracy–Widom distribution, that is,

$$\lim_{N\to\infty} \mathbb{P}\big(\gamma_0 N^{2/3}(\mu_1 - E_+) \le s\big) = F_1(s), \tag{2.19}$$

for all $s \in \mathbb{R}$, where $E_+ \equiv E_+(N)$ is given in (2.13) and $\gamma_0 = \gamma_0(N)$ is given in (2.17).

REMARK 2.8. For non-Gaussian X and general off-diagonal Σ, we can combine our results with edge universality results in [6, 31] to identify the Tracy–Widom distribution for the largest eigenvalues.

2.4. *Applications.* In this subsection we briefly discuss possible applications of our results to statistics. For a general overview of applications of random matrix theory to statistical inference we refer to the review [29], where the following application to signal detection problems is described.

Consider a signal-plus-noise vector

$$\mathbf{y} := D\mathbf{s} + \Sigma_0^{1/2}\mathbf{z} \tag{2.20}$$

of dimension M, where $\mathbf{s}$ is a k-dimensional real mean-zero signal vector with population covariance matrix S, D is a $M \times k$ real deterministic matrix which is of full column rank, $\mathbf{z}$ is an M-dimensional real or complex random vector and Σ_0 is an $M \times M$ deterministic positive-definite matrix. In many situations, $\mathbf{z}$ is assumed to be Gaussian. Assuming that the signal vector, $D\mathbf{s}$, and the noise vector, $\Sigma_0^{1/2}\mathbf{z}$, are independent the population covariance matrix, Σ, of $\mathbf{y}$ is given by

$$\Sigma = DSD^* + \Sigma_0. \tag{2.21}$$

A fundamental question is to detect signals from given data, for example, from independent samples, $(\mathbf{y}_i)_{i=1}^N$, of $\mathbf{y}$ and its associated sample covariance matrix Q. A first step in this analysis is to determine whether there is any signal present that can be detected at all. Once signals are detected, one is led to estimate k. More precisely, the first aim is to test, with general correlation noise $\Sigma_0^{1/2}\mathbf{z}$, whether there is no signal present, that is, to test the null hypothesis $k = 0$ against the alternative hypothesis $k > 1$. For details and results in the classical setting of large sample size N and low dimensionality M, see [30].

In the high dimensional setup, the described signal detection problem was considered in [7, 37], where the usability and performance of the largest eigenvalue as a test statistics was discussed in the presence of white Gaussian noise, that is, $\Sigma_0 = \mathbb{1}$ and Gaussian $\mathbf{z}$. In [38], the detection problem in presence of correlated Gaussian noise, that is, $\Sigma_0 \neq \mathbb{1}$ and Gaussian $\mathbf{z}$, was discussed at length. We also refer to [48] for further developments.

In the discussion above, it was implicitly assumed that Σ_0 is known a priori. For many real systems, however, Σ_0 is usually unknown. In particular, ζ_+, E_+ and γ_0 of, for example, (2.16), are unknown and the largest eigenvalue μ_1 may not directly be used as a test statistics. Following [40], we observe that E_+ and γ_0 are eliminated under the null hypothesis in the test statistics $R := (\mu_1 - \mu_2)/(\mu_2 - \mu_3)$, with μ_2, μ_3, the second, respectively, third largest eigenvalue of $\mathcal{Q}$. On the other hand, the limiting distribution of R is determined by the Tracy–Widom–Airy statistics under the null hypothesis as stated in Remark 2.6 above. In fact, in the complex setting the test statistics R was shown [40] to be asymptotically pivotal under the null hypothesis and one expects the same results to hold in the real setting. While there is no explicit formula for the limiting distribution of R, it may be effectively approximated by using numerics for extremal eigenvalues of GOE, respectively GUE, matrices.

3. Preliminaries.

3.1. *Notation.* We first introduce a notation for high-probability estimates which is suited for our purposes. A slightly different form was first used in [15].

DEFINITION 3.1. Let

$$X = \big(X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\big), \qquad Y = \big(Y^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\big)$$

be two families of nonnegative random variables where $U^{(N)}$ is a possibly N-dependent parameter set. We say that Y stochastically dominates X, uniformly in u, if for all (small) $\varepsilon > 0$ and (large) $D > 0$,

$$\sup_{u \in U^{(N)}} \mathbb{P}\big[X^{(N)}(u) > N^{\varepsilon} Y^{(N)}(u)\big] \le N^{-D}, \tag{3.1}$$

for sufficiently large $N \ge N_0(\varepsilon, D)$. If Y stochastically dominates X, uniformly in u, we write $X \prec Y$. If for some complex family X, we have $|X| \prec Y$ we also write $X = \mathcal{O}(Y)$.

The relation $\prec$ is a partial ordering: it is transitive and it satisfies the arithmetic rules of an order relation, for example, if $X_1 \prec Y_1$ and $X_2 \prec Y_2$ then $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1 X_2 \prec Y_1 Y_2$. Furthermore, the following property will be used on a few occasions: If $\Phi(u) \ge N^{-C}$ is deterministic, $Y(u)$ is a nonnegative random variable satisfying $\mathbb{E}[Y(u)]^2 \le N^{C'}$ for all u, and $Y(u) \prec \Phi(u)$ uniformly in u, then $\mathbb{E}[Y(u)] \prec \Phi(u)$, uniformly in u. This can be easily checked since

$$\mathbb{E}\big[|Y(u)|\mathbb{1}\big(|Y(u)| > N^{\varepsilon/2}\Phi\big)\big] \le \big(\mathbb{E}\big[|Y(u)|^2\big]\big)^{1/2}\big(\mathbb{P}\big[|Y(u)| > N^{\varepsilon/2}\Phi\big]\big)^{1/2} \le N^{-D},$$

for any (large) $D > 0$, and $\mathbb{E}[|Y(u)|\mathbb{1}(|Y(u)| \le N^{\varepsilon/2}\Phi)] \le N^{\varepsilon/2}\Phi(u)$, hence $\mathbb{E}[Y(u)] \le N^{\varepsilon}\Phi(u)$.

We use the symbols $O(\cdot)$ and $o(\cdot)$ for the standard big-O and little-o notation. The notation $O, o, \ll, \gg$, refer to the limit $N \to \infty$ unless otherwise stated. Here, $a \ll b$ means $a = o(b)$. We use c and C to denote positive constants that do not depend on N, usually with the convention $c \le C$. Their value may change from line to line. We write $a \sim b$, if there is $C \ge 1$ such that $C^{-1}|b| \le |a| \le C|b|$.

Finally, we use double brackets to denote index sets, that is,

$$[\![n_1, n_2]\!] := [n_1, n_2] \cap \mathbb{Z},$$

for $n_1, n_2 \in \mathbb{R}$.

3.2. *Local deformed Marchenko–Pastur law.* For small positive $\mathfrak{c}, \varepsilon$ and sufficiently large C_+, $(C_+ > E_+)$, we define the domain, $\mathcal{D}(\mathfrak{c}, \varepsilon)$, of the spectral parameter z by

$$\mathcal{D}(\mathfrak{c}, \varepsilon) := \{z = E + \mathrm{i}\eta \in \mathbb{C}^+ : E_+ - \mathfrak{c} \le E \le C_+, N^{\varepsilon}N^{-1} \le \eta \le 1\}.$$

Let $\kappa \equiv \kappa_E := |E - E_+|$. Then we have the following results.

LEMMA 3.2 (Theorem 3.1 in [6]). *Under Assumption 2.2, there is $c > 0$ such that*

$$\widehat{\rho}_{\mathrm{fc}}(E) \sim \sqrt{E_+ - E} \qquad \big(E \in [E_+ - 2c, E_+]\big). \tag{3.2}$$

The Stieltjes transform $\widehat{m}_{\mathrm{fc}}(z)$ of $\widehat{\rho}_{\mathrm{fc}}$ satisfies the following:

(i) *For $z \in \mathcal{D}(c, 0)$,*

$$\big|\widehat{m}_{\mathrm{fc}}(z)\big| \sim 1. \tag{3.3}$$

(ii) *For $z \in \mathcal{D}(c, 0)$,*

$$\operatorname{Im} \widehat{m}_{\mathrm{fc}}(z) \sim \begin{cases} \dfrac{\eta}{\sqrt{\kappa + \eta}}, & \text{if } E \ge E_+ + \eta, \\ \sqrt{\kappa + \eta}, & \text{if } E \in [E_+ - c, E_+ + \eta). \end{cases} \tag{3.4}$$

We introduce the z-dependent control parameter, $\Psi(z)$, by setting

$$\Psi \equiv \Psi(z) := \left(\frac{\operatorname{Im} \widehat{m}_{\mathrm{fc}}(z)}{N\eta}\right)^{1/2} + \frac{1}{N\eta}. \tag{3.5}$$

We remark that, for $z = E + \mathrm{i}\eta$ with $\kappa_E \le N^{-2/3+\varepsilon}$ and $\eta = N^{-2/3-\varepsilon}$, we have

$$\Psi \le C N^{-1/3+\varepsilon}.$$

Define the Green function $G_Q = ((G_Q)_{ij})$ by

$$G_Q(z) := (Q - z)^{-1} \qquad (z \in \mathbb{C}^+), \tag{3.6}$$

and denote its normalized trace by

$$m_Q(z) := \frac{1}{N} \mathrm{Tr} G_Q(z) \qquad (z \in \mathbb{C}^+). \tag{3.7}$$

Recall that μ_1 denotes the largest eigenvalue of the sample covariance matrix Q. We have the following local law from [6].

LEMMA 3.3 (Theorems 3.2 and 3.3 in [6]). *Under Assumption 2.2 we have, for any sufficiently small* $\varepsilon > 0$,

$$\big|m_Q(z) - \widehat{m}_{\mathrm{fc}}(z)\big| \prec \frac{1}{N\eta}, \qquad \max_{i,j}\big|(G_Q)_{ij}(z) - \delta_{ij}\widehat{m}_{\mathrm{fc}}(z)\big| \prec \Psi(z), \tag{3.8}$$

uniformly in z *on* $\mathcal{D}(c, \varepsilon)$, *where* c *is the constant in Lemma 3.2. Moreover, we have*

$$|\mu_1 - E_+| \prec N^{-2/3}, \tag{3.9}$$

where E_+ *is given in* (2.13).

3.3. *Density of states*. In this subsection we explain how the distribution of the largest eigenvalues of Q can be related to $m_Q(z)$ for appropriately chosen z. The arguments given here are small modifications of the methods presented in [21, 32, 42].

Recall the definition of the scaling factor γ_0 in (2.17). We set

$$T := \gamma_0 \Sigma, \tag{3.10}$$

and define

$$\widetilde{Q} := X^* T X. \tag{3.11}$$

We denote by $m_{\widetilde{Q}}$ the normalized trace of the Green function of $\widetilde{Q}$, that is,

$$m_{\widetilde{Q}}(z) := \frac{1}{N} \mathrm{Tr}(\widetilde{Q} - z)^{-1} \qquad (z \in C^+). \tag{3.12}$$

Let $\widetilde{\mu}_1 \ge \widetilde{\mu}_2 \ge \cdots \ge \widetilde{\mu}_N$ be the eigenvalues of $\widetilde{Q}$. Let $L_+ := \gamma_0 E_+$ and observe that from Lemma 3.3 we have

$$|\widetilde{\mu}_1 - L_+| \prec N^{-2/3}.$$

Thus, we may assume in (2.16) that $|s| \prec 1$.

Fix a small $\varepsilon > 0$ and let

$$E_* = L_+ + 2N^{-2/3+\varepsilon}.$$

We note that the choice of E_* guarantees that the probability of the event $\{\widetilde{\mu}_1 > E_*\}$ is negligible. For E satisfying

$$|E - L_+| \le N^{-2/3+\varepsilon}, \tag{3.13}$$

we let

$$\chi_E := \mathbb{1}_{[E, E_*]}.$$

We also define the Poisson kernel, θ_η, for $\eta > 0$,

$$\theta_\eta(x) := \frac{\eta}{\pi(x^2+\eta^2)} = \frac{1}{\pi} \operatorname{Im} \frac{1}{x - \mathrm{i}\eta}.$$

Introduce a smooth cutoff function $K : \mathbb{R} \to \mathbb{R}$ satisfying

$$K(x) = \begin{cases} 1, & \text{if } x \le 1/9, \\ 0, & \text{if } x \ge 2/9. \end{cases} \tag{3.14}$$

Let $\mathcal{N}(E_1, E_2)$ be the number of the eigenvalues in $(E_1, E_2]$, that is,

$$\mathcal{N}(E_1, E_2) := \big|\{\alpha : E_1 < \widetilde{\mu}_\alpha \le E_2\}\big|,$$

and define the density of states in the interval $(E_1, E_2]$ by

$$\mathfrak{n}(E_1, E_2) := \frac{1}{N} \mathcal{N}(E_1, E_2).$$

In order to estimate $\mathbb{P}(\widetilde{\mu}_1 \le E)$, we consider the following approximations:

$$\begin{aligned} \mathbb{P}(\widetilde{\mu}_1 \le E) &= \mathbb{E} K\big(\mathcal{N}(E, \infty)\big) \\ &\simeq \mathbb{E} K\big(\mathcal{N}(E, E_*)\big) \simeq \mathbb{E} K\bigg(N \int_E^{E_*} \operatorname{Im} m_{\widetilde{Q}}(y + \mathrm{i}\eta)\,\mathrm{d}y\bigg), \end{aligned} \tag{3.15}$$

with $\eta \sim N^{-2/3-\varepsilon'}$, for some small $\varepsilon' > 0$. The first approximation in (3.15) follows from Lemma 3.3, the rigidity of the eigenvalues, and the second from

$$\mathcal{N}(E, E_*) = \operatorname{Tr} \chi_E(\widetilde{Q}) \simeq \operatorname{Tr} \chi_E * \theta_\eta(\widetilde{Q}) = \frac{1}{\pi} N \int_E^{E_*} \operatorname{Im} m_{\widetilde{Q}}(y + \mathrm{i}\eta)\,\mathrm{d}y.$$

The following lemma shows that the approximations in (3.15) indeed hold.

LEMMA 3.4. *For $\varepsilon > 0$, let $\ell := \frac{1}{2} N^{-2/3-\varepsilon}$ and $\eta := N^{-2/3-9\varepsilon}$. Suppose that E satisfies (3.13). Recall that K is a smooth function satisfying (3.14). Then, for any sufficiently small $\varepsilon > 0$ and any (large) $D > 0$, we have*

$$\operatorname{Tr}\big(\chi_{E+\ell} * \theta_\eta(\widetilde{Q})\big) - N^{-\varepsilon} \le \mathcal{N}(E, \infty) \le \operatorname{Tr}\big(\chi_{E-\ell} * \theta_\eta(\widetilde{Q})\big) + N^{-\varepsilon}, \tag{3.16}$$

with high probability, and

$$\mathbb{E}K\big(\mathrm{Tr}\big(\chi_{E-\ell} * \theta_\eta(\widetilde{Q})\big)\big) \le \mathbb{P}(\widetilde{\mu}_1 \le E) \le \mathbb{E}K\big(\mathrm{Tr}\big(\chi_{E+\ell} * \theta_\eta(\widetilde{Q})\big)\big) + N^{-D}, \tag{3.17}$$

for any sufficiently large $N \ge N_0(\varepsilon, D)$.

PROOF. We may follow the proof of Corollary 4.2 of [42]. To prove the first part of the lemma, one sets $\ell_1 := N^{-2/3-3\varepsilon}$ and shows that

$$\big|\mathrm{Tr}\,\chi_E(\widetilde{Q}) - \mathrm{Tr}\,\chi_E * \theta_\eta(\widetilde{Q})\big| \le C\big(N^{-2\varepsilon} + \mathcal{N}(E-\ell_1, E+\ell_1)\big),$$

which corresponds to Lemma 4.1 of [42], by using Lemmas 3.2 and 3.3, the estimates on $|m_{\widetilde{Q}}(E+\mathrm{i}\ell) - \widehat{m}_{\mathrm{fc}}(E+\mathrm{i}\ell)|$ and $\mathrm{Im}\,\widehat{m}_{\mathrm{fc}}(E-\kappa+\mathrm{i}\ell)$, respectively. Then, by integrating $\mathrm{Tr}\,\chi_y * \theta_\eta(\widetilde{Q})$ over y on $[E-\ell, E]$, one can obtain the estimate

$$\mathrm{Tr}\,\chi_E(\widetilde{Q}) \le \mathrm{Tr}\,\chi_{E-\ell} * \theta_\eta(\widetilde{Q}) + CN^{-2\varepsilon} + C\ell^{-1}\ell_1\mathcal{N}(E-2\ell, E+\ell).$$

The term $\ell^{-1}\ell_1\mathcal{N}(E-2\ell, E+\ell)$ in the right-hand side inequality can easily be controlled by applying the local law, Lemma 3.3. This proves the second inequality of (3.16). The other inequality in (3.16) can be proved analogously.

If (3.16) holds, the condition $\widetilde{\mu}_1 \le E$ implies that $\mathrm{Tr}\,\chi_{E+\ell} * \theta_\eta(\widetilde{Q}) \le 1/9$. Thus, applying Markov inequality, we get the upper bound in (3.17). The lower bound in (3.17) can be obtained in a similar manner. □

4. Green function comparison and proof of the main result.

Having established Lemma 3.4, the proof of Theorem 2.4 directly follows from our main technical result: the Green function comparison theorem at the edge, Proposition 4.1 below. It compares the expectations of functions of the normalized traces of the Green functions of $\widetilde{Q}$ and X^*X. More precisely, we let

$$W := \sqrt{d}(1+\sqrt{d})^{-4/3}X^*X, \tag{4.1}$$

and introduce

$$m_W(z) := \frac{1}{N}\mathrm{Tr}(W-z)^{-1} \qquad \big(z \in \mathbb{C}^+\big).$$

It is well known that the distribution of the rescaled largest eigenvalue of W converges to the Tracy–Widom distribution; see [42].

Our main technical result is as follows. Recall that we write $L_+ = \gamma_0 E_+$, with E_+ given in (2.13) and with γ_0 given in (2.17).

PROPOSITION 4.1 (Green function comparison). *Let* $\varepsilon > 0$ *and set* $\eta = N^{-2/3-\varepsilon}$. *Denote by* M_+ *the upper edge of the Marchenko–Pastur law* ρ_{MP} *for* $W = \sqrt{d}(1+\sqrt{d})^{-4/3}X^*X$. *Let* $E_1, E_2 \in \mathbb{R}$ *satisfy* $E_1 < E_2$ *and*

$$|E_1|, |E_2| \le N^{-2/3+\varepsilon}. \tag{4.2}$$

Let $F:\mathbb{R}\to\mathbb{R}$ *be a smooth function satisfying*

$$\max_x \bigl|F^{(\ell)}(x)\bigr|\bigl(|x|+1\bigr)^{-C}\le C,\qquad \ell=1,2,3,4. \tag{4.3}$$

Then there exists a constant $\phi>0$ *such that, for any sufficiently large* N *and for any sufficiently small* $\varepsilon>0$, *we have*

$$\begin{aligned}&\biggl|\mathbb{E}F\biggl(N\int_{E_1}^{E_2}\operatorname{Im} m_{\widetilde{Q}}(x+L_++\mathrm{i}\eta)\,\mathrm{d}x\biggr)\\ &\qquad-\mathbb{E}F\biggl(N\int_{E_1}^{E_2}\operatorname{Im} m_W(x+M_++\mathrm{i}\eta)\,\mathrm{d}x\biggr)\biggr|\le N^{-\phi}.\end{aligned} \tag{4.4}$$

We outline the proof of Proposition 4.1 in the Appendices A, B and C.

REMARK 4.2. Proposition 4.1 can be extended as follows: Let $\varepsilon>0$ and set $\eta=N^{-2/3-\varepsilon}$. Let $E_0,E_1,\dots,E_k\in\mathbb{R}$ satisfy $E_1<E_2<\cdots<E_k$ and

$$|E_0|\le N^{-2/3+\varepsilon},\qquad |E_1-E_0|\le N^{-2/3+\varepsilon},\qquad \dots,$$
$$|E_k-E_0|\le N^{-2/3+\varepsilon}.$$

Let $F:\mathbb{R}^k\to\mathbb{R}$ be a smooth function satisfying

$$\max_x \bigl|F^{(\ell)}(x)\bigr|\bigl(|x|+1\bigr)^{-C}\le C,\qquad \ell=1,2,3,4.$$

Then there exists a constant $\phi>0$ such that, for any sufficiently large N and for any sufficiently small $\varepsilon>0$, we have

$$\begin{aligned}&\biggl|\mathbb{E}F\biggl(\biggl(N\int_{E_i}^{E_0}\operatorname{Im} m_{\widetilde{Q}}(x+L_++\mathrm{i}\eta)\,\mathrm{d}x\biggr)_{1\le i\le k}\biggr)\\ &\qquad-\mathbb{E}F\biggl(\biggl(N\int_{E_i}^{E_0}\operatorname{Im} m_W(x+M_++\mathrm{i}\eta)\,\mathrm{d}x\biggr)_{1\le i\le k}\biggr)\biggr|\le N^{-\phi}.\end{aligned}$$

The proof of this statement is similar to that of Proposition 4.1 and will be omitted.

Assuming the validity of Proposition 4.1, we now prove our main results.

PROOF OF THEOREM 2.4. We follow the proof of Theorem 1.1 of [42]. Let μ_1^W be the largest eigenvalue of W [see (4.1)] and denote by M_+ the upper edge of the rescaled Marchenko–Pastur law ρ_{MP}. We notice that the distribution of $N^{2/3}(\mu_1^W-M_+)$ converges to the Tracy–Widom law F_1. (See [22, 41, 42, 44].) Thus, in order to prove (2.16), it suffices to show that

$$\begin{aligned}\mathbb{P}\bigl[N^{2/3}(\mu_1^W-M_+)\le s\bigr]-N^{-\phi}&\le\mathbb{P}\bigl[N^{2/3}(\widetilde{\mu}_1-L_+)\le s\bigr]\\ &\le\mathbb{P}\bigl[N^{2/3}(\mu_1^W-M_+)\le s\bigr]+N^{-\phi},\end{aligned} \tag{4.5}$$

for some $\phi>0$.

Fix an s satisfying $|s| \prec 1$, and let $E := L_+ + sN^{-2/3}$. Let $\ell := \frac{1}{2}N^{-2/3-\varepsilon}$ and $\eta := N^{-2/3-9\varepsilon}$. For any sufficiently small $\varepsilon > 0$, we have from Lemma 3.4 that

$$\mathbb{P}(\widetilde{\mu}_1 \le E) \ge \mathbb{E}\big[K\big(\mathrm{Tr}\big(\chi_{E-\ell} * \theta_\eta(H)\big)\big)\big]. \tag{4.6}$$

From Proposition 4.1, we find that

$$\mathbb{E}\big[K\big(\mathrm{Tr}\big(\chi_{E-\ell} * \theta_\eta(H)\big)\big)\big] \ge \mathbb{E}\big[K\big(\mathrm{Tr}\big(\chi_{E-(L_+-M_+)-\ell} * \theta_\eta(W)\big)\big)\big] - N^{-\phi},$$

for some $\phi > 0$. Finally, we have from Corollary 4.2 of [42] that

$$\mathbb{E}\big[K\big(\mathrm{Tr}\big(\chi_{E-(L_+-M_+)-\ell} * \theta_\eta(W)\big)\big)\big] \ge \mathbb{P}\big(\mu_1^W \le E - (L_+ - M_+)\big) - N^{-\phi}.$$

Altogether, we have shown that

$$\mathbb{P}(\widetilde{\mu}_1 \le E) \ge \mathbb{P}\big(\mu_1^W \le E - (L_+ - M_+)\big) - 2N^{-\phi},$$

which proves the first inequality of (4.5). The second inequality can be proved similarly. □

PROOF OF COROLLARY 2.7. Let U be an $M \times M$ orthogonal matrix that diagonalizes Σ, that is, there exists an $M \times M$ real diagonal matrix D such that $\Sigma = U^* D U$. Then UX is a real random matrix with Gaussian entries, satisfying the assumptions in Definition 2.1. Thus, applying Theorem 2.4 with $X^*\Sigma X = (UX)^* D(UX)$, we get the desired result. □

5. Linearization of $\widetilde{Q}$. In this section we recall a well-known formalism that simplifies the computations in the proof of Proposition 4.1 considerably. Instead of working with the product matrices $\widetilde{Q} = X^*TX$ or $T^{1/2}XX^*T^{1/2}$, we may "linearize" the problem by introducing an $(N+M)\times(N+M)$ matrix H, whose respective entries are either $(x_{\alpha a})$, (t_α^{-1}), z or simply zero. The inverse of H is then related to the Green function of X^*TX, respectively, of $T^{1/2}XX^*T^{1/2}$, through Schur's complement formula or the Feshbach map. For similar applications in random matrix theory, see, for example, [1, 26].

The linearization of $\widetilde{Q}$ is established in Section 5.1. In Sections 5.2, 5.3 and 5.4 we collect useful technical results on the inverse of H.

5.1. *Schur complement.* Suppose that X and Σ satisfy the assumptions in Theorem 2.4. Let z be as in the previous section. We define an $(N+M)\times(N+M)$ matrix H by

$$H \equiv H(z) = \begin{pmatrix} -zI_N & X^* \\ X & -T^{-1} \end{pmatrix}, \tag{5.1}$$

where I_N is the identity matrix with size N and $T = \gamma_0\Sigma$.

We claim that $H(z)$ is invertible for $z \in \mathbb{C}^+$. To see this, assume that $\mathbf{v}$, $\mathbf{v} \neq 0$, is in the kernel of $H(z)$, $\mathrm{Im}\, z > 0$. Let P be an $N \times (N+M)$ matrix defined by

$$P = \begin{pmatrix} I_N & 0 \end{pmatrix},$$

and $\overline{P}$ an $M \times (N+M)$ matrix

$$\overline{P} = \begin{pmatrix} 0 & I_M \end{pmatrix},$$

where I_N and I_M are the identity matrices of size N and M, respectively. Writing $\mathbf{v}_P := P\mathbf{v}$ and $\mathbf{v}_{\overline{P}} := \overline{P}\mathbf{v}$, we must have

$$-z\mathbf{v}_P + X^*\mathbf{v}_{\overline{P}} = 0, \qquad X\mathbf{v}_P - T^{-1}\mathbf{v}_{\overline{P}} = 0.$$

Thus,

$$XX^*\mathbf{v}_{\overline{P}}/z = T^{-1}\mathbf{v}_{\overline{P}}.$$

Taking the inner product with $\mathbf{v}_{\overline{P}}$, we find that the right-hand side becomes

$$\langle \mathbf{v}_{\overline{P}}, T^{-1}\mathbf{v}_{\overline{P}}\rangle = \sum_{\alpha=1}^{M} (T^{-1})_{\alpha\alpha} |\mathbf{v}_{\overline{P}}(\alpha)|^2,$$

which is strictly positive unless $\mathbf{v}_{\overline{P}} = 0$, while the left-hand side becomes

$$\langle \mathbf{v}_{\overline{P}}, XX^*\mathbf{v}_{\overline{P}}\rangle/z = \|X^*\mathbf{v}_{\overline{P}}\|^2/z,$$

which is zero if $X^*\mathbf{v}_{\overline{P}} = 0$ or, since $z \in \mathbb{C}^+$, not real valued. This shows that $\mathbf{v}_{\overline{P}} = 0$, and since $-z\mathbf{v}_P + X^*\mathbf{v}_{\overline{P}} = 0$, this also shows that $\mathbf{v}_P = 0$. We thus get a contradiction allowing us to conclude that $\mathbf{v} = 0$ which shows that the kernel of $H(z)$, $z \in \mathbb{C}^+$, is trivial, *that is*, $H(z)$ is invertible for $z \in \mathbb{C}^+$.

We define the "Green function," G, of $H \equiv H(z)$ by

$$G(z) := H(z)^{-1} \qquad (z \in \mathbb{C}^+), \tag{5.2}$$

and the normalized traces, m and $\widetilde{m}$, of G by

$$m(z) := \frac{1}{N}\sum_{a=1}^{N} G_{aa}(z), \qquad \widetilde{m}(z) := \frac{1}{M}\sum_{\alpha=N+1}^{M+N} G_{\alpha\alpha}(z) \qquad (z \in \mathbb{C}^+). \tag{5.3}$$

Note that by Schur's complement formula we have

$$PG(z)P = \frac{1}{PH(z)P - PH(z)\overline{P}\frac{1}{\overline{P}H(z)\overline{P}}\overline{P}H(z)P} = \frac{1}{-zP + X^*TX}, \tag{5.4}$$

so that

$$G_{ab}(z) = \big[(\widetilde{Q} - z)^{-1}\big]_{ab},$$

for any $a, b \in [\![1, N]\!]$. In particular,

$$m(z) = m_{\widetilde{Q}}(z).$$

Also note that

$$z^{-1}\overline{P}G(z)\overline{P} = \frac{z^{-1}}{\overline{P}H\overline{P} - \overline{P}HP\frac{1}{PHP}PH\overline{P}} = \frac{1}{-zT^{-1} + XX^*}. \tag{5.5}$$

In the following, we use lowercase Roman letters for indices in $[\![1, N]\!]$, Greek letters for indices in $[\![N+1, M+N]\!]$ and uppercase Roman letters for indices in $[\![1, N+M]\!]$.

5.2. *Green function, minors and partial expectations.* Recall the definitions of the $(N+M)\times(N+M)$ matrix $H \equiv H(z)$ in (5.1) and of the Green function G in (5.2).

For $\mathbb{T} \subset [\![1, N+M]\!]$, we define the "matrix minor" $H^{(\mathbb{T})}$ by setting

$$(H^{(\mathbb{T})})_{\mathrm{AB}} := \mathbb{1}(\mathrm{A} \notin \mathbb{T})\mathbb{1}(\mathrm{B} \notin \mathbb{T}) H_{\mathrm{AB}} \qquad \big(A, B \in [\![1, N+M]\!]\big), \tag{5.6}$$

that is, the entries in the columns/rows indexed by $\mathbb{T}$ are replaced by zeros. The Green function $G^{(\mathbb{T})}(z)$ associated with $H^{(\mathbb{T})}$ is defined by

$$G^{(\mathbb{T})}_{\mathrm{AB}}(z) := \left(\frac{1}{H^{(\mathbb{T})}(z) - z}\right)_{\mathrm{AB}} \qquad \big(A, B \in [\![1, N+M]\!]\big). \tag{5.7}$$

We use the shorthand notation

$$\sum_a^{(\mathbb{T})} := \sum_{\substack{a=1\\ a\notin\mathbb{T}}}^{N}, \qquad \sum_{a\neq b}^{(\mathbb{T})} := \sum_{\substack{a=1,b=1\\ a\neq b, a,b\notin\mathbb{T}}}^{N}, \qquad \sum_{\alpha}^{(\mathbb{T})} := \sum_{\substack{\alpha=N+1\\ \alpha\notin\mathbb{T}}}^{N+M},$$

$$\sum_{\alpha\neq\beta}^{(\mathbb{T})} := \sum_{\substack{\alpha=N+1,\beta=N+1\\ \alpha\neq\beta,\alpha,\beta\notin\mathbb{T}}}^{N+M},$$

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

We further set

$$m^{(\mathbb{T})} := \frac{1}{N}\sum_a^{(\mathbb{T})} G^{(\mathbb{T})}_{aa}, \qquad \widetilde{m}^{(\mathbb{T})} := \frac{1}{M}\sum_\alpha^{(\mathbb{T})} G^{(\mathbb{T})}_{\alpha\alpha}.$$

Note that we use the normalizations N^{-1} and M^{-1} here since they are more convenient in computations.

Finally, we denote by $\mathbb{E}_a$, $\mathbb{E}_\alpha$ the partial expectation with respect to the variables $(x_{\alpha a})_{\alpha=M+1}^{N+M}$, respectively, $(x_{\alpha a})_{a=1}^{N}$.

5.3. *Green function identities.* The next lemma collects the main identities between the matrix elements of G and $G^{(\mathbb{T})}$.

LEMMA 5.1. *Let $G \equiv G(z)$, $z \in \mathbb{C}^+$, be defined in* (5.2). *Assume that the matrix T is diagonal. Then, for $a, b \in [\![1, N]\!]$, $\alpha, \beta \in [\![N+1, N+M]\!]$,* A, B, C $\in [\![1, N+M]\!]$, *the following identities hold*:

– *Schur complement/Feshbach formula*: *For any a and α*,

$$G_{aa} = \frac{1}{-z - \sum_{\alpha,\beta} x_{\alpha a} G^{(a)}_{\alpha\beta} x_{\beta a}},$$
$$G_{\alpha\alpha} = \frac{1}{-(T^{-1})_{\alpha\alpha} - \sum_{a,b} x_{\alpha a} G^{(\alpha)}_{ab} x_{\alpha b}}. \tag{5.8}$$

– *For* $a \neq b$,

$$G_{ab} = -G_{aa} \sum_{\alpha} x_{\alpha a} G^{(a)}_{\alpha b} = -G_{bb} \sum_{\beta} G^{(b)}_{a\beta} x_{\beta b}. \tag{5.9}$$

– *For* $\alpha \neq \beta$,

$$G_{\alpha\beta} = -G_{\alpha\alpha} \sum_{a} x_{\alpha a} G^{(\alpha)}_{a\beta} = -G_{\beta\beta} \sum_{b} G^{(\beta)}_{\alpha b} x_{\beta b}. \tag{5.10}$$

– *For any* a *and* α,

$$G_{a\alpha} = -G_{aa} \sum_{\beta} x_{\beta a} G^{(a)}_{\beta\alpha} = -G_{\alpha\alpha} \sum_{b} G^{(\alpha)}_{ab} x_{\alpha b}. \tag{5.11}$$

– *For* $\mathrm{A}, \mathrm{B} \neq \mathrm{C}$,

$$G_{\mathrm{AB}} = G^{(\mathrm{C})}_{\mathrm{AB}} + \frac{G_{\mathrm{AC}} G_{\mathrm{CB}}}{G_{\mathrm{CC}}}. \tag{5.12}$$

– *Ward identity*: *For any* a,

$$\sum_{b} |G_{ab}|^2 = \frac{\operatorname{Im} G_{aa}}{\eta}. \tag{5.13}$$

For the proof, see Lemma 4.2 in [20], Lemma 6.10 in [16] and equation (3.31) in [21].

5.4. *Local law for* H *at the edge*. Consider two families of random variables (X_i) and (Y_i), with $i \in [\![1, N]\!]$, satisfying

$$\mathbb{E} Z_i = 0, \qquad \mathbb{E}|Z_i|^2 = 1, \qquad \mathbb{E}|Z_i|^p \leq c_p \qquad (p \geq 3), \tag{5.14}$$

$Z_i = X_i, Y_i$, for all $p \in \mathbb{N}$ and some constants c_p, uniformly in $i \in [\![1, N]\!]$. The following lemma, taken from [17], provides useful large deviation estimates.

LEMMA 5.2. *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* (5.14). *Then we have the bounds*

$$\Big|\sum_{i} b_i X_i\Big| \prec \Big(\sum_{i} |b_i|^2\Big)^{1/2}, \tag{5.15}$$

$$\Big|\sum_{i} \sum_{j} a_{ij} X_i Y_j\Big| \prec \Big(\sum_{i,j} |a_{ij}|^2\Big)^{1/2}, \tag{5.16}$$

$$\Big|\sum_{i} \sum_{j} a_{ij} X_i X_j - \sum_{i} a_{ii}\Big| \prec \Big(\sum_{i,j} |a_{ij}|^2\Big)^{1/2}. \tag{5.17}$$

If the coefficients a_{ij} and b_i depend on an additional parameter u, then all of these estimates are uniform in u, that is, the threshold $N_0 = N_0(\varepsilon, D)$ in the definition of $\prec$ depends only on the family (c_p) from (5.14); *in particular, N_0 does not depend on u.*

From the large deviation estimates in Lemma 5.2 and the local law in Lemma 3.3, we obtain the following estimates.

LEMMA 5.3. *Let $G \equiv G(z)$, $z \in \mathbb{C}^+$, be defined in* (5.2). *Suppose that T is diagonal, that is, $T = \operatorname{diag}(t_\alpha)$. Then, under Assumption* 2.2, *the Green function G satisfies the following bounds uniformly in z on $\mathcal{D}(c, \varepsilon)$ (with $c, \varepsilon > 0$ as in Lemma* 3.3):

(i) *For any $\alpha \in [\![N+1, N+M]\!]$,*

$$|G_{\alpha\alpha}(z)| \prec 1, \qquad \operatorname{Im} G_{\alpha\alpha}(z) \prec \Psi. \tag{5.18}$$

(ii) *For any $a \in [\![1, N]\!]$ and $\alpha \in [\![N+1, N+M]\!]$,*

$$|G_{a\alpha}(z)| \prec \Psi. \tag{5.19}$$

(iii) *For any $\alpha, \beta \in [\![N+1, N+M]\!]$ with $\alpha \neq \beta$,*

$$|G_{\alpha\beta}(z)| \prec \Psi. \tag{5.20}$$

PROOF. We first claim that $|m - m^{(\alpha)}| \prec \Psi$. To prove this claim, we first let

$$\widetilde{\mathcal{Q}} := T^{1/2} X X^* T^{1/2}.$$

We notice that the normalized trace m of the Green function can be written in terms of $\widetilde{\mathcal{Q}}$ as

$$m(z) = m_{\widetilde{\mathcal{Q}}}(z) = \frac{1}{N}\left(\operatorname{Tr}(\widetilde{\mathcal{Q}} - z)^{-1} + \frac{N-M}{z}\right).$$

Next, we consider the minor $\widetilde{\mathcal{Q}}^{(\alpha)}$, which is obtained by removing all columns and rows of $\widetilde{\mathcal{Q}}$ indexed by α. Then

$$m^{(\alpha)}(z) = \frac{1}{N}\left(\operatorname{Tr}(\widetilde{\mathcal{Q}}^{(\alpha)} - z)^{-1} + \frac{N-M+1}{z}\right).$$

By Cauchy's eigenvalue interlacing property, we get

$$\left|\operatorname{Tr}(\widetilde{\mathcal{Q}} - z)^{-1} - \operatorname{Tr}(\widetilde{\mathcal{Q}}^{(\alpha)} - z)^{-1}\right| \le C\eta^{-1}.$$

(See Lemma 5.4 in [6] or Lemma 8.2 in [12].) This proves the desired claim.

From Schur's complement formula (5.8), we obtain that

$$\frac{1}{G_{\alpha\alpha}} = -t_\alpha^{-1} - \sum_{k,l} x_{\alpha k} G_{kl}^{(\alpha)} x_{\alpha l}.$$

Further, from the large deviation estimate (5.17) and the Ward identity (5.13), we find

$$\begin{aligned}\Big|m^{(\alpha)}-\sum_{k,l}x_{\alpha k}G^{(\alpha)}_{kl}x_{\alpha l}\Big| &\prec \frac{1}{N}\Big(\sum_{k,l}|G^{(\alpha)}_{kl}|^2\Big)^{1/2}\\ &=\sqrt{\frac{\operatorname{Im} m^{(\alpha)}}{N\eta}}\leq\sqrt{\frac{|m^{(\alpha)}-\widehat{m}_{\mathrm{fc}}|+\operatorname{Im}\widehat{m}_{\mathrm{fc}}}{N\eta}}\\ &\prec\sqrt{\frac{\Psi}{N\eta}}+\sqrt{\frac{\operatorname{Im}\widehat{m}_{\mathrm{fc}}}{N\eta}}\prec\Psi,\end{aligned}\tag{5.21}$$

uniformly in $z\in\mathcal{D}(c,\varepsilon)$.

Since $|\xi_+ + m|\prec\Psi$ and $t_\alpha^{-1}\geq\gamma_0(1+c)\xi_+$ for some $c>0$ [see (2.15)], we get $1\prec|G_{\alpha\alpha}|^{-1}$, hence $|G_{\alpha\alpha}|\prec 1$. Moreover, using once more Schur's complement formula (5.8), we find

$$\operatorname{Im} G_{\alpha\alpha}=\frac{\operatorname{Im}\sum_{k,l}x_{\alpha k}G^{(\alpha)}_{kl}x_{\alpha l}}{|-(T^{-1})_{\alpha\alpha}-\sum_{k,l}x_{\alpha k}G^{(\alpha)}_{kl}x_{\alpha l}|^2},$$

hence

$$|\operatorname{Im} G_{\alpha\alpha}|\prec\Big|\operatorname{Im}\sum_{k,l}x_{\alpha k}G^{(\alpha)}_{kl}x_{\alpha l}\Big|\prec\operatorname{Im} m^{(\alpha)}+\Big|m^{(\alpha)}-\sum_{k,l}x_{\alpha k}G^{(\alpha)}_{kl}x_{\alpha l}\Big|\prec\Psi,$$

uniformly in $\mathcal{D}(c,\varepsilon)$, where we used (5.21) and that $\operatorname{Im} m^{(\alpha)}\prec\operatorname{Im} m_{\mathrm{fc}}\sim\sqrt{\kappa+\eta}$. This proves statement (i).

From the Green function identity (5.11) and statement (i), we have

$$|G_{a\alpha}|=\Big|G_{\alpha\alpha}\sum_k x_{\alpha k}G^{(\alpha)}_{ak}\Big|\prec\Big|\sum_k x_{\alpha k}G^{(\alpha)}_{ak}\Big|,$$

where we used the local law of Lemma 3.3 and the fact that $|G_{\alpha\alpha}|\prec 1$. Further, it is obvious that the local law $|G^{(\alpha)}_{ka}|\prec\Psi$ holds, which can be proved in the same way as $|G_{ka}|\prec\Psi$ but without the αth column and the row. Thus, applying the large deviation estimate (5.15) and the local law $|G^{(\alpha)}_{ka}|\prec\Psi$, we get

$$|G_{a\alpha}|\prec\Big(\frac{1}{N}\sum_k|G^{(\alpha)}_{ka}|^2\Big)^{1/2}\prec\Psi,$$

uniformly in $z\in\mathcal{D}(c,\varepsilon)$, which proves statement (ii) of the lemma.

Similarly, we have from the Green function identity (5.10) and the large deviation estimate (5.15) that

$$|G_{\alpha\beta}|=\Big|G_{\alpha\alpha}\sum_k x_{\alpha k}G^{(\alpha)}_{k\beta}\Big|\prec\Big|\sum_k x_{\alpha k}G^{(\alpha)}_{k\beta}\Big|\prec\Big(\frac{1}{N}\sum_k|G^{(\alpha)}_{k\beta}|^2\Big)^{1/2}\prec\Psi,$$

where we used $|G_{k\beta}^{(\alpha)}| \prec \Psi$ [which is analogous to the statement (ii)] to get the last inequality. This proves statement (iii) of the lemma. □

We conclude this section by giving estimates on expectations of monomials of Green functions entries.

LEMMA 5.4. *Let $P \equiv P(z)$ be a monomial in the Green function entries $(G_{\mathrm{AB}}(z))$, with $z \in \mathcal{D}(c, \varepsilon)$, for some $\varepsilon, c > 0$. Then there exists a universal constant C, such that*

$$\mathbb{E}\big|P(z)\big|^2 \prec N^{Cn}, \tag{5.22}$$

where n is the degree of P. In particular, if $|P(z)| \prec \Psi(z)^k$, uniformly in $\mathcal{D}(c, \varepsilon)$, then $\mathbb{E}|P(z)| \prec \Psi(z)^k$, uniformly in $\mathcal{D}(c, \varepsilon)$. (See the paragraph after Definition 3.1.)

Moreover, the same conclusions hold with $G^{(\mathbb{T})}$ replacing G for any $\mathbb{T}$.

PROOF. First, we note that $|G_{ab}| \le \frac{1}{\eta}$, $a, b \in [\![1, N]\!]$, as follows from the self-adjointness of X^*TX and the spectral calculus.

Second, to bound $|G_{\alpha\beta}|$, $\alpha, \beta \in [\![N+1, M+N]\!]$, we recall that T^{-1} is a strictly positive operator by Assumption 2.2. Thus,

$$\mathrm{Im}\langle \mathbf{v}, (zT^{-1} - XX^*)\mathbf{v}\rangle = \eta\langle \mathbf{v}, T^{-1}\mathbf{v}\rangle \ge c\eta\|\mathbf{v}\|^2 \qquad \forall \mathbf{v} \in \mathbb{C}^M,$$

for some $c > 0$ independent of $\mathbf{v}$, where $\langle\cdot,\cdot\rangle$ denotes the canonical inner product in $\mathbb{C}^M$. Since $z^{-1}\overline{P}G(z)\overline{P} = (-zT^{-1} + XX^*)^{-1}$, $|z| > 0$, we get $|G_{\alpha\beta}| \le \frac{C|z|}{\eta}$.

Third, to bound $\mathbb{E}|G_{a\alpha}|^p$, $a \in [\![1, N]\!]$, $\alpha \in [\![N+1, N+M]\!]$, $p \ge 0$, we note that by (5.11) we have

$$|G_{a\alpha}| = |G_{aa}|\sum_{\beta}|x_{\beta a}G_{\beta\alpha}^{(a)}| \le \frac{C|z|}{\eta^2}N|x_{\beta a}|, \tag{5.23}$$

by the estimates above. From the moment bounds in (2.2), we then conclude that $\mathbb{E}|G_{\alpha a}|^p \le C^pN^{cp}$, where we also used that $\eta \gg N^{-1}$, $0 < |z| < C$ by assumption.

The lemma now easily follows from Hölder's inequality. □

In the rest of the paper we prove Proposition 4.1 with the formalism outlined in this section. The actual calculation will be done for the simple case $F' \equiv 1$; the proof for general F' is basically the same, though the computations are much longer for this case. The details for $F' \not\equiv 1$ can be found in [32].

6. Green function flow. The key idea of our proof of Proposition 4.1 is similar to the one of the proof of Proposition 5.2 in [32] for deformed Wigner matrices: We consider a continuous interpolation between the sample covariance matrices $\widetilde{Q}$ and W by introducing a time evolution that deforms T continuously to the identity. We then track the associated flow of the Green function for sufficiently long time. The outcome is an estimate on the time derivative of the Green function which is sufficiently accurate to prove Proposition 4.1.

6.1. *Preliminaries.* Suppose that $T = \gamma_0 \Sigma$ is diagonal, that is, $T = \operatorname{diag}(t_\alpha)$. We interpolate between $\Sigma = \operatorname{diag}(\sigma_\alpha)$ and the identity matrix $\mathbb{1}$ by introducing the time evolution $t \mapsto (\sigma_\alpha(t))$ defined by

$$\frac{1}{\sigma_\alpha(t)} = \mathrm{e}^{-t} \frac{1}{\sigma_\alpha(0)} + \bigl(1 - \mathrm{e}^{-t}\bigr), \qquad \Sigma(t) = \operatorname{diag}\bigl(\sigma_\alpha(t)\bigr) \qquad (t \ge 0). \tag{6.1}$$

From (2.12), it is natural to let $\xi_+(t)$ be the largest solution to

$$\frac{1}{M} \sum_\alpha \biggl(\frac{\sigma_\alpha(t)\xi_+(t)}{1 - \sigma_\alpha(t)\xi_+(t)} \biggr)^2 = d, \tag{6.2}$$

with $\xi_+(0) = \xi_+$. We then choose the scaling factor $\gamma \equiv \gamma(t)$ to be given by

$$\gamma(t) = \biggl(\frac{1}{N} \sum_\alpha \biggl(\frac{\sigma_\alpha(t)}{1 - \sigma_\alpha(t)\xi_+(t)} \biggr)^3 + \bigl(\xi_+(t)\bigr)^{-3} \biggr)^{-1/3}, \tag{6.3}$$

with $\gamma_0 = \gamma(0)$. (See also Remark 6.1.) We also introduce

$$\tau \equiv \tau(t) := \frac{\xi_+(t)}{\gamma(t)}. \tag{6.4}$$

For simplicity, we often omit the t-dependence in the notation for $T(t)$, $\gamma(t)$ and $\tau(t)$ in the following. Note that we have from (6.2), (6.3) and (6.4) that

$$\frac{1}{N} \sum_\alpha \biggl(\frac{1}{t_\alpha^{-1} - \tau} \biggr)^2 = \frac{1}{\tau^2}, \qquad \frac{1}{N} \sum_\alpha \biggl(\frac{1}{t_\alpha^{-1} - \tau} \biggr)^3 + \frac{1}{\tau^3} = 1. \tag{6.5}$$

In the following, we refer to the identities in (6.5) as "sum rules".

We let $z \equiv z(t)$ be time-dependent. Define the $(N+M) \times (N+M)$ matrix $H(t) \equiv H(z,t)$ by

$$H(z,t) = \begin{pmatrix} -z(t) I_N & X^* \\ X & -T^{-1}(t) \end{pmatrix},$$

with $T(t) = \gamma(t)\Sigma(t)$, $T(0) = \gamma_0 \Sigma$. We also let

$$G(z,t) := H(z,t)^{-1}, \qquad m(z,t) = \frac{1}{N} \sum_a G_{aa}(z,t) \qquad (z \in \mathbb{C}^+). \tag{6.6}$$

We now consider the evolution of the Green function $G \equiv G(t)$ under the evolution governed by (6.1). For the diagonal Green function entries G_{ii}, $i \in [\![1, N]\!]$, we get

$$\mathbb{E}\frac{\partial G_{ii}}{\partial t} = \dot{z}\sum_a \mathbb{E}[G_{ia}G_{ai}] - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}[G_{i\alpha}G_{\alpha i}]. \tag{6.7}$$

REMARK 6.1. Let $\widetilde{m}_{\mathrm{fc}}(z,t)$ be the solution to

$$\widetilde{m}_{\mathrm{fc}}(z,t) = \frac{1}{-z + \frac{1}{dM}\sum_\alpha \frac{t_\alpha}{t_\alpha \widetilde{m}_{\mathrm{fc}}(z,t)+1}} \qquad \big(z \in \mathbb{C}^+, t \geq 0\big)$$

such that $\operatorname{Im}\widetilde{m}_{\mathrm{fc}}(z,t) \geq 0$.

Setting $\widetilde{\rho}_{\mathrm{fc}}(E,t) := \lim_{\eta \searrow 0} \pi^{-1}\operatorname{Im}\widetilde{m}_{\mathrm{fc}}(E + \mathrm{i}\eta, t)$, we note that the rightmost point of the support of the measure $\widetilde{\rho}_{\mathrm{fc}}(t)$, denoted by $L_+ \equiv L_+(t)$, is given by $L_+ = \gamma E_+$, or equivalently,

$$L_+ = \frac{1}{\tau} + \frac{1}{dM}\sum_\alpha \frac{t_\alpha}{1 - t_\alpha \tau} = \frac{1}{\tau} + \frac{1}{N}\sum_\alpha \frac{1}{t_\alpha^{-1} - \tau}. \tag{6.8}$$

In fact, the rescaling by $\gamma(t)$ assures that

$$\widetilde{\rho}(E,t) = \frac{1}{\pi}\sqrt{L_+ - E}\big(1 + O(L_+ - E)\big) \qquad (t \geq 0),$$

as $E \nearrow L_+$, as may be checked by an explicit computation. In the framework of Remark 2.5, this choice of $\gamma(t)$ can be obtained by introducing $C_0(t)$ and $g_t(x)$, the extensions of C_0 and g_t in Remark 2.5, defined by

$$C_0(t) = \frac{1}{\pi}\left(\frac{2}{g_t''(\xi_+(t))}\right)^{1/2}, \qquad g_t(x) = \frac{1}{x} + \frac{1}{d}\int_{\mathbb{R}} \frac{E}{1 - Ex}\,\mathrm{d}\widetilde{\rho}(E,t).$$

6.2. *Proof of Proposition* 4.1. In this subsection we give the proof of Proposition 4.1, which is based on two technical lemmas, Lemmas 6.2 and 6.3 below. For simplicity, we choose $F' \equiv \mathbb{1}$. Recall the definition of the deterministic control parameter Ψ in (3.5).

The main ingredient of the proof of the Green function comparison theorem, Proposition 4.1, is the estimate $\operatorname{Im}\mathbb{E}[\partial_t G_{ii}(z)] = \mathcal{O}(M\Psi^5)$, for appropriately chosen z. (The naive size of $\mathbb{E}[\partial_t G_{ii}]$ is $\mathcal{O}(M\Psi^2)$ as one sees from (6.7).) Once we have established the estimate $\operatorname{Im}\mathbb{E}[\partial_t G_{ii}(z)] = \mathcal{O}(M\Psi^5)$, we can integrate it from $t = 0$ to $t = 2\log N$ to compare $\operatorname{Im} m_{\widetilde{Q}}$ with $\operatorname{Im} m|_{t=2\log N}$. The comparison between $\operatorname{Im} m|_{t=2\log N}$ and m_W can easily be done, since $\Sigma(t)$ is close enough to the identity at $t = 2\log N$.

To show that the imaginary part of (6.7) is much smaller than its naive size, we use, in a first step, the following "decoupling" lemma.

LEMMA 6.2. *Under the assumptions of Proposition 4.1 the following holds true. Let $z(t) \equiv z = L_+(t) + y + \mathrm{i}\eta$, with $L_+(t)$ as in (6.8), $y \in [-N^{-2/3+\varepsilon}, N^{-2/3+\varepsilon}]$ and $\eta = N^{-2/3-\varepsilon}$.*

Then there are z-dependent random variables X_{22}, X_{32}, X_{33}, X_{42}, X_{43}, X_{44} and X'_{44}, satisfying

$$X_{22} = \mathcal{O}(\Psi^2), \qquad X_{32}, X_{33} = \mathcal{O}(\Psi^3), \qquad X_{42}, X_{43}, X_{44}, X'_{44} = \mathcal{O}(\Psi^4),$$

such that

$$\begin{aligned}
&\mathbb{E}_\alpha[G_{i\alpha}G_{\alpha i}] \\
&\quad= \frac{1}{(t_\alpha^{-1}-\tau)^2}X_{22} - \frac{2}{(t_\alpha^{-1}-\tau)^3}X_{32} - \frac{2}{(t_\alpha^{-1}-\tau)^3}X_{33} + \frac{3}{(t_\alpha^{-1}-\tau)^4}X_{42} \\
&\qquad+ \frac{6}{(t_\alpha^{-1}-\tau)^4}X_{43} + \frac{12}{(t_\alpha^{-1}-\tau)^4}X_{44} + \frac{3}{(t_\alpha^{-1}-\tau)^4}X'_{44} + \mathcal{O}(\Psi^5),
\end{aligned} \tag{6.9}$$

uniformly in $t \ge 0$. The random variables above are explicitly given by

$$X_{22} = \frac{1}{N}\sum_s G_{is}G_{si}, \qquad X_{32} = (m+\tau)\frac{1}{N}\sum_s G_{is}G_{si},$$

$$X_{33} = \frac{1}{N^2}\sum_{r,s} G_{ir}G_{rs}G_{si}, \qquad X_{42} = (m+\tau)^2\frac{1}{N}\sum_s G_{is}G_{si},$$

$$X_{43} = (m+\tau)\frac{1}{N}\sum_{r,s} G_{ir}G_{rs}G_{si}, \qquad X_{44} = \frac{1}{N^3}\sum_{r,s,t} G_{ir}G_{rs}G_{st}G_{ti},$$

$$X'_{44} = \frac{1}{N^3}\sum_{r,s,t} G_{is}G_{si}G_{rt}G_{tr},$$

where $G \equiv G(z(t),t)$, $m \equiv m(z(t),t) = \frac{1}{N}\sum_s G_{ss}(z(t),t)$ and $\tau(t)$ is defined in (6.4).

We refer to Lemma 6.2 as a "decoupling" lemma, since on the right-hand side of (6.9) the Greek index α is, up to the error $\mathcal{O}(\Psi^5)$, decoupled from the Green functions which only have Roman indices as lower indices. Lemma 6.2 is proven in Appendix A.

Taking the time derivative of (6.8), we get

$$\begin{aligned}
\dot{z} = \dot{L}_+ &= -\frac{\dot{\tau}}{\tau^2} + \frac{1}{dM}\sum_\alpha \frac{t_\alpha^2\dot{\tau}}{(1-t_\alpha\tau)^2} + \frac{1}{dM}\sum_\alpha \frac{\partial_t t_\alpha}{1-t_\alpha\tau} \\
&\quad+ \frac{1}{dM}\sum_\alpha \frac{\tau t_\alpha(\partial_t t_\alpha)}{(1-t_\alpha\tau)^2}.
\end{aligned} \tag{6.10}$$

From (6.5), we observe that the first two terms on the right-hand side of (6.10) cancel. Thus, simplifying the last two terms in (6.10), we obtain

$$\dot{z} = \frac{1}{dM}\sum_{\alpha}\frac{\partial_t t_\alpha}{(1-t_\alpha\tau)^2} = \frac{1}{N}\sum_{\alpha}\frac{\partial_t t_\alpha}{t_\alpha^2}\frac{1}{(t_\alpha^{-1}-\tau)^2}. \tag{6.11}$$

Note that, by definition,

$$\partial_t\frac{1}{\sigma_\alpha(t)} = -\mathrm{e}^{-t}\frac{1}{\sigma_\alpha(0)} + \mathrm{e}^{-t} = 1 - \frac{1}{\sigma_\alpha(t)} \qquad (t \geq 0),$$

in particular, $\partial_t \sigma_\alpha^{-1}(t) = O(1)$, and

$$\frac{\partial_t t_\alpha}{t_\alpha^2} = -\partial_t\frac{1}{t_\alpha(t)} = \left(\frac{\partial_t\gamma(t)}{\gamma(t)}+1\right)\frac{1}{t_\alpha(t)} - \frac{1}{\gamma(t)} \qquad (t \geq 0).$$

(See the proof of Lemma 6.3 in Appendix C.) Moreover, from the definition of $\xi_+(t)$ in (6.2), it can be easily checked that $\partial_t\xi_+(t) = O(1)$, which also shows that $\partial_t\gamma(t) = O(1)$. Since $\gamma(t) \sim 1$, hence $t_\alpha(t) \sim 1$ as well, we find that $(\partial_t t_\alpha)/t_\alpha^2 = O(1)$.

Thus, plugging (6.9) into (6.7) we find

$$\begin{aligned}
\mathbb{E}\frac{\partial G_{ii}}{\partial t} &= \dot{z}\sum_a \mathbb{E}[G_{ia}G_{ai}] - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\frac{1}{(t_\alpha^{-1}-\tau)^2}\frac{1}{N}\sum_a \mathbb{E}[G_{ia}G_{ai}] \\
&\quad + \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}\left[\frac{2}{(t_\alpha^{-1}-\tau)^3}X_{32} + \frac{2}{(t_\alpha^{-1}-\tau)^3}X_{33}\right] \\
&\quad - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}\left[\frac{3}{(t_\alpha^{-1}-\tau)^4}X_{42} + \frac{6}{(t_\alpha^{-1}-\tau)^4}X_{43}\right] \\
&\quad - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}\left[\frac{12}{(t_\alpha^{-1}-\tau)^4}X_{44} + \frac{3}{(t_\alpha^{-1}-\tau)^4}X'_{44}\right] + \mathcal{O}(M\Psi^5).
\end{aligned} \tag{6.12}$$

Note that the first two terms in (6.12) cancel by (6.11) and that we have

$$\begin{aligned}
\mathbb{E}\frac{\partial G_{ii}}{\partial t} &= \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}\left[\frac{2}{(t_\alpha^{-1}-\tau)^3}X_{32} + \frac{2}{(t_\alpha^{-1}-\tau)^3}X_{33}\right] \\
&\quad - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}\left[\frac{3}{(t_\alpha^{-1}-\tau)^4}X_{42} + \frac{6}{(t_\alpha^{-1}-\tau)^4}X_{43}\right] \\
&\quad - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\mathbb{E}\left[\frac{12}{(t_\alpha^{-1}-\tau)^4}X_{44} + \frac{3}{(t_\alpha^{-1}-\tau)^4}X'_{44}\right] + \mathcal{O}(M\Psi^5).
\end{aligned} \tag{6.13}$$

To complete the proof of Proposition 4.1, we are going to show that the imaginary part of the right-hand side of (6.13) is of $\mathcal{O}(M\Psi^5)$ as is noted in the next lemma.

LEMMA 6.3. *Under the assumptions of Proposition 4.1 with the notation of Lemma 6.2, we have*

$$
\begin{aligned}
&\sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2} \operatorname{Im} \mathbb{E}\Big[\frac{2}{(t_\alpha^{-1}-\tau)^3} X_{32} + \frac{2}{(t_\alpha^{-1}-\tau)^3} X_{33}\Big] \\
&\quad - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2} \operatorname{Im} \mathbb{E}\Big[\frac{3}{(t_\alpha^{-1}-\tau)^4} X_{42}\Big] \\
&\quad - \sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2} \operatorname{Im} \mathbb{E}\Big[\frac{6}{(t_\alpha^{-1}-\tau)^4} X_{43} + \frac{12}{(t_\alpha^{-1}-\tau)^4} X_{44} + \frac{3}{(t_\alpha^{-1}-\tau)^4} X'_{44}\Big] \\
&\quad = \mathcal{O}(M\Psi^5),
\end{aligned} \tag{6.14}
$$

uniformly in $t \geq 0$.

We remark that the naive size of the right-hand side of (6.14) is $\mathcal{O}(M\Psi^3)$, but for our choice of γ the terms cancel up to errors of $\mathcal{O}(M\Psi^5)$. Similar to the discussion in [32], the sum rules in (6.5) have crucial roles in this cancellation mechanism. Lemma 6.3 is proven in the Appendix C.

PROOF OF PROPOSITION 4.1. For simplicity, we choose $F' \equiv \mathbb{1}$. From (6.13) and Lemma 6.3, we find that

$$
\mathbb{E}\Big[\operatorname{Im} \frac{\partial G_{ii}}{\partial t}\Big] = \mathcal{O}(\Psi^2). \tag{6.15}
$$

Integrating both sides of (6.15) from $t = 0$ to $t = 2\log N$, we obtain that

$$
\begin{aligned}
&\bigg|\mathbb{E}\Big[N \int_{E_1}^{E_2} \operatorname{Im} m(x + L_+ + \mathrm{i}\eta)\Big|_{t=0} \mathrm{d}x\Big] \\
&\quad - \mathbb{E}\Big[N \int_{E_1}^{E_2} \operatorname{Im} m(x + L_+ + \mathrm{i}\eta)\Big|_{t=2\log N} \mathrm{d}x\Big]\bigg| \leq N^{-1/3+C'\varepsilon},
\end{aligned} \tag{6.16}
$$

for some constant $C' > 0$.

At $t = \infty$, we have $\sigma_\alpha(\infty) = 1$, for all $\alpha \in [\![N+1, N+M]\!]$, hence by definition

$$
\xi_+(\infty) = \frac{\sqrt{d}}{1+\sqrt{d}}, \qquad \gamma(\infty) = \sqrt{d}(1+\sqrt{d})^{-4/3}.
$$

In particular,

$$
m(x + L_+ + \mathrm{i}\eta)|_{t=\infty} = m_W(x + M_+ + \mathrm{i}\eta).
$$

Let $T_{\mathfrak{f}} := 2\log N$. At $t = T_{\mathfrak{f}}$, we have $\sigma_\alpha(T_{\mathfrak{f}}) = 1 + O(N^{-2})$. Using the result at $t = \infty$, it can be easily seen that

$$
\gamma(T_{\mathfrak{f}}) = \gamma(\infty) + O(N^{-2}).
$$

Similarly, we also have that $z(T_{\mathfrak{f}}) = z(\infty) + O(N^{-2})$. Thus, the matrix $H(T_{\mathfrak{f}}) - H(\infty)$ is a diagonal matrix whose entries are $O(N^{-2})$.

Using the resolvent identity

$$G(T_{\mathfrak{f}}) - G(\infty) = -G(T_{\mathfrak{f}})\big(H(T_{\mathfrak{f}}) - H(\infty)\big)G(\infty), \tag{6.17}$$

we can now bound

$$\big|G_{ii}(T_{\mathfrak{f}}) - G_{ii}(\infty)\big| = \left|\sum_A -G_{iA}(T_{\mathfrak{f}})\big(H_{AA}(T_{\mathfrak{f}}) - H_{AA}(\infty)\big)G_{Ai}(\infty)\right| \prec N^{-5/3},$$

and we thus have

$$\begin{aligned}&\left|\mathbb{E}\left[N\int_{E_1}^{E_2}\operatorname{Im} m(x + L_+ + \mathrm{i}\eta)\Big|_{t=2\log N}\mathrm{d}x\right]\right.\\ &\qquad \left.- \mathbb{E}\left[N\int_{E_1}^{E_2}\operatorname{Im} m(x + L_+ + \mathrm{i}\eta)\Big|_{t=\infty}\mathrm{d}x\right]\right| \le N^{-4/3+C'\varepsilon}.\end{aligned} \tag{6.18}$$

Since $m(x + L_+ + \mathrm{i}\eta)|_{t=0} = m_{\widetilde{Q}}(x + L_+ + \mathrm{i}\eta)$, we get the desired result from (6.16) and (6.18). □

APPENDIX A: PROOF OF LEMMA 6.2

In this section we prove Lemma 6.2. We start expanding $\mathbb{E}[G_{i\alpha}G_{\alpha i}]$ in the random variables indexed by the Greek index α. The following expansion follows closely the expansions used in [32].

PROOF OF LEMMA 6.2. Using the formula for $G_{i\alpha}$ in (5.9), that is,

$$G_{i\alpha} = -G_{\alpha\alpha}\sum_k x_{\alpha k}G_{ik}^{(\alpha)},$$

we expand $G_{i\alpha}G_{\alpha i}$ in the lower index α as

$$G_{i\alpha}G_{\alpha i} = G_{\alpha\alpha}^2\sum_{k,l} G_{ik}^{(\alpha)}x_{\alpha k}x_{\alpha l}G_{li}^{(\alpha)}. \tag{A.1}$$

Note that, by Schur's complement formula (5.8),

$$G_{\alpha\alpha} = \frac{1}{h_{\alpha\alpha} - \sum_{p,q}^{(\alpha)} h_{\alpha p}G_{pq}^{(\alpha)}h_{q\alpha}} = \frac{1}{-t_\alpha^{-1} - \sum_{p,q} x_{\alpha p}G_{pq}^{(\alpha)}x_{\alpha q}}. \tag{A.2}$$

(The use of Roman letters p, q can be justified since $h_{\alpha p} = 0$ for $p \in [\![N+1, N+M]\!]$ and $p \neq \alpha$.)

We next expand $G_{\alpha\alpha}$ around $(-t_\alpha^{-1} + \tau)^{-1}$. (Note that $\limsup t_\alpha\tau < 1$, thus $t_\alpha^{-1} - \tau > c > 0$ for some constant c independent of N.) From the large deviation estimates in Lemma 5.2 and the Ward identity (5.13), we have

$$\left|\sum_{p,q} x_{\alpha p}G_{pq}^{(\alpha)}x_{\alpha q} + \tau\right| \prec \Psi. \tag{A.3}$$

Returning to (A.2), we thus have

$$G_{\alpha\alpha} = \frac{1}{-t_\alpha^{-1}+\tau} + \frac{1}{(-t_\alpha^{-1}+\tau)^2}\Big(\sum_{p,q} x_{\alpha p} G_{pq}^{(\alpha)} x_{\alpha q} + \tau\Big) + \frac{1}{(-t_\alpha^{-1}+\tau)^3}\Big(\sum_{p,q} x_{\alpha p} G_{pq}^{(\alpha)} x_{\alpha q} + \tau\Big)^2 + \mathcal{O}(\Psi^3),$$

respectively,

$$G_{\alpha\alpha}^2 = \frac{1}{(t_\alpha^{-1}-\tau)^2} - \frac{2}{(t_\alpha^{-1}-\tau)^3}\Big(\sum_{p,q} x_{\alpha p} G_{pq}^{(\alpha)} x_{\alpha q} + \tau\Big) + \frac{3}{(t_\alpha^{-1}-\tau)^4}\Big(\sum_{p,q} x_{\alpha p} G_{pq}^{(\alpha)} x_{\alpha q} + \tau\Big)^2 + \mathcal{O}(\Psi^3).$$

Hence, from the resolvent identity (A.1), obtain the following expansion of $G_{i\alpha}G_{\alpha i}$ in the lower index α,

$$\begin{aligned} G_{i\alpha}G_{\alpha i} &= \frac{1}{(t_\alpha^{-1}-\tau)^2}\sum_{s,t} G_{is}^{(\alpha)} x_{\alpha s} x_{\alpha t} G_{ti}^{(\alpha)} \\ &\quad - \frac{2}{(t_\alpha^{-1}-\tau)^3}\Big(\sum_{p,q}^{(\alpha)} x_{\alpha p} G_{pq}^{(\alpha)} x_{\alpha q} + \tau\Big)\sum_{s,t} G_{is}^{(\alpha)} x_{\alpha s} x_{\alpha t} G_{ti}^{(\alpha)} \\ &\quad + \frac{3}{(t_\alpha^{-1}-\tau)^4}\Big(\sum_{p,q} x_{\alpha p} G_{pq}^{(\alpha)} x_{\alpha q} + \tau\Big)^2 \sum_{s,t} G_{is}^{(\alpha)} x_{\alpha s} x_{\alpha t} G_{ti}^{(\alpha)} + \mathcal{O}(\Psi^5). \end{aligned}$$

Taking the partial expectation $\mathbb{E}_\alpha$ we get

$$\begin{aligned} &\mathbb{E}_\alpha[G_{i\alpha}G_{\alpha i}] \\ &\quad = \frac{1}{(t_\alpha^{-1}-\tau)^2}\frac{1}{N}\sum_s G_{is}^{(\alpha)} G_{si}^{(\alpha)} \\ &\qquad - \frac{2}{(t_\alpha^{-1}-\tau)^3}(m^{(\alpha)}+\tau)\frac{1}{N}\sum_s G_{is}^{(\alpha)} G_{si}^{(\alpha)} \\ &\qquad - \frac{4}{(t_\alpha^{-1}-\tau)^3}\frac{1}{N^2}\sum_{s,t} G_{is}^{(\alpha)} G_{st}^{(\alpha)} G_{ti}^{(\alpha)} \\ &\qquad + \frac{3}{(t_\alpha^{-1}-\tau)^4}(m^{(\alpha)}+\tau)^2\frac{1}{N}\sum_s G_{is}^{(\alpha)} G_{si}^{(\alpha)} \\ &\qquad + \frac{12}{(t_\alpha^{-1}-\tau)^4}(m^{(\alpha)}+\tau)\frac{1}{N^2}\sum_{s,t} G_{is}^{(\alpha)} G_{st}^{(\alpha)} G_{ti}^{(\alpha)} \end{aligned} \tag{A.4}$$

$$
\begin{aligned}
&+\frac{6}{(t_\alpha^{-1}-\tau)^4}\frac{1}{N^3}\sum_{s,p,q} G_{is}^{(\alpha)}G_{si}^{(\alpha)}G_{pq}^{(\alpha)}G_{qp}^{(\alpha)} \\
&+\frac{24}{(t_\alpha^{-1}-\tau)^4}\frac{1}{N^3}\sum_{s,p,q} G_{is}^{(\alpha)}G_{sp}^{(\alpha)}G_{pq}^{(\alpha)}G_{qi}^{(\alpha)}+\mathcal{O}(\Psi^5).
\end{aligned}
$$

In a next step, we expand (A.4) in the upper index α by using the resolvent formula (5.12), that is,

$$
G_{is}^{(\alpha)}=G_{is}-\frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}. \tag{A.5}
$$

In other words, using (A.5), we can remove the upper index α from the Green functions entries in (A.4) at the expense of higher order terms containing α as a lower index in the Green function entries. We obtain for the first term in (A.4) that

$$
\begin{aligned}
G_{is}^{(\alpha)}G_{si}^{(\alpha)}&=G_{is}G_{si}-\frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}G_{si}-G_{is}^{(\alpha)}\frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}} \\
&=G_{is}G_{si}-\frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}G_{si}^{(\alpha)}-G_{is}^{(\alpha)}\frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}}-\frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}\frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}}.
\end{aligned} \tag{A.6}
$$

We stop expanding the first term on the right-hand side of (A.6), since it does not contain the index α, and we set

$$
X_{22}:=\frac{1}{N}\sum_s G_{is}G_{si}. \tag{A.7}
$$

Using (5.11), the partial expectation of the second term on the right-hand side of (A.6) can be expanded in the lower index α to get

$$
\begin{aligned}
\mathbb{E}_\alpha\left[\frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}G_{si}^{(\alpha)}\right]&=\mathbb{E}_\alpha\left[G_{\alpha\alpha}\sum_{k,l}G_{ik}^{(\alpha)}x_{\alpha k}x_{\alpha l}G_{ls}^{(\alpha)}G_{si}^{(\alpha)}\right] \\
&=-\frac{1}{t_\alpha^{-1}-\tau}\frac{1}{N}\sum_k G_{ik}^{(\alpha)}G_{ks}^{(\alpha)}G_{si}^{(\alpha)} \\
&\quad+\frac{1}{(t_\alpha^{-1}-\tau)^2}(m^{(\alpha)}+\tau)\frac{1}{N}\sum_k G_{ik}^{(\alpha)}G_{ks}^{(\alpha)}G_{si}^{(\alpha)} \\
&\quad+\frac{2}{(t_\alpha^{-1}-\tau)^2}\frac{1}{N^2}\sum_{k,l}G_{ik}^{(\alpha)}G_{kl}^{(\alpha)}G_{ls}^{(\alpha)}G_{si}^{(\alpha)}+\mathcal{O}(\Psi^5).
\end{aligned} \tag{A.8}
$$

Expanding the first term in the right-hand side of (A.8) further using (5.12), we get

$$
\begin{aligned}
G_{ik}^{(\alpha)}G_{ks}^{(\alpha)}G_{si}^{(\alpha)}&=G_{ik}G_{ks}G_{si}-\frac{G_{i\alpha}G_{\alpha k}}{G_{\alpha\alpha}}G_{ks}G_{si} \\
&\quad-G_{ik}^{(\alpha)}\frac{G_{k\alpha}G_{\alpha s}}{G_{\alpha\alpha}}G_{si}-G_{ik}^{(\alpha)}G_{ks}^{(\alpha)}\frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}}.
\end{aligned} \tag{A.9}
$$

We stop expanding the first term on the right-hand side of (A.9), since it does no more contain the index a, and we let

$$X_{33} := \frac{1}{N^2} \sum_{k,s} G_{ik} G_{ks} G_{si}. \tag{A.10}$$

Expanding the remaining terms on the right-hand side of (A.9) in the lower index α using (5.11), we obtain

$$\begin{aligned} \mathbb{E}_\alpha\left[\frac{G_{i\alpha}G_{\alpha k}}{G_{\alpha\alpha}} G_{ks} G_{si}\right] &= -\frac{1}{t_\alpha^{-1} - \tau} \mathbb{E}_\alpha\left[\sum_{l,m} G_{il}^{(\alpha)} x_{\alpha l} x_{\alpha m} G_{mk}^{(\alpha)} G_{ks} G_{si}\right] + \mathcal{O}(\Psi^5) \\ &= -\frac{1}{t_\alpha^{-1} - \tau} \frac{1}{N} \sum_l G_{il}^{(\alpha)} G_{lk}^{(\alpha)} G_{ks} G_{si} + \mathcal{O}(\Psi^5) \\ &= -\frac{1}{t_\alpha^{-1} - \tau} \frac{1}{N} \sum_l G_{il} G_{lk} G_{ks} G_{si} + \mathcal{O}(\Psi^5) \end{aligned}$$

and, similarly,

$$\mathbb{E}_\alpha\left[G_{ik}^{(\alpha)} \frac{G_{k\alpha}G_{\alpha s}}{G_{\alpha\alpha}} G_{si}\right] = -\frac{1}{t_\alpha^{-1} - \tau} \frac{1}{N} \sum_l G_{ik} G_{kl} G_{ls} G_{si} + \mathcal{O}(\Psi^5)$$

respectively,

$$\mathbb{E}_\alpha\left[G_{ik}^{(\alpha)} G_{ks}^{(\alpha)} \frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}}\right] = -\frac{1}{t_\alpha^{-1} - \tau} \frac{1}{N} \sum_l G_{ik} G_{ks} G_{sl} G_{li} + \mathcal{O}(\Psi^5).$$

Thus, setting

$$X_{44} := \frac{1}{N^3} \sum_{k,l,s} G_{ik} G_{kl} G_{ls} G_{si}, \tag{A.11}$$

we have

$$\mathbb{E}_\alpha\left[\frac{1}{N^2} \sum_{k,s} G_{ik}^{(\alpha)} G_{ks}^{(\alpha)} G_{si}^{(\alpha)}\right] = X_{33} + \frac{3}{t_\alpha^{-1} - \tau} X_{44} + \mathcal{O}(\Psi^5). \tag{A.12}$$

Next, we consider the $\mathcal{O}(\Psi^4)$ terms on the right-hand side of (A.8). Let

$$X_{43} := (m + \tau) \frac{1}{N^2} \sum_{k,s} G_{ik} G_{ks} G_{si}. \tag{A.13}$$

Then, we have for the second term on the right-hand side of (A.8) that

$$\left(m^{(\alpha)} + \tau\right) \frac{1}{N^2} \sum_{k,s} G_{ik}^{(\alpha)} G_{ks}^{(\alpha)} G_{si}^{(\alpha)} = X_{43} + \mathcal{O}(\Psi^5). \tag{A.14}$$

The last term on the right-hand side of (A.8) is simply estimated by

$$\frac{1}{N^3}\sum_{k,l,s} G_{ik}^{(\alpha)} G_{kl}^{(\alpha)} G_{ls}^{(\alpha)} G_{si}^{(\alpha)} = X_{44} + \mathcal{O}(\Psi^5). \tag{A.15}$$

In sum, we find

$$\begin{aligned}&\mathbb{E}_\alpha\biggl[\frac{1}{N}\sum_s \frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}} G_{si}^{(\alpha)}\biggr] \\ &\quad = -\frac{1}{t_\alpha^{-1}-\tau} X_{33} + \frac{1}{(t_\alpha^{-1}-\tau)^2} X_{43} - \frac{1}{(t_\alpha^{-1}-\tau)^2} X_{44} + \mathcal{O}(\Psi^5).\end{aligned} \tag{A.16}$$

Similarly, we also have

$$\begin{aligned}&\mathbb{E}_\alpha\biggl[\frac{1}{N}\sum_s G_{is}^{(\alpha)} \frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}}\biggr] \\ &\quad = -\frac{1}{t_\alpha^{-1}-\tau} X_{33} + \frac{1}{(t_\alpha^{-1}-\tau)^2} X_{43} - \frac{1}{(t_\alpha^{-1}-\tau)^2} X_{44} + \mathcal{O}(\Psi^5).\end{aligned} \tag{A.17}$$

For the last term in (A.6), we obtain

$$\frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}\frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}} = G_{\alpha\alpha}^2 \sum_{k,l,p,q} G_{ik}^{(\alpha)} x_{\alpha k} x_{\alpha l} G_{ls}^{(\alpha)} G_{sp}^{(\alpha)} x_{\alpha p} x_{\alpha q} G_{qi}^{(\alpha)}.$$

Hence, denoting

$$X'_{44} := \frac{1}{N^3}\sum_{k,l,s} G_{is}G_{si}G_{kl}G_{lk}, \tag{A.18}$$

we find

$$\begin{aligned}&\mathbb{E}_\alpha\biggl[\frac{1}{N}\sum_s \frac{G_{i\alpha}G_{\alpha s}}{G_{\alpha\alpha}}\frac{G_{s\alpha}G_{\alpha i}}{G_{\alpha\alpha}}\biggr] \\ &\quad = \frac{2}{(t_\alpha^{-1}-\tau)^2} X_{44} + \frac{1}{(t_\alpha^{-1}-\tau)^2} X'_{44} + \mathcal{O}(\Psi^5).\end{aligned} \tag{A.19}$$

Thus, from (A.6), (A.16), (A.17) and (A.19) we obtain

$$\begin{aligned}&\mathbb{E}_\alpha\biggl[\frac{1}{(t_\alpha^{-1}-\tau)^2}\frac{1}{N}\sum_s G_{is}^{(\alpha)} G_{si}^{(\alpha)}\biggr] \\ &\quad = \frac{1}{(t_\alpha^{-1}-\tau)^2} X_{22} + \frac{2}{(t_\alpha^{-1}-\tau)^3} X_{33} \\ &\qquad - \frac{2}{(t_\alpha^{-1}-\tau)^4} X_{43} - \frac{1}{(t_\alpha^{-1}-\tau)^4} X'_{44} + \mathcal{O}(\Psi^5),\end{aligned} \tag{A.20}$$

which completes the expansion of the first term in (A.4). The calculation and the result coincide with those in the deformed Wigner case in [32], except the sign of the X_{33} term. The discrepancy is due to the sign difference in the coefficient $(t_\alpha^{-1} - \tau)^{-1}$.

Adapting the expansion procedure of [32], we conclude, with the definitions

$$\begin{aligned} X_{32} &:= (m+\tau)\frac{1}{N}\sum_s G_{is}G_{si}, \\ X_{42} &:= (m+\tau)^2\frac{1}{N}\sum_s G_{is}G_{si}, \end{aligned} \tag{A.21}$$

that

$$\begin{aligned} &\mathbb{E}_\alpha[G_{i\alpha}G_{\alpha i}] \\ &\quad= \frac{1}{(t_\alpha^{-1}-\tau)^2}X_{22} - \frac{2}{(t_\alpha^{-1}-\tau)^3}X_{32} \\ &\qquad - \frac{2}{(t_\alpha^{-1}-\tau)^3}X_{33} + \frac{3}{(t_\alpha^{-1}-\tau)^4}X_{42} \\ &\qquad + \frac{6}{(t_\alpha^{-1}-\tau)^4}X_{43} + \frac{12}{(t_\alpha^{-1}-\tau)^4}X_{44} + \frac{3}{(t_\alpha^{-1}-\tau)^4}X'_{44} + \mathcal{O}(\Psi^5). \end{aligned} \tag{A.22}$$

This shows (6.9), and hence completes the proof of Lemma 6.2. □

Before we move on to the next section, we introduce some more notation. For $k \in \mathbb{N}$, let

$$A_k := \frac{1}{N}\sum_\rho \frac{1}{(t_\rho^{-1}-\tau)^k}. \tag{A.23}$$

We remark that from (6.5), we have

$$A_2 = \tau^{-2}, \qquad A_3 + \tau^{-3} = 1. \tag{A.24}$$

Finally, averaging (A.22) over α, we have in this notation

$$\begin{aligned} \frac{1}{N}\sum_\alpha \mathbb{E}_\alpha[G_{i\alpha}G_{\alpha i}] &= A_2X_{22} - 2A_3(X_{32}+X_{33}) \\ &\qquad + 3A_4(X_{42}+2X_{43}+4X_{44}+X'_{44}) + \mathcal{O}(\Psi^5). \end{aligned} \tag{A.25}$$

This concludes the current appendix.

APPENDIX B: OPTICAL THEOREMS

In this section we establish the following "optical theorem."

LEMMA B.1. *Under the assumptions of Proposition 4.1 with the notation of Lemma 6.2, we have*

$$
\begin{aligned}
&2\mathbb{E}[X_{32}+X_{33}]-\frac{1}{N} \\
&\quad = 3(A_4-\tau^{-4})\mathbb{E}[X_{42}+2X_{43}+4X_{44}+X'_{44}]+\mathcal{O}(\Psi^5),
\end{aligned} \tag{B.1}
$$

uniformly in $t \geq 0$.

Lemma B.1 is an example of what we call optical theorems: optical theorems assure that the expectations of certain linear combinations of the random variables introduced in Lemma 6.2 are smaller than their naive sizes obtained from power counting using the local laws in Lemmas 3.2 and 5.3. Such estimates were key technical inputs in the proof of edge universality for deformed Wigner matrices in [32]. As in [32], the optical theorems used in this paper are obtained by combining expansions of random variables, for example, X_{22} or X_{33}, with the sum rules in (6.5). In the rest of this appendix, we derive the required optical theorems.

The proof of Lemma B.1 is given in Section B.4 based on estimates obtained in Sections B.1, B.2 and B.3.

B.1. Optical theorem from X_{22}. To derive the first optical theorem, we consider

$$
\sum_s G_{is}G_{si} = G_{ii}^2 + \sum_s^{(i)} G_{is}G_{si}. \tag{B.2}
$$

Similar to the expansion of $G_{\alpha\alpha}$, we now expand G_{ss} around $-\tau$. We notice that

$$
\left|\tau^{-1} - z - \sum_{\gamma,\delta} x_{\gamma s} G_{\gamma\delta}^{(s)} x_{\delta s}\right| \prec \Psi,
$$

which can be checked from (6.8) and the estimate

$$
\left|G_{\alpha\alpha} - \frac{1}{-t_\alpha^{-1}+\tau}\right| \prec \Psi.
$$

Thus, using Schur's complement formula (5.8), we obtain the following expansion of G_{ss} in the lower index s:

$$
\begin{aligned}
G_{ss} &= \frac{1}{h_{ss} - \sum_{\gamma,\delta}^{(s)} h_{\gamma s} G_{\gamma\delta}^{(s)} h_{s\delta}} = \frac{1}{-\tau^{-1}+\tau^{-1}-z-\sum_{\gamma,\delta} x_{\gamma s} G_{\gamma\delta}^{(s)} x_{\delta s}} \\
&= -\tau - \tau^2\Big(\tau^{-1} - z - \sum_{\gamma,\delta} x_{\gamma s} G_{\gamma\delta}^{(s)} x_{\delta s}\Big) \\
&\quad - \tau^3\Big(\tau^{-1} - z - \sum_{\gamma,\delta} x_{\gamma s} G_{\gamma\delta}^{(s)} x_{\delta s}\Big)^2 + \mathcal{O}(\Psi^3).
\end{aligned}
$$

Using the resolvent formula (5.9) we therefore get the following expansion of $G_{is}G_{si}$ in the lower index s, for $s \neq i$:

$$\begin{aligned} G_{is}G_{si} &= G_{ss}^2 \sum_{\rho,\sigma} G_{i\rho}^{(s)} x_{\rho s} x_{\sigma s} G_{\sigma i}^{(s)} \\ &= \tau^2 \sum_{\rho,\sigma} G_{i\rho}^{(s)} x_{\rho s} x_{\sigma s} G_{\sigma i}^{(s)} \\ &\quad + 2\tau^3 \Big(\tau^{-1} - z - \sum_{\gamma,\delta} x_{\gamma s} G_{\gamma\delta}^{(s)} x_{\delta s}\Big) \sum_{\rho,\sigma} G_{i\rho}^{(s)} x_{\rho s} x_{\sigma s} G_{\sigma i}^{(s)} \\ &\quad + 3\tau^4 \Big(\tau^{-1} - z - \sum_{\gamma,\delta} x_{\gamma s} G_{\gamma\delta}^{(s)} x_{\delta s}\Big)^2 \sum_{\rho,\sigma} G_{i\rho}^{(s)} x_{\rho s} x_{\sigma s} G_{\sigma i}^{(s)} + \mathcal{O}(\Psi^5). \end{aligned}$$

Taking the partial expectation $\mathbb{E}_s$, we obtain, for $s \neq i$,

$$\begin{aligned} \mathbb{E}_s[G_{is}G_{si}] &= \frac{\tau^2}{N} \sum_{\rho} G_{i\rho}^{(s)} G_{\rho i}^{(s)} + \frac{2\tau^3}{N} \Big(\tau^{-1} - z - \frac{\widetilde{m}^{(s)}}{d}\Big) \sum_{\rho} G_{i\rho}^{(s)} G_{\rho i}^{(s)} \\ &\quad - \frac{4\tau^3}{N^2} \sum_{\rho,\sigma} G_{i\rho}^{(s)} G_{\rho\sigma}^{(s)} G_{\sigma i}^{(s)} \\ &\quad + \frac{3\tau^4}{N} \Big(\tau^{-1} - z - \frac{\widetilde{m}^{(s)}}{d}\Big)^2 \sum_{\rho} G_{i\rho}^{(s)} G_{\rho i}^{(s)} \\ &\quad - \frac{12\tau^4}{N^2} \Big(\tau^{-1} - z - \frac{\widetilde{m}^{(s)}}{d}\Big) \sum_{\rho,\sigma} G_{i\rho}^{(s)} G_{\rho\sigma}^{(s)} G_{\sigma i}^{(s)} \\ &\quad + \frac{6\tau^4}{N^3} \sum_{\rho,\sigma,\gamma} G_{i\rho}^{(s)} G_{\rho i}^{(s)} G_{\sigma\gamma}^{(s)} G_{\gamma\sigma}^{(s)} \\ &\quad + \frac{24\tau^4}{N^3} \sum_{\rho,\sigma,\gamma} G_{i\rho}^{(s)} G_{\rho\gamma}^{(s)} G_{\gamma\sigma}^{(s)} G_{\sigma i}^{(s)} + \mathcal{O}(\Psi^5). \end{aligned} \tag{B.3}$$

Using the resolvent formula (5.5) to remove the upper indices s in (B.3), we get, for $s \neq i$,

$$\begin{aligned} \mathbb{E}_s[G_{is}G_{si}] &= \frac{\tau^2}{N} \sum_{\rho} G_{i\rho} G_{\rho i} + \frac{2\tau^3}{N} \Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big) \sum_{\rho} G_{i\rho} G_{\rho i} \\ &\quad - \frac{2\tau^3}{N^2} \sum_{\rho,\sigma} G_{i\rho} G_{\rho\sigma} G_{\sigma i} + \frac{3\tau^4}{N} \Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)^2 \sum_{\rho} G_{i\rho} G_{\rho i} \\ &\quad - \frac{6\tau^4}{N^2} \Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big) \sum_{\rho,\sigma} G_{i\rho} G_{\rho\sigma} G_{\sigma i} \end{aligned} \tag{B.4}$$

$$+\frac{12\tau^4}{N^3}\sum_{\rho,\sigma,\gamma} G_{i\rho}G_{\rho\sigma}G_{\sigma\gamma}G_{\gamma i}$$
$$+\frac{3\tau^4}{N^3}\sum_{\rho,\sigma,\gamma} G_{i\rho}G_{\gamma\sigma}G_{\sigma\gamma}G_{\rho i}+\mathcal{O}(\Psi^5).$$

We next expand all terms on the right-hand side of (B.4) except the first one to change Greek indices into Roman indices. Recall from (A.23) that

$$A_k=\frac{1}{N}\sum_{\rho}\frac{1}{(t_\rho^{-1}-\tau)^k}. \tag{B.5}$$

The last two terms on the right-hand side of (B.4) are easy to convert. For example,

$$\begin{aligned}G_{i\rho}G_{\rho\sigma}G_{\sigma\gamma}G_{\gamma i}&=G_{i\rho}G_{\rho\sigma}G^{(\rho)}_{\sigma\gamma}G^{(\rho)}_{\gamma i}+\mathcal{O}(\Psi^5)\\&=\frac{1}{(t_\rho^{-1}-\tau)^2}\sum_{j,k}G^{(\rho)}_{ij}x_{\rho j}x_{\rho k}G^{(\rho)}_{k\sigma}G^{(\rho)}_{\sigma\gamma}G^{(\rho)}_{\gamma i}+\mathcal{O}(\Psi^5),\end{aligned}$$

which shows that

$$\mathbb{E}_\rho[G_{i\rho}G_{\rho\sigma}G_{\sigma\gamma}G_{\gamma i}]=\frac{1}{(t_\rho^{-1}-\tau)^2}\frac{1}{N}\sum_j G_{ij}G_{j\sigma}G_{\sigma\gamma}G_{\gamma i}+\mathcal{O}(\Psi^5). \tag{B.6}$$

Repeating the argument once more, we also find, using (6.5), that

$$\begin{aligned}\mathbb{E}\biggl[\frac{12\tau^4}{N^3}\sum_{\rho,\sigma,\gamma}G_{i\rho}G_{\rho\sigma}G_{\sigma\gamma}G_{\gamma i}\biggr]&=\frac{12\tau^4}{N^3}A_2^3\mathbb{E}\biggl[\sum_{j,k,l}G_{ij}G_{jk}G_{kl}G_{li}\biggr]+\mathcal{O}(\Psi^5)\\&=12\tau^{-2}\mathbb{E}[X_{44}]+\mathcal{O}(\Psi^5).\end{aligned} \tag{B.7}$$

Similarly,

$$\mathbb{E}\biggl[\frac{3\tau^4}{N^3}\sum_{\rho,\sigma,\gamma}G_{i\rho}G_{\rho i}G_{\gamma\sigma}G_{\sigma\gamma}\biggr]=3\tau^{-2}\mathbb{E}[X'_{44}]+\mathcal{O}(\Psi^5). \tag{B.8}$$

The other fourth-order terms in (B.4) require more treatment. We first consider

$$\begin{aligned}\tau^{-1}-z-\frac{\widetilde{m}}{d}&=\tau^{-1}-z-\frac{1}{N}\sum_\beta G_{\beta\beta}\\&=\tau^{-1}-z+\frac{1}{N}\sum_\beta\frac{1}{t_\beta^{-1}-\tau}\\&\quad-\frac{1}{N}\sum_\beta\frac{1}{(t_\beta^{-1}-\tau)^2}\biggl(\sum_{p,q}x_{\beta p}G^{(\beta)}_{pq}x_{\beta q}+\tau\biggr)+\mathcal{O}(\Psi^2)\end{aligned}$$

$$= (L_+ - z) - \frac{1}{N}\sum_\beta \frac{1}{(t_\beta^{-1} - \tau)^2}\Big(\sum_{p,q} x_{\beta p} G^{(\beta)}_{pq} x_{\beta q} + \tau\Big) + \mathcal{O}(\Psi^2)$$

$$= -\frac{1}{N}\sum_\beta \frac{1}{(t_\beta^{-1} - \tau)^2}\Big(\sum_{p,q} x_{\beta p} G^{(\beta)}_{pq} x_{\beta q} + \tau\Big) + \mathcal{O}(\Psi^2).$$

We then obtain for the fifth term on the right-hand side of (B.4) that

$$\begin{aligned} &\mathbb{E}\Big[-\frac{6\tau^4}{N^2}\Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)\sum_{\rho,\sigma} G_{i\rho}G_{\rho\sigma}G_{\sigma i}\Big] \\ &\quad = \mathbb{E}\Big[\frac{6\tau^4}{N^2}\Big(\frac{1}{N}\sum_\beta \frac{1}{(t_\beta^{-1}-\tau)^2}\Big)(m^{(\beta)} + \tau)\sum_{\rho,\sigma} G^{(\beta)}_{i\rho}G^{(\beta)}_{\rho\sigma}G^{(\beta)}_{\sigma i}\Big] + \mathcal{O}(\Psi^5) \\ &\quad = 6\tau^{-2}\mathbb{E}\Big[(m+\tau)\frac{1}{N^2}\sum_{k,l} G_{ik}G_{kl}G_{ki}\Big] + \mathcal{O}(\Psi^5) \\ &\quad = 6\tau^{-2}\mathbb{E}[X_{43}] + \mathcal{O}(\Psi^5). \end{aligned} \tag{B.9}$$

Similarly, we have for the fourth term on the right-hand side of (B.4) that

$$\mathbb{E}\Big[\frac{3\tau^4}{N}\Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)^2 \sum_\rho G_{i\rho}G_{\rho i}\Big] = 3\tau^{-2}\mathbb{E}[X_{42}] + \mathcal{O}(\Psi^5). \tag{B.10}$$

This completes the discussion of the fourth-order terms in (B.4).

We move on to the third-order terms on the right-hand side of (B.4). Adapting the expansion method above, we note that

$$\begin{aligned} &\frac{1}{N}\Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)\sum_\rho G_{i\rho}G_{\rho i} \\ &\quad = -\frac{1}{N^2}\sum_\beta \frac{1}{(t_\beta^{-1}-\tau)^2}\Big(\sum_{p,q} x_{\beta p}G^{(\beta)}_{pq}x_{\beta q} + \tau\Big)\sum_\rho G_{i\rho}G_{\rho i} \\ &\qquad + \frac{1}{N^2}\sum_\beta \frac{1}{(t_\beta^{-1}-\tau)^3}\Big(\sum_{p,q} x_{\beta p}G^{(\beta)}_{pq}x_{\beta q} + \tau\Big)^2\sum_\rho G_{i\rho}G_{\rho i} \\ &\qquad + \frac{L_+ - z}{N}\sum_\rho G_{i\rho}G_{\rho i} + \mathcal{O}(\Psi^5). \end{aligned} \tag{B.11}$$

Taking the partial expectation $\mathbb{E}_\beta$, we get for the summand in the first term on the right-hand side of (B.11) that

$$\begin{aligned} &\mathbb{E}_\beta\Big[\frac{1}{(t_\beta^{-1}-\tau)^2}\Big(\sum_{p,q} x_{\beta p}G^{(\beta)}_{pq}x_{\beta q} + \tau\Big)G_{i\rho}G_{\rho i}\Big] \\ &\quad = \frac{1}{(t_\beta^{-1}-\tau)^2}(m^{(\beta)} + \tau)G^{(\beta)}_{i\rho}G^{(\beta)}_{\rho i} \end{aligned} \tag{B.12}$$

$$+\mathbb{E}_\beta\biggl[\frac{1}{(t_\beta^{-1}-\tau)^2}\biggl(\sum_{p,q}x_{\beta p}G_{pq}^{(\beta)}x_{\beta q}+\tau\biggr)\frac{G_{i\beta}G_{\beta\rho}}{G_{\beta\beta}}G_{\rho i}^{(\beta)}\biggr]$$
$$+\mathbb{E}_\beta\biggl[\frac{1}{(t_\beta^{-1}-\tau)^2}\biggl(\sum_{p,q}x_{\beta p}G_{pq}^{(\beta)}x_{\beta q}+\tau\biggr)G_{i\rho}^{(\beta)}\frac{G_{\rho\beta}G_{\beta i}}{G_{\beta\beta}}\biggr]+\mathcal{O}(\Psi^5).$$

Expanding the first term on the right-hand side of (B.12) with respect to the upper index β, we find

$$\begin{aligned}\mathbb{E}\biggl[&\frac{1}{(t_\beta^{-1}-\tau)^2}(m^{(\beta)}+\tau)G_{i\rho}^{(\beta)}G_{\rho i}^{(\beta)}\biggr]\\
&=\mathbb{E}\biggl[\frac{1}{(t_\beta^{-1}-\tau)^2}(m+\tau)G_{i\rho}G_{\rho i}\biggr]\\
&\quad+\mathbb{E}\biggl[\frac{2}{(t_\beta^{-1}-\tau)^3}(m+\tau)\sum_k G_{ik}G_{k\rho}G_{\rho i}\biggr]\\
&\quad+\mathbb{E}\biggl[\frac{1}{(t_\beta^{-1}-\tau)^3}\sum_{k,l}G_{kl}G_{lk}G_{i\rho}G_{\rho i}\biggr]+\mathcal{O}(\Psi^5).\end{aligned}\tag{B.13}$$

Similarly, we get for the expectation of the second term on the right-hand side of (B.12) that

$$\begin{aligned}\mathbb{E}\biggl[&\frac{1}{(t_\beta^{-1}-\tau)^2}\biggl(\sum_{p,q}x_{\beta p}G_{pq}^{(\beta)}x_{\beta q}+\tau\biggr)\frac{G_{i\beta}G_{\beta\rho}}{G_{\beta\beta}}G_{\rho i}^{(\beta)}\biggr]\\
&=-\mathbb{E}\biggl[\frac{1}{(t_\beta^{-1}-\tau)^3}(m+\tau)\sum_k G_{ik}G_{k\rho}G_{\rho i}\biggr]\\
&\quad-\mathbb{E}\biggl[\frac{2}{(t_\beta^{-1}-\tau)^3}\frac{1}{N^2}\sum_{k,l}G_{ik}G_{kl}G_{l\rho}G_{\rho i}\biggr]\\
&\quad+\mathcal{O}(\Psi^5)\end{aligned}\tag{B.14}$$

and for the expectation of the third term on the right-hand side of (B.12) that

$$\begin{aligned}\mathbb{E}\biggl[&\frac{1}{(t_\beta^{-1}-\tau)^2}\biggl(\sum_{p,q}x_{\beta p}G_{pq}^{(\beta)}x_{\beta q}+\tau\biggr)G_{i\rho}^{(\beta)}\frac{G_{\rho\beta}G_{\beta i}}{G_{\beta\beta}}\biggr]\\
&=-\mathbb{E}\biggl[\frac{1}{(t_\beta^{-1}-\tau)^3}(m+\tau)\sum_k G_{ik}G_{k\rho}G_{\rho i}\biggr]\\
&\quad-\mathbb{E}\biggl[\frac{2}{(t_\beta^{-1}-\tau)^3}\frac{1}{N^2}\sum_{k,l}G_{ik}G_{kl}G_{l\rho}G_{\rho i}\biggr]\\
&\quad+\mathcal{O}(\Psi^5).\end{aligned}\tag{B.15}$$

Thus, from (B.12), (B.13), (B.14) and (B.15) we find

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{(t_\beta^{-1}-\tau)^2}\Big(\sum_{p,q} x_{\beta p} G^{(\beta)}_{pq} x_{\beta q}+\tau\Big) G_{i\rho}G_{\rho i}\Big] \\
&\quad = \mathbb{E}\Big[\frac{1}{(t_\beta^{-1}-\tau)^2}(m+\tau) G_{s\rho}G_{\rho s}\Big] \\
&\qquad - \mathbb{E}\Big[\frac{4}{(t_\beta^{-1}-\tau)^3}\frac{1}{N^2}\sum_{k,l} G_{ik}G_{kl}G_{l\rho}G_{\rho i}\Big] \\
&\qquad + \mathbb{E}\Big[\frac{1}{(t_\beta^{-1}-\tau)^3}\sum_{k,l} G_{kl}G_{lk}G_{i\rho}G_{\rho i}\Big] \\
&\qquad + \mathcal{O}(\Psi^5).
\end{aligned}
\tag{B.16}
$$

We next remove the Greek index ρ on the right-hand side of (B.16). We note that

$$
\begin{aligned}
&(m+\tau)G_{i\rho}G_{\rho i} \\
&\quad = G_{\rho\rho}^2 (m^{(\rho)}+\tau)\sum_{k,l} G^{(\rho)}_{ik} x_{\rho k} x_{\rho l} G^{(\rho)}_{li} \\
&\qquad + \frac{1}{N}\sum_j G_{j\rho}G_{\rho j}G_{\rho\rho}\sum_{k,l} G^{(\rho)}_{ik} x_{\rho k} x_{\rho l} G^{(\rho)}_{li}.
\end{aligned}
$$

Thus, taking the partial expectation $\mathbb{E}_\rho$, we find

$$
\begin{aligned}
&\mathbb{E}_\rho\big[(m+\tau)G_{i\rho}G_{\rho i}\big] \\
&\quad = \frac{1}{(t_\rho^{-1}-\tau)^2}(m^{(\rho)}+\tau)\frac{1}{N}\sum_k G^{(\rho)}_{ik}G^{(\rho)}_{ki} \\
&\qquad - \frac{2}{(t_\rho^{-1}-\tau)^3}(m^{(\rho)}+\tau)^2\frac{1}{N}\sum_k G^{(\rho)}_{ik}G^{(\rho)}_{ki} \\
&\qquad - \frac{4}{(t_\rho^{-1}-\tau)^3}(m^{(\rho)}+\tau)\frac{1}{N^2}\sum_{k,l} G^{(\rho)}_{ik}G^{(\rho)}_{kl}G^{(\rho)}_{li} \\
&\qquad - \frac{2}{(t_\rho^{-1}-\tau)^3}\frac{1}{N^3}\sum_{j,k,l} G^{(\rho)}_{ik}G^{(\rho)}_{kj}G^{(\rho)}_{jl}G^{(\rho)}_{li} \\
&\qquad - \frac{1}{(t_\rho^{-1}-\tau)^3}\frac{1}{N^3}\sum_{j,k,l} G^{(\rho)}_{ik}G^{(\rho)}_{ki}G^{(\rho)}_{jl}G^{(\rho)}_{lj} + \mathcal{O}(\Psi^5).
\end{aligned}
\tag{B.17}
$$

Expanding the right-hand side of (B.17) with respect to the upper index ρ, we obtain

$$\begin{aligned}&\mathbb{E}\big[(m+\tau)G_{i\rho}G_{\rho i}\big]\\ &\quad=\frac{1}{(t_\rho^{-1}-\tau)^2}\mathbb{E}[X_{32}]\\ &\qquad-\frac{1}{(t_\rho^{-1}-\tau)^3}\big(2\mathbb{E}[X_{42}]+2\mathbb{E}[X_{43}]+2\mathbb{E}[X_{44}]\big)+\mathcal{O}\big(\Psi^5\big).\end{aligned}\tag{B.18}$$

We thus have for the first term on the right-hand side of (B.11) that

$$\begin{aligned}&\mathbb{E}\biggl[-\frac{1}{N^2}\sum_\beta\frac{1}{(t_\beta^{-1}-\tau)^2}\biggl(\sum_{p,q}x_{\beta p}G^{(\beta)}_{pq}x_{\beta q}+\tau\biggr)\sum_\rho G_{i\rho}G_{\rho i}\biggr]\\ &\quad=-\tau^{-4}\mathbb{E}[X_{32}]+\tau^{-2}A_3\big(2\mathbb{E}[X_{42}]+2\mathbb{E}[X_{43}]\\ &\qquad+6\mathbb{E}[X_{44}]-\mathbb{E}\big[X'_{44}\big]\big)+\mathcal{O}\big(\Psi^5\big).\end{aligned}\tag{B.19}$$

The fourth-order terms in (B.11) can easily be handled: we have

$$\begin{aligned}&\mathbb{E}\biggl[\frac{1}{N^2}\sum_\beta\frac{1}{(t_\beta^{-1}-\tau)^3}\biggl(\sum_{p,q}x_{\beta p}G^{(\beta)}_{pq}x_{\beta q}+\tau\biggr)^2\sum_\rho G_{i\rho}G_{\rho i}\biggr]\\ &\quad=\tau^{-2}A_3\big(\mathbb{E}[X_{42}]+2\mathbb{E}\big[X'_{44}\big]\big)+\mathcal{O}\big(\Psi^5\big),\end{aligned}\tag{B.20}$$

respectively,

$$\mathbb{E}\biggl[\frac{L_+-z}{N}\sum_\rho G_{i\rho}G_{\rho i}\biggr]=\tau^{-2}(L_+-z)\mathbb{E}[X_{22}]+\mathcal{O}\big(\Psi^5\big).\tag{B.21}$$

We thus obtain from (B.11), (B.19), (B.20) and (B.21) that

$$\begin{aligned}&\mathbb{E}\biggl[\frac{2\tau^3}{N}\biggl(\tau^{-1}-z-\frac{\widetilde{m}}{d}\biggr)\sum_\rho G_{i\rho}G_{\rho i}\biggr]\\ &\quad=-2\tau^{-1}\mathbb{E}[X_{32}]+2\tau A_3\big(3\mathbb{E}[X_{42}]+2\mathbb{E}[X_{43}]+6\mathbb{E}[X_{44}]\\ &\qquad+\mathbb{E}\big[X'_{44}\big]\big)+2\tau(L_+-z)\mathbb{E}[X_{22}]+\mathcal{O}\big(\Psi^5\big).\end{aligned}\tag{B.22}$$

The third term on the right-hand side of (B.4), which is also $\mathcal{O}(\Psi^3)$, can be expanded in a similar manner: We begin with

$$G_{i\rho}G_{\rho\sigma}G_{\sigma i}=G_{i\rho}G_{\rho\sigma}G^{(\rho)}_{\sigma i}+G_{i\rho}G_{\rho\sigma}\frac{G_{i\rho}G_{\rho\sigma}}{G_{\rho\rho}}.\tag{B.23}$$

The second term on the right-hand side of (B.23) can easily be controlled: we have

$$\frac{1}{N^2}\sum_{\rho,\sigma}\mathbb{E}\biggl[G_{i\rho}G_{\rho\sigma}\frac{G_{i\rho}G_{\rho\sigma}}{G_{\rho\rho}}\biggr]=-\tau^{-2}A_3\big(2\mathbb{E}[X_{44}]+\mathbb{E}\big[X'_{44}\big]\big)+\mathcal{O}\big(\Psi^5\big).$$

Taking the partial expectation $\mathbb{E}_\rho$, we have

$$\begin{aligned}
&\mathbb{E}_\rho[G_{i\rho}G_{\rho\sigma}G_{\sigma i}^{(\rho)}] \\
&\quad = \frac{1}{(t_\rho^{-1}-\tau)^2}\frac{1}{N}\sum_k G_{ik}^{(\rho)}G_{k\sigma}^{(\rho)}G_{\sigma i}^{(\rho)} \\
&\qquad - \frac{2}{(t_\rho^{-1}-\tau)^3}\frac{1}{N}(m^{(\rho)}+\tau)\sum_k G_{ik}^{(\rho)}G_{k\sigma}^{(\rho)}G_{\sigma i}^{(\rho)} \\
&\qquad - \frac{4}{(t_\rho^{-1}-\tau)^3}\frac{1}{N^2}\sum_{k,l} G_{ik}^{(\rho)}G_{kl}^{(\rho)}G_{l\sigma}^{(\rho)}G_{\sigma i}^{(\rho)} + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.24}$$

Thus, expanding with respect to the upper index ρ, we obtain

$$\begin{aligned}
\mathbb{E}_\rho[G_{i\rho}G_{\rho\sigma}G_{\sigma s}^{(\rho)}] = {} & \frac{1}{(t_\rho^{-1}-\tau)^2}\frac{1}{N}\sum_k G_{ik}G_{k\sigma}G_{\sigma i} \\
& - \frac{2}{(t_\rho^{-1}-\tau)^3}\frac{1}{N}(m+\tau)\sum_k G_{ik}G_{k\sigma}G_{\sigma i} \\
& - \frac{1}{(t_\rho^{-1}-\tau)^3}\frac{1}{N^2}\sum_{k,l} G_{ik}G_{kl}G_{l\sigma}G_{\sigma i} + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.25}$$

Repeating the same procedure with σ instead of ρ, we eventually find

$$\begin{aligned}
&\mathbb{E}\left[\frac{2\tau^3}{N^2}\sum_{\rho,\sigma} G_{s\rho}G_{\rho\sigma}G_{\sigma s}\right] \\
&\quad = 2\tau^{-1}\mathbb{E}[X_{33}] - 2\tau A_3\big(4\mathbb{E}[X_{43}] + 6\mathbb{E}[X_{44}] + 2\mathbb{E}[X'_{44}]\big) + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.26}$$

We conclude from (A.25), (B.4), (B.7), (B.8), (B.9), (B.10), (B.22) and (B.26) that

$$\begin{aligned}
&\frac{1}{N}\sum_\alpha \mathbb{E}[G_{i\alpha}G_{\alpha i}] \\
&\quad = \frac{1}{N^2}\sum_s^{(i)}\sum_\rho \mathbb{E}[G_{i\rho}G_{\rho i}] + \frac{\tau^{-2}}{N}\mathbb{E}[G_{ii}^2] + 2\tau^{-1}(L_+ - z)\mathbb{E}[X_{22}] \\
&\qquad - 2(A_3+\tau^{-3})\mathbb{E}[X_{32}+X_{33}] \\
&\qquad + 3(A_4+\tau^{-4}+2\tau^{-1}A_3)\mathbb{E}[X_{42}+2X_{43}+4X_{44}+X'_{44}] + \mathcal{O}(\Psi^5).
\end{aligned}$$

Since

$$\frac{1}{N^2}\sum_s^{(i)}\sum_\rho \mathbb{E}[G_{i\rho}G_{\rho i}] = \frac{1}{N}\sum_\rho \mathbb{E}[G_{i\rho}G_{\rho i}] + \mathcal{O}(\Psi^5),$$

we obtain the relation

$$
\begin{aligned}
&2(A_3+\tau^{-3})\mathbb{E}[X_{32}+X_{33}]\\
&\quad=3(A_4+\tau^{-4}+2\tau^{-1}A_3)\mathbb{E}[X_{42}+2X_{43}+4X_{44}+X'_{44}]\\
&\qquad+\frac{\tau^{-2}}{N}\mathbb{E}[G_{ii}^2]+2\tau^{-1}(L_+-z)\mathbb{E}[X_{22}]+\mathcal{O}(\Psi^5).
\end{aligned}\tag{B.27}
$$

Recalling that

$$
G_{ii}^2=\tau^2+2\tau^3\Big(\tau^{-1}-z-\sum_{\gamma,\delta}x_{\gamma i}G_{\gamma\delta}^{(i)}x_{\delta i}\Big)+\mathcal{O}(\Psi^2),
$$

we find

$$
\begin{aligned}
\mathbb{E}[G_{ii}^2]&=\tau^2+2\tau^3\mathbb{E}\Big[\tau^{-1}-z-\frac{\widetilde{m}}{d}\Big]+\mathcal{O}(\Psi^2)\\
&=\tau^2-2\tau\mathbb{E}[m+\tau]+\mathcal{O}(\Psi^2).
\end{aligned}\tag{B.28}
$$

Thus, plugging (B.28) into (B.27) and recalling from (6.5) that $A_3+\tau^{-3}=1$, we find

$$
\begin{aligned}
&2\mathbb{E}[X_{32}+X_{33}]-\frac{1}{N}\\
&\quad=3(A_4+\tau^{-4}+2\tau^{-1}A_3)\mathbb{E}[X_{42}+2X_{43}+4X_{44}+X'_{44}]\\
&\qquad-\frac{2\tau^{-1}}{N}\mathbb{E}[m+\tau]+2\tau^{-1}(L_+-z)\mathbb{E}[X_{22}]+\mathcal{O}(\Psi^5).
\end{aligned}\tag{B.29}
$$

The identity (B.29) is the optical theorem derived from X_{22}. We remark that the second and third term on the right-hand side of (B.29) are both $\mathcal{O}(\Psi^4)$. In Section B.3 we show that they can be written as linear combinations of X_{42}, X_{43}, X_{44} and X'_{44}.

B.2. Optical theorems from X_{32} and X_{33}. In a next step, we derive further optical theorems using the ideas presented in Section B.1. We start by considering

$$
X_{32}=(m+\tau)\frac{1}{N}\sum_s G_{is}G_{si}=(m+\tau)\frac{1}{N}\sum_s^{(i)}G_{is}G_{si}+(m+\tau)\frac{1}{N}G_{ii}^2.\tag{B.30}
$$

To estimate the first term on the very right-hand side of (B.30), we consider, for $s\neq i$,

$$
(m+\tau)G_{is}G_{si}=\big(m^{(s)}+\tau\big)G_{is}G_{si}+\frac{1}{N}\sum_j^{(s)}\frac{G_{js}G_{sj}}{G_{ss}}G_{is}G_{si}+\mathcal{O}(\Psi^5).\tag{B.31}
$$

We expand the first term on the right-hand side of (B.31) with respect to the lower index s to get

$$
\begin{aligned}
&(m^{(s)}+\tau)G_{is}G_{si}\\
&\quad=(m^{(s)}+\tau)G_{ss}^2\sum_{\rho,\sigma}G_{i\rho}^{(s)}x_{\rho s}x_{\sigma s}G_{\sigma i}^{(s)}\\
&\quad=\tau^2(m^{(s)}+\tau)\sum_{\rho,\sigma}G_{i\rho}^{(s)}x_{\rho s}x_{\sigma s}G_{\sigma i}^{(s)}\\
&\qquad+2\tau^3(m^{(s)}+\tau)\Big(\tau^{-1}-z-\sum_{\gamma,\delta}x_{\gamma s}G_{\gamma\delta}^{(s)}x_{\delta s}\Big)\sum_{\rho,\sigma}G_{i\rho}^{(s)}x_{\rho s}x_{\sigma s}G_{\sigma i}^{(s)}\\
&\qquad+\mathcal{O}(\Psi^5).
\end{aligned}
$$

Taking the partial expectation $\mathbb{E}_s$, we obtain

$$
\begin{aligned}
&\mathbb{E}_s\big[(m^{(s)}+\tau)G_{is}G_{si}\big]\\
&\quad=\frac{\tau^2}{N}(m^{(s)}+\tau)\sum_{\rho}G_{i\rho}^{(s)}G_{\rho i}^{(s)}\\
&\qquad+\frac{2\tau^3}{N}(m^{(s)}+\tau)\Big(\tau^{-1}-z-\frac{\widetilde{m}^{(s)}}{d}\Big)\sum_{\rho}G_{i\rho}^{(s)}G_{\rho i}^{(s)}\\
&\qquad-\frac{4\tau^3}{N^2}(m^{(s)}+\tau)\sum_{\rho,\sigma}G_{i\rho}^{(s)}G_{\rho\sigma}^{(s)}G_{\sigma i}^{(s)}+\mathcal{O}(\Psi^5).
\end{aligned}
$$

Since

$$
\begin{aligned}
&\frac{\tau^2}{N}(m^{(s)}+\tau)\sum_{\rho}G_{i\rho}^{(s)}G_{\rho i}^{(s)}\\
&\quad=\frac{\tau}{N}(m+\tau)\sum_{\rho}G_{i\rho}G_{\rho i}\\
&\qquad+\frac{2\tau}{N}(m+\tau)\sum_{\rho}G_{is}G_{s\rho}G_{\rho i}+\frac{\tau}{N^2}\sum_{j}\sum_{\rho}G_{js}G_{sj}G_{i\rho}G_{\rho i}+\mathcal{O}(\Psi^5)
\end{aligned}
$$

and since

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{\tau^3}{N^2}(m+\tau)\sum_{\rho,\sigma}G_{i\rho}G_{\rho\sigma}G_{\sigma i}\Big]\\
&\quad=\mathbb{E}\Big[\frac{\tau}{N^2}(m+\tau)\sum_{k}\sum_{\sigma}G_{ik}G_{k\sigma}G_{\sigma i}\Big]+\mathcal{O}(\Psi^5),
\end{aligned}
$$

we obtain

$$\mathbb{E}\bigg[(m^{(s)}+\tau)\frac{1}{N}\sum_{s}^{(i)} G_{is}G_{si}\bigg] = \mathbb{E}\bigg[\frac{\tau^2}{N}(m+\tau)\sum_{\rho} G_{i\rho}G_{\rho i}\bigg] - \tau^{-1}\mathbb{E}[2X_{42}+2X_{43}-X'_{44}] + \mathcal{O}(\Psi^5).$$

Moreover, we have that

$$\mathbb{E}\bigg[\frac{1}{N}\sum_{j}^{(s)} \frac{G_{js}G_{sj}}{G_{ss}} G_{is}G_{si}\bigg] = -\tau^{-1}\mathbb{E}[2X_{44}+X'_{44}] + \mathcal{O}(\Psi^5).$$

We thus find the relation

$$\mathbb{E}[X_{32}] - N^{-1}\mathbb{E}[(m+\tau)G_{ii}^2] = \tau^2\mathbb{E}\bigg[(m+\tau)\frac{1}{N}\sum_{\rho} G_{i\rho}G_{\rho i}\bigg] - 2\tau^{-1}\mathbb{E}[X_{42}+X_{43}+X_{44}] + \mathcal{O}(\Psi^5).$$

Applying (B.18), we obtain

$$\begin{aligned} &\mathbb{E}[X_{32}] - N^{-1}\mathbb{E}[(m+\tau)G_{ii}^2] \\ &\quad = \mathbb{E}[X_{32}] - 2(\tau^2 A_3 + \tau^{-1})\mathbb{E}[X_{42}+X_{43}+X_{44}] + \mathcal{O}(\Psi^5). \end{aligned} \tag{B.32}$$

Further, since

$$N^{-1}\mathbb{E}[(m+\tau)G_{ii}^2] = \tau^2 N^{-1}\mathbb{E}[m+\tau] + \mathcal{O}(\Psi^5),$$

we obtain from (B.32) the identity

$$N^{-1}\mathbb{E}[m+\tau] = 2(A_3+\tau^{-3})\mathbb{E}[X_{42}+X_{43}+X_{44}] + \mathcal{O}(\Psi^5), \tag{B.33}$$

which is the optical theorem derived from X_{32}.

We next derive the optical theorem obtained from

$$X_{33} = \frac{1}{N^2}\sum_{k,s} G_{ik}G_{ks}G_{si}. \tag{B.34}$$

Since the contributions to the sums in (B.34) from the cases $i = k$ or $s = k$ are negligible [of $\mathcal{O}(\Psi^2)$], we assume that $i, s \neq k$. Expanding the summand in (B.34) with respect to the lower index k, we get

$$G_{ik}G_{ks}G_{si} = G_{kk}^2 \sum_{\rho,\sigma} G_{i\rho}^{(k)} x_{\rho k} x_{\sigma k} G_{\sigma s}^{(k)} G_{si}^{(k)} + G_{ik}G_{ks}\frac{G_{sk}G_{ki}}{G_{kk}}. \tag{B.35}$$

Taking the partial expectation $\mathbb{E}_k$, we find for the first term on the right-hand side of (B.35) that

$$\mathbb{E}\bigg[G_{kk}^2\sum_{\rho,\sigma}G_{i\rho}^{(k)}x_{\rho k}x_{\sigma k}G_{\sigma s}^{(k)}G_{si}^{(k)}\bigg]$$
$$=\mathbb{E}\bigg[\frac{\tau^2}{N}\sum_{\rho}G_{i\rho}^{(k)}G_{\rho s}^{(k)}G_{si}^{(k)}\bigg]+\mathbb{E}\bigg[\frac{2\tau^3}{N}\Big(\tau^{-1}-z-\frac{\widetilde{m}^{(k)}}{d}\Big)\sum_{\rho}G_{i\rho}^{(k)}G_{\rho s}^{(k)}G_{si}^{(k)}\bigg]$$
$$-\mathbb{E}\bigg[\frac{4\tau^3}{N^2}\sum_{\rho,\sigma}G_{i\rho}^{(k)}G_{\rho\sigma}^{(k)}G_{\sigma s}^{(k)}G_{si}^{(k)}\bigg]+\mathcal{O}(\Psi^2).$$

Expanding further with respect to the upper index k, we thus find from (B.35) that

$$\begin{aligned}\mathbb{E}[X_{33}]&=\mathbb{E}\bigg[\frac{\tau^2}{N^2}\sum_{s}\sum_{\rho}G_{i\rho}G_{\rho s}G_{si}\bigg]\\&\quad-\tau^{-1}\mathbb{E}\big[2X_{43}+3X_{44}+X_{44}'\big]+\mathcal{O}(\Psi^5).\end{aligned}\tag{B.36}$$

Expanding the summand in the first term on the right of (B.36) with respect the index ρ, we get

$$\mathbb{E}[X_{33}]=\mathbb{E}[X_{33}]-\big(\tau^2A_3+\tau^{-1}\big)\mathbb{E}\big[2X_{43}+3X_{44}+X_{44}'\big]+\mathcal{O}(\Psi^5),\tag{B.37}$$

that is, recalling $A_3+\tau^{-3}=1$ [see (A.24)],

$$\tau^{-2}\mathbb{E}\big[2X_{43}+3X_{44}+X_{44}'\big]=\mathcal{O}(\Psi^5),\tag{B.38}$$

which is the optical theorem derived from X_{33}.

B.3. Optical theorem from mX_{22}. We return to the concluding remarks of Section B.1. In the present subsection we show that the terms $(L_+-z)\mathbb{E}[X_{22}]$ and $N^{-1}\mathbb{E}[m+\tau]$, both appearing in (B.29), can be decomposed into linear combinations of X_{42}, X_{43}, X_{44} and X_{44}'. The latter term, $N^{-1}\mathbb{E}[m+\tau]$, can be handled by (B.33), while the former needs to be dealt with the optical theorem obtained from mX_{22}. Recall that

$$mX_{22}=\frac{1}{N^2}\sum_{a,s}G_{aa}G_{is}G_{si}.\tag{B.39}$$

Expanding the summand on the right-hand side of (B.39) in the index a, we get

$$\begin{aligned}mX_{22}&=\frac{1}{N^2}\sum_{a\neq s}\big(G_{aa}G_{is}^{(a)}G_{si}^{(a)}+G_{is}^{(a)}G_{sa}G_{ai}+G_{ia}G_{as}G_{si}\big)+\mathcal{O}(\Psi^5)\\&=\frac{1}{N^2}\sum_{a\neq s}\Big(G_{aa}G_{is}^{(a)}G_{si}^{(a)}-\frac{G_{ia}G_{as}}{G_{aa}}G_{sa}G_{ai}\Big)+2X_{33}+\mathcal{O}(\Psi^5).\end{aligned}\tag{B.40}$$

Expanding the second summand of the first term on the right-hand side of (B.40) with respect the lower index a, we get

$$\mathbb{E}[mX_{22}] = \mathbb{E}\bigg[\frac{1}{N^2}\sum_{a\neq s} G_{aa}G_{is}^{(a)}G_{si}^{(a)}\bigg] + \tau^{-1}\mathbb{E}[2X_{44} + X_{44}'] + 2\mathbb{E}[X_{33}] + \mathcal{O}(\Psi^5). \tag{B.41}$$

We expand the summand of the first term on the right of (B.41) further in the lower index a to find

$$\begin{aligned}\mathbb{E}_a[G_{aa}G_{is}^{(a)}G_{si}^{(a)}] &= -\tau G_{is}^{(a)}G_{si}^{(a)} - \tau^2\bigg(\tau^{-1} - z - \frac{\widetilde{m}^{(a)}}{d}\bigg)G_{is}^{(a)}G_{si}^{(a)} \\ &\quad - \tau^3\bigg(\tau^{-1} - z - \frac{\widetilde{m}^{(a)}}{d}\bigg)^2 G_{is}^{(a)}G_{si}^{(s)} \\ &\quad - \frac{2\tau^3}{N^2}\sum_{\gamma,\delta} G_{\gamma\delta}^{(a)}G_{\delta\gamma}^{(a)}G_{is}^{(a)}G_{si}^{(a)} + \mathcal{O}(\Psi^5).\end{aligned} \tag{B.42}$$

Expanding the first term on the right-hand side of (B.42) with respect the upper index a, we get

$$G_{is}^{(a)}G_{si}^{(a)} = G_{is}G_{si} - G_{is}^{(a)}\frac{G_{sa}G_{ai}}{G_{aa}} - \frac{G_{ia}G_{as}}{G_{aa}}G_{si}^{(a)} - \frac{G_{ia}G_{as}}{G_{aa}}\frac{G_{sa}G_{ai}}{G_{aa}}. \tag{B.43}$$

We stop expanding the first term on the right-hand side of (B.43) which will eventually, after averaging over s, become X_{22}. For the second term on the right-hand side of (B.43), we have

$$\begin{aligned}\mathbb{E}_a\bigg[G_{is}^{(a)}\frac{G_{sa}G_{ai}}{G_{aa}}\bigg] &= -\frac{\tau}{N}\sum_{\gamma} G_{is}^{(a)}G_{s\gamma}^{(a)}G_{\gamma i}^{(a)} \\ &\quad - \frac{\tau^2}{N}\bigg(\tau^{-1} - z - \frac{\widetilde{m}^{(a)}}{d}\bigg)\sum_{\gamma} G_{is}^{(a)}G_{s\gamma}^{(a)}G_{\gamma i}^{(a)} \\ &\quad + \frac{2\tau^2}{N^2}\sum_{\gamma,\delta} G_{is}^{(a)}G_{s\gamma}^{(a)}G_{\gamma\delta}^{(a)}G_{\delta i}^{(a)} + \mathcal{O}(\Psi^5).\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\tau}{N^2}\sum_{a\neq s} G_{is}^{(a)}\frac{G_{sa}G_{ai}}{G_{aa}}\Big] \\
&= -\mathbb{E}\Big[\frac{\tau^2}{N^3}\sum_{a\neq s}\sum_{\gamma} G_{is}^{(a)}G_{s\gamma}^{(a)}G_{\gamma i}^{(a)}\Big] + \tau^{-1}\mathbb{E}[X_{43}+2X_{44}] + \mathcal{O}(\Psi^5) \\
&= -\mathbb{E}\Big[\frac{\tau^2}{N^2}\sum_{s}\sum_{\gamma} G_{is}G_{s\gamma}G_{\gamma i}\Big] + \tau^{-1}\mathbb{E}[X_{43}-X_{44}] + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.44}$$

Following the calculation in (B.36)–(B.37), we obtain from (B.44) that

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\tau}{N^2}\sum_{a\neq s} G_{is}^{(a)}\frac{G_{sa}G_{ai}}{G_{aa}}\Big] \\
&= -\mathbb{E}[X_{33}] + \tau^2 A_3\mathbb{E}\big[2X_{43}+3X_{44}+X_{44}'\big] \\
&\quad + \tau^{-1}\mathbb{E}[X_{43}-X_{44}] + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.45}$$

The third term on the right-hand side of (B.43) can be expanded in a similar manner. In sum, we get

$$\begin{aligned}
&-\mathbb{E}\Big[\frac{\tau}{N^2}\sum_{a\neq s} G_{is}^{(a)}G_{si}^{(a)}\Big] \\
&= -\tau\mathbb{E}[X_{22}] - 2\mathbb{E}[X_{33}] \\
&\quad + 2\tau^2 A_3\mathbb{E}\big[2X_{43}+3X_{44}+X_{44}'\big] \\
&\quad + \tau^{-1}\mathbb{E}\big[2X_{43}+X_{44}'\big] + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.46}$$

We next consider the second term on the right-hand side of (B.42). We note that

$$\begin{aligned}
&\Big(\tau^{-1} - z - \frac{\widetilde{m}^{(a)}}{d}\Big)G_{is}^{(a)}G_{si}^{(a)} \\
&= \Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)G_{is}G_{si} \\
&\quad + \tau^{-1}\Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)G_{ia}G_{as}G_{si} \\
&\quad + \tau^{-1}\Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big)G_{is}G_{sa}G_{ai} \\
&\quad - \frac{\tau^{-1}}{N}\sum_{\gamma} G_{\gamma a}G_{a\gamma}G_{is}G_{si} + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.47}$$

We expand the first term on the right-hand side of (B.47) similar to (B.11) to get

$$
\begin{aligned}
&\Big(\tau^{-1} - z - \frac{\widetilde{m}}{d}\Big) G_{is} G_{si} \\
&\quad = -\frac{1}{N} \sum_{\beta} \frac{1}{(t_\beta^{-1} - \tau)^2} \Big(\sum_{p,q} x_{\beta p} G_{pq}^{(\beta)} x_{\beta q} + \tau \Big) G_{is} G_{si} \\
&\qquad + \frac{1}{N} \sum_{\beta} \frac{1}{(t_\beta^{-1} - \tau)^3} \Big(\sum_{p,q} x_{\beta p} G_{pq}^{(\beta)} x_{\beta q} + \tau \Big)^2 G_{is} G_{si} \\
&\qquad + (L_+ - z) G_{is} G_{si} + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.48}
$$

Taking the partial expectation $\mathbb{E}_\beta$ and proceeding as in (B.12)–(B.15) we find for the first term on the right-hand side of (B.48) that

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{\tau^2}{N^3} \sum_{i,s} \sum_{\beta} \frac{1}{(t_\beta^{-1} - \tau)^2} \Big(\sum_{p,q} x_{\beta p} G_{pq}^{(\beta)} x_{\beta q} + \tau \Big) G_{is} G_{si} \Big] \\
&\quad = \mathbb{E}[X_{32}] + \tau^2 A_3 \mathbb{E}[X'_{44} - 4X_{44}] + \mathcal{O}(\Psi^5).
\end{aligned}
$$

We thus have

$$
\begin{aligned}
&\mathbb{E}\Big[-\frac{\tau^2}{N^2} \sum_{a \neq s} \Big(\tau^{-1} - z - \frac{\widetilde{m}^{(a)}}{d}\Big) G_{is}^{(a)} G_{si}^{(a)} \Big] \\
&\quad = \mathbb{E}[X_{32}] - \tau^2 A_3 \mathbb{E}[X_{42} + 4X_{44} + X'_{44}] \\
&\qquad - \tau^2 (L_+ - z) \mathbb{E}[X_{22}] \\
&\qquad + \tau^{-1} \mathbb{E}[2X_{43} + X'_{44}] + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.49}
$$

From (B.42), (B.46) and (B.49) we find for the first term on the right-hand side of (B.41) that

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{N^2} \sum_{a \neq s} G_{aa} G_{is}^{(a)} G_{si}^{(a)} \Big] \\
&\quad = -\tau \mathbb{E}[X_{22}] - 2\mathbb{E}[X_{33}] + 2\tau^2 A_3 \mathbb{E}[2X_{43} + 3X_{44} + X'_{44}] \\
&\qquad + \tau^{-1} \mathbb{E}[2X_{43} + X'_{44}] \\
&\qquad + \mathbb{E}[X_{32}] - \tau^2 A_3 \mathbb{E}[X_{42} + 4X_{44} + X'_{44}] - \tau^2 (L_+ - z) \mathbb{E}[X_{22}] \\
&\qquad + \tau^{-1} \mathbb{E}[2X_{43} + X'_{44}] \\
&\qquad - \tau^{-1} \mathbb{E}[X_{42} + 2X'_{44}] \\
&\qquad + \tau^{-1} \mathbb{E}[2X_{44} + X'_{44}] + 2\mathbb{E}[X_{33}] + \mathcal{O}(\Psi^5).
\end{aligned} \tag{B.50}
$$

Plugging (B.50) into (B.41) we finally find

$$\begin{aligned}\mathbb{E}[mX_{22}] &+ \tau^2(L_+ - z)\mathbb{E}[X_{22}] \\ &= -\tau\mathbb{E}[X_{22}] + \mathbb{E}[X_{32}] \\ &\quad + (\tau^2 A_3 + \tau^{-1})\mathbb{E}[-X_{42} + 4X_{43} + 2X_{44} + X'_{44}] + \mathcal{O}(\Psi^5).\end{aligned}$$

Since $X_{32} = (m+\tau)X_{22}$ by definition, we obtain

$$(L_+ - z)\mathbb{E}[X_{22}] = (A_3 + \tau^{-3})\mathbb{E}[-X_{42} + 4X_{43} + 2X_{44} + X'_{44}] + \mathcal{O}(\Psi^5), \tag{B.51}$$

which is the optical theorem obtained from mX_{22}.

B.4. Proof of Lemma B.1. In this subsection we prove Lemma B.1 based on the optical theorems derived in Sections B.1, B.2 and B.3.

PROOF OF LEMMA B.1. For simplicity set

$$X_3 := 2(X_{32} + X_{33}), \qquad X_4 := 3(X_{42} + 2X_{43} + 4X_{44} + X'_{44}). \tag{B.52}$$

From (B.29), (B.33) and (B.51), we have

$$\begin{aligned}\mathbb{E}[X_3] - N^{-1} &= (A_4 + \tau^{-4} + 2\tau^{-1}A_3)\mathbb{E}[X_4] - 2\tau^{-1}N^{-1}\mathbb{E}[m+\tau] \\ &\quad + 2\tau^{-1}(L_+ - z)\mathbb{E}[X_{22}] + \mathcal{O}(\Psi^5),\end{aligned}$$

hence,

$$\begin{aligned}\mathbb{E}[X_3] - N^{-1} &= (A_4 + \tau^{-4} + 2\tau^{-1}A_3)\mathbb{E}[X_4] \\ &\quad - \tau^{-1}\mathbb{E}[6X_{42} - 4X_{43} - 2X'_{44}] + \mathcal{O}(\Psi^5).\end{aligned} \tag{B.53}$$

Subtracting 8-times (B.38) from (B.53), we obtain

$$\begin{aligned}\mathbb{E}[X_3] - N^{-1} &= (A_4 + \tau^{-4} + 2\tau^{-1}A_3)\mathbb{E}[X_4] \\ &\quad - \tau^{-1}6\mathbb{E}[X_{42} + 2X_{43} + 4X_{44} + X'_{44}] + \mathcal{O}(\Psi^5) \\ &= (A_4 + \tau^{-4} + 2\tau^{-1}A_3 - 2\tau^{-1})\mathbb{E}[X_4] + \mathcal{O}(\Psi^5).\end{aligned}$$

Using $A_3 + \tau^{-3} = 1$ [see (A.24)], we conclude that

$$\mathbb{E}[X_3] - N^{-1} = (A_4 - \tau^{-4})\mathbb{E}[X_4] + \mathcal{O}(\Psi^5). \tag{B.54}$$

This proves (B.1) and completes the proof of Lemma B.1. □

APPENDIX C: PROOF OF LEMMA 6.1

In this last section we prove Lemma 6.3.

PROOF OF LEMMA 6.3. In a first step of the proof of (6.14), we express $(\partial_t t_\alpha)/t_\alpha^2$ in terms of γ and $\dot\gamma$.

From the time evolution of $\Sigma = \operatorname{diag}(\sigma_\alpha)$ in (6.1), we have

$$\partial_t \frac{1}{\sigma_\alpha(t)} = -\mathrm{e}^{-t}\frac{1}{\sigma_\alpha(0)} + \mathrm{e}^{-t} = 1 - \frac{1}{\sigma_\alpha(t)} \qquad (t \ge 0).$$

Since $t_\alpha = \gamma\sigma_\alpha$ by the definition of T, we get

$$\frac{\partial_t t_\alpha}{t_\alpha^2} = -\partial_t \frac{1}{t_\alpha(t)} = \left(\frac{\dot\gamma}{\gamma} + 1\right)\frac{1}{t_\alpha(t)} - \frac{1}{\gamma} \qquad (t \ge 0).$$

Recalling the definitions of (A_k) in (A.23) and that $A_2 = \tau^{-2}$, we then obtain, dropping for simplicity the t-dependence from the notation,

$$\frac{1}{N}\sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^2}\frac{1}{(t_\alpha^{-1}-\tau)^3} = \left(\frac{\dot\gamma}{\gamma}+1\right)\tau^{-2} + \left(\frac{\dot\gamma}{\gamma}+1\right)\tau A_3 - \frac{1}{\gamma}A_3, \tag{C.1}$$

respectively,

$$\frac{1}{N}\sum_\alpha \frac{\partial_t t_\alpha}{t_\alpha^3}\frac{1}{(t_\alpha^{-1}-\tau)^4} = \left(\frac{\dot\gamma}{\gamma}+1\right)A_3 + \left(\frac{\dot\gamma}{\gamma}+1\right)\tau A_4 - \frac{1}{\gamma}A_4. \tag{C.2}$$

Using the short-hand notation

$$X_3 = 2(X_{32} + X_{33}), \qquad X_4 = 3\big(X_{42} + 2X_{43} + 4X_{44} + X'_{44}\big) \tag{C.3}$$

[see (B.52)], we observe that (6.14) is proven, once we have established that

$$\begin{aligned}&\big[(\dot\gamma+\gamma)\tau^{-2} + (\dot\gamma\tau + \gamma\tau - 1)A_3\big]\operatorname{Im}\mathbb{E}[X_3]\\ &\quad = \big[(\dot\gamma+\gamma)A_3 + (\dot\gamma\tau+\gamma\tau-1)A_4\big]\operatorname{Im}\mathbb{E}[X_4] + \mathcal{O}(\Psi^5).\end{aligned} \tag{C.4}$$

Combining the following lemma with Lemma B.1, it is straightforward to assure the validity (C.4).

LEMMA C.1. *Let γ and τ be defined in* (6.3) *and* (6.4). *Then we have*

$$(\dot\gamma+\gamma)\tau^{-2} + (\dot\gamma\tau+\gamma\tau-1)A_3 = \gamma\big(\tau^{-2}A_4 - A_3^2\big), \tag{C.5}$$

$$(\dot\gamma+\gamma)A_3 + (\dot\gamma\tau+\gamma\tau-1)A_4 = \gamma\big(\tau^{-2}A_4 - A_3^2\big)\big(A_4 - \tau^{-4}\big). \tag{C.6}$$

Assuming the correctness of Lemma C.1, we can recast (C.4) as

$$\begin{aligned}&\gamma\big(\tau^{-2}A_4 - A_3^2\big)\operatorname{Im}\mathbb{E}[X_3]\\ &\quad = \gamma\big(\tau^{-2}A_4 - A_3^2\big)\big(A_4 - \tau^{-4}\big)\operatorname{Im}\mathbb{E}[X_4] + \mathcal{O}(\Psi^5).\end{aligned} \tag{C.7}$$

Since $\mathbb{E}[X_3]-1/N=(A_4-\tau^{-4})\mathbb{E}[X_4]+\mathcal{O}(\Psi^5)$, by the optical theorem (B.1), we see that (C.7), respectively (C.4), indeed hold true. This in turn proves, by the discussion above, the claim in (6.14), that is, Lemma 6.3.

It remains to prove Lemma C.1:

PROOF OF LEMMA C.1. First, we differentiate the sum rule

$$\frac{1}{N}\sum_\alpha\left(\frac{1}{t_\alpha^{-1}-\tau}\right)^2=\frac{1}{\tau^2}$$

[see (6.5)] with respect to t to find

$$\frac{\dot\tau}{\tau^3}=\frac{1}{N}\sum_\alpha\frac{\partial_t t_\alpha^{-1}-\dot\tau}{(t_\alpha^{-1}-\tau)^3}=-\dot\tau A_3-\gamma^{-1}A_3+\left(\frac{\dot\gamma}{\gamma}+1\right)\frac{1}{N}\sum_\alpha\frac{t_\alpha^{-1}}{(t_\alpha^{-1}-\tau)^3},$$

which yields

$$(A_3+\tau^{-3})\dot\tau=\gamma^{-1}\big[(\dot\gamma+\gamma)(\tau^{-2}+\tau A_3)-A_3\big].\tag{C.8}$$

Using $A_3+\tau^{-3}=1$, we hence get

$$\dot\tau=\gamma^{-1}\big[(\dot\gamma+\gamma)\tau-A_3\big].\tag{C.9}$$

Similarly, differentiating the sum rule

$$\frac{1}{N}\sum_\alpha\left(\frac{1}{t_\alpha^{-1}-\tau}\right)^3+\frac{1}{\tau^3}=1$$

[see (6.5)] with respect to t we find

$$(A_4-\tau^{-4})\dot\tau=\gamma^{-1}\big[(\dot\gamma+\gamma)(A_3+\tau A_4)-A_4\big].$$

Combination with (C.9) yields

$$(A_4-\tau^{-4})\big[(\dot\gamma+\gamma)\tau-A_3\big]=(\dot\gamma+\gamma)(A_3+\tau A_4)-A_4,$$

hence

$$\dot\gamma+\gamma=\tau^{-4}A_3-A_3A_4+A_4=\tau^{-4}A_3+\tau^{-3}A_4.\tag{C.10}$$

Thus, we can write the left-hand side of (C.5) as

$$\begin{aligned}&(\tau^{-4}A_3+\tau^{-3}A_4)\tau^{-2}+(\tau^{-3}A_3+\tau^{-2}A_4-1)A_3\\&\quad=(\tau^{-3}A_3+\tau^{-2}A_4)-A_3\\&\quad=(\tau^{-3}A_3+\tau^{-2}A_4)-A_3(A_3+\tau^{-3})\\&\quad=(\tau^{-2}A_4-A_3^2).\end{aligned}\tag{C.11}$$

This proves (C.5). Similarly, we have for the left-hand side of (C.6)

$$
\begin{aligned}
&(\tau^{-4}A_3+\tau^{-3}A_4)A_3+(\tau^{-3}A_3+\tau^{-2}A_4-1)A_4\\
&\quad=(\tau^{-4}A_3+\tau^{-3}A_4)A_3+(\tau^{-3}A_3+\tau^{-2}A_4-(A_3+\tau^{-3})^2)A_4\\
&\quad=\tau^{-4}A_3^2+\tau^{-2}A_4^2-A_3^2A_4-\tau^{-6}A_4\\
&\quad=(A_4-\tau^{-4})(\tau^{-2}A_4-A_3^2).
\end{aligned}
\tag{C.12}
$$

This proves (C.6), and hence completes the proof of Lemma C.1. □

Having proven Lemma C.1, we can complete the proof of Lemma 6.3. □

Acknowledgments. We thank Horng-Tzer Yau for numerous discussions and remarks. We are grateful to Ben Adlam, Jinho Baik, Zhigang Bao, Paul Bourgade, László Erdős, Iain Johnstone and Antti Knowles for comments. We are also grateful to the anonymous referee for carefully reading our manuscript and suggesting several improvements.

REFERENCES

[1] ANDERSON, G. W. (2013). Convergence of the largest singular value of a polynomial in independent Wigner matrices. *Ann. Probab.* **41** 2103–2181. MR3098069

[2] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[3] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

[4] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575

[5] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

[6] BAO, Z., PAN, G. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864

[7] BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inform. Theory* **57** 2400–2419. MR2809098

[8] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** 1–53. MR3183577

[9] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). On the principal components of sample covariance matrices. Available at arXiv:1404.0788.

[10] BLOEMENDAL, A. and VIRÁG, B. (2011). Limits of spiked random matrices II. Available at arXiv:1109.3704.

[11] BLOEMENDAL, A. and VIRÁG, B. (2013). Limits of spiked random matrices I. *Probab. Theory Related Fields* **156** 795–825. MR3078286

[12] BORDENAVE, C. and GUIONNET, A. (2013). Localization and delocalization of eigenvectors for heavy-tailed random matrices. *Probab. Theory Related Fields* **157** 885–953. MR3129806

[13] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976

[14] EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592

[15] ERDŐS, L., KNOWLES, A. and YAU, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922

[16] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770

[17] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** 1–58. MR3068390

[18] ERDŐS, L., SCHLEIN, B., YAU, H.-T. and YIN, J. (2012). The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1–46. MR2919197

[19] ERDŐS, L., YAU, H.-T. and YIN, J. (2011). Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.* **2** 15–81. MR2847916

[20] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427

[21] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147

[22] FELDHEIM, O. N. and SODIN, S. (2010). A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.* **20** 88–123. MR2647136

[23] FÉRAL, D. and PÉCHÉ, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *J. Math. Phys.* **50** 073302, 33. MR2548630

[24] FORRESTER, P. J. (1993). The spectrum edge of random matrix ensembles. *Nuclear Phys. B* **402** 709–728. MR1236195

[25] FORRESTER, P. J. (2013). Probability densities and distributions for spiked and general variance Wishart β-ensembles. *Random Matrices Theory Appl.* **2** 1350011, 19. MR3149441

[26] GIRKO, V. L. (1984). The circular law. *Teor. Veroyatn. Primen.* **29** 669–679. MR0773436

[27] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991

[28] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[29] JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I* 307–333. Eur. Math. Soc., Zürich. MR2334195

[30] KAY, S. (1998). *Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory*. Prentice Hall, New York.

[31] KNOWLES, A. and YIN, J. (2014). Anisotropic local laws for random matrices. Available at arXiv:1410.3516.

[32] LEE, J. O. and SCHNELLI, K. (2015). Edge universality for deformed Wigner matrices. *Rev. Math. Phys.* **27** 1550018, 94. MR3405746

[33] LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434

[34] MARCHENKO, V. A. and PASTUR, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR Sb.* **1** 457–483.

[35] MEHTA, M. L. (1991). *Random Matrices*, 3rd ed. Academic Press, New York.

[36] MO, M. Y. (2012). Rank 1 real Wishart spiked model. *Comm. Pure Appl. Math.* **65** 1528–1638. MR2969495

[37] NADAKUDITI, R. R. and EDELMAN, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Trans. Signal Process.* **56** 2625–2638. MR1500236

[38] NADAKUDITI, R. R. and SILVERSTEIN, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE J. Sel. Top. Signal Process* **4** 468–480.

[39] ONATSKI, A. (2008). The Tracy–Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.* **18** 470–490. MR2398763

[40] ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070

[41] PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670

[42] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978

[43] SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541

[44] SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444

[45] TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449

[46] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246

[47] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[48] VINOGRADOVA, J., COUILLET, R. and HACHEM, W. (2013). Statistical inference in large antenna arrays under unknown noise pattern. *IEEE Trans. Signal Process.* **61** 5633–5645. MR3130031

[49] VOICULESCU, D. V., DYKEMA, K. J. and NICA, A. (1992). *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups. CRM Monograph Series* **1**. Amer. Math. Soc., Providence, RI. MR1217253

[50] WANG, K. (2012). Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices Theory Appl.* **1** 1150005, 24. MR2930383

DEPARTMENT OF MATHEMATICAL SCIENCES
KAIST
291 DAEHAK-RO, YUSEONG-GU
DAEJEON 34141
REPUBLIC OF KOREA
E-MAIL: jioon.lee@kaist.edu

IST AUSTRIA
AM CAMPUS 1
3400 KLOSTERNEUBURG
AUSTRIA
E-MAIL: kevin.schnelli@ist.ac.at